УНІВЕРСИТЕТ ДЕРЖАВНОЇ ФІСКАЛЬНОЇ СЛУЖБИ УКРАЇНИ

СЕРІЯ “НА ДОПОМОГУ СТУДЕНТУ УДФСУ”
Заснована 2016 року

Н. Новицька,
Т. Паянок,
Н. Параниця,
С. Богдан

АНАЛІЗ ДАНИХ ЗА ДОПОМОГОЮ R

Навчально-методичний посібник

Ірпінь
2021

УДК 519.22.35(075.8)
ББК 22.1727я73
А 64

Рекомендовано до друку Вченою радою
Університету державної фіскальної служби України
(*протокол № 15 від 24 грудня 2020 р.*)

Рецензенти:

Скрипник А.В., д.е.н., професор, професор кафедри економічної кібернетики Національного університету біоресурсів і природокористування України;

Лісовий А.В., д.е.н., професор, завідувач кафедри аудиту та економічного аналізу Університету державної фіскальної служби України.

Автори:

Новицька Н.В., к.е.н., доцент; Паянок Т.М., к.е.н., доцент; Параниця Н.В., к.е.н., доцент; Богдан С.В., асистент кафедри

А 64 **Аналіз даних за допомогою R** : навчально-методичний посібник / Н. Новицька, Т. Паянок, Н. Параниця, С. Богдан. – Ірпінь : Університет державної фіскальної служби України, 2021. – 324 с. – (Серія "На допомогу студенту УДФСУ"; т. 87).

ISBN 978-966-337-624-0

Навчальний посібник спрямований на формування комплексу знань, практичних навичок і вмінь студентів щодо використання прикладної програми R у процесі аналізу даних. Видання містить перелік практичних завдань різного рівня складності, тестів та питань до самоконтролю здобувачів вищої освіти.

Призначений для викладачів і здобувачів вищої освіти, а також усіх, хто цікавиться питаннями використання прикладних програм в аналізі даних для використання в управлінській діяльності.

УДК 159.9(075.8)
ББК 88.5я73

© Новицька Н., Паянок Т., Параниця Н.,
Богдан С., 2021
ISBN 978-966-337-624-0
© Університет ДФС України, 2021

ЗМІСТ

ПЕРЕДМОВА.....6

ІНСТРУКЦІЯ З ІНСТАЛЯЦІЇ ТА ВСТАНОВЛЕННЯ R та R-STUDIO.....9

ЧАСТИНА 1. НАВЧАЛЬНО-МЕТОДИЧНИЙ ПРАКТИКУМ.....13

ТЕМА 1. ПРОГРАМНИЙ ПРОДУКТ R ЯК ІНСТРУМЕНТ АНАЛІЗУ ДАНИХ.....14

1.1. Основні відомості про систему та програмний додаток R.....15

1.2. Дані та операції над ними, основи управління даними в системі R.....23

1.3. Методи та інструменти управління даними в середовищі R.....34

Питання для самоперевірки та контролю знань.....38

Тестові завдання для самоконтролю та перевірки знань.....39

ТЕМА 2. ПРЕДСТАВЛЕННЯ ТА ВІЗУАЛІЗАЦІЯ ДАНИХ В СЕРЕДОВИЩІ R.....42

2.1. Основи роботи з діаграмами в середовищі R.....42

2.2. Базові діаграми та методи їх побудови і конфігурації в системі R.....46

2.3. Візуалізація даних та типи складних діаграм у середовищі R.....57

Питання для самоперевірки та контролю знань.....64

Тестові завдання для самоконтролю та перевірки знань.....65

ТЕМА 3. СТАТИСТИЧНА ОБРОБКА ТА АНАЛІЗ ДАНИХ В СЕРЕДОВИЩІ R 68
3.1. Описова статистика та таблиці частот і спряженості в системі R 69
3.2. Кореляційний та коваріаційний аналізи в середовищі R 79
3..3 Статистична перевірка гіпотез за допомогою тестів Стьюдента в системі R 84
Питання для самоперевірки та контролю знань 88
Тестові завдання для самоконтролю та перевірки знань 88

ТЕМА 4. МЕТОДИКА ПРОВЕДЕННЯ РЕГРЕСІЙНОГО АНАЛІЗУ В R 91
4.1. Основи регресійного аналізу в середовищі R 92
4.2. Оцінка придатності моделі за результатами регресійного аналізу в системі R 102
4.3. Коригування регресійних моделей у цілях їх удосконалення в середовищі R 112
Питання для самоперевірки та контролю знань 122
Тестові завдання для самоконтролю та перевірки знань ... 123

ТЕМА 5. МЕТОДИКА ПРОВЕДЕННЯ ДИСПЕРСІЙНОГО АНАЛІЗУ В R 126
5.1. Однофакторний дисперсійний аналіз в середовищі R 126
5.2. Методика здійснення однофакторного коваріаційного аналізу в системі R 132
5.3. Багатофакторний дисперсійний аналіз у середовищі R 135
Питання для самоперевірки та контролю знань 140
Тестові завдання для самоконтролю та перевірки знань ... 140

ЧАСТИНА 2. ЛАБОРАТОРНИЙ ПРАКТИКУМ 143
ЛАБОРАТОРНА РОБОТА 1. Побудова, дослідження та аналіз лінійної багатофакторної економетричної моделі .. 144
ЛАБОРАТОРНА РОБОТА 2. Побудова та дослідження нелінійної багатофакторної економетричної моделі на прикладі неокласичної виробничої функції Кобба-Дугласа та її використання для аналізу та прогнозування процесу виробництва 178
ЛАБОРАТОРНА РОБОТА 3. Побудова та дослідження лінійної багатофакторної економетричної моделі з мільтиколіанерністю 204
ЛАБОРАТОРНА РОБОТА 4. Побудова та дослідження лінійної багатофакторної економетричної моделі з гетероскедастичністю 232
ЛАБОРАТОРНА РОБОТА 5. Побудова, дослідження лінійної багатофакторної економетричної моделі з автокореляцією залишків та моделі з інструментальними змінними 247

ЧАСТИНА 3. ТЕСТОВИЙ ПРАКТИКУМ 297

ТЕРМІНОЛОГІЧНИЙ СЛОВНИК 318

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ 321

ПЕРЕДМОВА

Важливим інструментом та методом сучасних аналітиків є програми для аналітичних розрахунків та обробки даних. Ефективне програмне забезпечення для аналізу даних – запорука якості аналітичного дослідження і надійності, придатності та ефективності отриманих результатів. У діяльності сучасних фахівців з аналізу обов'язковою вимогою до компетентностей висувається саме володіння інформаційними системами і технологіями аналізу даних. Одним із сучасних та потужних інструментів аналізу даних є середовище R.

У навчально-методичному посібнику «Аналіз даних за допомогою R» викладено теоретико-методичні та практичні аспекти одного з найпотужніших, професійних і сучасних засобів роботи з числами, починаючи від простих обчислювальних завдань і закінчуючи статистичною обробкою великих масивів даних – мовою R. R – це вільнопоширюване програмне середовище з відкритим кодом і мова програмування для статистичної обробки даних і роботи з графікою. R можна застосовувати у будь-яких задачах з даними. Середовище покликане розв'язувати завдання математичної статистики, первинного аналізу даних і математичного моделювання. За допомогою R можна підготувати дані для дослідження, яке може бути здійснено за допомогою реалізованих у різних функціях статистичних методів, а потім вивести отримані результати для подальшого аналізу.

Навчально-методичний посібник «Аналіз даних за допомогою R» *призначений* для використання в освітньому процесі підготовки здобувачів вищої освіти галузі знань 07 «Управління та адміністрування», спеціальності 071 «Облік і оподаткування» в межах викладання дисциплін «Економетрика», «Прикладна економетрика», «Аналіз даних за допомогою R» тощо.

Метою навчально-методичного посібника «Аналіз даних за допомогою R» є поглиблене вивчення теоретичних та практичних

основ застосування мови програмування в середовищі R для комплексного аналізу масиву даних.

Основними завданнями навчального посібника є такі:

– систематизувати основні відомості про систему та програмний додаток R;

– визначити особливості даних та операції над ними, основ управління даними в R;

– розглянути методи та інструменти управління даними в середовищі R;

– вивчити основи роботи з діаграмами в середовищі R;

– дослідити базові діаграми та методи їх побудови і конфігурації в системі R;

– застосувати візуалізацію даних та визначити типи складних діаграм у середовищі R;

– дослідити описову статистику та таблиці частот і спряженості в системі R;

– здійснити кореляційний та коваріаційний аналізи в середовищі R;

– провести статистичну перевірку гіпотез за допомогою тестів Стьюдента в системі R;

– вивчити основи регресійного аналізу в середовищі R;

– здійснити оцінку придатності моделі за результатами регресійного аналізу в R;

– забезпечити коригування регресійних моделей у цілях їх вдосконалення в середовищі R;

– провести однафакторний дисперсійний аналіз у середовищі R;

– систематизувати методику здійснення однофакторного коваріаційного аналізу в R;

– здійснити багатофакторний дисперсійний аналіз в середовищі R.

Сформована мета та завдання навчального посібника «Аналіз даних за допомогою R» покликані на отримання таких *результатів*, що дозволять:

– уміти розвивати та підвищувати свій загальнокультурний і професійний рівень, самостійно освоювати нові методи роботи та

знання щодо комплексного бачення сучасних проблем економіки та управління;

– володіти інноваційними технологіями, обґрунтовувати вибір та пояснювати застосовування нової методики підготовки і надання облікової інформації для потреб управління суб'єктом господарювання;

– збирати, оцінювати та аналізувати фінансові та нефінансові дані для формування релевантної інформації в цілях прийняття управлінських рішень;

– обґрунтовувати інноваційні підходи до інформаційного забезпечення системи контролю використання ресурсного потенціалу суб'єктів господарювання та органів державного сектору з урахуванням стратегії розвитку бізнесу;

– обґрунтовувати вибір і порядок застосування управлінських інформаційних технологій для обліку, аналізу, аудиту та оподаткування в системі прийняття управлінських рішень з метою їх оптимізації;

– готувати й обґрунтовувати висновки задля консультування власників, менеджменту суб'єкта господарювання та інших користувачів інформації у сфері обліку, аналізу, контролю, аудиту, оподаткування;

– вміти проєктувати, планувати і проводити пошукові і розвідувальні роботи, здійснювати їх інформаційне, методичне, матеріальне, фінансове та кадрове забезпечення.

У навчально-методичному посібнику «Аналіз даних за допомогою R» викладено теоретичний матеріал за питаннями тем, а також питання для самоконтролю після вивчення тем, тестові завдання закритого типу, лабораторні завдання для самоконтролю знань; систематизовано глосарій; список використаних та рекомендованих джерел.

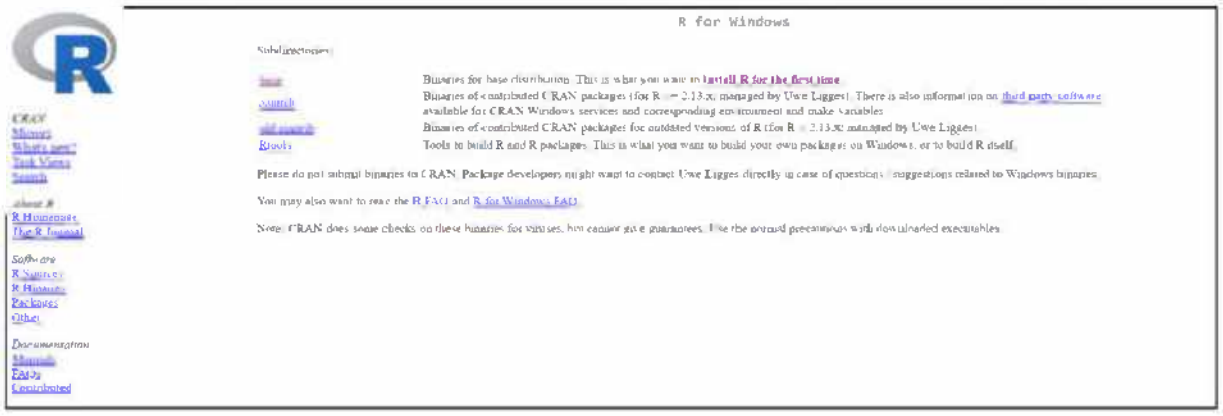
Авторський колектив

ІНСТРУКЦІЯ З ІНСТАЛЯЦІЇ ТА ВСТАНОВЛЕННЯ R та R-STUDIO

1. Перейти на онлайн-ресурс: **http://lib.stat.cmu.edu/R/CRAN/**



2. Обираємо та натискаємо пункт **install R for the first time**:



3. Обираємо та натискаємо пункт **Download R 4.0.3 for Windows**, після цього починається завантаження інсталяційного пакета програми:



4. Встановіть RStudio за посиланням:

http://www.rstudio.com/products/rstudio/download/

Rstudio – це зручна графічна оболонка до R. Не потрібно плутати Rstudio з R-studio – платною програмою для відновлення даних.

5. Налаштуйте Rstudio.

Запустіть RStudio. У **розділі Tools---Global Options---Sweave--- "Weave .Rnw files using"** виберіть **knitr**.

6. Якщо ім'я користувача введене українськими або російськими буквами, потрібно створити спеціальну папку, де R буде зберігати пакети. Створіть папку для установки пакетів без українських та російських букв, наприклад C: / Rlib. Відкрийте в меню Пуск "Зміна змінних середовища поточного користувача"(просто в пошук досить набрати цю фразу).

У розділі "Змінні середовища користувача для..." натисніть кнопку "Створити". В імені змінної напишіть R_LIBS (саме великими буквами), а в значенні змінної вкажіть шлях до створеної на 1 кроці папці(C:/Rlib).

Після цього запустіть RStudio і перевірте, що R знає, куди йому ставити пакети. Для цього виконайте в R команду: .libPaths(). Вона повинна вказати шлях до папки C: / Rlib.

Замість пунктів 2–4 можна запустити командний рядок Windows (Пуск – Виконати – набрати "cmd" – натиснути "ok") і в ній виконати: setx R_LIBS C:/Rlib.

7. Встановіть усі необхідні для курсу пакети R. Для цього відкрийте RStudio, скопіюйте та вставте в робоче середовище такі команди:

```
    install.packages("tidyverse")   # колекція
пакетів від Hadley Wickham
    install.packages("gdata")
    install.packages("dplyr")# МНК
    install.packages("psych")#          описові
статистики
    install.packages("lmtest")    #    тести    в
лінійних моделях
    install.packages("sjPlot")#
```

```
    install.packages("gdata")#
    install.packages("sgof")#
    install.packages("ggplot2")#
    install.packages("foreign")#
    install.packages("hexbin")#
    install.packages("car") тести
    install.packages("sandwich") # оцінки кова-
ріаційної матриці робасті до гетероскедастич-
ності
    install.packages("erer") # добірка пакетів
для емпіричних досліджень
    install.packages("GGally") # матриця діаграм
розсіювання    install.packages("lattice")    #
конкурент ggplot2
    install.packages("vcd") # мозаїчний графік
install.packages("hexbin") # графів з шести-
кутників
    install.packages("sjPlot") # візуалізація
результатів МНК
    install.packages("psych")# тести
    install.packages("glmnet") # LASSO install.-
packages("HSAUR")#
    install.packages("sgof")#
    install.packages("car") тести
    install.packages("RODBC")# для доступу до
файлів Excel
    install.packages("Hmisc") # імпорт даних в
форматі SPSS
    install.packages("vcd") # графіки
    install.packages ("MASS") # узагальнений МНК
    install.packages ("AER") # інструментальні
змінні
    install.packages ("ivpack") # додаткові
опції для інструментальних змінних
```

Поставте курсор послідовно на кожну команду та натисніть клавішу Ctrl+Enter. Для цього необхідно з'єднання з інтернетом. Якщо на екрані з'являються червоні написи, це не означає, що в назвах пакетів є помилки. Ознакою помилки є тільки явне повідомлення **Error**. Тепер ви встановили всі необхідні пакети для виконання лабораторної роботи.

8. Увімкніть антивірус.

ЧАСТИНА 1

Навчально-методичний практикум

ТЕМА 1. Програмний продукт R як інструмент аналізу даних

ТЕМА 2. Представлення та візуалізація даних у середовищі R

ТЕМА 3. Статистична обробка та аналіз даних у середовищі R

ТЕМА 4. Методика проведення регресійного аналізу в R

ТЕМА 5. Методика проведення дисперсійного аналізу в R

ТЕМА 1
ПРОГРАМНИЙ ПРОДУКТ R ЯК ІНСТРУМЕНТ АНАЛІЗУ ДАНИХ

Мета: *систематизувати основні відомості про систему та програмний додаток R; визначити особливості даних та операції над ними, основ управління даними в R; розглянути методи та інструменти управління даними в середовищі R.*

План

1.1. Основні відомості про систему та програмний додаток R.

1.2. Дані та операції над ними, основи управління даними в системі R.

1.3. Методи та інструменти управління даними в середовищі R.

Ключові поняття та терміни: *мова програмування R, RStudio, робочий простір, поточна робоча директорія, набір даних, фактори, списки, математичні функції, статистичні функції, функції розподілу, циклічні конструкції, транспонування, агрегування.*

Основні функції та програмні коди: `help.start(), help(), help.search(), example(), apropos(), data(), getwd(), setwd(), ls(), rm(), help(), history(), savehistory(), loadhistory(), save.image(), save(), load(), q(), rownames(), factor (), object(), data frame(), vector(), matrix(), array(), data frame(), list(), edit(), abs(), sqrt(), ceiling(), floor(), mean(), median() , sd(), var(), mad(), range(), sum(), min(), max(), beta, binom, Cauchy ,chisq, exp, f, norm, pois, for (),while(),if-else(),ifelse(), switch(), t(), by().`

1.1. Основні відомості про систему та програмний додаток R

R – це мова програмування і середовище для статистичних обчислень і графічного аналізу, подібна з мовою S, спочатку розробленою у 1976 рік Bell Labs. Початок функціонування – 1993 року, коли двоє молодих новозеландських учених Росс Іхака і Роберт Джентльмен анонсували свою нову розробку, яку назвали R. Вони розробили безкоштовну вільну реалізацію з легко-розширюваною модульною архітектурою. Незабаром виникла розподілена система зберігання і поширення пакетів до R, відома під абревіатурою «CRAN». Це програма для аналізу даних з відкритим кодом.

Мова R останнім часом стала активно конкурувати з комерційними статистичними пакетами, такими як SAS, SPSS, Stata.

Серед користувачів мови R такі компанії, як: Boeing, Google, Pfizer, Merck, Bank of America, the InterContinental Hotels Group і Shell.

Ключова особливість мови R – це її відкритість, що дає можливість фахівцям самостійно створювати власні алгоритми. По суті, це framework для розробки. Сьогодні існує величезна кількість (більше 7000) готових, безкоштовних пакетів для R.

У січні 2015 року фірма Microsoft викупила компанію Revolution Analytics, яка була заснована 2007 року ученими Єльського університету. Крім безпосередньої участі в розробці R, компанія відома своїми продуктами для розробників на мові програмування R, призначеного для вирішення завдань у сфері статистики та аналізу даних, що сьогодні особливо актуально в контексті обробки Big Data.

Відомо, що Microsoft і раніше застосовувала вільну мову R, а ця покупка, очевидно, свідчить про інтерес софтверного гіганта до цього напряму. У січні 2016 року Microsoft оголосила про появу мови R у Visual Studio.

У СУБД Oracle для аналізу великих обсягів даних використовується надбудова Oracle Enterprise R, яка є частиною опції Oracle Advanced Analytics. Oracle R Enterprise спеціально

розроблений, щоб справлятися з величезними масивами даних, для цього він використовує масштабованість і паралелізм СУБД Oracle, а також вбудовані функції СУБД для оптимізації роботи скриптів, написаних на R.

R застосовується скрізь, де необхідна робота з даними. Найбільше його застосовують для статистичного аналізу – від обчислення середніх до часових рядів.

Переваги мови R полягають у:

– наявності безкоштовного кросплатформенного середовища для простого написання програм;

– багатому арсеналі статистичних методів. У R реалізовані складні статистичні процедури, які ще недоступні в інших програмах;

– якісній векторній графіці, за допомогою якої реалізовані найрізноманітніші і потужні методи візуалізації даних з доступних;

– отриманні даних з різних джерел у придатному для використання вигляді. R може імпортувати дані з різних джерел, включаючи текстові файли, системи управління базами даних, інші статистичні програми і спеціалізовані сховища даних. R може також записувати дані у форматах усіх цих систем;

– існування понад 7000 перевірених прикладних пакетів;

– встановленні на різні операційні системи, включаючи Windows, Unix, Linux і Mac OS.

У R є і чимало недоліків. Найголовніший з них – це складність у вивченні програми. Команд багато, вводити їх треба вручну, запам'ятати все важко, а звичної системи меню немає. Тому часом важко знайти, як саме застосовувати ту чи іншу методику аналізу. Зручним засобом обчислень у R є RStudio.

RStudio являє собою безкоштовне інтегроване середовище розробки для R. Завдяки ряду своїх особливостей цей програмний продукт спрощує роботу з R і робить її досить зручною.

Другий недолік R – порівняна повільність. Деякі функції, особливо такі, що використовують цикли, і види об'єктів, особливо списки і таблиці даних, працюють у десятки разів повільніше, ніж їх аналоги в комерційних пакетах. Але цей недолік

виправляється розробниками, хоча і повільно. Нові версії R здатні здійснювати паралельні обчислення, адже створюються оптимізовані варіанти підпрограм, що працюють набагато швидше, пам'ять у R використовується все ефективніше, а замість циклів рекомендується застосовувати векторизовані обчислення.

R – це чутлива до регістру клавіатури інтерпретована мова програмування. Можливий варіант введення команд по одній у відповідь на запит на введення команди (>), або запускати набір команд з вихідного файлу. Типи даних у середовищі R різноманітні: вектори, матриці, таблиці даних і списки (сукупність декількох об'єктів).

Основний функціонал програми реалізується за допомогою вбудованих і створюваних користувачем функцій, і всі об'єкти (набори даних) зберігаються в пам'яті програми до завершення інтерактивної сесії. Основні функції доступні за замовчуванням. Решта функцій містяться в пакетах, які можуть активуватися в ході роботи з програмою в міру необхідності.

Командні рядки складаються з функцій і присвоєнь. R використовує символ <- для присвоєнь замість звичайного знаку рівності. Наприклад, командний рядок

`x <- rnorm(3)` створює об'єкт типу вектор, який називається `x` і містить три випадкові елементи нормального розподілу.

Коментарі передує символом #. Будь-який текст, який йде після #, ігнорується програмою.

Для запуску програми в OS Windows слід запустити R зі стартового меню. В операційній системі Mac треба відкрити програму через директорію додатків. В Linux необхідно набрати «R» в командному рядку терміналу. Будь-яка із цих дій запустить середовище R (рис. 1.1).



Рис. 1.1. Стартове вікно середовища R

Для освоєння інтерфейсу середовища R пропонуємо виконати такий приклад.

Приклад. Здійснюється дослідження величини мобілізованого податку з доходів фізичних осіб протягом 10 років (табл. 1.1).

Таблиця 1.1

Величини податкових надходжень та середньомісячної заробітної плати

Середньомісячна заробітна плата, грн	Величина податкових надходжень, млн грн
1806	957
2122	1090
2448	1246
2822	1399
3094	1469
3129	1684
3984	2246
4887	3313
6709	4465
8106	5415

Спочатку слід здійснити введення даних про середньомісячну заробітну плату та величину мобілізованого податку з доходів фізичних осіб у вигляді векторів, використовуючи функцію `c()`, яка перетворює аргументи у вектор або список. Потім слід обчислити середнє арифметичне і стандартне відхилення для значень мобілізованого податку, а також коефіцієнт кореляції між середньомісячною заробітною платою і мобілізованим податком, який графічно зображує взаємозв'язок між цими двома змінними для візуального вивчення тренду.

Функція `q()`, як показано в наведеному нижче програмному коді, завершує сесію і дозволяє вийти з програми.

```
> salary <- c(1806,2122,2448,2822,3094,-
3129,3984,4887,6709,8106)
> tax <- c(957,1090,1246,1399,1469,-
1684,2246,3313,4465,5415)
> mean(tax)
[1] 2328.4
> sd(tax)
[1] 1551.771
> cor(salary,tax)
[1] 0.9949681
> plot(salary,tax)
> q()
```

З програмного коду видно, що середня величина мобілізованого податку за останні 10 років становить 2328,4 млн грн, а стандартне відхилення дорівнює 1551,771 млн грн і що існує сильний лінійний взаємозв'язок між величиною середньомісячної заробітної плати за роками (коефіцієнт кореляції дорівнює 0,9949). Цей взаємозв'язок також видно на діаграмі розсіювання на рис. 1.2.

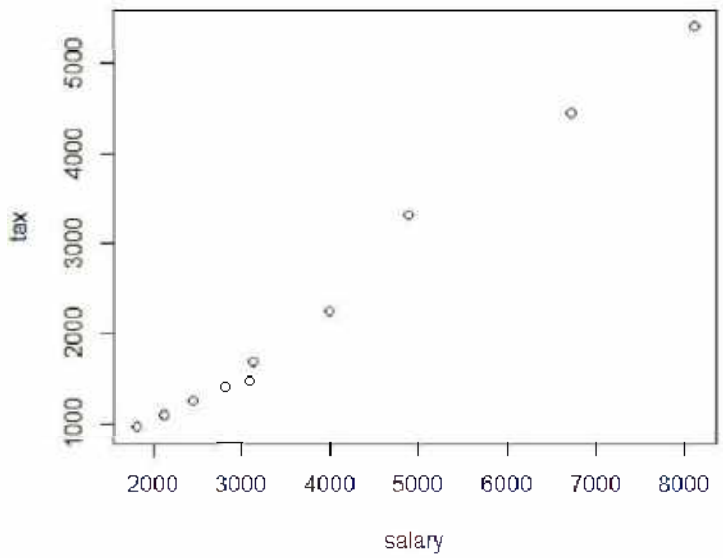


Рис. 1.2. Діаграма розсіювання величини мобілізованого податку з доходів фізичних осіб (tax, у млн грн) залежно від величини середньомісячної заробітної плати (salary, у грн)

Зрозуміло, що зі зростанням величини середньомісячної заробітної плати зростатиме й мобілізований податок з доходів фізичних осіб.

R володіє великими довідковими матеріалами. Вбудована система допомоги містить детальні роз'яснення, посилання на літературу та приклади для кожної функції зі встановлених пакетів. Довідку можна викликати за допомогою функцій, перерахованих у табл. 1.2.

Таблиця 1.2

Довідка та її виклик у середовищі R

Функція	**Дія**
`help.start(функція)`	загальна довідка
`help(функція)` або `?функція`	довідка за функцією *функція*
`help.search("функція")` або `??функція`	пошук у довідці записів, що містять *функція*
`example(функція)`	приклади використання функції *функція*
`apropos("функція", mode="function")`	список всіх доступних функцій, в назві яких є *функція*
`data()`	список демонстраційних даних, що містяться в завантажених пакетах

Функція `help.start()` відкриває вікно браузера з переліком доступних посібників різного рівня складності, що часто ставляться, і посилань на джерела. Функція `RSiteSearch()` здійснює пошук на задану тему в онлайн-довідниках і архівах розсилок і представляє результати в вікні браузера. Функція `vignette()` викликає список вступних статей у форматі PDF.

Безпосередню роботу в програмі R користувач здійснює у робочому просторі.

Робочий простір – це поточне робоче середовище R в пам'яті комп'ютера, яке включає в себе будь-які створені користувачем

об'єкти (вектори, матриці, функції, таблиці даних або списки). У кінці кожної сесії можна зберегти робочий простір, і він автоматично завантажиться при наступному запуску програми. Команди інтерактивно вводяться у відповідь на запит до їх введення. Можна використовувати стрілки вгору і вниз для переміщення між введеними раніше командами. Це дозволяє викликати попередню команду, відредагувати її і знову виконати її, натиснувши клавішу Enter.

Поточна робоча директорія – це та директорія, де знаходяться файли даних і куди за замовчуванням зберігаються результати. Функція `getwd()` дозволяє дізнатися, яка директорія в даний момент є робочою. Можна призначити робочу директорію за допомогою функції `setwd()`. Якщо з'являється необхідність імпортувати файл, який знаходиться не в робочій директорії, потрібно написати повний шлях до нього. Завжди слід заносити в лапки назви файлів і директорій.

Деякі стандартні команди для управління робочим простором наведені в табл. 1.3.

Таблиця 1.3

Функції управління робочим простором у R

Функція	**Дія**
`getwd()`	виведення на екран назву поточної робочої директорії
`setwd("директорія")`	призначення *директорія* поточною робочою директорією
`ls()`	виведення на екран переліку об'єктів у поточному робочому просторі
`rm("перелік об'єктів")`	видалити один або декілька об'єктів
`help(options)`	довідка про доступні опції
`history(n)`	вивести на екран останні *n* команд
`savehistory("файл")`	зберегти історію команд у *файл*
`loadhistory("файл")`	завантажати історію команд
`save.image("файл")`	зберегти робочий простір у *файл*

Продовження таблиці 1.3

`save("перелік об'єктів", file="файл")`	зберегти визначені об'єкти у файл
`load("файл")`	завантажити збережений робочий простір у поточну сесію
`q()`	вийти з програми

За замовчуванням запуск R починає інтерактивну сесію, де введення здійснюється з клавіатури, а результати виводяться на екран. Однак можна також запустити команди зі скрипту (файлу, який містить функції R), а виведення можливе безпосередньо в різні пристрої.

Введення. Функція `source ("файл")` запускає скрипт. Якщо не прописаний шлях до файлу, мається на увазі, що він знаходиться в поточній робочій директорії. Наприклад, команда `source("myscript.R")` запускає серію команд R, які записані в файлі `myscript.R`. Прийнято, щоб файли скриптів мають розширення `.R`, але це не обов'язкова умова.

Текстове виведення. Функція `sink("файл")` виводить усі результати у файл з назвою `файл`. За замовчуванням, якщо цей файл вже існує, нова версія записується замість попередньої. Параметр `append=TRUE` дозволяє додавати новий текст у файл, а не записувати його замість старого тексту. Параметр `split=TRUE` дозволяє виводити результати і на екран, і в текстовий файл. Виконання команди `sink()` без аргументів відновить виведення результатів тільки на екран.

Графічне виведення. Хоча команда `sink()` управляє виведенням тексту, вона не робить ніякого впливу на виведення графіки. Для управління виводу зображень слід використовувати одну з функцій: `pdf("файл.pdf")` – для виведення у формат ***.pdf*, `png("файл.png")` та `jpeg("файл.jpeg")` – для виведення у формат ***.png* та ***.jpeg* відповідно.

Функція `dev.off()` направляє виведення графіки назад на екран.

1.2. Дані та операції над ними, основи управління даними в системі R

Перший етап будь-якого аналізу даних – створення набору даних, в якому міститься інформація для вивчення, у потрібному форматі. У R це завдання складається з таких аспектів:

– вибору типу даних;

– введення або імпорту даних в обраному форматі.

Набір даних – це, як правило, прямокутний масив даних, в якому ряди відповідають спостереженням, а стовпці – ознакам.

Приклад. У табл. 1.4 поданий набір даних про здійснені аудиторські перевірки.

Таблиця 1.4

Вихідні дані здійснених аудиторських перевірок

Установа, в якій проводився аудит	Дата надання висновку	Кількість порушень	Форма власності	Вид висновку
1	02/01/2020	12	Type1	positive
2	04/05/2020	14	Type2	conditionally positive
3	07/07/2020	5	Type1	negative
4	10/11/2020	16	Type1	positive

Institution – порядковий номер установи, в якій проводився аудит, відповідно до плану перевірки; Date – дата надання аудиторського висновку: місяць / день / рік; Violation – кількість виявлених порушень; Property – форма власності установи, в якій проводився аудит (Type1 – державна, Type2 – приватна); Opinion – вид аудиторського висновку (positive – позитивний; conditionally positive – умовно позитивний; negative – негативний).

Представники різних професій по-різному називають рядки і стовпці в наборі даних. Статистики називають їх спостереженнями (observation) і змінними (variable), аналітики, які працюють з базами даних, говорять про записи (record) і поля (field), а ті, хто працює в сфері пошуку в даних прихованої інформації (data mining) і машинного навчання (machine learning), називають їх

зразками (example) і властивостями (attribute). У контексті вивчення аналізу даних будемо застосовувати поняття «спостереження» і «змінна».

Потрібно розрізняти структуру набору даних і типи даних, які його складають. У наборі даних, наведеному в табл. 1.4, Institution – це назва рядка або спостереження, Date – змінна в форматі дати, Violation – безперервна (або кількісна – continuous) змінна, Property – номінальна (nominal) змінна і Opinion – це порядкова (або шкальна – ordered) змінна.

R працює з різними структурами даних, включаючи скаляри, вектори, масиви даних, таблиці даних і списки. Таблиця 1.4 буде прочитана в R як таблиця даних. Така велика різноманітність підтримуваних структур дає мові R велику гнучкість у роботі з даними.

Типи даних у R бувають числовими (numeric), текстовими (character), логічними (TRUE / FALSE, істина / хибність), комплексними (уявне число) і необробленими (байти). Змінні Institution, Date і Violation будуть прочитані R як числові, а Property і Opinion – як текстові. Крім того, необхідно буде задати в програмі, що Institution – це назви спостережень, що в Date записані дати і що Property і Opinion – це порядкова і номінальна змінні відповідно. В R імена рядків називаются **rownames**, а категоріальні (номінальні та порядкові) змінні – **factor** (фактор).

R працює з різними структурами даних, включаючи скаляри, вектори, матриці, масиви даних, таблиці даних і списки. Вони різняться за типами даних, способом створення, складністю пристрою, а також способом позначати і витягувати їх окремі елементи.

У R об'єктом (**object**) називається все, що може бути подано у вигляді змінних, включаючи константи, різні типи даних, функції і навіть діаграми. У об'єктів є вид (визначає, в якому вигляді об'єкт зберігається в пам'яті) і клас (який вказує загальним функціям типу `print`, як ним управляти).

Таблиця даних (**data frame**) – це тип структури даних у R, аналогічний тому виду, в якому зберігаються дані в звичайних статистичних програмах (наприклад, у SAS, SPSS). Стовпці – це

змінні, а рядки – це спостереження. В одній таблиці даних можуть міститися змінні різних типів (наприклад, числові і текстові). Таблиці даних – це основний тип структури даних.

Фактори – це номінальні або порядкові змінні. У R вони зберігаються і обробляються особливим чином.

Вектори (**vector**) – це одномірні масиви даних, які можуть містити числові, текстові або логічні значення. Для створення вектора застосовується функція об'єднання `c()`.

Вектори можуть бути різних типів:

```
a <- c(1,6,18,14,28,39)
b <- c("податок на додану вартість","податок з доходів фізичних осіб","податок на прибуток підприємства")
c <- c(TRUE, TRUE, TRUE, FALSE, TRUE, FALSE)
```

Тут *a* – числовий вектор, *b* – текстовий вектор, *c* – логічний вектор. Зверніть увагу на те, що всі елементи вектора повинні бути одного типу (числові, текстові або логічні). Не можна змішувати дані різних типів в одному векторі.

Скаляри – це вектори, що складаються з одного елемента, наприклад `f <- 3, g <- "військовий збір" h <- TRUE`. Вони використовуються для позначення констант.

Окремі елементи вектора можна викликати за допомогою числового вектора, що складається з номерів елементів і замкнутого в квадратні дужки. Наприклад, `a[c(2,4]` позначає другий і четвертий елементи вектора a. Ось ще приклади:

```
a <- c(1,6,18,14,28,39)
> a[3]
[1] 18
> a[c(2,4,6)]
[1] 6 14 39
> a[3:5]
[1] 18 14 28
```

Знак двокрапки в останньому прикладі використаний для створення послідовності чисел.

Матриця (**matrix**) – це двовимірний масив даних, в якому кожен елемент має однаковий тип (числовий, текстовий або

логічний). Матриці створюють за допомогою функції matrix. Загальний формат такий:

```
mymatrix <- matrix(вектор,nrow=кількість_-
рядків, ncol=кількість_стовпців,-
byrow=логічне_значення,dimnames=list-
(текст_вектор_назви_строк,текст_вектор_назви_ст
овпців))
```

де vector містить елементи матриці, nrow і ncol, які визначають число рядків і стовпців у матриці, а dimnames містить назви рядків і стовпців (їх вказувати не обов'язково), які зберігаються у вигляді текстових векторів. Параметр byrow визначає, як повинна заповнюватися матриця – по рядках (byrow=TRUE) або за стовпцями (byrow=FALSE). За замовчуванням матриця заповнюється за стовпцями. Наведений нижче програмний код ілюструє застосування функції matrix.

Приклад. Необхідно створити матрицю розміром 3 × 6, яка відображає дані статистичного дослідження величини доходів (у тис. грн) трьох домогосподарств за останні півроку.

```
> y <- matrix(1:18,nrow=3,ncol=6)
> y
[,1] [,2] [,3] [,4] [,5] [,6]
[1,] 1 4 7 10 13 16
[2,] 2 5 8 11 14 17
[3,] 3 6 9 12 15 18
> cells <-c(25,39,17,52,101,69,78,20,13,95,-
52,25,67,71,16,18, 98,71)
> rnames <-c("Домогосподарство 1",
"Домогосподарство 2", "Домогосподарство 3")
> cnames <-c("січень","лютий","березень",-
"квітень","травень", "червень")
> mymatrix <-matrix(cells,nrow=3,ncol=6,-
byrow=TRUE,dimnames= list(rnames,cnames))
> mymatrix
січень лютий березень квітень травень
червень
```

```
Домогосподарство 1 25 39 17 52 101 69
Домогосподарство 2 78 20 13 95 52 25
Домогосподарство 3 67 71 16 18 98 71
```

Спочатку створюємо матрицю розміром 3 × 6. Потім задаємо назви рядків і стовпців, заповнюємо матрицю по рядках. Можна позначати рядки, стовпці і елементи матриці за допомогою індексів і квадратних дужок. Наприклад, `X[i,]` позначає і-й рядок матриці X, `X[,j]` – позначає її j-ий стовпець, а `X[i,j]` відповідає елементу цієї матриці, розташованому на перетині цього рядка і цього стовпчика. Як індекси можна використовувати числові вектори, щоб позначити одразу кілька рядків або стовпців.

Масиви даних (**array**) схожі з матрицями, але можуть мати більше двох вимірів. Масиви даних створюються за допомогою функції `array` за таким зразком:

```
myarray <- array (vector, dimensions, dimnames)
```

де `vector` містить самі дані, `dimensions` – це числовий вектор із зазначенням розмірності для кожного вимірювання, а `dimnames` – це необов'язковий список назв вимірювань. Нижче, як приклад, поданий програмний код, за допомогою якого створений тривимірний масив чисел.

Приклад. Необхідно побудувати масив розміром 2 × 3 × 6, що відображає дані статистичного дослідження величини доходів (у тис. грн) вибірково двох домогосподарств з трьох областей за останні півроку.

```
> dim1 <-c("Домогосподарство 1",-
"Домогосподарство 2")
> dim2 <-c("Київська обл.","Житомирська
обл.","Львівська обл.")
> dim3 <-c("січень", "лютий", "березень",
"квітень", "травень", "червень")
> cells <-c(25,39,17,52,101,69,78,20,13,95,
52,25,67,71,16,18,98,71,45,49,27,56,111,79,88,
21, 15, 95,52, 25,66,71,16,19,98,72)
```

```
> z <-array(cells,c(2,3,6), dimnames=list
(dim1,dim2,dim3))
> z
, , січень

Київська обл. Житомирська обл. Львівська
обл.
Домогосподарство 1 25 17 101
Домогосподарство 2 39 52 69

, , лютий

Київська обл. Житомирська обл. Львівська
обл.
Домогосподарство 1 78 13 52
Домогосподарство 2 20 95 25

, , березень

Київська обл. Житомирська обл. Львівська
обл.
Домогосподарство 1 67 16 98
Домогосподарство 2 71 18 71

, , квітень

Київська обл. Житомирська обл. Львівська
обл.
Домогосподарство 1 45 27 111
Домогосподарство 2 49 56 79

, , травень

Київська обл. Житомирська обл. Львівська
обл.
Домогосподарство 1 88 15 52
```

```
Домогосподарство 2 21 95 25

, , червень

Київська обл. Житомирська обл. Львівська обл.
Домогосподарство 1 66 16 98
Домогосподарство 2 71 19 72
```

Очевидно, що масиви даних – це просто розширені матриці. Вони можуть бути корисні при написанні програм для реалізації нових статистичних методів. Як і в матрицях, всі елементи масиву даних повинні мати однаковий тип. Система позначень елементів тут така ж, як для матриць. Таблиця даних (**data frame**) – це більш широко використовуваний, порівняно з матрицею, об'єкт, оскільки різні стовпці можуть містити різні типи даних (числовий, текстовий тощо). Таблиця даних – це найбільш часто використовувана структура даних у R.

Набір даних про аудиторські перевірки (табл. 1.4) складається з числових і текстових даних. Ці дані потрібно подати у вигляді таблиці даних, а не матриці, оскільки тут є дані різних типів. Таблиця даних створюється за допомогою функції `data.frame()`:

```
mydata <- data.frame(col1,col2,col3)
```

де `col1,col2,col3` – це вектори будь-якого типу (текстового, числового чи логічного), які стануть стовпцями таблиці. Назви кожному стовпцю можна привласнити за допомогою функції `names()`.

Приклад. Побудувати вихідні дані здійснених аудиторських перевірок на основі табл. 1.4.

```
> institution <-c(1,2,3,4)
> violation <-c(12,14,5,16)
> property <-c("Type1","Type2", "Type1", "Type1")
> opinion <-c("positive","conditionally positive","negative", "positive")
```

```
> auditdata <-data.frame(institution,
violation, property,opinion)
> auditdata
 institution violation property opinion
1 1 12 Type1 positive
2 2 14 Type2 conditionally positive
3 3 5 Type1 negative
4 4 16 Type1 positive
```

Кожен стовпчик повинен містити дані тільки одного типу, при цьому в одній таблиці даних можуть бути стовпчики з даними різного типу. Існує кілька способів позначити елементи таблиці даних. Можна використовувати індекси або вказувати номери стовпців.

Функція `attach()` додає зазначену таблицю даних до шляху пошуку R.

Якщо потрібно створити об'єкти, які будуть існувати поза конструкцією `with()`, слід використовувати спеціальний символ присвоєння `<<-` замість звичайного (`<-`). Цей прийом дозволить зберегти створений об'єкт у робочому просторі поза конструкцією `with()`.

У R назви рядків можуть бути призначені за допомогою параметра `rowname` функції створення таблиці даних.

```
auditdata <- data.frame(institution,
violation, property, opinion, row.names=
institution)
```

призначає `institution` змінною, яка буде використовуватися для позначення рядків при виведенні даних і створення діаграм у R.

Змінні бувають номінальними, порядковими або безперервними. Номінальні змінні – це категоріальні дані, які неможливо впорядкувати. Змінна `property` – це приклад номінальних даних. Навіть якщо позначимо Type 1 (тип 1) одиницею, а Type 2 (тип 2) – двійкою, все одно ці цифри не можна буде порівнювати в термінах «більше – менше». Порядкові дані можна впорядкувати, але не оцінити кількісно. Змінна `opinion` –

приклад порядкових даних. Безперервні змінні можуть приймати будь-яке значення в межах певного діапазону. Їх значення можна впорядкувати і збагнути, наскільки одне з них більше іншого.

Кількість виявлених помилок, виражена в одиницях, є безперервною змінною і може приймати такі значення, як 16 або 36, а також будь-які значення між цими двома.

Категоріальні (номінальні і порядкові) дані називаються у R **факторами**. Фактори важливі в R, оскільки вони визначають, як дані будуть проаналізовані і графічно представлені.

Функція `factor()` зберігає категоріальні дані у вигляді вектора з цілих чисел в діапазоні від одного до k (де k – число унікальних значень категоріальної змінної) і у вигляді внутрішнього вектора з ланцюга символів (вихідних значень змінної), відповідних цим цілим числам.

Списки – це найскладніший тип даних у R. Фактично список – це упорядкований набір об'єктів (компонентів). Список може об'єднувати різні (можливо, не пов'язані між собою) об'єкти під одним ім'ям. Наприклад, список може являти собою поєднання векторів, матриць, таблиць даних і навіть інших списків. Список можна створити за допомогою функції `list()`:

```
mylist <- list(об'єкт 1, об'єкт 2, …)
```

де об'єкти – це будь-які структури даних. Об'єктам у списку можна задати ім'я. У вигляді списку можна зберігати будь-яке число об'єктів.

Можна позначати елементи списку, вказавши їх номер або назву всередині подвійних квадратних дужок. Списки – це важливий тип структури даних у R з двох причин. По-перше, вони дозволяють упорядкувати і викликати на екран розрізнену інформацію. По-друге, результати виконання багатьох команд є списками. У цьому випадку користувач отримує з таких списків потрібну інформацію.

Тепер, коли є структури даних, потрібно наповнити їх даними. Зазвичай аналітики стикаються з даними, які надходять з різних джерел і в різних форматах. Завдання полягає в тому, щоб

імпортувати дані в програму, проаналізувати їх і представити звіт про результати. У R реалізовані різні способи імпорту даних.

У R можна вводити дані з клавіатури, імпортувати з текстових файлів, з Microsoft Excel і Access, з поширених статистичних програм, спеціалізованих форматів, а також з різних систем управління базами даних. Оскільки ніколи не можна вгадати, звідки можуть надійти дані, розглянемо найпоширеніші джерела по порядку.

Найпростіший спосіб введення даних – це введення з клавіатури. Функція `edit()` відкриває текстовий редактор, куди можна ввести свої дані. Алгоритм реалізації зазначеної дії такий:

1. Створити порожню таблицю даних (або матрицю), вказавши назви і типи змінних.

2. Відкрити текстовий редактор із цим об'єктом, ввести дані і зберегти результат у вигляді об'єкта з даними.

Приклад. Створити таблицю вихідних даних щодо основних засобів підприємства із інформацією, що відображатиме термін корисного використання, групу основних засобів та залишкову вартість об'єкта основного засобу.

```
> mydata <-data.frame(term=numeric(0), group=character(0), residualvalue=numeric(0))
> mydata <-edit(mydata)
```

Присвоєння типу `term=numeric(0)` створюють порожню (без даних) змінну заданого типу. Функція `edit()` працює з копією об'єкта.

Після того, як закрити текстовий редактор, результати зберігаються у вигляді обраного об'єкта (у цьому випадку об'єкт `mydata`). Повторне введення функції `mydata <- edit (mydata)` дозволяє редагувати введені дані і додавати нові. Скорочена версія команди `mydata <- edit (mydata)` – це `fix (mydata)`.

Цей метод корисний для невеликих обсягів даних. Для наборів даних більшого розміру слід використовувати інші методи: імпорт даних з існуючих текстових файлів, електронних

таблиць Excel, статистичних програм або систем управління базами даних.

Імпорт даних з текстових файлів можливий за допомогою команди `read.table()`, функції, яка зберігає дані у вигляді таблиці. Приклад використання функції:

```
mydataframe <-read.table(file,header=
логічне_значення, sep="роздільник", row.names=
"назва")
```

де `file` – файл документа, `header` – логічне значення, що визначає, чи містить перший рядок назви змінних, `sep` – вказує, яким символом розділені елементи даних.

Кращий спосіб прочитати файл у форматі Excel – це зберегти його у форматі текстового файлу з роздільниками й імпортувати в R, як це описано вище. У Windows для доступу до файлів Excel також можна використовувати пакет RODBC. У першому рядку електронної таблиці повинні міститися назви змінних (стовпців). Перш за все необхідно завантажити та встановити пакет RODBC: `install.packages("RODBC")`.

Код, призначений для імпорту даних:

```
library(RODBC)
channel <- odbcConnectExcel ("myfile.xls")
mydataframe <- sqlFetch(channel,"mysheet")
odbcClose(channel)
```

де `myfile.xls` – це файл Excel, `mysheet` – це назва потрібного листа з робочої книги Excel, `channel` – це допоміжний об'єкт RODBC, створений функцією `odbcConnectExcel()`, і `mydataframe` – це вихідна таблиця даних. Цей пакет можна також використовувати для імпорту даних з Microsoft Access. Подробиці вказані у файлі довідки: `help(RODBC)`.

Набори даних у форматі SPSS можна імпортувати в R за допомогою функції `read.spss()` з пакета `foreign`. Інший варіант – використовувати функцію `spss.get()` з пакета Hmisc. Це інтерфейсна функція, яка автоматично визначає багато

параметрів функції `read.spss()`, що робить перетворення файлів більш простим і ефективним.

Перш за все необхідно завантажити та встановити пакет Hmisc: `install.packages("Hmisc")`.

Код, призначений для імпорту даних:

```
library(Hmisc)
mydataframe <-spss.get("mydata.sav", use.value.labels=TRUE)
```

де `mydata.sav` – це файл даних SPSS, який потрібно імпортувати, параметр `use.value.labels=TRUE` означає, що змінні з підписаними значеннями будуть перетворені у фактори з відповідними рівнями, а `mydataframe` – це вихідна таблиця даних.

1.3. Методи та інструменти управління даними в середовищі R

Під час аналізу даних необхідно створювати нові змінні і перетворювати існуючі. Це досягається за допомогою присвоєнь у такому форматі:

```
змінна <- вираз
```

Термін вираз означає оператори і функції. Арифметичні оператори, які використовуються при складанні формул в R:

– додавання (+);

– віднімання (-);

– добуток (*);

– ділення (/);

– піднесення до степеню (^ або **);

– знаходження залишку від ділення (`x%%y`) тощо.

Приклад. Маємо дані таблиці `sales` за продажами продукції протягом перших двох кварталів (`x1` та `x2`), необхідно створити нову змінну `sumx`, яка є сумою цих двох змінних, і нову змінну `meanx` – середнє арифметичне цих двох змінних. Якщо використати програмний код

```
sumx <- x1+x2
```

```
meanx <- (x1+x2)/2
```

отримаємо повідомлення про помилку, оскільки R не розпізнає те, що x1 і x2 – це частина таблиці даних sales. Якщо використати програмний код:

```
sumx <- sales$x1+sales$x2
meanx <- (sales$x1+sales$x2)/2
```

то він буде виконаний, але в результаті вийде вихідна таблиця даних (sales) і два окремих вектори (sumx і meanx). Однак це не те, що вимагалося в умові задачі. Умовою поставлено визначити нові змінні частиною вихідної таблиці даних.

Мету прикладу покликаний досягти наступний програмний код:

```
attach(mydata)
mydata$sumx <- x1+x2
mydata$meanx <- (x1 + x2)/2
detach(mydata)
```

При перекодуванні нові значення змінної визначаються, виходячи із значень цієї та / або інших змінних. Наприклад, це стосується таких аспектів операцій:

– перетворення безперервних даних у категоріальні;

– заміна помилкових даних правильними значеннями;

– створення змінної типу «відповідає / не відповідає».

Для перекодування даних можна використовувати один із знаків логічних операцій, тобто виразів, які повертають значення TRUE / FALSE (істина / хибність):

– менше за (<);

– менше або дорівнює (<=);

– більше за (>);

– більше або дорівнює (>=);

– тотожно дорівнює (==);

– не дорівнює (! =) тощо.

Якщо необхідно перейменувати назви змінних, можна змінити їх в інтерактивному режимі або за допомогою програмного коду. Для зміни назви змінної можна використовувати команду:

1) `fix(leadership)` – для виклику інтерактивного редактора;

2) `rename()` – для зміни за допомогою програмного коду.

Дати зазвичай вводяться в R у вигляді текстових рядків, а потім переводяться у формат дати і зберігаються в числовому вигляді. Для цієї операції використовується функція `as.Date()` у форматі `as.Date(x,"початковий вигляд")`, де x – це текстові дані, а `початковий вигляд` вказує, в якому форматі представлені дати.

Типові **математичні функції** у R:

– модуль – `abs(x)`;

– корінь квадратний – `sqrt(x)`;

– найменше цілочислове значення – `ceiling(x)`;

– найбільше цілочислове значення – `floor(x)`;

– тригонометричні функції – `cos(x),sin(x),tan(x), acos(x),asin(x), atan(x)`;

– логарифмічні функції – `log(x,base=n),log(x), log10(x)`;

– експоненційна функція – `exp(x)` тощо.

Зазначені математичні функції застосовуються для перетворення даних.

Найпоширеніші **статистичні функції**, що застосовуються в R:

– середнє арифметичне значення – `mean(x)`;

– медіанне значення – `median(x)`;

– стандартне відхилення – `sd(x)`;

– дисперсія – `var(x)`;

– абсолютне відхилення медіани – `mad(x)`;

– розмах значень – `range(x)`;

– сума значень – `sum(x)`;

– мінімальне значення – `min(x)`;

– максимальне значення – `max(x)` тощо.

Функції розподілу зазвичай використовуються для створення штучних даних із відомими параметрами і для

обчислення ймовірностей для створених користувачами статистичних функцій.

У програмі R функції розподілу мають вигляд:

– β-розподіл – `beta`;

– біноміальний розподіл – `binom`;

– розподіл Коші – `Cauchy`;

– χ^2-розподіл – `chisq`;

– експоненційний розподіл – `exp`;

– F-розподіл – розподіл Фішера – `f`;

– нормальний розподіл – `norm`.

– розподіл Пуассона – `pois` тощо.

Одна з властивостей функцій R – це те, що їх можна застосовувати до різних типів об'єктів (скаляри, вектори, матриці, масиви і таблиці даних).

Циклічні конструкції багаторазово виконують одну і ту ж команду або набір команд, поки не буде виконана задана умова. До таких конструкцій відносять `for` і `while`.

Конструкція `for`. Циклічна конструкція `for` повторно виконує певну команду, поки значення змінної буде міститися в послідовності `seq`. Синтаксис команди такий:

```
for (var in seq) statement
```

Конструкція `while`. Циклічна конструкція `while` повторно виконує команду, поки задана умова не перестане бути істинною. Загальний вигляд застосування конструкції такий:

```
while (cond) statement
```

Цикли в R можуть бути неефективними і займати багато часу, коли маємо справу з великими обсягами даних. За будь-якої можливості краще замість циклів використовувати вбудовані числові і текстові функції в R разом з функціями сімейства `apply`.

Виконання за умови означає, що команди діють, тільки якщо виконується певна умова. До таких конструкцій відносять `if-else`, `ifelse` і `switch`.

Конструкція `if-else`. Керуюча конструкція `if-else` виконує команду, якщо правильно задано умову. Як опція, інша

команда може виконуватися, якщо задана умова виявиться правильною. Синтаксис такий:

```
if (cond) твердження
if (cond) твердження 1 else твердження 2
```

Конструкція `ifelse`. Конструкція `ifelse` – компактна і векторизована версія конструкції `ifelse`. Синтаксис такий:

```
ifelse (cond, твердження 1, твердження 2)
```

Перша команда виконується, якщо умова `cond` істинна. Якщо умова хибна – виконується друга команда. Слід використовувати цю конструкцію, якщо потрібно розділити дані на дві категорії або оперувати векторами.

Конструкція `switch`. Конструкція `switch` вибирає команди залежно від значення, яке приймає вираз. Синтаксис такий:

```
switch(expr,…)
```

де три крапки означає команди, що відповідають можливим значенням `expr`. Найпростіше зрозуміти, як працює ця конструкція, на прикладі наведеного нижче програмного коду.

Зокрема в R доступні функції транспонування та агрегації.

Транспонування (переворот таблиці на 90°) – напевно, найпростіший спосіб реструктуризації даних. Слід використовувати функцію `t()`, щоб транспонувати матрицю або таблицю даних. В останньому випадку назви рядків стають іменами змінних (стовпців). У R досить просто **агрегувати** дані, використовуючи задану функцію і одну або більше змінних як критерій `(by)`.

Питання для самоперевірки та контролю знань

1. Охарактеризуйте основні відомості про систему та програмний додаток R.

2. У чому полягають переваги мови R?

3. Назвіть основні структурні елементи стартового вікна середовища R.

4. Яким чином здійснюється виклик довідки в середовищі R?

5. Назвіть основні функції управління робочим простором у R.

6. Охарактеризуйте основні дані, операції та дії, що здійснюються в середовищі R.

7. Назвіть основні методи та інструменти управління даними в середовищі R.

8. Перелічіть математичні, статистичні функції, що використовуються в R.

9. Які існують функції розподілу в середовищі R?

10. Охарактеризуйте циклічні конструкції в R, а також дії транспонування та агрегації.

Тестові завдання для самоконтролю та перевірки знань

1. Перевагами мови програмування R виступають такі:

а) наявність безкоштовного кросплатформенного середовища для простого написання програм;

б) якісна векторна графіка, за допомогою якої реалізовані найрізноманітніші і потужні методи візуалізації даних з доступних;

в) встановлення на різні операційні системи, включаючи Windows, Unix, Linux і Mac OS;

г) усі відповіді є правильними.

2. Поточне робоче середовище R в пам'яті комп'ютера, яке включає в себе будь-які створені користувачем об'єкти (вектори, матриці, функції, таблиці даних або списки), – це:

а) робоча директорія;

б) робочий простір;

в) довідка;

г) поточна сесія.

3. Директорія в середовищі R, де знаходяться файли даних і куди за замовчуванням зберігаються результати, – це:

а) поточна робоча директорія;

б) стаціонарна робоча директорія;

в) хмарна робоча директорія;

г) база даних R.

4. Усе, що може бути подано у вигляді змінних, включаючи константи, різні типи даних, функції і навіть діаграми в середовищі R, – це:

а) вектор;

б) набір даних;

в) об'єкт;

г) фактор.

5. Одномірні масиви даних у середовищі R, які можуть містити числові, текстові або логічні значення, – це:

а) набір даних;

б) вектор;

в) об'єкт;

г) фактор.

6. Двовимірний масив даних у середовищі R, в якому кожен елемент має однаковий тип (числовий, текстовий або логічний), – це:

а) фактор;

б) вектор;

в) матриця;

г) об'єкт.

7. Категоріальні (номінальні і порядкові) дані в середовищі R мають назву:

а) списку;

б) векторів;

в) таблиць;

г) масивів даних.

8. У середовищі R для операцій над даними використовуються функції:

а) математичні;

б) статистичні;

в) розподілу;

г) усі відповіді є правильними.

9. Конструкції, що багаторазово виконують одну і ту саму команду або набір команд, поки не буде виконана задана умова, називаються:

а) умовні;

б) циклічні;

в) розподільні;

г) формалізовані.

10. Дія, що дозволяє здійснити поворот таблиці (матриці) на 90° для зміни місцями стовпців та рядків, називається:

а) декомпозиція;

б) агрегація;

в) транспонування;

г) циклування.

ТЕМА 2
ПРЕДСТАВЛЕННЯ ТА ВІЗУАЛІЗАЦІЯ ДАНИХ У СЕРЕДОВИЩІ R

Мета: *вивчити основи роботи з діаграмами в середовищі R; дослідити базові діаграми та методи їх побудови і конфігурації в системі R; застосувати візуалізацію даних та визначити типи складних діаграм у середовищі R.*

План

2.1. Основи роботи з діаграмами в середовищі R.

2.2. Базові діаграми та методи їх побудови і конфігурації у системі R.

2.3. Візуалізація даних та типи складних діаграм в середовищі R.

Ключові поняття та терміни: *стовпчасті діаграми, стовпчаста діаграма із накопиченням, стовпчаста діаграма з угрупованням, спінограма, кругові діаграми, гістограми, ящикова діаграма, точкові діаграми, діаграми розсіювання, матриці діаграм розсіювання, бульбашкова діаграма, матриці кореляції, корелограми.*

Основні функції та програмні коди: `barplot()`, `spine()`, `pie ()`, `hist()`, `boxplot()`, `dotchart()`, `plot()`, `pairs()`, `symbols()`, `corrgram()`.

2.1. Основи роботи з діаграмами в середовищі R

R – це зручна програма для побудови діаграм. У стандартній інтерактивній сесії створюється діаграма, вводячи по одній команді і додаючи елементи діаграми, поки не виходить бажаний вигляд діаграми.

У R діаграми зазвичай створюються в такому інтерактивному стилі (рис. 2.1).

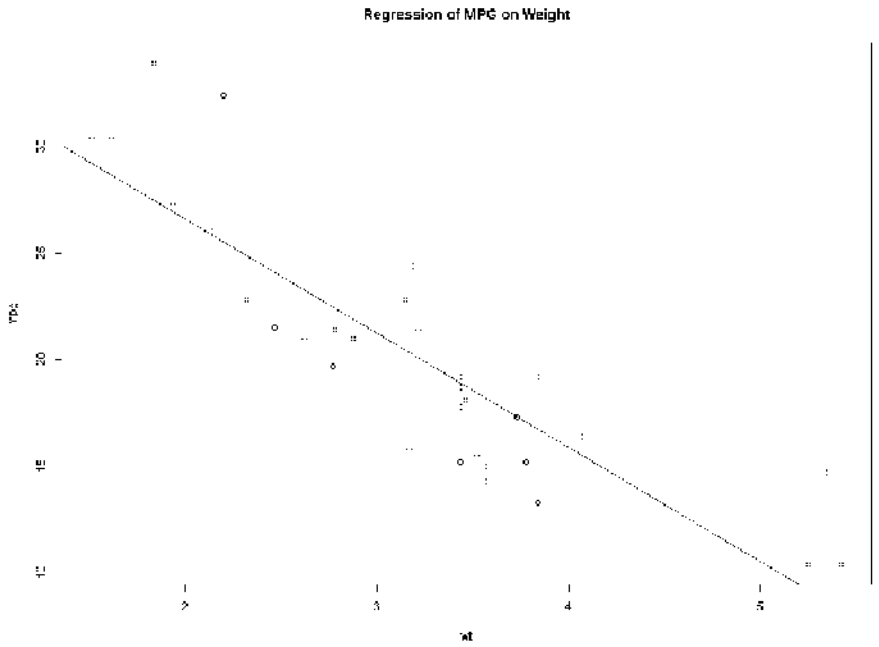


Рис. 2.1. Створення діаграми в R

Для побудови діаграми використовувався програмний код:

```
> attach(mtcars)
> plot(wt,mpg)
> abline(lm(mpg~wt))
> title("Regression of MPG on Weight")
> detach(mtcars)
```

Діаграми можна зберігати за допомогою програмного коду або меню графічного, призначеного для користувача інтерфейсу. Для збереження діаграми за допомогою коду слід розмістити команди побудови діаграми між командами, які призначають місце введення і виведення. Наприклад, наступний програмний код дозволяє зберегти діаграму в форматі PDF під назвою `mygraph.pdf` в поточній робочій директорії:

```
pdf("mygraph.pdf")
attach(mtcars)
plot(wt / mpg)
abline(lm (mpg ~ wt))
title("Regression of MPG on Weight")
detach(mtcars)
dev.off()
```

Крім `pdf()` можна використовувати функції `win.metafile()`, `png()`, `jpeg()`, `bmp()`, `tiff()`, `xfig()` і `postscript()`, щоб зберігати діаграми в інших форматах. Формат Windows metafile доступний тільки під Windows.

Спосіб збереження діаграм за допомогою графічного користувацького інтерфейсу різниться залежно від оперативної системи. Під Windows при активованому графічному пристрої потрібно вибрати в меню файл / зберегти як, а потім в діалозі вибрати потрібний формат графічного файлу і директорію для збереження. Під Mac потрібно вибрати File / Save as в рядку меню, коли активовано графічний пристрій Quartz. У цьому випадку діаграму можна зберегти в єдиному форматі – PDF. В Unix діаграми можна зберігати тільки за допомогою коду.

Нова діаграма, яка створюється за допомогою команди високого рівня, такої як `plot()`, `hist()` або `boxplot()`, зазвичай замінює попередню діаграму. Для того, щоб створити більше однієї діаграми і мати доступ до кожної з них, є кілька способів.

По-перше, можна відкрити новий графічний пристрій, перед тим як створювати нову діаграму:

```
dev.new ()
команди для побудови діаграми 1
dev.new ()
команди для побудови діаграми 2
тощо
```

Кожна нова діаграма буде з'являтися в останньому відкритому вікні.

По-друге, можна отримати доступ до кількох діаграм одразу через призначений для користувача інтерфейс. В операційній системі Mac можна переміщатися між діаграмами в будь-який час за допомогою кнопок Back і Forward в меню Quartz. У Windows ця операція складається з двох етапів. Після того, як відкрите перше вікно графіки, виберіть у меню Історія команд / Запис. Потім

використовуйте пункти меню Попередній і Наступний для переміщення між створеними діаграмами.

По-третє, можна використовувати функції `dev.new()`, `dev.next()`, `dev.prev()`, `dev.set()` і `dev.off()` для одночасного відкриття декількох вікон графіки і вибору необхідної діаграми. Ці команди працюють під усіма операційними системами.

R створює діаграми при мінімальних витратах зусиль користувача. Однак можна використовувати графічні параметри, щоб призначати шрифти, кольори, типи ліній, створювати осі, допоміжні лінії і анотації. Подібна гнучкість відкриває широкі можливості оптимізації діаграм.

Приклад. Подані в таблиці дані характеризують ступінь впливу кількості внесених добрив на урожайність господарських культур – пшениці та ячменю.

Таблиця 2.1

Вплив кількості внесених добрив на урожайність культур

Кількість внесених добрив, кг	Урожайність пшениці, ц	Урожайність ячменю, ц
10	220	210
20	290	260
30	380	340
40	580	430
50	870	580

Вносимо дані в систему R за допомогою програмного коду:

```
fertilizer <- c(10,20,30,40,50)
cropcapacity_wheat <- c(220,290,380,580,870)
cropcapacity_barley <- c(210,260,340,430,580)
```

Простий лінійний графік, який зображує залежність урожайності пшениці від кількості добрив, можна створити так:

```
plot (fertilizer, cropcapacity_wheat, type = "b")
```

`plot()` – це функція загального призначення, яка будує діаграми в R (те, що вийде в результаті застосування, залежить від типу об'єкта, до якого функція застосована). У цьому випадку

`plot(x, y, type = "b")` має у своєму розпорядженні `x` на горизонтальній осі, а `y` – на вертикальній, зображує точки з координатами `(x;y)` і з'єднує їх лініями. Параметр `type="b"` означає, що на графіку повинні бути показані і точки, і лінії.

Отриманий графік зображено на рис. 2.2.

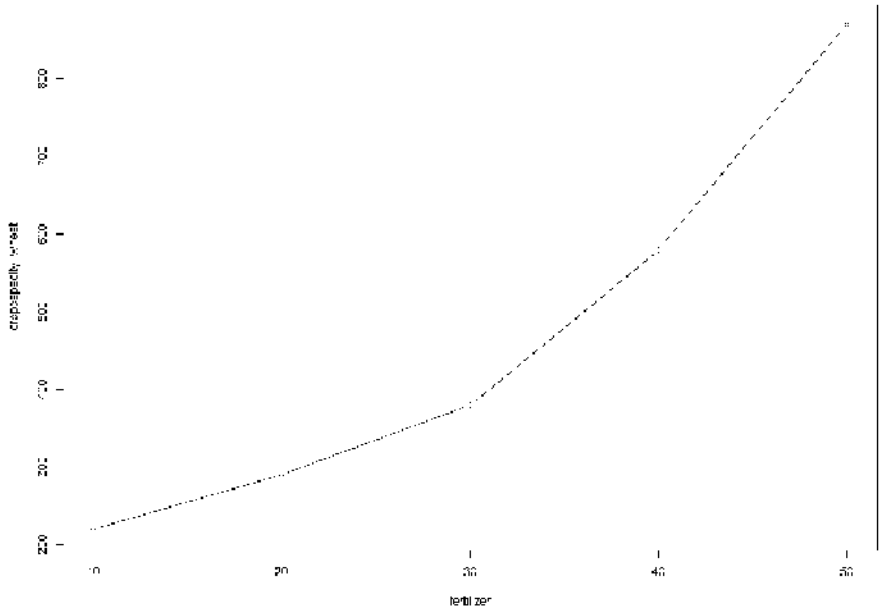


Рис. 2.2. Графік залежності урожайності пшениці від кількості внесених добрив

Для присвоєння назви діаграмі використовується функція `title()`.

2.2. Базові діаграми та методи їх побудови і конфігурації в системі R

Стовпчасті діаграми (bar plot) відображають розподіл (частоту різних значень) категоріальної змінної у вигляді вертикальних або горизонтальних стовпчиків. Найпростіший спосіб побудувати таку діаграму – застосувати функцію `barplot()`:

```
barplot(height)
```

де `height` – це вектор або матриця.

Якщо змінна `height` – вектор, його значення визначають висоту стовпців і створюється вертикальна стовпчаста діаграма.

Додавання параметра `horiz=TRUE` дозволяє отримати горизонтальну стовпчасту діаграму. Також можна додати параметри, що створюють написи. Параметр `main` визначає назву діаграми, тоді як параметри `xlab` і `ylab` додають підписи по осі x і осі y відповідно.

У наведених нижче прикладах продемонструємо результати дослідження нового методу лікування артриту. Відповідні дані містяться в таблиці Arthritis, поширюваної з пакетом vcd. Оскільки цей пакет не входить у базову версію R, слід переконатися, що перед першим використанням завантажено і встановлено його командою `install.packages("vcd")`.

Значення змінної `counts` можна графічно відобразити у вигляді вертикальної або горизонтальної стовпчастої діаграми. Програмний код для виконання цього завдання наведено нижче, а результат зображений на рис. 2.3–2.4.

```
> library(vcd)
> counts <-table(Arthritis$Improved)
> counts
None Some Marked
42 14 28
```

Тут видно, що здоров'я 28 пацієнтів помітно поліпшилося (Marked), 14 пацієнтам стало трохи краще (Some) і 42 людини так і не одужали (None).

```
> barplot(counts,
+ main="Вертикальна стовпчаста діаграма",
+ xlab="Покращення",ylab="Частота")
barplot(counts,
+ main="Горизонтальна стовпчаста діаграма",
+ xlab="Частота",ylab="Покращення",
+ horiz=TRUE)
```

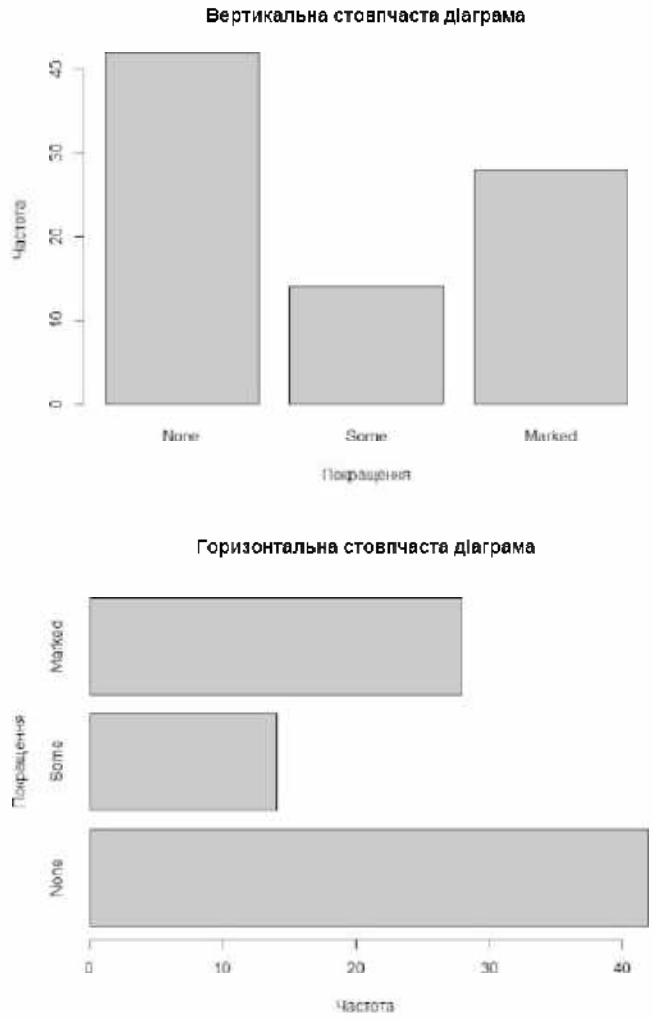


Рис. 2.3–2.4. Стовпчасті діаграми в R

Якщо `height` – матриця, а не вектор, для неї буде побудована **стовпчаста діаграма із накопиченням** (stacked bar plot) або **стовпчаста діаграма з угрупованням** (grouped bar plot). Якщо `beside=FALSE` (за замовчуванням), то кожній змінній матриці відповідає стовпець діаграми, а значення змінної задають висоту «нашарованих» один на одного частин цього стовпчика. Якщо `beside=TRUE`, то кожній змінній відповідає група стовпців діаграми, а значення змінної відображені в розташованих поруч стовпцях цієї групи.

Розглянемо зведені в таблицю дані про реакцію пацієнта на різні способи лікування:

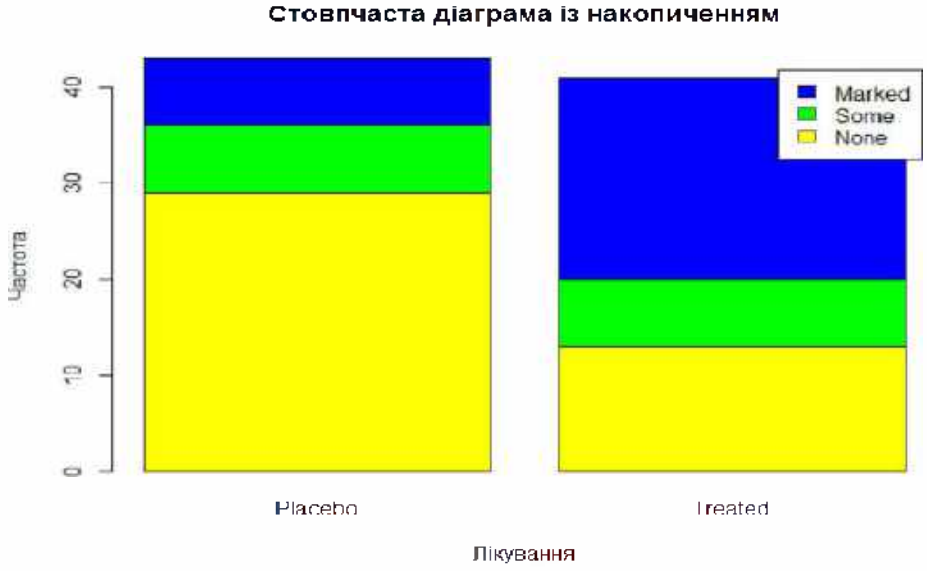
```
> library(vcd)
> counts <- table (Arthritis $ Improved, Arthritis $ Treatment)
```

```
> counts
 Placebo Treated
 None 29 13
 Some 7 7
 Marked 7 21
```

Ці результати можна подати у вигляді стовпчастої діаграми, із накопиченням або з угрупованням. Утворені діаграми зображені на рис. 2.5–2.6.

```
> barplot(counts,
+ main="Стовпчаста діаграма із накопиченням",
+ xlab="Лікування",ylab="Частота",
+ col=c("yellow","green","blue"),
+ legend=rownames(counts))
```

```
> barplot(counts,
+ main="Стовпчаста діаграма із групуванням",
+ xlab="Лікування",ylab="Частота",
+ col=c("yellow","green","blue"),
+ legend=rownames(counts),beside=TRUE)
```



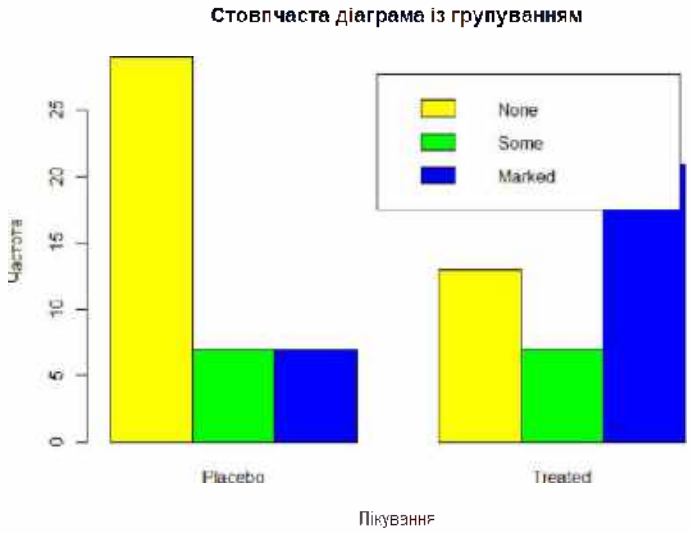


Рис. 2.5–2.6. Стовпчасті діаграми із накопиченням та групуванням у R

Перша команда дозволяє отримати складову стовпчасту діаграму із накопиченням, а друга – стовпчасту діаграму з групуванням. Також було додано опцію `col`, щоб розфарбувати стовпці в різні кольори. Параметр `legend` виводить підписи до стовпців в умовні позначення.

Розглянемо спеціалізований варіант стовпчастих діаграм, який називається **спінограмою**. В цьому випадку масштаб стовпчастої діаграми з вертикальним розташуванням стовпців змінюється так, що висота кожного результуючого стовпця стає рівною 1, а висоти його складових відображають пропорції. Спінограми створюються за допомогою команди `spine()` з пакета `vcd`. Наступний програмний код дозволяє отримати просту спінограму:

```
< library (vcd)
< attach (Arthritis)
< counts <- table (Treatment, Improved)
< spine (counts, main = "Спінограма")
< detach (Arthritis)
```

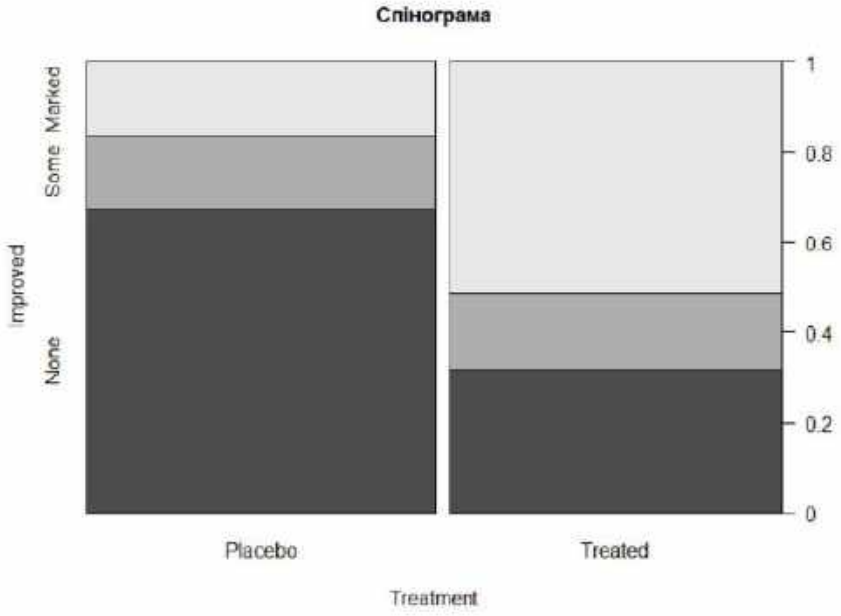


Рис. 2.7. Спінограма в середовищі R

Результат показаний на рис. 2.7. Зрозуміло, що пацієнтів, стан яких помітно покращився, більше серед тих, хто отримував справжні ліки, а не плацебо.

Незважаючи на те, що **кругові діаграми** широко використовуються в бізнес-аналітиці, вони критикуються більшістю статистиків. Вони рекомендують застосовувати стовпчасті або точкові діаграми замість кругових діаграм, оскільки користувачам легше порівнювати довжини, ніж площі й обсяги. Саме з таких міркувань можливості побудови кругових діаграм у R досить сильно обмежені порівняно з іншими статистичними програмами.

Кругові діаграми створюються за допомогою функції:

```
pie (x, labels)
```

де `x` – це позбавлений негативних значень числовий вектор, який визначає площу кожного сегмента діаграми, а `labels` – текстовий вектор, що містить підписи для сегментів.

У поданому нижче програмному коді містяться приклади, а результат представлений на рис. 2.8–2.9.

```
> par(mfrow=c(2,2))
> slices <-c(10,12,4,16,8)
```

```
    > lblis <-c("США","Англія", "Австралія",
"Німеччина","Франція")
    > pie(slices,labels=lblis,
    + main="Проста кругова діаграма")

    > pct <-round(slices/sum(slices)*100)
    > lblis2 <-paste(lblis,"",pct,"%",sep="")
    >  pie(slices,labels=lblis2,  col=rainbow
(lenght(lblis2)),
    + main="Кругова діаграма з відсотками")
```



Рис. 2.8–2.9. Кругові діаграми в середовищі R

Спочатку задається параметр діаграми так, що всі чотири елементи об'єднані на одній панелі. Для другої кругової діаграми перетворюється розмір вибірок у відсотки і додається ця інформація до підписів сегментів. Крім того, сегменти цієї діаграми розфарбовані з використанням функції `rainbow()`. У нашому випадку `rainbow(length(lblis2)` означає `rainbow(5)`, що задає п'ять кольорів для діаграми.

Гістограми графічно відображають розподіл значень безперервних змінних, розділяючи діапазон значень на задане число відрізків по осі x і відображаючи частоту значень всередині кожного відрізка на осі y. Гістограму можна створити за допомогою команди:

```
hist(x)
```

де x – це числовий вектор. Опція `freq=FALSE` дозволяє побудувати гістограму на основі щільності ймовірності, а не частот значень. Параметр `breaks` визначає число відрізків. За замовчуванням усі відрізки мають однакову довжину. У програмному коді наведені приклади створення варіантів гістограм:

```
> par(mfrow=c(2,2))
> hist(mtcars$mpg)

> hist(mtcars$mpg,
+ breaks=12,
+ col="red",
+ xlab="Витрати палива",
+ main="Кольорова гістограма")
```

Результати відображені на рис. 2.10.

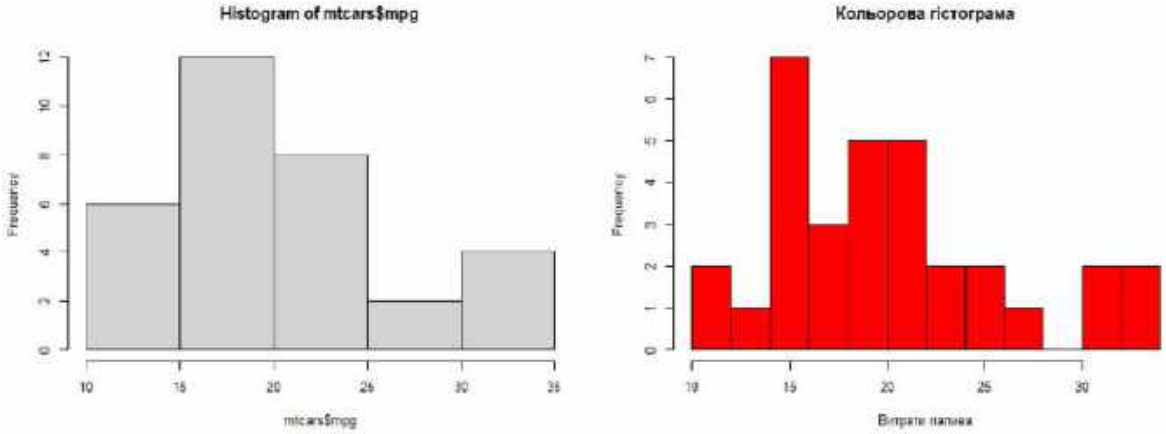


Рис. 2.10. Гістограми в середовищі R

Перша гістограма – гістограма за замовчуванням, за якої ніякі параметри функції не вказуються. У цьому випадку діапазон значень розбивається на п'ять відрізків, заголовок і підписи по осях створюються автоматично. Для другої гістограми дані більш інформативні та привабливі, а також зазначено, що діапазон значень потрібно розбити на 12 відрізків і зафарбувати стовпчики червоним кольором.

Ящикова діаграма (box plot) ілюструє розподіл значень неперервної змінної, відображаючи п'ять параметрів: мінімум, нижній квартиль (25-й перцентиль), медіану (50-й перцентиль),

верхній квартиль (75-й перцентиль) і максимум. На цій діаграмі також можуть бути відображені ймовірні «аномальні точки» (значення, що виходять за діапазон в ± 1,5 міжквартильного розмаху, різниці верхньої і нижньої квартилей). Наприклад, команда

```
boxplot(mtcars$mpg,main="Ящик з вусами",
ylab="Витрати палива")
```

дозволяє побудувати діаграму, показану на рис. 2.11.

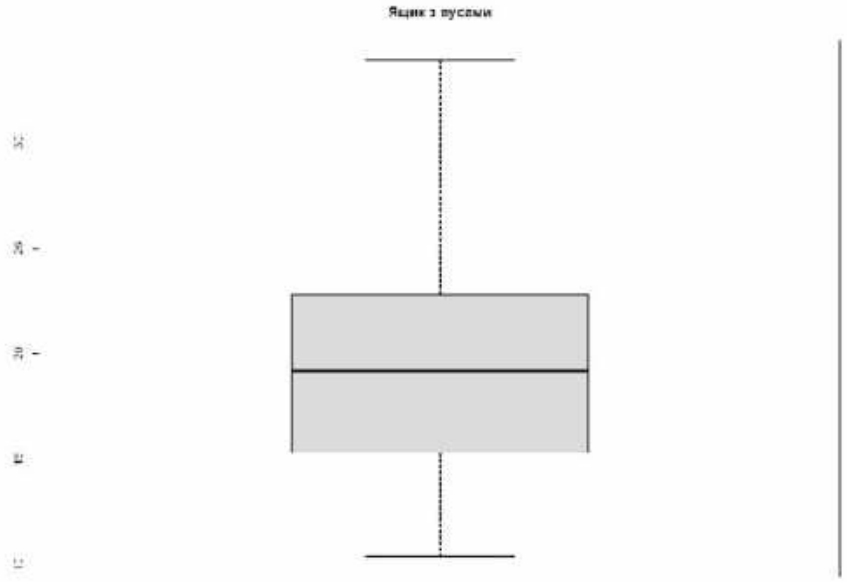


Рис. 2.11. Ящикова діаграма («ящик з вусами») в R

За замовчуванням кожен «вус» прямує до мінімального або максимального значення, яке не виходить за межі 1,5 міжквартильного розмаху. Значення, що виходять за ці межі, позначаються точками і їх не показано.

Наприклад, у нашій вибірці даних про машини медіана витрат палива становить 19,2 милі на галон, 50 % значень потрапляють у діапазон між 15,3 і 22,8, найменше значення дорівнює 10,4, а найбільше – 33,9. Ці значення можна отримати за допомогою команди `boxplot.stats(mtcars$mpg)` виводить значення статистик, які використовувалися для побудови діаграми. Тут не спостерігається «аномальних точок» і є невеликі відхилення розподілу в позитивну сторону (верхній «вус» довший нижнього).

Ящикові діаграми можна побудувати для окремих змінних або для груп змінних. Загальний вигляд команди такий:

```
boxplot (formula, data=dataframe)
```

де `formula` – це формула, а `dataframe` позначає таблицю даних (або список), де містяться дані. Прикладом формули може бути вираз `y~A`, де для кожного значення категоріальної змінної `A` буде побудована окрема діаграма розмахів для числової змінної `y`. Формула `y~A*B` дозволить отримати окремі діаграми розмахів для всіх комбінацій значень змінної `y`, заданих категоріальним змінними `A`.

Параметр `varwidth=TRUE` дозволяє отримати діаграму, на якій ширина «ящиків» буде пропорційна квадратному кореню з розміру вибірки. У цьому випадку слід використовувати параметр `horizontal=TRUE`, щоб поміняти осі місцями. У розміщеному нижче програмному коді повертаємося до порівняння витрат палива у автомобілів з чотирма, шістьма і вісьмома циліндрами, але тепер уже за допомогою розташованих поруч діаграм розмахів.

```
> boxplot(mpg~cyl,data=mtcars,
+ main="Дані про витрати палива",
+ xlab="Кількість циліндрів",
+ ylab="Витрати палива")
```

Діаграму подано на рис. 2.12.

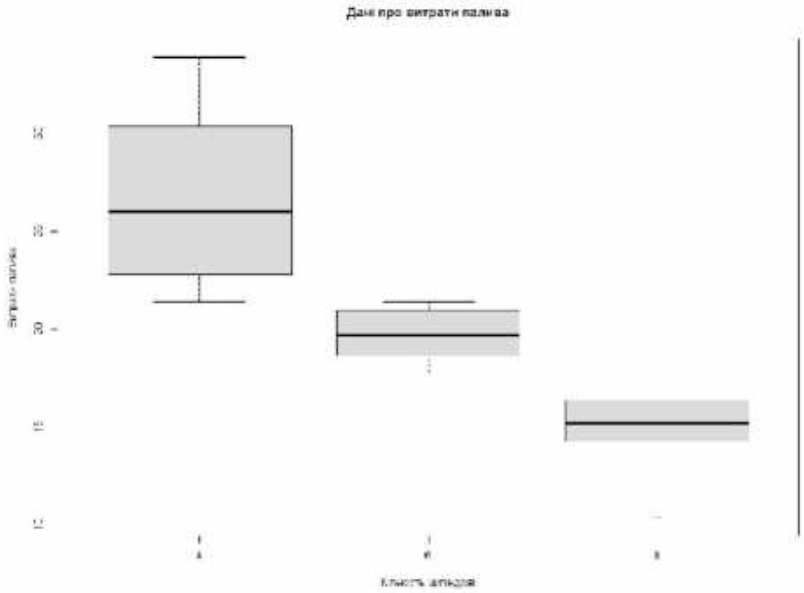


Рис. 2.12. Ящикові діаграми (діаграми з вусами) в R

На рис. 2.12 можна побачити, що автомобілі з різним числом циліндрів розрізняються за витратами палива. Також зрозуміло, що розподіл значень витрат палива у машин з шістьма циліндрами більш симетричний порівняно з чотирма і восьмициліндровими машинами. Для автомобілів з чотирма циліндрами характерний більший розкид (і позитивне зрушення розподілу) значень витрати палива. У групі з восьмициліндровими двигунами є аномальні точки.

Точкові діаграми (dot plot) дозволяють зобразити безліч названих значень на простій горизонтальній шкалі. Ця діаграма створюється за допомогою функції `dotchart()`, яка застосовується в такому форматі:

```
dotchart(x,labels=)
```

де `x` – це числовий вектор, а параметр `labels` задає вектор, у якому містяться підписи до кожної точки. Можна додати параметр `groups` для призначення фактора, що визначає групування елементів вектора `x`. У цьому випадку параметр `gcolor` визначає колір підписів для різних груп, а параметр `cex` визначає розмір підписів. Приклад для набору даних `mtcars`:

```
>
dotchart(mtcars$mpg,labels=row.names(mtcars),ce
x=8,
> main="Витрати палива в залежності від марки
автівки",
> xlab="Міль на галон")
```

Діаграма на рис. 2.13 дозволяє побачити витрати палива для кожної марки автомобілів на одній і тій же горизонтальній осі. Зазвичай точкові діаграми стають більш цікавими, коли значення відсортовані, а кожній групі відповідають свої символ і колір.

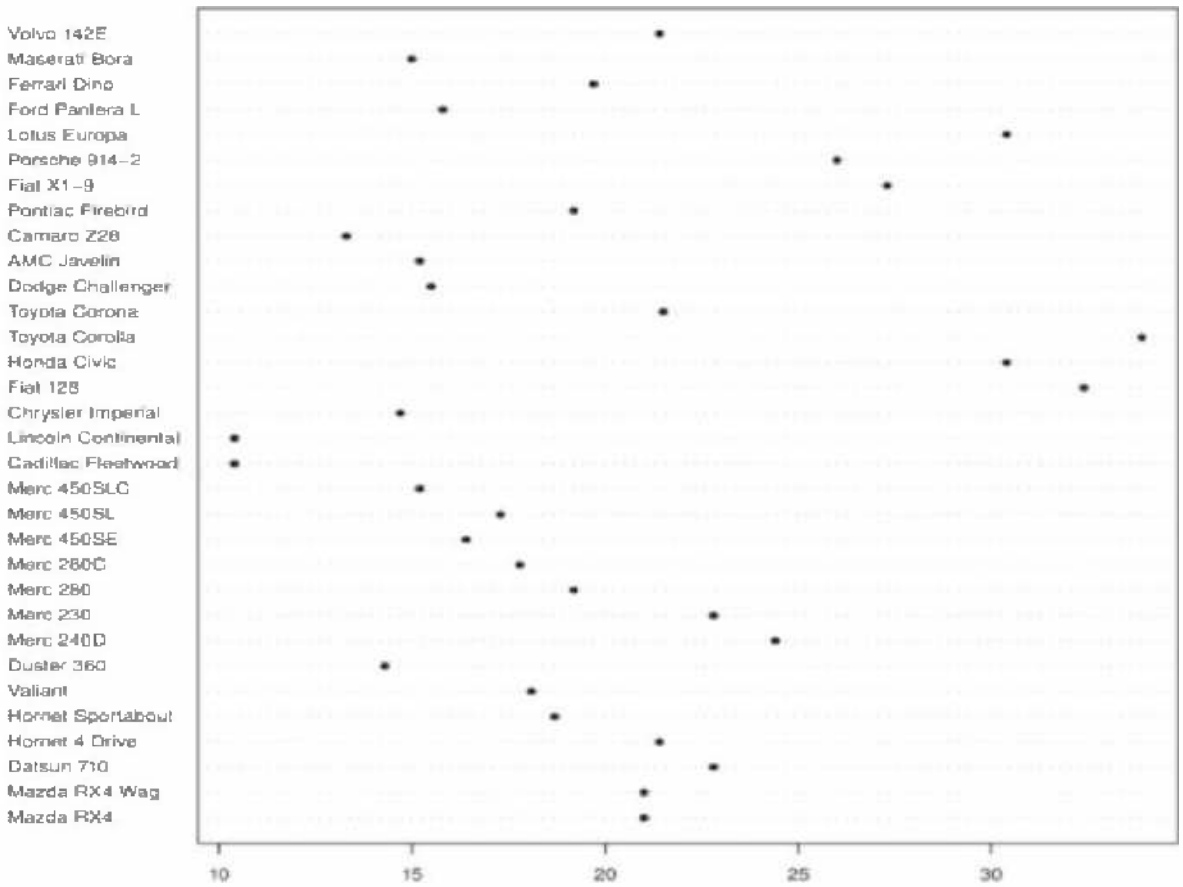


Рис. 2.13. Точкова діаграма в R

2.3. Візуалізація даних та типи складних діаграм у середовищі R

Діаграми розсіювання ілюструють взаємозв'язок між двома безперервними змінними. Розглянемо приклад зображення взаємозв'язку між однією парою змінних (x і y).

Основна функція для створення діаграм розсіювання в R – це `plot(x, y)`, де x і y – це числові вектори, що визначають x і y координати точок на діаграмі. Приклад поданий у програмному коді:

```
> attach(mtcars)
> plot(wt, mpg, main="Проста діаграма розсіювання залежності витрат палива від ваги автівки",
xlab="Вага автівки (тисячі фунтів)",
ylab="Міль на галлон", pch=19)
```

```
> abline(lm(mpg~wt), col="red", lwd=2,
lty=1)
> lines(lowess(wt,mpg), col="blue", lwd=2,
lty=2)
```

Отриману діаграму подано на рис. 2.14.

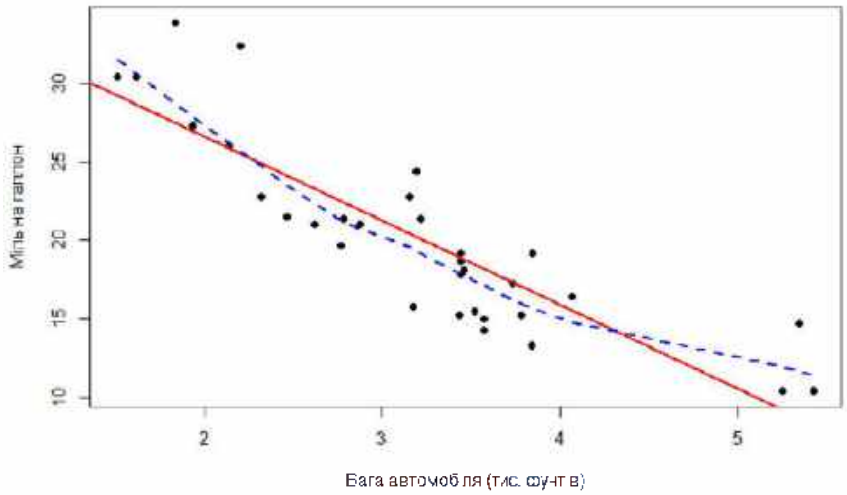


Рис. 2.14. Проста діаграма розсіювання залежності витрат палива від ваги автівки

Команди в програмному коді завантажують таблицю даних `mtcars` і створюють базову версію діаграми розсіювання, на якій спостереження позначені заповненими кружками. Як і очікувалося, витрати палива зменшуються зі збільшенням ваги машини, хоча цей зв'язок не ідеально лінійний. Функція `abline()` використана для додавання апроксимуючої регресійної лінії (linear line of best fit), а згладжуюча крива додана за допомогою функції `lowess()`. Згладжуюча крива – це непараметрична апроксимуюча крива, заснована на локально зваженій поліноміальній регресії.

У R існують як мінімум чотири корисних функції для створення **матриць діаграм розсіювання**. Базовий варіант матриць діаграми розсіювання можна отримати за допомогою функцій `pairs()`. Побудувати цю матрицю для змінних `mpg,disp,drat` і `wt` можна за допомогою такого програмного коду:

```
pairs(~ mpg+disp+drat+wt, data = mtcars,
```

```
    main = "Базовий варіант матриці діаграм
розсіювання")
```

Усі змінні, перераховані праворуч від знаку ~ потрапляють на діаграму. Вона подана на рис. 2.15.

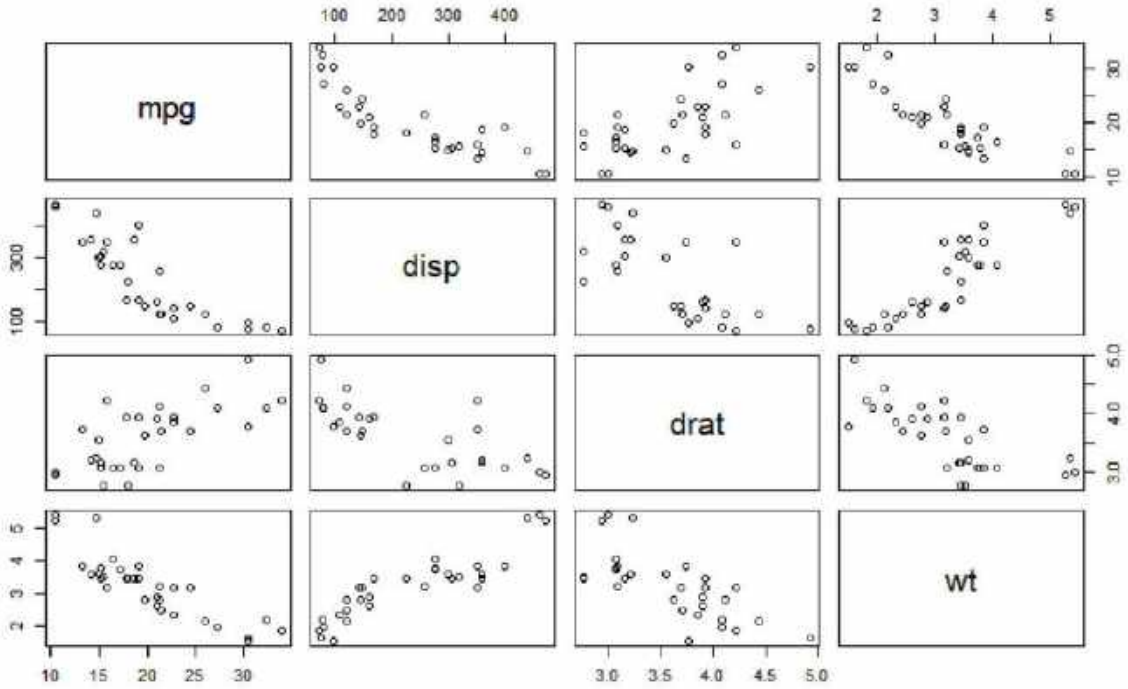


Рис. 2.15. Базовий варіант матриці діаграм розсіювання

На цій діаграмі відображені взаємозв'язки між усіма парами змінних. Наприклад, діаграма розсіювання для `mpg` та `disp` розташована на перетині рядка і стовпця з назвами цих змінних. Шість діаграм розсіювання зверху від головної діагоналі точно такі ж, як і знизу. Вони так розташовані просто для зручності. Змінюючи параметри команди, можна відображати тільки верхній або тільки нижній трикутник.

Також у R доступна функція зробити двомірну діаграму розсіювання і використовувати розмір точок для відображення значень третьої змінної. У результаті вийде так звана **бульбашкова діаграма** (bubble plot).

Цю діаграму можна створити за допомогою функції `symbols()`. Цю функцію можна використовувати для зображення кіл, квадратів, зірок, «термометрів» і діаграм

розсіювання із заданими координатами (x, y). Кола зображуються за допомогою команди:

```
symbols (x, y, circle = radius)
```

де x, y і radius – це вектори, які задають x і y-координати і діаметри кіл відповідно.

Якщо необхідно, щоб значення третьої змінної були пропорційні площі, а не радіусу кіл, то потрібно буде використовувати таку команду:

```
symbols (x, y, circle = sqrt (z/pi)),
```

де z – це третя змінна, значення якої слід відобразити.

Застосуємо цю команду до набору даних mtcars, відобразивши значення ваги автомобілів по осі x, витрат палива – по осі y, а об'єм двигуна – за допомогою радіуса кіл. Такі команди

```
> attach(mtcars)
> r<-sqrt(disp/pi)
>
symbols(wt,mpg,circle=r,inches=0.30,fg="white",
bg="lightblue",                    main="Бульбашкова
діаграма",ylab="Витрати палива (міль на
галлон)",xlab="Вага автомобіля (тис. фунтів")
> text(wt,mpg,rownames(mtcars),cex=0.6)
> detach(mtcars)
```

дозволяють отримати бульбашкову діаграму (рис. 2.16).

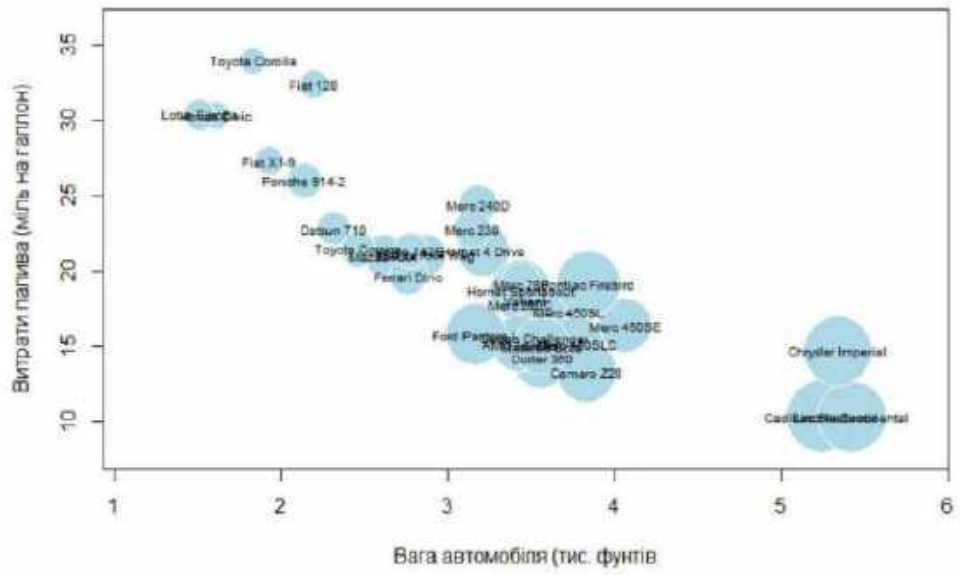


Рис. 2.16. Бульбашкова діаграма

Параметр `inches` – це масштабний коефіцієнт, який можна використовувати для визначення розміру кіл (за замовчуванням радіус найбільшого кола дорівнює одному дюйму). Функція `text()` не обов'язкова. Тут вона використовується для того, щоб підписати назви автомобілів. З діаграми видно, що збільшення витрати палива пов'язано зі зменшенням як ваги машин, так і об'єму двигуна.

У цілому фахівці, залучені до роботи над програмою R, прагнуть уникати бульбашкових діаграм з причин, через які вони не люблять кругових діаграм. Людям, як правило, важче порівнювати об'єми, ніж відстані. Однак бульбашкові діаграми вельми популярні в бізнес-аналітиці.

Матриці кореляції – це один з основних елементів багатовимірної статистики. Які змінні з розглянутих сильно корелюють одна з одною, а які – ні? Чи існують кластери змінних, які пов'язані між собою певним способом? Зі збільшенням числа змінних відповісти на такі питання стає все складніше. **Корелограми** – це порівняно новий спосіб візуалізації кореляційних матриць.

Розглянемо кореляції між змінними в наборі даних `mtcars`. У ньому є 11 змінних, кожна з яких відповідає якій-небудь характеристиці 32 автомобілів. Можна обчислити кореляційну матрицю за допомогою таких команд:

```
> options(digits=2)
> cor(mtcars)
mpg cyl disp hp drat wt qsec vs
mpg 1.00 -0.85 -0.85 -0.78 0.681 -0.87 0.419
0.66
cyl -0.85 1.00 0.90 0.83 -0.700 0.78 -0.591
-0.81
disp -0.85 0.90 1.00 0.79 -0.710 0.89 -0.434
-0.71
hp -0.78 0.83 0.79 1.00 -0.449 0.66 -0.708 -
0.72
drat 0.68 -0.70 -0.71 -0.45 1.000 -0.71 0.091
0.44
```

```
    wt -0.87 0.78 0.89 0.66 -0.712 1.00 -0.175 -
0.55
    qsec 0.42 -0.59 -0.43 -0.71 0.091 -0.17 1.000
0.74
    vs 0.66 -0.81 -0.71 -0.72 0.440 -0.55 0.745
1.00
    am 0.60 -0.52 -0.59 -0.24 0.713 -0.69 -0.230
0.17
    gear 0.48 -0.49 -0.56 -0.13 0.700 -0.58
-0.213 0.21
    carb -0.55 0.53 0.39 0.75 -0.091 0.43 -0.656
-0.57
    am gear carb
    mpg 0.600 0.48 -0.551
    cyl -0.523 -0.49 0.527
    disp -0.591 -0.56 0.395
    hp -0.243 -0.13 0.750
    drat 0.713 0.70 -0.091
    wt -0.692 -0.58 0.428
    qsec -0.230 -0.21 -0.656
    vs 0.168 0.21 -0.570
    am 1.000 0.79 0.058
    gear 0.794 1.00 0.274
    carb 0.058 0.27 1.000
```

Які зі змінних найбільше пов'язані між собою? Які змінні порівняно незалежні? Чи є тут якась структура? Не так-то легко відповісти на такі питання, просто дивлячись на цю кореляційну матрицю і не витративши багато часу і зусиль. Цю саму кореляційну матрицю можна подати графічно за допомогою функції `corrgram()` з пакету `corrgram`.

Формат застосування функції `corrgram()` такий:

```
corrgram (x, order =, panel =, text.panel =,
diag.panel =)
```

де `x` – це таблиця даних з одним спостереженням на рядок. Якщо `order = TRUE`, то порядок змінних змінюється згідно з

результатами аналізу головних компонент кореляційної матриці. Таке сортування змінних допомагає виявити закономірності в кореляційній структурі.

Параметр `panel` визначає вид діаграми (крім головної діагоналі – її властивості задаються окремо). Замість нього можна використовувати параметри `lower.panel` і `upper.panel`, щоб окремо визначати вид нижньої і верхньої (щодо головної діагоналі) половин діаграми. Параметри `text.panel` і `diag.panel` відносять до головної діагоналі.

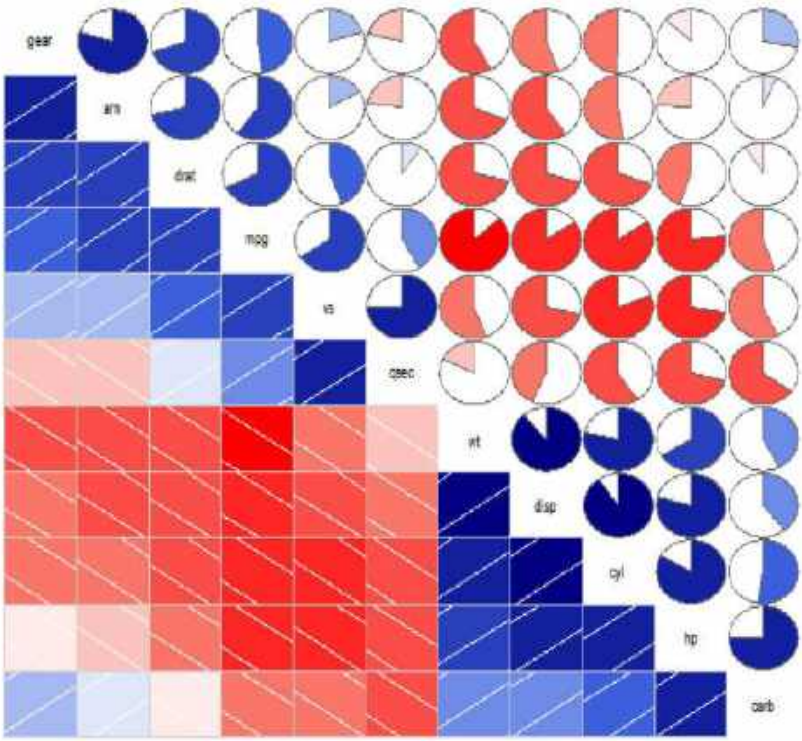


Рис. 2.17. Корелограма кореляцій між змінними

Щоб зрозуміти, що означає ця діаграма, почнемо з нижнього трикутника комірок (комірки нижче головної діагоналі). За замовчуванням блакитний колір і штрихування з лівого нижнього кута до правого верхнього свідчить про позитивну кореляцію між двома змінними, на перетині яких знаходиться ця група. Навпаки, червоний колір і штрихування з верхнього лівого кута до правого нижнього відповідають негативній кореляції. Чим темніший і насичений колір, тим сильніше кореляція. Слабкі, близькі до нуля корсляції будуть подані «вицвілими» комірками. На такій діаграмі

порядок рядків і стовпців був автоматично змінений за результатами аналізу головних компонент так, щоб змінні з подібною кореляційною структурою формували кластери.

Можна бачити, що змінні `gear`, `am`, `drat` і `mpg` позитивно корелюють один з одним. Також помітно, що змінні `wt`, `disp`, `cyl`, `hp` і `carb` теж позитивно корелюють між собою. Однак змінні з першої групи негативно корелюють зі змінними другої групи. Очевидно і те, що кореляція між змінними `carb` і `am` слабка; це ж справедливо для таких пар змінних, як `vs` і `gear`, `vs` і `am`, `drat` і `qsec`.

На верхньому трикутнику діаграми та сама інформація представлена у вигляді кругових діаграм. Тут кольори мають таке ж значення, а сила кореляції виражена в розмірі закрашеного сегмента кругової діаграми. Сегменти, відповідні позитивним кореляціям, починаються від положення «12 годин» і заповнюють коло за годинниковою стрілкою. Сегменти, відповідні негативним кореляціям, заповнюють коло проти годинникової стрілки.

Питання для самоперевірки та контролю знань

1. Які найпростіші засоби візуалізації даних та їх взаємозв'язків існують у R?

2. Що розуміють під стовпчастими діаграмами та за допомогою якої функції можна побудувати стовпчасту діаграму? Що є аргументами цієї функції?

3. У чому полягає різниця між вертикальною та горизонтальною стовпчастою діаграмою? Чим відрізняється функція їх побудови?

4. Чим відрізняються стовпчаста діаграма із накопиченням та стовпчаста діаграма з угрупованням?

5. Що розуміють під спінограмою? У чому полягає її призначення як засобу візуалізації даних?

6. Назвіть особливості побудови кругових діаграм у R. У чому полягає аналітичність кругових діаграм та за допомогою яких функцій можна ускладнювати діаграми?

7. У чому полягає призначення побудови діаграм? Що є аргументами функції побудови гістограми?

8. Що розуміють під ящиковою діаграмою? Охарактеризуйте її основні елементи та аналітичне призначення.

9. Назвіть особливості та призначення простої діаграми розсіювання залежності та матриць діаграм розсіювання.

10. У чому полягає аналітичне призначення бульбашкової діаграми та корелограми? Чим вони відрізняються?

Тестові завдання для самоконтролю та перевірки знань

1. Тип діаграм, що відображають розподіл (частоту різних значень) категоріальної змінної у вигляді вертикальних або горизонтальних стовпчиків, – це:

а) кругові діаграми;

б) спінограми;

в) стовпчасті діаграми;

г) ящикові діаграми.

2. Якщо дані задані не у форматі вектора, а у форматі матриці, то доцільно будувати:

а) стовпчасту діаграму із накопиченням або угрупованням;

б) спінограму;

в) ящикову діаграму;

г) корелограму.

3. Діаграма, при побудові якої масштаб стовпчастої діаграми з вертикальним розташуванням стовпців змінюється так, що висота кожного результуючого стовпця стає рівною 1, а висоти його складових відображають пропорції, – це:

а) ящикова діаграма;

б) корелограма;

в) спінограма;

г) бульбашкова діаграма.

4. Один із статистичних графіків у формі кола, який поділений на сегменти, що ілюструють чисельне співвідношення, – це:

а) кругові діаграми;

б) корелограми;

в) бульбашкові діаграми;

г) ящикові діаграми.

5. Діаграми, що графічно відображають розподіл значень безперервних змінних, розділяючи діапазон значень на задане число відрізків по осі x і відображаючи частоту значень всередині кожного відрізка на осі y, – це:

а) кругові діаграми;

б) гістограми;

в) корелограми;

г) бульбашкові діаграми.

6. Діаграма, що ілюструє розподіл значень неперервної змінної, відображаючи п'ять параметрів: мінімум, нижній квартиль, медіану, верхній квартиль і максимум, – це:

а) спінограма;

б) гістограма;

в) ящикова діаграма;

г) бульбашкова діаграма.

7. Діаграми, що дозволяють зобразити безліч названих значень на простій горизонтальній шкалі, – це:

а) точкові діаграми;

б) ящикові діаграми;

в) бульбашкові діаграми;

г) кругові діаграми.

8. Діаграми, що ілюструють взаємозв'язок між двома безперервними змінними, – це:

а) ящикові діаграми;

б) діаграми розсіювання;

в) ящикові діаграми;

г) корелограми.

9. Двомірна діаграма розсіювання, з якою прийнято використовувати розмір точок для відображення значень третьої змінної, – це:

а) ящикові діаграми;

б) бульбашкові діаграми;

в) корелограми;

г) діаграми розсіювання.

10. Діаграма, що використовується як засіб візуалізації зображення статистики кореляції, – це:

а) ящикова діаграма;

б) корелограма;

в) спінограма;

г) бульбашкова діаграма.

ТЕМА 3
СТАТИСТИЧНА ОБРОБКА ТА АНАЛІЗ ДАНИХ У СЕРЕДОВИЩІ R

***Мета*:** *дослідити описову статистику та таблиці частот і спряженості в системі R; здійснити кореляційний та коваріаційний аналізи в середовищі R; провести статистичну перевірку гіпотез за допомогою тестів Стьюдента в системі R.*

План

3.1. Описова статистика та таблиці частот і спряженості в системі R.

3.2. Кореляційний та коваріаційний аналізи в середовищі R.

3.3. Статистична перевірка гіпотез за допомогою тестів Стьюдента в системі R.

***Ключові поняття та терміни*:** *середнє значення, медіана, мода, стандартне відхилення, дисперсія, мінімальне та максимальне значення змінної, розмах, ексцес та коефіцієнт асиметрії, таблиці частот і таблиці спряженості, χ-квадрат, тест Фішер, тест Кохран-Мантеля-Хензеля, коефіцієнти кореляції, коефіцієнти Пірсона, Спірмена, Кенделла, тест Стьюдента, тест Манна-Уітні, тест Вілкоксона.*

***Основні функції та програмні коди*:** `summary()`, `apply()`, `sapply()`, `fivenum()`,`aggregate()`, `by()`, `table ()`, `xtabs()`, `prop.table()`, `margin.table()`, `addmargins()`, `ftable()`, `table()`, `chisq.test()`, `fisher.test()`, `mantelhaen.test ()`, `cor()`, `cov()`, `wilcox.test()`.

3.1. Описова статистика та таблиці частот і спряженості в системі R

Розглянемо числові характеристики центральної тенденції, розкиду і тип розподілу значень безперервних змінних. Здійснимо це на прикладі кількох змінних з набору даних про характеристики різних марок автомобілів (Motor Trend Car Road Test, mtcars). Як показники візьмемо витрати палива (у милях на галон – miles per gallon, `mpg`), потужності (кінських сил – horsepower, `hp`) і вазі (weight, `wt`).

```
> vars <- c("mpg", "hp", "wt")
> head(mtcars[vars])
mpg hp wt
Mazda RX4 21.0 110 2.620
Mazda RX4 Wag 21.0 110 2.875
Datsun 710 22.8 93 2.320
Hornet 4 Drive 21.4 110 3.215
Hornet Sportabout 18.7 175 3.440
Valiant 18.1 105 3.460
```

Для початку розглянемо **описові статистики** для всіх 32 марок автомобілів у цілому. Потім обчислимо описові статистики окремо для кожного типу коробки передач (`am`) і числа циліндрів (`cyl`). Тип коробки передач – це дихотомічна змінна, яка приймає значення 0 (автоматична коробка) і 1 (механічна коробка), число циліндрів може бути рівним 4, 5 і 6.

У базовій версії реалізована функція `summary()`, яка дозволяє обчислити описові статистики. Приклад міститься у програмному коді.

```
> summary(mtcars[vars])
mpg hp wt
Min. :10.40 Min. : 52.0 Min. :1.513
1st Qu.:15.43 1st Qu.: 96.5 1st Qu.:2.581
Median :19.20 Median :123.0 Median :3.325
Mean :20.09 Mean :146.7 Mean :3.217
3rd Qu.:22.80 3rd Qu.:180.0 3rd Qu.:3.610
Max. :33.90 Max. :335.0 Max. :5.424
```

Функція `summary()` дозволяє обчислити мінімум, максимум, квартилі і середнє для числових змінних або частоти значень для факторів і логічних векторів. Можна також використовувати функції `apply()` або `sapply()`, щоб обчислити будь-яку описову статистику. Формат застосування функції `sapply()` такий:

```
sapply (x, F, options)
```

де `x` – це таблиця даних (або матриця), а `F` – довільна функція.

Якщо опції (options) вказані, вони належать до функції F. Зазвичай у цьому випадку використовуються такі функції, як `mean`, `sd`, `var`, `min`, `max`, `median`, `length`, `range` і `quantile`. Функція `fivenum()` обчислює п'ять описових статистик Тьюкі (Tukey) (мінімум, нижній квартиль, медіана, верхній квартиль, максимум).

Зазначимо, але в базовій версії програми немає функцій для характеристики асиметрії і ексцесу розподілу значень, але такі функції можна створити самостійно. Приклад в наступному програмному коді демонструє, як обчислити кілька описових статистик, включаючи асиметрію і ексцес.

```
> mystats <-function(x, na.omit=FALSE){
 if(na.omit)
 x <- x[!is.na(x)]
 m <- mean(x)
 n <- lenght(x)
 s <- sd (x)
 skew <- sum((x-m)^3/s^3)/n
 kurt <- sum((x-m)^4/s^4)/n-3

return(c(n=n,mean=m,stdev=s,skew=skew,kurtosis=
kurt))
 }
> sapply(mtcars[vars],mystats)
 mpg hp wt
n 32.001 32.000 32.0000
mean 20.091 146.688 3.2172
```

```
stdev 6.027 68.563 0.9785
skew 0.611 0.726 0.4231
kurtosis -0.373 -0.136 -0.0227
```

Для автомобілів середні витрати палива становлять 20,1 миль на галон зі стандартним відхиленням 6.0. Максимум кривої розподілу значень зміщений вправо (+ 0,61) і знаходиться трохи нижче, ніж це буває в разі нормального розподілу (–0,37). Це буде найкраще помітно при графічному зображенні даних. Врахуйте, що якщо необхідно б було видалити рядки з пропущеними значеннями, то потрібно написати такий програмний код:

```
sapply (mtcars[vars],mystats,na.omit=TRUE).
```

При порівнянні груп об'єктів або спостережень зазвичай звертають увагу на значення описових статистик для кожної групи, а не для всієї вибірки в цілому. Знову ж таки, в R реалізовано кілька способів досягнення мети. Обчислимо описові статистики для кожного типу коробки передач. Під час обчислення описових статистик для окремих груп даних можна використовувати функцію `aggregate()`, як це показано в програмному коді:

```
> aggregate(mtcars[vars], by=list(am=
mtcars$am), mean)
am mpg hp wt
1  0 17.1 160 3.77
2  1 24.4 127 2.41
> aggregate(mtcars[vars],
by=list(am=mtcars$am), sd)
am mpg hp wt
1 0 3.83 53.9 0.777
```

Потрібно звернути увагу на використання виразу `list(am=mtcars$am)`. Якщо записати `list(mtcars$am)`, то стовпчик `am` був би названий не `am`, а `Group. 1`. У цьому випадку було використано знак присвоєння, щоб підписати стовпець більш змістовно. Якщо маємо більше однієї змінної, можна використовувати код на зразок цього:

by=list(name1=groupvar1, name2 = groupvar2, ..., groupvarN).

Функцію aggregate() можна використовувати для обчислення тільки однієї описової статистики за один раз, такої як середнє, стандартне відхилення і подібних до них. Для одночасного обчислення декількох статистик потрібно використовувати функцію by(). Формат її застосування такий:

```
by (data, I, F)
```

де data – це таблиця даних або матриця, I – фактор або список чинників, який визначає групи, і F – довільна функція. Наведений нижче програмний код містить приклад.

```
> dstats <- function(x)(c(mean=mean(x), sd=sd(x)))
> by(mtcars[vars], mtcars$am, dstats)
mtcars$am: 0
mean.mpg mean.hp mean.wt sd.mpg sd.hp sd.wt
 17.147  160.263    3.769  3.834 53.908 0.777
-------------------------------------------
mtcars$am: 1
mean.mpg mean.hp mean.wt sd.mpg sd.hp sd.wt
24.392  126.846    2.411  6.167 84.062 0.617
```

Далі розглянемо **таблиці частот і таблиці спряженості** для категоріальних змінних, а також ознайомимся з тестами на незалежність, заходами зв'язку. Будемо використовувати команди, реалізовані в базовій версії R. У поданих нижче прикладах A, B і C позначають категоріальні змінні.

Для прикладу використаємо набір даних Arthritis з пакета vcd. Ці дані опубліковані Kock & Edward і являють собою результати клінічного дослідження нового способу лікування ревматоїдного артриту подвійним сліпим методом. Ось перші кілька спостережень:

```
> library(vcd)
> head(Arthritis)
ID Treatment Sex Age Improved
1 57 Treated Male 27 Some
```

```
2 46 Treated Male 29 None
3 77 Treated Male 30 None
4 17 Treated Male 32 Marked
5 36 Treated Male 46 Marked
6 23 Treated Male 58 Marked
```

Спосіб лікування (`Treatment`: плацебо – `Placebo` і ліки – `Treated`), стать (`sex`: чоловіча – `Male` і жіноча – `Female`) і ступінь поліпшення стану хворих (`Improved`: `None` – немає, `Some` – деякий, `Marked` – помітний) – це категоріальні чинники. Створимо таблиці частот і спряженості для цих даних.

У R реалізовано кілька методів створення таблиць частот і спряженості. Найважливіші функції такі:

`table (var1, var2, ..., varN)` – створює N-мірну таблицю спряженості для N категоріальних змінних (факторів);

`xtabs(formula, data)` – створює N-мірну таблицю спряженості на основі формули і матриці або таблиці даних;

`prop.table(table, margins)` – являє значення таблиці у вигляді часток від значень крайнього поля таблиці, що задається параметром `margins`;

`margin.table(table, margins)` – підсумовує значення таблиці по рядках або стовпцях (визначається параметром `margins`);

`addmargins(table, margins)` – обчислює описові статистики margins (суми за замовчуванням) для таблиці;

`ftable(table)` – створює компактну «плоску» таблицю спряженості.

Частоти значень однієї змінної можна розрахувати за допомогою функції `table()`. Наприклад:

```
> mytable <- with(Arthritis, table (Improved))
> mytable
Improved
None    Some     Marked
42   14  28
```

Частоти значень однієї змінної можна розрахувати за допомогою функції `table()`:

```
> prop.table(mytable)
Improved
None    Some     Marked
0.500    0.167    0.333
```

Або у відсотки, використовуючи `prop.table()*100`:

```
> prop.table(mytable)*100
Improved
None    Some     Marked
50.0    16.7     33.3
```

Звідси видно, що у половини пацієнтів після лікування відбулося певне поліпшення стану здоров'я (16,7 + 33,3).

У разі побудови таблиць для двох змінних використовують функцію `table ()`:

```
mytable <- table (A, B),
```

де A – це змінна, значенням якої відповідають рядки таблиці спряженості, а B – стовпці. Як альтернатива можна скористатися функцією `xtabs()`, яка дозволяє при створенні таблиці спряженості вводити дані у вигляді формули. Ось формат її застосування:

```
mytable <- xtabs (~ A + B, data = mydata)
```

де mydata – це матриці або таблиця даних. У загальному випадку змінні, для яких будується таблиця спряженості, знаходяться у правій частині формули (тобто праворуч від знака ~) і розділяються знаками +. Якщо змінна знаходиться у лівій частині формули, передбачається, що це вектор з частотами значень.

Для набору даних Arthritis вийде така таблиця:

```
> Mytable <- xtabs (~ Treatment + Improved,
data = Arthritis)
> mytable
 Improved
Treatment  None    Some    Marked
Placebo    29      7       7
Treated    13      7       21
```

За допомогою команд `margin.table()` і `prop.table()` можна розрахувати відповідно частоти і частки від цілого за стовпцями або рядками. При підсумовуванні й обчисленні частки в рядках отримаємо:

```
> margin.table (mytable, 1)
Treatment
Placebo Treated
43         41
> prop.table(mytable, 1)
Improved
Treatment  None    Some    Marked
Placebo    0.674   0.163   0.163
Treated    0.317   0.171   0.512
```

Індекс 1 належить до першої змінної у форматі застосування функції `table()`. Розглядаючи отриману таблицю, можна помітити, що у 51 % пацієнтів, які отримували справжні ліки, настало помітне поліпшення. Для порівняння: такий ефект наставив тільки у 16 % приймаючих плацебо пацієнтів.

Обчислення сум і часток від цілого для стовпців виглядає так:

```
> margin.table(mytable, 2)
Improved
None    Some    Marked
42       14       28
> prop.table(mytable, 2)
Improved
Treatment  None    Some    Marked
Placebo    0.690   0.500   0.250
Treated    0.310   0.500   0.750
```

У цьому випадку індекс 2 належить до другої змінної у форматі застосування функції `table()`.

Частки від цілого для елементів таблиці розраховуються так:

```
> prop.table(mytable)
Improved
Treatment  None    Some    Marked
Placebo    0.3452 0.0833  0.0833
Treated    0.1548 0.0833  0.2500
```

Можна використовувати функцію addmargins() для додавання сум по стовпцях або рядках:

```
addmargins(mytable)
Improved
Treatment  None    Some    Marked      Sum
Placebo     29       7         7       43
Treated     13       7        21       41
Sum         42      14        28       84
> addmargins(prop.table(mytable))
Improved
Treatment  None    Some    Marked      Sum
Placebo   0.3452   0.0833  0.0833      0.5119
Treated   0.1548   0.0833  0.2500      0.4881
Sum       0.5000   0.1667 0.3333       1.0000
```

Під час використання функції addmargins() за замовчуванням обчислюються суми і для стовпців, і для рядків таблиці. На противагу цьому команда

```
> addmargins(prop.table(mytable, 1), 2)
Improved
Treatment  None    Some    Marked      Sum
Placebo    0.674   0.163   0.163       1.000
Treated    0.317   0.171   0.512       1.000
```

дозволяє обчислити суму тільки для рядків. Аналогічно команда

```
addmargins(prop.table(mytable, 2), 1)
Improved
Treatment  None    Some    Marked
Placebo    0.690   0.500   0.250
Treated    0.310   0.500   0.750
Sum        1.000   1.000   1.000
```

обчислює суми тільки для стовпців. З цієї таблиці видно, що 25 % пацієнтів, чий стан помітно покращився після лікування, отримували плацебо.

Зокрема в R реалізовано кілька способів перевірки незалежності категоріальних даних. Далі розглянемо три тести: χ-квадрат, Фішера і Кохран-Мантеля-Хензеля.

Тест χ-квадрат. Застосувати функцію `chisq.test()` можна стосовно двомірної таблиці, щоб за допомогою тесту **χ-квадрат** (chi-square test) перевірити незалежність змінних, які складають рядки і стовпці. Приклад наведено в програмному коді:

```
> library(vcd)
> mytable <- xtabs(~Treatment+Improved, data=Arthritis)
> chisq.test(mytable)

Pearson's Chi-squared test data: mytable

data: mytable
X-squared = 13.1, df = 2, p-value = 0.001463
> mytable <- xtabs(~Improved+Sex, data=Arthritis)
> chisq.test(mytable)

Pearson's Chi-squared test

data: mytable
X-squared = 4.84, df = 2, p-value = 0.0889

Warning message:
In chisq.test(mytable) : Chi-squared approximation may be incorrect.
```

Результати тесту свідчать про те, що існує зв'язок між типом лікування і ступенем поліпшення стану пацієнтів ($p < 0,01$). А ось між статтю пацієнтів і ступенем поліпшення їх стану зв'язку немає ($p > 0.05$). Значення статистичної помилки першого роду (p-value) – це ймовірність того, що в генеральній сукупності залежність між даними змінними відсутня. Оскільки ця ймовірність досить мала, відхиляємо гіпотезу про незалежність результату лікування від його типу. Ймовірність у другому випадку не така вже й мала, є підстави припускати, що спосіб лікування і стать пацієнтів не залежать одне від одного. Попередження (warning message) в програмному коді з'являється тому, що в одному з шести

елементів таблиці спряженості (деяке поліпшення стану у чоловіків) очікуване значення менше п'яти, що може призвести до недійсності апроксимації χ-квадрат.

Точний тест Фішера (Fisher's exact test) можна провести за допомогою команди `fisher.test()`. Цей тест перевіряє нульову гіпотезу про незалежність стовпців і рядків у таблиці спряженості. Формат застосування цієї функції такий:

```
fisher.test (mytable)
```

де `mytable` – це двомірна таблиця. Наприклад:

```
> mytable <- xtabs (~ Treatment + Improved,
data = Arthritis)
> fisher.test (mytable)
data: mytable
p-value = 0.001393
alternative hypothesis: two.sided
```

На відміну від багатьох статистичних програм, у R функцію `fisher.test()` можна застосовувати до будь-якої таблиці спряженості (з двома і більше стовпцями і рядками), а не тільки до таблиці розмірності 2x2.

Тест Кохрана-Мантеля-Хензель. Функція `mantelhaen.test()` дозволяє провести χ-квадрат тест Кохрана-Мантеля-Хензель (Cochran-Mantel-Haenszel) щодо нульової гіпотези про те, що дві номінальні змінні умовно незалежні при кожному значенні третьої змінної. Наведений нижче програмний код дає можливість перевірити гіпотезу про те, що змінні Treatment і Improved незалежні при кожному значенні змінної Sex. При проведенні тесту передбачається, що тристороння взаємодія (Treatment x Improved x Sex) відсутня.

```
> Mytable <- xtabs (~ Treatment + Improved +
Sex, data = Arthritis)
> Mantelhaen.test (mytable
data: mytable
Cochran-Mantel-Haenszel M^2 = 14.6, df = 2,
p-value = 0.0006647
```

Отриманий результат дозволяє вважати, що незалежно від статі пацієнтів між способом лікування і його результатом є виражений зв'язок.

3.2. Кореляційний та коваріаційний аналізи в середовищі R

Коефіцієнти кореляції використовуються для опису зв'язків між кількісними змінними. Знак коефіцієнта (+ або -) свідчить про напрям зв'язку (позитивний чи негативний), а величина коефіцієнта показує силу зв'язку (варіює від 0 – немає зв'язку до 1 – абсолютно передбачуваний взаємозв'язок).

Розглянемо різні коефіцієнти кореляції поряд з тестами на достовірність. У подальшому будемо використовувати набір даних state.x77, що входить у базову версію R. У ньому містяться дані про чисельність, доходи, відсоток неписьменного населення, середню тривалістю життя, рівень злочинності і частки людей із середньою освітою для 50 штатів США в 1977 році. Там також є дані про температурний режим і про площі штатів. Потрібно набрати `help(state.x77)`, щоб дізнатися більше про ці дані. На додаток до базової версії програми нам знадобляться пакети `psych` і `ggm`.

У R можна розраховувати різні коефіцієнти кореляції, включаючи коефіцієнти Пірсона, Спірмена, Кенделла, часткові тощо. Розглянемо їх.

Лінійний коефіцієнт кореляції Пірсона (Pearson product moment correlation) відображає ступінь лінійного зв'язку між двома кількісними змінними. Коефіцієнт рангової кореляції Спірмана (Spearman's Rank Order correlation) – міра зв'язку між двома ранжованими змінними. Тау Кенделла (Kendall's Tau) – також непараметричний показник рангової кореляції.

Функція `cor()` дозволяє обчислити всі три коефіцієнти, а функція `cov()` розраховує коваріації. У цих функцій є багато опцій, але спрощений формат обчислення кореляцій такий:

```
cor (x, use =, method =)
```

Опції можна охарактеризувати так:

– X – матриця або таблиця даних;

– use= – спрощує роботу з пропущеними даними. Може приймати такі значення: all.obs (передбачається, що пропущені значення відсутні; їх наявність викличе повідомлення про помилку), everything (будь-яка кореляція, що включає рядок з пропущеним значенням, що не буде обчислюватися – позначається як missing), complete.obs (враховуються тільки рядки без пропущених значень) і pairwise.complete.obs (враховуються усі повні спостереження для кожної пари змінних окремо);

– method= – визначає тип коефіцієнта кореляції. Можливі значення – pearson, spearman або kendall.

Значення опцій за замовчуванням: use="everything" і method="person". Приклад поданий у наведеному нижче програмному коді.

```
> states <- state.x77[,1:6]
> cov(states)
 Population Income Illiteracy
Population     19931683.7588     571229.7796
292.8679592
Income 571229.7796 377573.3061 -163.7020408
Illiteracy 292.8680 -163.7020 0.3715306
Life Exp -407.8425 280.6632 -0.4815122
Murder 5663.5237 -521.8943 1.5817755
HS Grad -3551.5096 3076.7690 -3.2354694
Life Exp Murder HS Grad
Population   -407.8424612   5663.523714   -
3551.509551
Income 280.6631837 -521.894286 3076.768980
Illiteracy -0.4815122 1.581776 -3.235469
Life Exp 1.8020204 -3.869480 6.312685
Murder -3.8694804 13.627465 -14.549616
HS Grad 6.3126849 -14.549616 65.237894
```

Перша команда виводить на екран таблицю дисперсій і коваріацій. Друга – таблицю корсляційних косфіцієнтів Пірсона,

третя – кореляційних коефіцієнтів Спірмена. Можна побачити, наприклад, що між середнім доходом і часткою людей з середньою освітою існує сильна позитивна кореляція, а між середньою тривалістю життя і часткою неписьменного населення – сильна негативна кореляція

Потрібно звернути увагу на те, що за замовчуванням виходить квадратна таблиця (наведені поєднання всіх змінних з усіма). Також можна вивести на екран прямокутну таблицю, як у наведеному прикладі:

```
> x <- states[,c("Population", "Income", "Illiteracy", "HS Grad")]
> y <- states[,c("Life Exp","Murder")]
> cor(x,y)
            Life Exp Murder
Population -0.06805195 0.3436428
Income 0.34025534 -0.2300776
Illiteracy -0.58847793 0.7029752
HS Grad 0.58221620 -0.4879710
```

Цей спосіб застосування функції особливо зручний, коли необхідно визначити зв'язок між двома наборами змінних. Потрібно врахувати, що ці результати не дозволяють дізнатися, чи відрізняються достовірно коефіцієнти кореляції від нуля (іншими словами, чи можна на підставі наявної вибірки впевнено стверджувати, що коефіцієнт кореляції для генеральної сукупності відрізняється від нуля). Для цього потрібно використовувати тести на достовірність.

Як перевірити статистичну значущість обчислених коефіцієнтів кореляції? Стандартна нульова гіпотеза – це відсутність зв'язку (тобто коефіцієнт кореляції для генеральної сукупності дорівнює нулю). Для перевірки значимості окремих кореляційних коефіцієнтів Пірсона, Спірмена і Кенделла можна використовувати функцію `cor.test()`. Спрощений формат її використання такий:

```
cor.test (x, y, alternative =, method =)
```

де `x` і `y` – це змінні, кореляція між якими досліджується, опція `alternative` визначає тип тесту (`"two.side"`, `"less"` або

"greater"), опція method задає тип кореляції ("pearson", "kendall" або "spearman"). Слід використовувати опцію alternative="less" для перевірки гіпотези про те, що в генеральній сукупності коефіцієнт кореляції менше нуля, а опцію alternative="greater" – для перевірки того, що він більше нуля. За замовчуванням alternative = "two.side" (перевіряється гіпотеза про те, що коефіцієнт кореляції в генеральній сукупності не дорівнює нулю). Приклад наведено в наступному програмному коді:

```
> cor.test(states[,3],states[,5])

    Pearson's product-moment correlation

    data: states[, 3] and states[, 5]
    t = 6.8479, df = 48, p-value = 1.258e-08
    alternative hypothesis: true correlation is
not equal to 0
    95 percent confidence interval:
    0.5279280 0.8207295
    sample estimates:
    cor
    0.7029752
```

Тут нульова гіпотеза полягає у тому, що коефіцієнт кореляції Пірсона між середньою тривалістю життя і рівнем злочинності дорівнює нулю. Якщо цей коефіцієнт для генеральної сукупності дорівнює нулю, то його значення для випадкової вибірки дорівнюватиме 0,703 рідше, ніж в одному випадку з 10 мільйонів (це і означає p-value = 1,258e-08). З огляду на те, наскільки мала ймовірність, відхиляємо нульову гіпотезу і приймемо альтернативну – про те, що значення цього коефіцієнта для генеральної сукупності не дорівнює нулю.

На жаль, за допомогою функції cor.test() одночасно можна перевірити значущість тільки одного коефіцієнта кореляції. Зате в пакеті psych є функція corr.test(), яка дозволяє зробити більше. З її допомогою можна обчислити коефіцієнти

кореляції Пірсона, Спірмена і Кенделла між декількома змінними і оцінити їх достовірність. Приклад наведено в наступному програмному коді:

```
> library(psych)
> corr.test(states,use="complete")
Call:corr.test(x = states, use = "complete")
Correlation matrix
Population Income Illiteracy Life Exp
Population 1.00 0.21 0.11 -0.07
Income 0.21 1.00 -0.44 0.34
Illiteracy 0.11 -0.44 1.00 -0.59
Life Exp -0.07 0.34 -0.59 1.00
Murder 0.34 -0.23 0.70 -0.78
HS Grad -0.10 0.62 -0.66 0.58
Murder HS Grad
Population 0.34 -0.10
Income -0.23 0.62
Illiteracy 0.70 -0.66
Life Exp -0.78 0.58
Murder 1.00 -0.49
HS Grad -0.49 1.00
Sample Size
[1] 50
Probability values
Population Income Illiteracy Life Exp
Population 0.00 0.59 1.00 1.0
Income 0.15 0.00 0.01 0.1
Illiteracy 0.46 0.00 0.00 0.0
Life Exp 0.64 0.02 0.00 0.0
Murder 0.01 0.11 0.00 0.0
HS Grad 0.50 0.00 0.00 0.0
Murder HS Grad
Population 0.10 1
Income 0.54 0
Illiteracy 0.00 0
Life Exp 0.00 0
```

```
Murder 0.00 0
HS Grad 0.00 0
```

Опція `use=` може приймати значення `"pairwise"` або `"complete"` (для попарного або порядкового видалення пропущених значень відповідно). Значення опції `method=` бувають такими: `"pearson"` (за замовчуванням), `"spearman"` або `"kendall"`. З наведеного прикладу видно, що коефіцієнт кореляції між чисельністю населення і часткою людей із середньою освітою (-0,10) не відрізняється від нуля (p = 0,5)

3.3. Статистична перевірка гіпотез за допомогою тестів Стьюдента в системі R

Якщо результат виражений у вигляді категоріальної змінної, можна використовувати методи, які були описані раніше. У цьому питанні зосередимося на порівнянні груп, де досліджується безперервна змінна, значення якої розподілені нормально. Як приклад використаємо набір даних `uscrime`, що входить у пакет `MASS`. Ці дані містять інформацію про вплив карального законодавства на рівень злочинності в 47 штатах США в 1960 році. Будемо досліджувати змінні `Prob` (ймовірність потрапити до в'язниці), `U1` (рівень безробіття для міських жителів чоловічої статі у віці від 14 до 24 років) і `U2` (цей же показник для чоловіків у віці 35–39 років). Категоріальна змінна `So` (яка вказує, чи відноситься штат до групи південних штатів) буде використана як групувальна. Дані були масштабовані авторами дослідження.

Чи правда, що ймовірність опинитися у в'язниці після скоєння злочину вище в південних штатах? Щоб відповісти на це питання, потрібно порівняти ймовірність позбавлення волі в південних штатах та інших. Для перевірки гіпотези про рівність двох середніх значень можна використовувати тест Стьюдента для незалежних вибірок. У цьому випадку мається на увазі, що ці дві групи не залежать одна від одної і дані походять з нормальних розподілів. Функцію можна використовувати у двох форматах. Або в такому:

```
t.test (y ~ x, data)
```

де y – це числова змінна, а x – дихотомічна змінна, або в такому:

```
t.test (y1, y2)
```

де y1 і y2 – це числові вектори (аналізовані значення для кожної з груп). Необов'язковий аргумент data призначає матрицю або таблицю даних, в якій містяться дані. На відміну від більшості статистичних програм, в R за замовчуванням не мається на увазі рівність дисперсій і використовується трансформація ступенів свободи Уелча (Welch degrees of freedom modification). Вказати на рівність дисперсій можна за допомогою параметра var.equal=TRUE. За замовчуванням використовується двостороння альтернативна гіпотеза (про те, що середні значення розрізняються, але неважливо, в який бік). Можна використовувати опції alternative="less" або alternative="greater", щоб перевірити наявність відмінностей у певному напрямі. За допомогою наведеного програмного коду порівнюється ймовірність потрапити у в'язницю в південних (група 1) та інших (група 0) штатах за допомогою двостороннього тесту, не припускаючи рівності дисперсій.

```
> library(MASS)
> t.test(Prob~So,data=UScrime)

    Welch Two Sample t-test

    data: Prob by So
    t = -3.8954, df = 24.925, p-value =
    0.0006506
    alternative hypothesis: true difference in
means is not equal to 0
    95 percent confidence interval:
    -0.03852569 -0.01187439
    sample estimates:
    mean in group 0 mean in group 1
    0.03851265 0.06371269
```

На підставі цього тесту можна відкинути гіпотезу про рівність шансів потрапити до в'язниці в південних та інших штатах ($p < 0{,}001$).

Якщо дві групи незалежні, можна використовувати тест рангових сум Вілкоксона (Wilcoxon rank sum test), більш широко відомий як тест Манна-Уітні (Mann-Whitney U test), щоб дізнатися, чи відбуваються спостереження з одного і того ж розподілу ймовірностей (іншими словами, чи дійсно ймовірність отримати більш високі значення вище в одній вибірці, ніж в іншій). Цю функцію можна застосовувати в такому форматі:

```
wilcox.test(y ~ x, data)
```

де `y` – це числова змінна, а `x` – дихотомічна, або в такому:

```
wilcox.test (y1, y2)
```

де `y1` і `y2` – це досліджувані змінні для кожної групи. Необов'язковий аргумент `data` призначає матрицю або таблицю даних, в якій містяться досліджувані змінні. За замовчуванням це двосторонній тест. Можна додати параметр `exact`, щоб обчислити точний критерій (`exact test`), і параметр `alternative="less"` або `alternative="greater"`, щоб перевірити відповідну односторонню гіпотезу.

Якщо використовувати тест Вілкоксона для вирішення питання про ймовірність потрапляння до в'язниці, отримаємо такий результат:

```
> with(UScrime,by(Prob,So,median))
So: 0
[1] 0.038201
------------------------------------
So: 1
[1] 0.055552
> wilcox.test(Prob~So,data=UScrime)

Wilcoxon rank sum exact test

data: Prob by So
W = 81, p-value = 8.488e-05
```

```
    alternative hypothesis: true location shift
is not equal to 0
```

Знову можна відхилити гіпотезу про рівність можливостей опинитися у в'язниці в південних порівняно зі штатами ($p < 0,01$).

Тест Вілкоксона – це непараметрична альтернатива тесту Стьюдента для залежних вибірок. Його потрібно застосовувати в тих випадках, коли групи залежні одна від одної і не спостерігається нормальний розподіл значень. Формат застосування функції такий самий, як і в попередньому прикладі, тільки потрібно додати параметр `paired=TRUE`. Застосуємо цей тест для вирішення питання про безробіття:

```
    > sapply(UScrime[c("U1","U2")],median)
    U1 U2
    92 34
    >
with(UScrime,wilcox.test(U1,U2,paired=TRUE))

    Wilcoxon signed rank test with continuity
correction

    data: U1 and U2
    V = 1128, p-value = 2.464e-09
    alternative hypothesis: true location shift
is not equal to 0
```

Знову приходимо до того ж висновку, що і в разі тесту Стьюдента для залежних змінних.

У цьому випадку параметричні тести Стьюдента і їх непараметричні аналоги дали однакові результати. Коли умови застосування тестів Стьюдента виконуються, параметричні тести мають велику потужність (ймовірність виявити існуючі відмінності вище) порівняно з непараметричними. Непараметричні тести доцільно застосовувати, коли умови застосування параметричних тестів істотно порушуються (наприклад, для порядкових змінних).

Питання для самоперевірки та контролю знань

1. Які показники можливо обчислити при здійсненні описової статистики в середовищі R?

2. Чи можливо обчислити значення асиметрії та ексцесу в середовищі? Яким чином?

3. Які показники описової статистики можливо обчислити за допомогою моделі описової статистики Тьюкі?

4. Для яких типів змінних доцільно будувати таблиці частот і таблиці спряженості?

5. За допомогою яких тестів можливо перевірити незалежність категоріальних даних?

6. У чому полягає різниця між тестами χ-квадрат, Фішера і Кохран-Мантеля-Хепзеля?

7. Охарактеризуйте призначення кореляційного та коваріаційного аналізів у системі R?

8. Чим відрізняються коефіцієнти Пірсона, Спірмена, Кенделла? У чому полягає їх призначення?

9. За допомогою яких тестів можливо перевірити гіпотези в середовищі R?

10. У чому полягає різниця між тестами Стьюдента та Вілкоксона? Охарактеризуйте кожен із зазначених тестів.

Тестові завдання для самоконтролю та перевірки знань

1. Характеристика положення для вибіркового розподілу є складовою описової статистики і має назву:

а) середнього значення;

б) медіани;

в) моди;

г) стандартного відхилення.

2. Обчислити п'ять описових статистик (мінімум, нижній квартиль, медіана, верхній квартиль, максимум) можливо за допомогою функції:

а) Фішера;

б) Стьюдента;

в) Тьюкі;

г) Вілкоксона.

3. У базовій версії програми R відсутня можливість обчислення таких описових статистик:

а) моди та медіани;

б) асиметрії і ексцеса;

в) мінімуму та максимуму;

г) верхнього та нижнього квартиля.

4. Метод статистичного опису даних, що є функцією, яка, насамперед, визначає для кожного показника ідеальне значення, так як ця величина зазвичай вже виміряна, – це:

а) таблиця спряженості;

б) таблиця частот;

в) таблиця описової статистики;

г) таблиця кореляцій.

5. Таблиця у форматі матриці, яка показує розподіл частоти змінних, – це:

а) таблиця частот;

б) таблиця описової статистики;

в) таблиця кореляцій;

г) таблиця спряженості.

6. Метод статистичної оцінки гіпотез, за якого нульова гіпотеза правильна, або будь-які, в яких це так асимптотично, тобто, що вибірковий розподіл може бути зроблений для апроксимації розподілу як завгодно близько, роблячи розмір вибірки досить великим, – це:

а) χ-квадрат;

б) Фішера;

в) Кохран-Мантеля-Хензеля;

г) Стьюдента.

7. Тест статистичної значущості, який використовується в аналізі таблиць спряженості для вибірок маленьких розмірів, – це:

а) Фішера;

б) Кохран-Мантеля-Хензеля;

в) Стьюдента;

г) χ-квадрат.

8. Показник кореляції (лінійної залежності) між двома змінними X та Y, який набуває значень від −1 до +1 включно, – це:

а) коефіцієнт рангової кореляції Спірмена;

б) коефіцієнт кореляції Кенделла;

в) коефіцієнт кореляції Пірсона;

г) коефіцієнт кореляції Стьюдента.

9. Непараметрична міра статистичної залежності між двома змінними, яка оцінює, наскільки можна описати відношення між змінними за допомогою монотонної функції, – це:

а) коефіцієнт кореляції Пірсона;

б) коефіцієнт рангової кореляції Спірмена;

в) коефіцієнт кореляції Стьюдента;

г) коефіцієнт кореляції Кенделла.

10. Загальна назва для класу методів статистичної перевірки гіпотез (статистичних критеріїв), заснованих на розподілі Стьюдента:

а) критерії Стьюдента;

б) критерій Кенделла;

в) критерій Фішера;

г) критерій Пірсона.

ТЕМА 4
МЕТОДИКА ПРОВЕДЕННЯ РЕГРЕСІЙНОГО АНАЛІЗУ В R

***Мета*:** *вивчити основи регресійного аналізу в середовищі R; здійснювати оцінку придатності моделі за результатами регресійного аналізу в R; забезпечити коригування регресійних моделей у цілях їх вдосконалення в середовищі R.*

План

4.1. Основи регресійного аналізу в середовищі R.

4.2. Оцінка придатності моделі за результатами регресійного аналізу в системі R.

4.3. Коригування регресійних моделей у цілях їх вдосконалення в середовищі R.

***Ключові поняття та терміни*:** *регресійна модель, регресія, регресійні коефіцієнти, стандартні відхилення, t-критерій Стьюдента, довірчий інтервал, діаграма розсіювання, квадратична регресія, діаграма ядерної оцінки функції щільності і графік-щітка, мультиколінеарність, гетероскедастичність, гомоскедастичність, критерій Дарбіна-Уотсона, фактор інфляції дисперсії, викиди.*

***Основні функції та програмні коди*:** `lm()`, `coefficients()`, `confint()`, `fitted()`, `residuals()`, `anova()`, `vcov()`, `AIC()`, `plot()`, `predict()`, `summary()`, `confint()`, `qqPlot()`, `durbinWatsonTest()`, `crPlots()`, `ncvTest()`, `spreadLevelPlot()`, `outlierTest()`, `avPlots()`, `influencePlot()`, `scatterplot()`, `scatterplotMatrix()`, `vif()`.

4.1. Основи регресійного аналізу в середовищі R

У R основна функція для побудови регресійних моделей – це `lm()`. Ця функція входить до пакета dplyr, перед її використанням необхідно запустити команду `library(dplyr)`.

Формат застосування функції `lm` такий:

```
myfit <- lm (formula, data)
```

де `formula` описує вид моделі, яку потрібно побудувати, а `data` – це таблиця з даними, які використовуються для створення моделі. Отриманий об'єкт (`myfit` у цьому випадку) – це список, який містить велику інформацію про побудовану модель. Формула зазвичай записується в такому вигляді:

```
Y ~ X1+X2+...+Xk
```

де ~ відокремлює залежну змінну зліва від незалежних змінних (розділених знаками +) справа. Для різних змін цієї формули можна використовувати інші символи.

Нижче перераховані кілька корисних функцій для виконання простого або множинного регресійного аналізу. Кожна з них застосовується до об'єкта, створеного за допомогою функції `lm()`, щоб отримати додаткову інформацію про побудовану модель.

– `summary()` – показує детальну інформацію про побудовану модель;

– `coefficients()` – перераховує параметри моделі (вільний член і регресійні коефіцієнти);

– `confint()` – обчислює довірчі інтервали для параметрів моделі (за замовчуванням 95 %);

– `fitted()` – виводить на екран передбачені значення згідно з побудованою моделлю;

– `residuals()` – показує залишки для побудованої моделі;

– `anova()` – створює таблицю ANOVA (дисперсійного аналізу) для побудованої моделі або таблицю ANOVA, що порівнює дві або більше моделей;

– `vcov()` – виводить коваріаційну матрицю для параметрів моделі;

– AIC() – обчислює інформаційний критерій Акаіке (Akaike's Information Criterion);

– plot() – створює діагностичні діаграми для оцінки адекватності моделі;

– predict() – використовує побудовану модель для знаходження прогнозу залежної змінної залежно від прогнозних значень факторів.

Коли у регресійній моделі є одна залежна і одна незалежна змінна, такий підхід називається простою лінійною регресією. У випадку наявності лише однієї залежної змінної, але в модель входять її ступені (наприклад, X, X^2, X^3), це називається поліноміальною регресією. Якщо є більше однієї незалежної змінної, називаємо це множинною регресією.

Розглянемо, як використовуються функції, перераховані вище, на простому прикладі. Набір даних women, що створено з базовою версією програми, містить дані про зріст і вагу 15 жінок у віці від 30 до 39 років. Завдання – передбачити вагу за значеннями зростання, що допоможе виявити жінок з надмірною або недостатньою вагою. Регресійний аналіз проводиться за допомогою такого програмного коду:

```
library(dplyr)
fit <-lm(weight~height,data=women)
summary(fit)
```

У консолі відобразиться релультат економетричного моделювання

```
Call:
lm(formula = weight ~ height, data = women)

Residuals:
Min 1Q Median 3Q Max
-1.7333 -1.1333 -0.3833 0.7417 3.1167

Coefficients:
Estimate Std. Error t value
(Intercept) -87.51667 5.93694 -14.74
height 3.45000 0.09114 37.85
```

```
    Pr(>|t|)
    (Intercept) 1.71e-09 ***
    height 1.09e-14 ***
    ---
    Signif. codes:
    0 '***' 0.0\01 '**' 0.01 '*' 0.05 '.' 0.1 '
' 1

    Residual standard error: 1.525 on 13 degrees
of freedom
    Multiple R-squared: 0.991,   Adjusted   R-
squared: 0.9903
    F-statistic: 1433 on 1 and 13 DF, p-value:
1.091e-14
```

Частину таблиці вихідних даних можна побачити за допомогою команди:

```
    women$weight
```

У консолі можна побачити запитувану вагу жінок:

```
    [1] 115 117 120 123 126 129 132 135 139 142
146
    [12] 150 154 159 164
    > fitted(fit)
    1 2 3 4 5
    112.5833 116.0333 119.4833 122.9333 126.3833
    6 7 8 9 10
    129.8333 133.2833 136.7333 140.1833 143.6333
    11 12 13 14 15
    147.0833 150.5333 153.9833 157.4333 160.8833
    > residuals(fit)
    1 2 3 4
    2.41666667 0.96666667 0.51666667 0.06666667
    5 6 7 8
```

```
    -0.38333333   -0.83333333   -1.28333333   -
1.73333333
     9 10 11 12
    -1.18333333   -1.63333333   -1.08333333   -
0.53333333
     13 14 15
     0.01666667 1.56666667 3.11666667

    plot(women$height,women$weight, xlab="Зріст
(дюймів)", ylab="Вага (фунтів)")
    > abline(fit)
```

З отриманих результатів випливає, що рівняння для передбачення ваги щодо зростання має такий вигляд:

$$\widehat{Weight} = -87{,}52 + 3{,}45 \cdot \widehat{Height}$$

Оскільки нульове зростання неможливе, не будемо намагатися інтерпретувати вільний член. У цьому випадку це просто константа. Зі стовпця `Pr (>|t|)` зрозуміло, що коефіцієнт регресії (3,45) статистично значимо відрізняється від нуля ($p < 0{,}001$) і означає, що на кожен дюйм зростання очікується збільшення ваги на 3,45 фунта. Множинний коефіцієнт детермінації (0,991) означає, що модель пояснює 99,1 % дисперсії значень ваги. Цей коефіцієнт є квадратом коефіцієнта кореляції між реальними і передбаченими значеннями. Стандартну помилку залишків (1,53 фунта) можна інтерпретувати як усереднену помилку передбачення ваги по зростанню з використанням цієї моделі. P-значення дозволяє перевірити, чи прогнозують незалежні змінні значення залежної краще.

З метою візуалізації реальних і передбачених значень, а також залишків (різниці між передбаченими і реальними значеннями) побудовано графік. Очевидно, що найбільші залишки характерні для най нижчих і високих жінок, що також можна бачити на діаграмі (рис. 4.1).

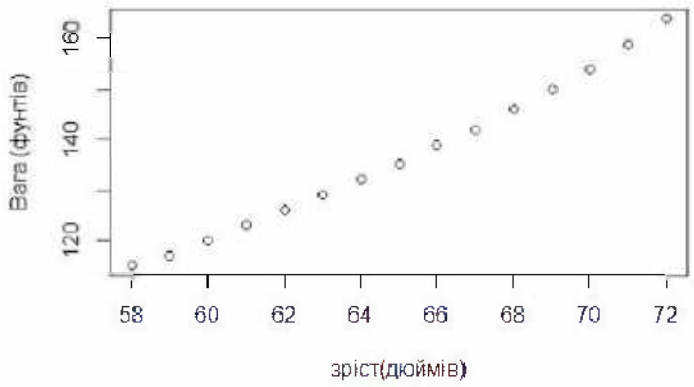


Рис. 4.1. Діаграма розсіювання за умовою задачі

На діаграмі помітно, що можна поліпшити прогноз, використовуючи лінію з одним вигином. Наприклад, квадратична модель може краще відповідати даним. Поліноміальна регресія дозволяє прогнозувати залежну змінну з незалежної, якщо цей зв'язок має вигляд полінома певного ступеня.

З рис. 4.1 випливає, що точність передбачення можна поліпшити, якщо використовувати квадратичне регресійне рівняння (тобто включити в нього X^2).

Підігнати модель у вигляді квадратичного рівняння можна за допомогою такого виразу:

```
> fit2 <-lm(weight~height+I (height^2), data=women)
```

Тут потрібно пояснити нове позначення: `I(height^2)`. `height^2` додає в рівняння зведений у квадрат зріст. Функція `I` інтерпретує вміст дужок як регулярний вираз R. Це потрібно, оскільки оператор `^` має у формулах особливе значення, яке потрібно нівелювати в цьому випадку.

У програмному коді наведені результати побудови квадратичного рівняння:

```
> fit2 <-lm(weight~height+I(height^2), data=women)
> summary(fit2)
```

```
    Call:
    lm(formula = weight ~ height + I(height^2),
data = women)

    Residuals:
    Min 1Q Median 3Q Max
    -0.50941 -0.29611 -0.00941 0.28615 0.59706

    Coefficients:
    Estimate Std. Error t value Pr(>|t|)
    (Intercept) 261.87818 25.19677 10.393 2.36e-
07
    height -7.34832 0.77769 -9.449 6.58e-07
    I(height^2) 0.08306 0.00598 13.891 9.32e-09

    (Intercept) ***
    height ***
    I(height^2) ***
    ---
    Signif. codes:
    0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Residual standard error: 0.3841 on 12 degrees
of freedom
    Multiple R-squared: 0.9995,  Adjusted  R-
squared: 0.9994
    F-statistic: 1.139e+04 on 2 and 12 DF, p-
value: < 2.2e-16
    > plot(women$height,women$weight)
    > lines(women$height,fitted(fit2))
```

За результатами цього аналізу регресійне рівняння набуває вигляду:

$$\widehat{Weight} = 261{,}88 + 7{,}35 \cdot \widehat{Height} + 0{,}083 \cdot \widehat{Height}^2$$

Щоб порівняти дві економетричні моделі, можна використати функцію `mtable()` за допомогою такого виразу:

```
mtable(fit,fit2)
```

У програмному коді наведені результати порівняння:

```
Calls:
fit: lm(formula = weight ~ height, data = women)
fit2: lm(formula = weight ~ height + I(height^2), data = women)

=========================================
fit fit2
-----------------------------------------
(Intercept) -87.517*** 261.878***
(5.937) (25.197)
height 3.450*** -7.348***
(0.091) (0.778)
I(height^2) 0.083***
(0.006)
-----------------------------------------
R-squared 0.991 0.999
N 15 15
```

Обидва регресійних коефіцієнти виявляються значущими на рівні $p < 0{,}0001$. Частка поясненої дисперсії в цьому випадку підвищилася до 99,9%. Значимість квадратного члена ($t = 13{,}89$, $p < 0{,}001$) вказує на те, що його включення в модель поліпшило її адекватність. Якщо подивитися на діаграму для другої моделі (рис. 4.2), то можна побачити, що крива краще описує реальні дані.

Якщо існує більше однієї незалежної змінної, проста лінійна регресія перетворюється у множинну лінійну регресію, а хід обчислень стає більш складним. З технічної точки зору, поліноміальна регресія – це окремий випадок множинної регресії. При квадратичній регресії є дві незалежні змінні (X і X^2), а при кубічній регресії – три незалежні змінні (X, X^2, X^3). Розглянемо більш загальний приклад.

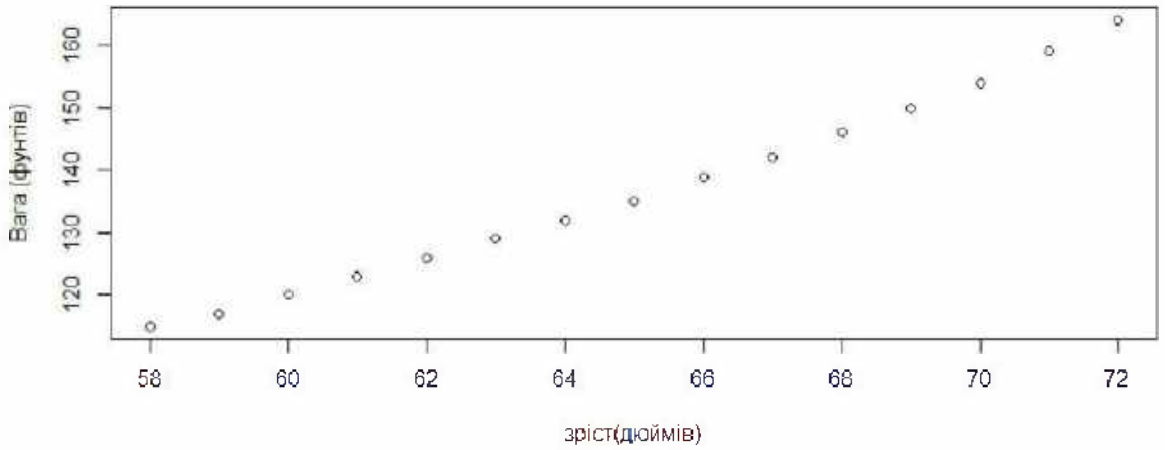


Рис. 4.2. Квадратична регресія за умови задачі

Для цього прикладу використовуємо набір даних `state.x77`, який представлений з базовою версією програми. Потрібно дослідити зв'язок між рівнем злочинності і іншими характеристиками для кожного штату, включаючи чисельність населення, рівень неписьменності, середній дохід і температуру, нижче нуля (середнє число днів з мінусовою температурою).

Оскільки функція `lm()` працює з таблицею даних (а набір даних `state.X77` зберігається у вигляді матриці), можна спростити це за допомогою такого програмного коду:

```
states <- as.data.frame(state.x77 [,c ("Murder", "Population", "Illiteracy", "Income","Frost")])
```

Цей код створює таблицю даних, яка називається `states` і містить потрібні нам змінні. Будемо використовувати цю таблицю в подальшому аналізі.

Важливий перший крок у множинній регресії – дослідження парних взаємозв'язків між змінними. Парні кореляції обчислюються за допомогою функції `cor()`, а діаграми розсіювання створюються за допомогою функції `scatterplotMatrix()` з пакета `car`:

```
    states <- as.data.frame(state.x77 [,c
("Murder", "Population", "Illiteracy",
"Income", "Frost")])
    cor(states)
```

У програмному коді наведений результат моделювання:

```
    Murder Population Illiteracy
    Murder 1.0000000 0.3436428 0.7029752
    Population 0.3436428 1.0000000 0.1076224
    Illiteracy 0.7029752 0.1076224 1.0000000
    Income -0.2300776 0.2082276 -0.4370752
    Frost -0.5388834 -0.3321525 -0.6719470

    Income Frost
    Murder -0.2300776 -0.5388834
    Population 0.2082276 -0.3321525
    Illiteracy -0.4370752 -0.6719470
    Income 1.0000000 0.2262822
    Frost 0.2262822 1.0000000

    library(car)
    scatterplotMatrix(states,spread=FALSE,lty.s
mooth=2)
```

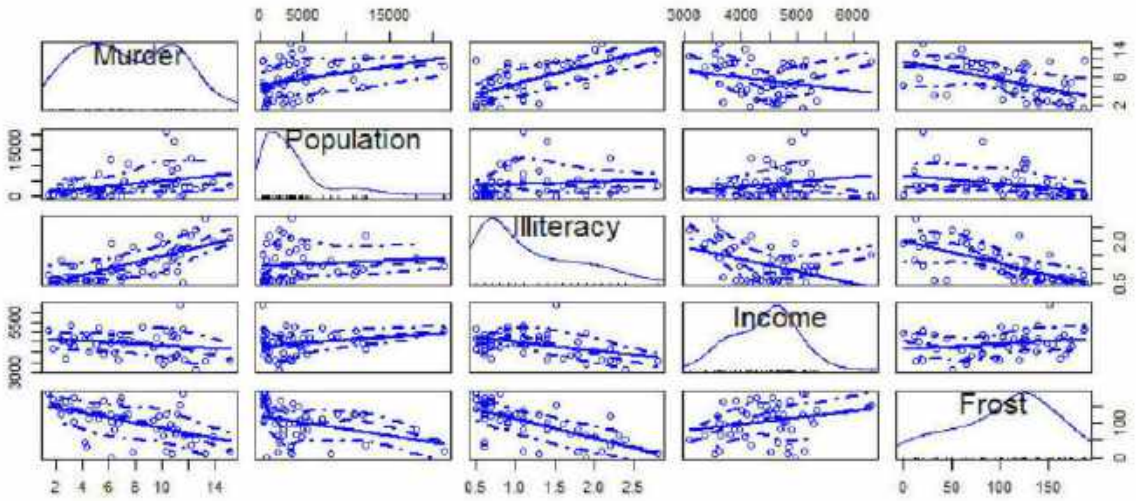


Рис. 4.3. Діаграма ядерної оцінки функції щільності і графік-щітка

За замовчуванням функція створить діаграми розсіювання для всіх пар змінних з накладенням згладжування (loess) за допомогою кривої і регресійної прямої. На головній діагоналі представлені діаграми щільності і графіки-щітки для кожної змінної.

Можна побачити, що рівень злочинності має бімодальний розподіл, а розподіл кожної незалежної змінної тією чи іншою мірою асиметричний. Рівень злочинності зростає разом зі зростанням населення і рівнем неписьменності, а падає – зі збільшенням доходу і морозу на вулиці. Водночас більш холодні штати характеризуються нижчим рівнем неписьменності і чисельності населення, а також високими доходами.

Побудуємо тепер множинну регресійну модель за допомогою функції lm():

```
fit<-lm(Murder~Population+Illiteracy+
Income+Frost, data=states)
summary(fit)
```

У програмному коді наведений результат моделювання:

```
Call:
lm(formula = Murder ~ Population + Illiteracy
+ Income + Frost,
data = states)

Residuals:
Min 1Q Median 3Q Max
-4.7960 -1.6495 -0.0811 1.4815 7.6210

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.235e+00 3.866e+00 0.319 0.7510
Population 2.237e-04 9.052e-05 2.471 0.0173 *
Illiteracy 4.143e+00 8.744e-01 4.738 2.19e-
05 ***
Income 6.442e-05 6.837e-04 0.094 0.9253
Frost 5.813e-04 1.005e-02 0.058 0.9541
---
```

```
    Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
0.05 '.' 0.1 ' ' 1

    Residual standard error: 2.535 on 45 degrees
of freedom
    Multiple R-squared: 0.567,    Adjusted    R-
squared: 0.5285
    F-statistic: 14.73 on 4 and 45 DF, p-value:
9.133e-08
```

Рівняння регресії виглядає так:

$$\widehat{\text{Murder}} = 1.235 + 0.0002 \cdot \widehat{\text{Population}} + 4.143 \cdot \widehat{\text{Illiteracy}} + 0.0001 \cdot \widehat{\text{Income}} + 0.0006 \cdot \widehat{\text{Frost}}$$

Якщо існує більше однієї незалежної змінної, то регресійні коефіцієнти показують, наскільки збільшиться значення залежної змінної при зміні даної незалежної змінної на одиницю за умови, що всі інші незалежні змінні залишаться незмінними. Наприклад, регресійний коефіцієнт для рівня неписьменності дорівнює 4,14. Це свідчить про те, що збільшення неписьменності на 1 % пов'язане зі збільшенням рівня злочинності на 4,14 % при постійних значеннях чисельності населення, доходу і низької температури на вулиці. Цей коефіцієнт статистично значимо відрізняється від нуля при $p < 0{,}0001$. З іншого боку, регресійний коефіцієнт для морозу на вулиці статистично не відрізняється від нуля. Це говорить про те, що мороз і рівень злочинності не пов'язані між собою лінійно при незмінних значеннях усіх інших змінних. Взяті разом незалежні змінні можуть пояснити 57 % дисперсії рівня злочинності по штатах.

4.2. Оцінка придатності моделі за результатами регресійного аналізу в системі R

Для регресійного аналізу було використано функцію `lm()` для побудови регресійної МНК-моделі і функцію `summary()`, щоб отримати протокол моделі: оцінки параметрів і загальні статистики, що характеризують її. Ступінь впевненості у

параметрах регресії залежить від того, наскільки модель задовольняє обмеженням МНК. Хоча функція `summary()`, наведена в програмному коді вище, описує модель, вона нічого не повідомляє про те, наскільки модель відповідає статистичним обмеженням.

Помилки в даних або не правильно визначені взаємозв'язки між залежними і незалежними змінними можуть привести до того, що у результаті отримаємо значною мірою неточну модель. З одного боку, можна вирішити, що між залежною і незалежною змінними немає зв'язку, в той час як насправді він існує. З іншого – можна прийти до висновку, що залежна і незалежна змінна пов'язані, а насправді це буде не так. Також можна отримати модель, яка, будучи застосованою до реальних даних, робить неадекватні прогнози зі значними і непотрібними помилками.

Розглянемо результат застосування функції `confint()` до набору даних `states` для вирішення завдання множинної регресії.

```
    fit<-lm(Murder~Population+Illiteracy+-
Income+Frost,data=states)
    confint(fit)
    2.5 % 97.5 %
    (Intercept) -6.552191e+00 9.0213182149
    Population 4.136397e-05 0.0004059867
    Illiteracy 2.381799e+00 5.9038743192
    Income -1.312611e-03 0.0014414600
    Frost -1.966781e-02 0.0208304170
```

Отриманий результат свідчить про те, що можна бути на 95 % впевненими в тому, що інтервал [2,38, 5,90] містить реальне значення, на яке змінюється рівень злочинності, при зміні рівня неписьменності на 1 %. Крім того, оскільки довірчий інтервал для морозу на вулиці (`Frost`) містить нуль, можна зробити висновок, що зміни цього параметра не пов'язані з рівнем злочинності при постійних значеннях інших змінних.

Однак впевненість у цих результатах сильна настільки, наскільки виконуються статистичні допущення, що лежать в основі моделі.

Набір методів під назвою «оцінка придатності регресійних моделей» надає необхідні інструменти для оцінки регресійної моделі і може допомогти виявити і усунути проблеми. Розглянемо більш нові, вдосконалені методи, реалізовані в пакеті `car`:

– `qqPlot()` – діаграма порівняння квантилів;

– `dwt()`– тест Дарбіна-Уотсона (Durbin-Watson test) на автокореляцію залишків;

– `bgtest()` тест Бреуша-Годфрі (Breusch-Godfrey serial correlation LM-test) на автокореляцію залишків;

– `gqtest ()` тест Голдфельда-Квандта (Goldfeld-Quandt test) на гетероскедастичність;

– `crPlots()` – діаграма компонент і залишків;

– `ncvTest()` – тест на неоднорідність дисперсії залишків.

– `spreadLevelPlot()` – діаграма для виявлення неоднорідності дисперсії залишків (spread-level plot);

– `outlierTest()` – тест Бонферроні на викиди;

– `avPlots()` – діаграми доданих змінних;

– `influencePlot()` – діаграма впливу спостережень на регресію;

– `scatterplot()` – удосконалена діаграма розсіювання;

– `scatterplotMatrix()` – удосконалена матриця діаграм розсіювання;

– `vif()` – дисперсійно-інфляційний фактор.

Функція `qqPlot()` – це більш акуратний метод перевірки припущення про нормальність порівняно з функцією `plot()` з базовою версією програми. Вона зображує зв'язок між залишками Стьюдента (також званими виключеними залишками Стьюдента, або залишками, обчисленими методом послідовного виключення значень – jackknifed residuals) і квантиль розподілу Стьюдента з n-p ступенями свободи, де n – це обсяг вибірки, а p – число параметрів регресії (включаючи вільний член). Програмний код такий:

```
> library(car)
>fit<-lm(Murder~Population+Illiteracy+
Income+ Frost,data=states)
>       qqPlot(fit,labels=row.names(states),
id.method="identify", simulate=TRUE, main="Q-Q
Plot")
Nevada Rhode Island
28 39
```

Функція `qqPlot` створить графік розподілу ймовірностей, наведений на рис. 4.4.

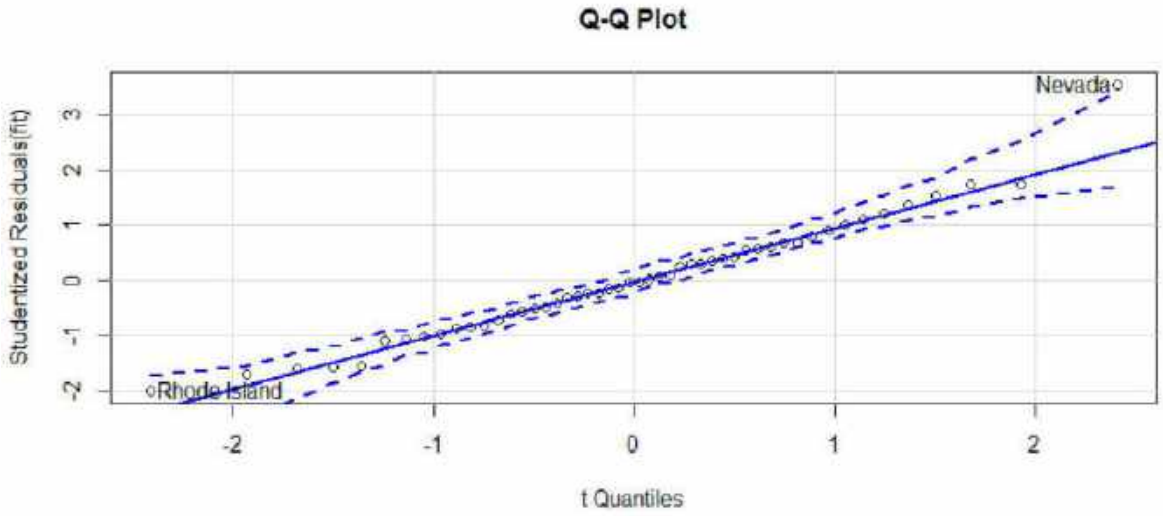


Рис. 4.4. Графічна перевірка нормальності розподілу

Опція `id.method="identify"` робить діаграму інтерактивною – після того як графік намальований, клацання миші по точках графіка підписують їх значення, задані опцією `labels`. Натискання клавіші Esc, вибір команди Stop з випадаючого графічного меню або клацання правою кнопкою миші на графік вимкне інтерактивний режим. У цьому випадку підписано Неваду. Якщо `simulate=TRUE`, то на графіку також зображуються 95 % довірчі інтервали з використанням параметричних бутстреп-методів.

За винятком Невади, всі точки потрапляють близько до лінії і знаходяться всередині довірчих меж, що свідчить про нормальність розподілу. Але дійсно потрібно поглянути на Неваду. У неї є великий позитивний залишок (порівняно з

передбаченим значенням), що показує, що модель недооцінює рівень злочинності в цьому штаті. А саме:

```
states["Nevada",]      Murder      Population
Illiteracy Income Frost
```

У консолі відобразиться результат:

```
Nevada 11.5 590 0.5 5149 188
fitted(fit)["Nevada"]
```

У консолі відобразиться результат:

```
Nevada
3.878958
residuals(fit)["Nevada"]
```

У консолі відобразиться результат:

```
Nevada
7.621042
rstudent(fit)["Nevada"]
```

У консолі відобразиться результат:

```
Nevada
3.542929
```

Тут видно, що рівень злочинності дорівнює 11,5 %, а модель прогнозує 3,9 %. А чому ж тоді Неваді рівень злочинності вищий, ніж це передбачено чисельністю населення, рівнем доходу, неграмотності і морозів.

Для повноти викладу наведемо ще один спосіб візуалізації залишків. Для цього побудуємо наступний програмний код. Функція `residplot()` створює **гістограму залишків Стьюдента** з накладанням кривої нормального розподілу, кривої ядерної оцінки функції щільності. Для використання цієї функції пакет `car` не потрібен.

```
> residplot<-function(fit,nbreaks=10) {
+ z<-rstudent(fit)
+ hist(z,breaks=nbreaks,freq=FALSE)
+ rug(jitter(z),col="brown")
+ curve(dnorm(x,mean=mean(z),sd=sd(z)),
+ add=TRUE,col="blue",lwd=2)
+ lines(density(z)$x,density(z)$y,
+ col="red",lwd=2,lty=2)
```

```
    + legend("topright",
    +  legend=c("Крива нормального розподілу",
"Ядерна оцінка щільності"),
    + lty=1:2,col=c("blue","red"),cex=.7)
    + }
    > residplot(fit)
```

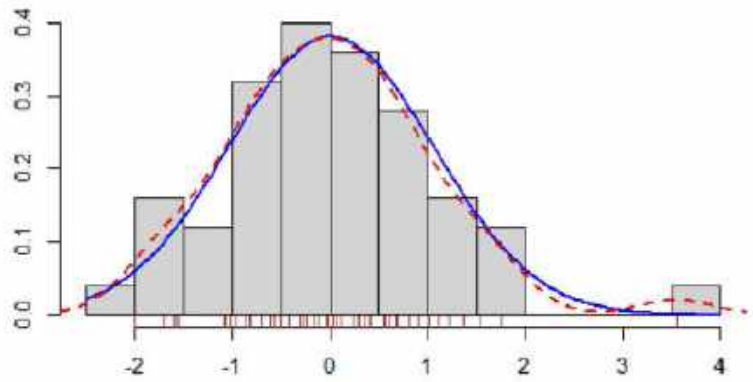


Рис. 4.5. Розподіл залишків Стьюдента за допомогою функції `residplot()`

Як можна спостерігати, залишки досить добре відповідають нормальному розподілу, за винятком великого викиду. Хоча Q-Q-діаграма, можливо, більш інформативна, завжди легше оцінити асиметрію розподілу за гістограмою або графіком функції щільності, ніж за ймовірнісною діаграмою.

Як вказувалося раніше, кращий спосіб зрозуміти, чи не пов'язані між собою значення залежної змінної (і, таким чином, залишки), – це знання про спосіб збору даних. Наприклад, для часових рядів часто властива автокореляція – спостереження, зроблені в короткі відрізки часу, більш сильно корелюють один з одним, ніж рознесені в часі спостереження. У пакеті `car` реалізована функція для проведення **тесту Дарбіна-Уотсона** на наявність **автокореляції в залишках**. Цей тест можна використовувати до завдання з множинної регресії за допомогою такого програмного коду:

```
dwt (fit)
```

У консолі відобразиться результат:

```
lag Autocorrelation D-W Statistic p-value
1 -0.2006929 2.317691 0.218
```

```
Alternative hypothesis: rho != 0
```

Високе значення ймовірності статистичної помилки першого роду (p = 0,218) свідчить про відсутність автокореляції і, отже, про незалежність залишків. Значення інтервалу (lag, у цьому випадку 1) свідчить про перевірку гіпотези про автокореляцію першого порядку, а саме: кожне значення порівнюється з сусіднім, наступним за ним значенням у наборі даних. Однак, будучи призначеним для часових рядів, цей тест не так добре підходить для інших типів даних. Зауважуємо, що функція `dwt(fit)` для обчислення p використовує бутстреп-метод. Якщо не додати опцію `simulate=FALSE`, результати тесту будуть відрізнятися.

Звернемо увагу на проблему, яка не пов'язана безпосередньо зі статистичними припущеннями, але все одно важлива, оскільки дозволяє інтерпретувати результати множинної регресії.

Альтернативним тестом для перевірки наявності автокореляції залишків є тест Бреуша-Годфрі, який дозволяє перевіряти наявність автокореляції зилишків не тільки першого порядку, а й вищих. Синтаксис функції такий:

```
bgtest(fit,order = 1)
```

У програмному коді наведений результат розрахунків:

```
Breusch-Godfrey test for serial correlation
of order up to 1
data: fit
LM test = 2.2707, df = 1, p-value = 0.1318
```

```
#критичне значення хі-квадрат критерію
qchisq(0.95,1)
```

У програмному коді наведений результат розрахунків:

```
[1] 3.841459
```

Оскільки підтвердилась альтернативна гіпотеза за тестом Дарбіна-Уотсона, можна зробити висновок про наявність у моделі автокореляції залишків. Цей висновок підтверджує тест Бреуша-Годфрі. Оскільки $BG < \chi_1^2$ (2.2707<3.8415) в моделі відсутня автокореляція залишків першого порядку з ймовірністю 95 %.

Важливою проблемою при побудові економетричної моделі може бути наявність гетероскедастичності. Для цього необхідно провести діагностику моделі на основі тесту Гольдфельда-

Квандта. З цією метою використовуємо функцію gqtest, аргументами якої є назва моделі, змінна, що зумовлює гетероскедастичність, наприклад ВВП: `order.by = ~Population`, вихідний масив інформації: data=state і кількість спостережень, які необхідно вилучити fraction = 0.25. Приклад написання команди:

```
gqtest(fit, order.by = ~ Population, data=states, fraction = 0.25)
```

```
Goldfeld-Quandt test
```

У програмному коді наведений результат розрахунків:

```
data: fit
GQ = 0.65771, df1 = 14, df2 = 13, p-value = 0.7767
alternative hypothesis: variance increases from segment 1 to 2
```

Далі необхідно знайти критичне значення F-критерію Фішера для певного рівня ймовірності та ступенів свободи df1, df2. Приклад команди:

```
qf(0.95,14,13)
```

У програмному коді наведений результат розрахунків:

```
[1] 2.553619
```

Порівняємо значення критерію $R^*_{\text{Population}}$ (BG) (0.65771) з критичним (табличним) значенням F-критерію при $\gamma_1 = 14$ і $\gamma_2 = 13$ ступенях свободи і рівні значущості $\alpha = 0{,}05$ $F_{(\alpha - 0{,}05)\ 7\ 6} = 2.553619$. Оскільки $R^*_{\text{Population}} < F_{(\alpha = 0{,}05)\ 14\ 13}$, з ймовірністю 95 %, дисперсія залишків у моделі не пропорційна до значення пояснювальної змінної `Population` (за цією змінною гетероскедастичності немає).

Тести для змінної моделі `Income`:

```
gqtest(fit, order.by = ~ Income, data= states, fraction = 0.25)
```

У програмному коді наведений результат розрахунків:

```
Goldfeld-Quandt test

data: fit
```

```
    GQ = 4.305, df1 = 14, df2 = 13, p-value =
0.006184
    alternative hypothesis: variance increases
from segment 1 to 2
```

Тест для змінної моделі `Frost`:

```
    gqtest(fit, order.by = ~ Frost, data=states,
fraction = 0.25)
```

У програмному коді наведений результат тестування:

```
    Goldfeld-Quandt test

    data: fit
    GQ = 2.8118, df1 = 14, df2 = 13, p-value =
0.03537
    alternative hypothesis: variance increases
from segment 1 to 2
```

Порівняємо значення критеріїв R^*_{Income} (BG) R^*_{Frost} (4.305, 2.8118) з критичним (табличним) значенням F-критерію при $\gamma_1 = 14$ і $\gamma_2 = 13$ ступенях свободи і рівні значущості $\alpha = 0{,}05$ $F_{(\alpha = 0{,}05)\ 7\ 6} = 2.553619$. Оскільки $R^*_{\text{Income}} > F_{(\alpha = 0{,}05)\ 14\ 13}$ та $R^*_{\text{Frost}} > F_{(\alpha - 0{,}05)\ 14\ 13}$ з ймовірністю 95 %, дисперсія залишків у моделі пропорційна до значень пояснювальних змінних `Income` та `Frost` (за цими змінними гетероскедастичності в моделі є).

Оскільки деякі фактори можуть повністю корелюватися з точністю до помилки округлення, коефіцієнт регресії вимірює вплив однієї незалежної змінної на залежну при постійних значеннях іншої змінної. Це означає, що варто аналізувати залежність між екзогенними змінними. Цей феномен має назву **мультиколінеарності**. Він призводить до великих довірчих інтервалів для параметрів моделі й ускладнює інтерпретацію окремих регресійних коефіцієнтів.

Мультиколінеарність можна виявити за допомогою статистики, званої **дисперційно-інфляційним коефіцієнтом**. Квадратний корінь з цієї статистики для будь-якої незалежної змінної вказує на ступінь збільшення довірчого інтервалу для

параметра регресії даної змінної порівняно з моделлю без корельованих незалежних змінних (звідси назва статистики). Фактор інфляції дисперсії можна обчислити за допомогою функції `vif()` з пакета `car`. Зазвичай значення квадратного кореня цієї статистики, що перевищують 2, вказують на наявність мультиколінеарності. Програмний код наведено нижче:

```
library(car)
vif(fit)
```

У програмному коді наведений результат тестування:

```
Population Illiteracy Income Frost
1.245282 2.165848 1.345822 2.082547
> sqrt(vif(fit))>2 # проблема?
Population Illiteracy Income Frost
FALSE FALSE FALSE FALSE
```

Отриманий результат свідчить, що в нашому випадку проблема мультиколінеарності відсутня.

Також у регресійній моделі може існувати проблема викидів. **Викиди** – це спостереження, які погано прогнозуються моделлю. Вони характеризуються великими позитивними або негативними залишками. Позитивні залишки свідчать про те, що модель недооцінює залежну змінну, негативні – ознака переоцінки.

У пакеті `car` також реалізований статистичний тест на викиди. Функція `outlierTest()` обчислює значення ймовірності статистичної помилки першого роду з поправкою Бонферроні для найбільшого залишку Стьюдента:

```
library(car)
outlierTest(fit)
```

У програмному коді наведений результат розрахунків:

```
rstudent unadjusted p-value Bonferroni p
Nevada 3.542929 0.00095088 0.047544
```

Звідси видно, що Невада визначена як викид (p = 0,048). Зауважимо, що ця функція проводить тест для одного найбільшого (позитивного або негативного) залишку на значимість його як викиду. Якщо тест не дав відчутного результату, то в даних немає

викидів. А якщо результат тесту значимий, потрібно видалити викид і провести тест заново, щоб дізнатися, чи є інші викиди.

4.3. Коригування регресійних моделей у цілях їх удосконалення в середовищі R

Видалення викидів може поліпшити відповідність набору даних вимогам нормальності. Впливові спостереження також часто видаляють, оскільки вони занадто сильно впливають на результати. Після видалення найбільшого викиду або впливового спостереження модель підбирається заново. Якщо після цього все одно залишаються викиди або впливові спостереження, процес повторюється, поки не буде досягнуто допустиму відповідність моделі даних. Потрібно бути обережним, коли приймається рішення про видалення спостережень. Іноді можна зрозуміти, що спостереження – це викид через помилки при зборі даних, або тому, що протокол дослідження не був дотриманий, або тестований порушив інструкції. У таких випадках видалення спостережень є абсолютно обґрунтованим. В інших випадках незвичайні спостереження можуть виявитися найцікавішими об'єктами в отриманих даних. Розуміння причин, за якими це спостереження відрізняється від інших, може значно наблизити до з'ясування досліджуваної проблеми і вирішення інших питань, про які можна і не замислюватися. Деякі найбільш серйозні досягнення можуть бути обумовлені інтуїтивною прозорливістю, коли помічаєте те, що не відповідає очікуванням.

Коли моделі не відповідають вимогам нормальності, лінійності або гомоскедастичності, трансформація однієї або більше змінних може поліпшити або виправити ситуацію. Перетворення зазвичай полягають у заміні змінної Y на змінну Y^{λ}. Часто використовувані значення λ і їх інтерпретація такі: $1/Y^2$, $1/Y$, $\log(Y)$, Y^2 тощо.

Якщо Y – пропорція, то часто використовується логіт-перетворення `[ln(Y/1-Y)]`.

Якщо модель не відповідає вимогам нормальності, зазвичай намагаються перетворити залежну змінну. За допомогою функції powerTransform() з пакета car можна оцінити за методом максимальної правдоподібності величину λ, зведення в яку, швидше за все, нормалізує змінну X^{λ}. У наведеному нижче програмному коді цей метод застосований до набору даних states.

```
library(car)
summary(powerTransform(states$Murder))
```

У програмному коді наведений результат розрахунків:

```
bcPower Transformation to Normality
Est Power Rounded Pwr Wald Lwr Bnd
states$Murder 0.6055 1 0.0884
Wald Upr Bnd
states$Murder 1.1227

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)
LRT df pval
LR test, lambda = (0) 5.665991 1 0.017297

Likelihood ratio test that no transformation is needed
LRT df pval
LR test, lambda = (1) 2.122763 1 0.14512
```

З результату застосування функції випливає, що змінну Murder можна нормалізувати, замінивши її на $Murder^{0,6}$. Оскільки 0,6 близьке до 0,5, можна спробувати наблизити змінну до нормальності перетворенням методом вилучення квадратного кореня. Однак у цьому випадку гіпотеза про λ=1 не може бути відкинута (p = 0,145), тому необхідність такого перетворення змінної неочевидна.

У випадку, коли вимоги лінійності не виконуються, зазвичай допомагає перетворення незалежних змінних. Для оцінки методом найбільшої правдоподібності у тій мірі, в яку потрібно звести

незалежні змінні для більшої відповідності моделі вимогу лінійності, можна використовувати функцію `boxTidwell()` з пакета `car`. Ось приклад перетворень Боксу-Тідвелл (Box-Tidwell) для моделі, яка прогнозує рівень злочинності в штатах за чисельністю населення та рівнем неписьменності:

```
library(car)
boxTidwell(Murder~Population+Illiteracy,data=states)
```

У програмному коді наведений результат розрахунків:

```
MLE of lambda Score Statistic (z)
Population 0.86939 -0.3228
Illiteracy 1.35812 0.6194
Pr(>|z|)
Population 0.7468
Illiteracy 0.5357

iterations = 19
```

З отриманого результату випливає, що для отримання більшої лінійності варто спробувати перетворення $\texttt{Population}^{0,87}$ і $\texttt{Illiteracy}^{1,36}$. Однак результати тесту для змінних `Population` (P = -75) і `Illiteracy` (p = 0.54) свідчать, що ні одну з них не потрібно перетворювати.

Нарешті, перетворення залежної змінної можуть допомогти в разі гетероскедастичності (непостійній дисперсії залишків). У програмному коді було показано, що функція `spreadLevelPlot()` з пакета `car` дозволяє зрозуміти, в який ступінь потрібно звести залежну змінну, щоб збільшити гомоскедастичність. Знову ж таки, в прикладі з набором даних `states` вимога сталості дисперсії помилок виконується, і ніякі перетворення не потрібні.

Зміна кількості змінних, що входять у модель, буде впливати на ступінь її відповідності даним. Іноді додавання важливої змінної може виправити багато з тих проблем, які були зазначені вище. Видалення сумнівних змінних може привести до аналогічного ефекту.

Видалення змінних – це особливо важливий засіб при боротьбі з мультиколінеарністю. Якщо єдина задача – це прогнозування, тоді мультиколінеарність – не проблема. Однак якщо треба інтерпретувати окремі незалежні змінні, тоді потрібно якось справлятися з мультиколінеарністю. Найбільш поширений підхід – видалити одну зі змінних, через які спостерігається даний феномен (тобто одну зі змінних зі значенням квадратного кореня з фактора інфляції дисперсії більше двох). Альтернативний підхід – використовувати гребневу регресію, різновид множинної регресії, розроблену для застосування у випадках мультиколінеарності.

Вирішення проблеми мультиколінеарності також можливе через застосування методу гребеневої або рідж-регресії (ridge regression). Це один з методів регуляризації, який виконує певні „коригування" відхилень від нормального розподілу залишків для забезпечення надійності і ефективності оцінок параметрів регресії. Цей метод передбачає введення штрафу за невиправдано великі значення оцінок, причому його величина пропорційна величині параметра λ, за допомогою якого можна налаштовувати алгоритм підгонки моделі й отримати більш стійке рішення. Метод рідж-регресії (ridge regression) реалізує класичну L2-регуляризацію Тихонова.

Для знаходження оптимального регуляційного параметра λ доцільно використовувати узагальнену перехресну перевірку на основі зовнішнього критерію крос-валідації GCV (generalized cross-validation).

Відпровідні функції містяться в пакеті glmnet. Першим кроком є завантаження бібліотеки з цим пакетом:

```
library("glmnet") # LASSO + ridge
```

Далі вводимо вектор залежної змінної через таку команду:

```
y <- states$Murder
```

Після цього задаємо матрицю ознак-факторів (без стовпчика одиниць) через таку команду:

```
X0 <- model.matrix(data=states, Murder~0+
Population + Illiteracy + Income + Frost)
```

Знаходимо корсляційну матрицю ознак-факторів:

```
cor(X0)
```

У програмному коді наведений результат розрахунків:

```
    Population Illiteracy Income Frost
    Population 1.0000000 0.1076224 0.2082276 -
0.3321525
    Illiteracy 0.1076224 1.0000000 -0.4370752 -
0.6719470
    Income    0.2082276    -0.4370752    1.0000000
0.2262822
    Frost    -0.3321525    -0.6719470    0.2262822
1.0000000
```

Приклад застосування методу крос-валідації для визначення оптимального значення лямбда наведена нижче:

```
cv <- cv.glmnet(X0,y,alpha=1)
```

Залежність крос-валідаційної суми квадратів залишків від лямбда можна побачити на графіку, який можна знайти за допомогою функції:

```
plot(cv)
```

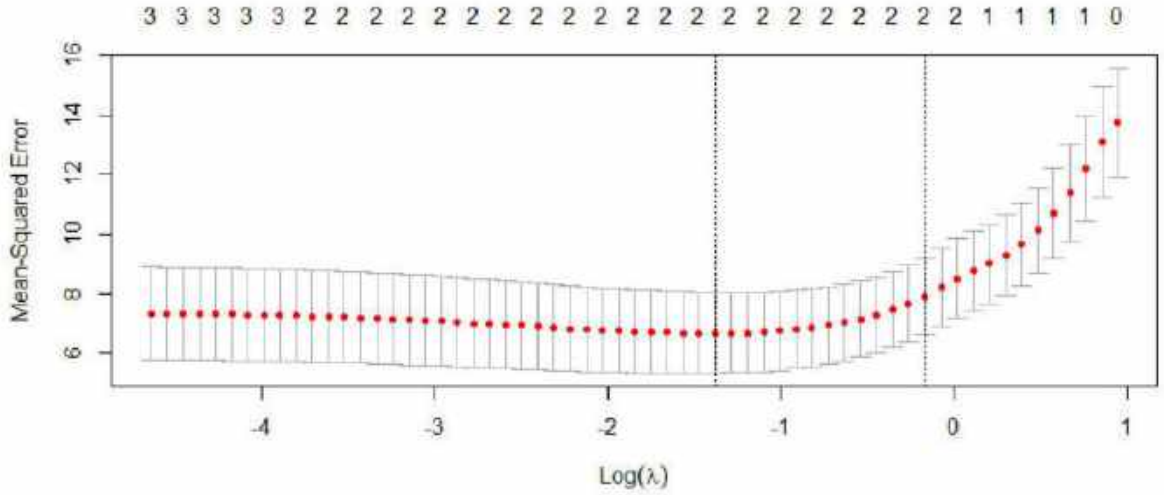


Рис. 4.6. Залежність значення параметра регуляризації λ від критерію крос-валідації (GCV)

Знаходження лямбди, яка мінімізує крос-валідаційну суму квадратів залишків:

```
cv$lambda.min
```

У програмному коді наведений результат розрахунків:

```
[1] 0.2509918
```

На основі цього методу було визначено, що оптимальне значення параметра регуляризації становить λ = 0,251.

Оцінку рідж-регресії можна знайти за допомогою функції:

```
m_rr <- glmnet(X0,y,alpha=0, lambda= cv$lambda.min)
```

При цьому коефіцієнти моделі, які відповідають мінімальній лямбда, можна знайти так:

```
coef(m_rr,s="cv$lambda.min")
```

Результат розрахунків відобразиться у консолі:

```
5 x 1 sparse Matrix of class "dgCMatrix"
 1
(Intercept) 2.853759e+00
Population 2.047603e-04
Illiteracy 3.625219e+00
Income -5.007404e-05
Frost -3.490755e-03
```

У результаті застосування методу рідж регресії рівняння регресії виглядає так:

$$\widehat{\text{Murder}} = 2.8538 + 0.0002 \cdot \widehat{\text{Population}} + 3.6252 \cdot \widehat{\text{Illiteracy}} - 0.0001 \cdot \widehat{\text{Income}} - 0.0035 \cdot \widehat{\text{Frost}}$$

При побудові економетричних моделй важливою проблемою є гетероскедастичність. За наявності гетероскедастичності оцінки параметрів, отримані МНК, як правило, залишаються незміщеними, обґрунтованими, але неефективними, буде зростати дисперсія оцінок параметрів моделі, яка приводить до збільшення їх стандартних похибок; довірчі інтервали оцінок параметрів моделі також будуть більшими. Як наслідок, F та t-критерії дають неточні результати.

Економетрична модель, якій притаманна гетероскедастичність, є узагальненою моделлю, і для оцінювання її параметрів потрібно скористатися узагальненим методом найменших квадратів. Цей метод реалізовано в пакеті MASS, тому вводимо команду:

```
library("MASS")
```

Будуємо економетричну модель узагальненим методом найменших квадратів, виходячи з гіпотези про те, що дисперсія залишків пропорційна до зміни доходу. Для цього використаємо функцію `rlm()`:

```
fit2<-rlm(Murder~Population+Illiteracy+
Income+Frost,data=states, weight=1/Income)
```

Викликаємо протокол економетричної моделі:

```
summary(fit2)
```

У програмному коді наведно результат розрахунків:

```
Call: rlm(formula = Murder ~ Population +
Illiteracy + Income + Frost,
data = states, weights = 1/Income)
Residuals:
Min 1Q Median 3Q Max
-0.066115   -0.024747   -0.000836   0.022823
0.114906

Coefficients:
Value Std. Error t value
(Intercept) 3.3931 3.9691 0.8549
Population 0.0002 0.0001 2.6146
Illiteracy 3.9024 0.8601 4.5374
Income -0.0003 0.0007 -0.4219
Frost -0.0038 0.0100 -0.3770

Residual standard error: 0.03536 on 45
degrees of freedom
```

У результаті застосування узагальненого методу найменших квадратів, на підставі гіпотези про те, що дисперсія залишків пропорційна до зміни доходу, рівняння регресії виглядає так:

$$\widehat{\text{Murder}} = 3{,}3931 + 0.0002 \cdot \widehat{\text{Population}} + 3.9024 \cdot \widehat{\text{Illiteracy}} - 0.0003 \cdot \widehat{\text{Income}} - 0.0038 \cdot \widehat{\text{Frost}}$$

У випадку наявності в моделі лагових змінних одним з методів її усунення є введення лагових змінних. Зокрема, значення залежної змінної попереднього періоду. Хоча представлена у прикладі модель не містить автокореляції залишків, продемонструємо на її прикладі можливість застосування такої техніки.

Виконаємо формування лагових змінних

```
states$Murder.lag<- with(states, c(NA, Murder[-length(Murder)]))
```

Зауважимо, що для Алабами ці значення не визначено через відсутність значення змінної Murder).

Подивитись на утворений фрейм можна через використання функції: head(states). У програмному коді наведно результат розрахунків:

```
Murder Population Illiteracy Income Frost Murder.lag
Alabama 15.1 3615 2.1 3624 20 NA
Alaska 11.3 365 1.5 6315 152 15.1
Arizona 7.8 2212 1.8 4530 15 11.3
Arkansas 10.1 2110 1.9 3378 65 7.8
California 10.3 21198 1.1 5114 20 10.1
Colorado 6.8 2541 0.7 4884 166 10.3
```

Будуємо економетричну модель залежності рівня злочинності в штатах за чисельністю населення, рівнем неписьменності та значення залежної змінної лагового періоду за допомогою функції lm(), яку викликаємо її протокол за допомогою функції summary(fit3).

```
fit3<-lm(Murder~Population+ Illiteracy+ Murder.lag,data=states)
summary(fit3)
```

У програмному коді наведно результат розрахунків:

```
Call:
lm(formula = Murder ~ Population + Illiteracy + Murder.lag, data = states)

Residuals:
Min 1Q Median 3Q Max
```

```
    -4.825 -1.359 -0.179 1.432 7.021

    Coefficients:
    Estimate Std. Error t value Pr(>|t|)
    (Intercept)    2.801e+00    1.035e+00    2.708
0.00955 **
    Population 2.262e-04 7.731e-05 2.926 0.00537
**
    Illiteracy 3.889e+00 5.805e-01 6.699 2.85e-
08 ***
    Murder.lag   -1.381e-01   9.294e-02   -1.486
0.14424
    ---
    Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
0.05 '.' 0.1 ' ' 1

    Residual standard error: 2.398 on 45 degrees
of freedom
    (1 observation deleted due to missingness)
    Multiple R-squared: 0.5736,   Adjusted    R-
squared: 0.5452
    F-statistic: 20.18 on 3 and 45 DF, p-value:
1.958e-08
```

Хоча в цьому випадку включення до моделі значення залежної змінної лагового періоду як ознаки-фактора не має економічного сенсу, така техніка дозволяє продемонструвати метод позбавлення автокореляції залишків.

Також можна припустити, що рівень неписьменності залежить від доходу населення. З метою перевірки цієї гіпотези доцільно використати метод інструментальних змінних. Економетричне моделювання проводимо за допомогою функції `ivreg ()`, яка знаходиться у бібліотеках пакетах `AER` та `ivpack`. Приклад виконання:

```
    library("AER") # інструментальні змінні
```

```
    library("ivpack") # додаткові опції для
інструментальних змінних
    fit5 <- ivreg(Murder ~ Illiteracy |Income,
data=states)
    summary(fit5)
```

У консолі відобразится результат:

```
    Call:
    ivreg(formula = Murder ~ Illiteracy | Income,
data = states)

    Residuals:
    Min 1Q Median 3Q Max
    -5.3924 -1.8627 -0.1608 1.8167 6.2580

    Coefficients:
    Estimate Std. Error t value Pr(>|t|)
    (Intercept) 3.648 1.758 2.075 0.0433 *
    Illiteracy 3.188 1.466 2.175 0.0346 *
    ---
    Signif. codes: 0 '***' 0.001 '**' 0.01 '*'
0.05 '.' 0.1 ' ' 1

    Residual standard error: 2.733 on 48 degrees
of freedom
    Multiple R-Squared: 0.463,    Adjusted    R-
squared: 0.4518
    Wald test: 4.732 on 1 and 48 DF, p-value:
0.03457
```

Щоб порівняти дві моделі, що побудовані МНК та методом інструментальних змінних, використаємо функцію `mtable()`:

```
compar<- mtable(fit4, fit5)
compar
```

У програмному коді наведено результат розрахунків:

```
fit4: lm(formula = Murder ~ Illiteracy, data
= states)
fit5: ivreg(formula = Murder ~ Illiteracy |
Income, data = states)

=======================================
fit4 fit5
---------------------------------------
(Intercept) 2.397** 3.648*
(0.818) (1.758)
Illiteracy 4.257*** 3.188*
(0.622) (1.466)
---------------------------------------
R-squared 0.494
N 50 50
adj. R-squared 0.452
=======================================
Significance: *** = p < 0.001;
** = p < 0.01;
* = p < 0.05
```

Можна побачити, що значимість оцінок параметрів знизилась, що означає, що МНК має кращі результати, ніж метод інструментальних змінних.

Питання для самоперевірки та контролю знань

1. У чому полягає призначення регресійного аналізу? Які параметри регресійного рівняння можна отримати?

2. Які види регресійних моделей можливо побудувати в системі R? У чому полягають особливості окремих видів регресії?

3. Яким чином можна інтерпретувати регресійні коефіцієнти лінійної та квадратичної регресійної моделі?

4. Охарактеризуйте призначення діаграми ядерної оцінки функції щільності і графік-щітка.

5. За допомогою яких критеріїв можна здійснити оцінку придатності моделі за результатами регресійного аналізу?

6. Охарактеризуйте особливості застосування критерію Дарбіна-Уотсона. У чому полягає його призначення?

7. Для чого будується гістограма залишків Стьюдента? Яким чином її можна інтерпретувати?

8. За допомогою якого тесту можна здійснити оцінку викидів у регресійній моделі?

9. Назвіть методи перетворення змінних у регресійних моделях. Охарактеризуйте кожен з них.

10. У якому випадку доцільно застосовувати видалення змінних у регресійних моделях?

Тестові завдання для самоконтролю та перевірки знань

1. Регресійний аналіз – метод, який дозволяє:

а) встановити залежність між двома змінними у вигляді рівняння;

б) отримати кількісну оцінку тісноти зв'язку між змінними;

в) провести класифікацію вихідних даних;

г) правильна відповідь відсутня.

2. Кореляційний аналіз – метод, який дозволяє:

а) провести класифікацію вихідних даних;

б) встановити залежність між двома змінними у вигляді рівняння;

в) отримати кількісну оцінку тісноти зв'язку між змінними;

г) правильна відповідь відсутня.

3. Рівняння парної регресії доповнюється коефіцієнтом парної кореляції, тому що:

а) нсобхідно знати тісноту зв'язку в лінійній формі;

б) це потрібно для отримання оцінок коефіцієнтів регресії;

в) це необхідно для розрахунку величини залишкової дисперсії;

г) правильна відповідь відсутня.

4. Рівняння регресії – це:

а) формула статистичного зв'язку між змінними;

б) рівняння тісноти зв'язку між змінними;

в) рівняння математичного сподівання випадкової величини;

г) різновид прогресії.

5. Відповідно до змісту регресії фактична величина залежної змінної складається з:

а) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишків;

б) теоретичного значення залежної змінної, знайденого за рівнянням регресії, скоригованого на величину стандартної помилки;

в) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишкової дисперсії;

г) правильна відповідь відсутня.

6. Як називається різниця між фактичним і розрахованим значеннями залежної змінної:

а) різниця;

б) білий шум;

в) залишки;

г) гетероскедастичність.

7. Класичний підхід до оцінювання коефіцієнтів регресії заснований на:

а) методі найменших квадратів;

б) графічній оцінці;

в) методі максимальної правдоподібності;

г) кореляційному аналізі.

8. Суть методу найменших квадратів полягає у:

а) мінімізації суми квадратів відхилень фактичних значень результативної ознаки від теоретичних;

б) мінімізації суми квадратів коефіцієнтів регресії;

в) мінімізації суми квадратів значень залежної змінної;

г) правильна відповідь відсутня.

9. Оцінка параметра β показує:

а) на скільки % зміниться результат при зміні фактора на 1 %;

б) на скільки одиниць зміниться фактор при зміні результату на 1 одиницю;

в) на скільки одиниць зміниться результат при зміні фактора на 1 одиницю;

г) у скільки разів зміниться результат при зміні фактора на одну одиницю.

10. Якщо за однією і тією ж вибіркою розраховані регресії у на х і х на у, то чи збігаються в цьому випадку лінії регресії:

а) так, оскільки оцінки параметрів регресії однакові;

б) так, оскільки оцінки параметрів регресії різні;

в) ні, оскільки оцінки параметрів регресії різні;

г) правильна відповідь відсутня.

ТЕМА 5
МЕТОДИКА ПРОВЕДЕННЯ ДИСПЕРСІЙНОГО АНАЛІЗУ В R

Мета: *провести однофакторний дисперсійний аналіз в середовищі R; систематизувати методику здійснення однофакторного коваріаційного аналізу в R; здійснити багатофакторний дисперсійний аналіз в середовищі R.*

План

5.1. Однофакторний дисперсійний аналіз в середовищі R.

5.2. Методика здійснення однофакторного коваріаційного аналізу в системі R.

5.3. Багатофакторний дисперсійний аналіз в середовищі R.

Ключові поняття та терміни: *дисперсія, дисперсійний аналіз, однофакторний дисперсійний аналіз, однофакторний коваріаційний аналіз з однією коваріатою, двофакторний дисперсійний аналіз, двофакторний коваріаційний аналіз з двома коваріатами.*

Основні функції та програмні коди: `aov()`, `y~A`, `y~x+A`, `y~A*B`, `y~x1+x2+A*B`, `y^B+A`, `y~A+Error(Об'єкт/A)`, `y~B*W+Error(Об'єкт/W)`, `plotmeans()`, `glht()`.

5.1. Однофакторний дисперсійний аналіз в середовищі R

Хоча методології ANOVA і регресії розвивалися окремо, з функціональної точки зору, це два окремих випадки загальної лінійної моделі. Можемо аналізувати моделі ANOVA за допомогою тієї ж функції `lm()`, яка використовувалася для регресійного аналізу в попередній темі. Проте в цій темі спочатку доцільно розглянути функцію `aov()`. Результати застосування

цих функцій рівнозначні, але `aov()` представляє результати в більш звичному форматі.

Синтаксис застосування функції `aov()` такий: `aov(формула,data=таблиця_даних)`. Нижче перераховані спеціальні символи, які можуть бути використані в формулах. У цій таблиці залежна змінна позначена як y, а літери A, B і C відповідають факторам.

– `y~A` – однофакторний дисперсійний аналіз;

– `y~x+A` – однофакторний коваріаційний аналіз з однією коваріатою;

– `y~A*B` – двофакторний дисперсійний аналіз;

– `y~x1+x2+A*B` – двофакторний коваріаційний аналіз з двома коваріатами;

– `y^B+A` – випадковий блоковий;

– `y~A+Error(Об'єкт/A)` – однофакторний дисперсійний аналіз для залежних змінних;

– `y~B*W+Error(Об'єкт/W)` – дисперсійний аналіз з повторними змінами з одним внутрішньогруповим фактором (W) і одним міжгруповим фактором (B).

Наведені формули для декількох найбільш затребуваних варіантів дослідження. Малі літери – це кількісні змінні, великі літери – це групуючі фактори, а об'єкт – це змінна, яка містить унікальні ідентифікатори для об'єктів.

Порядок членів у формулі важливий, якщо є більше одного фактора і дизайн незбалансований або є коваріати. Якщо одна з цих двох умов дотримана, змінні з правого боку рівняння будуть корельовані одна з одною. У такому випадку не існує однозначного способу розділити їх вплив на залежну змінну.

Наприклад, при двофакторному дисперсійному аналізі з нерівним числом спостережень для різних комбінацій факторів модель `y~A*B` не дасть того ж результату, що і модель `y~B*A`.

За замовчуванням у R використовується послідовний підхід (I типу) для обчислення ефектів ANOVA. Перша модель може бути сформульована як `y~A+B+A:B`. Підсумкова таблиця дисперсійного аналізу в R буде оцінювати:

– вплив A на y;

– вплив B на y при постійних значеннях A;

– взаємодію A і B при постійних значеннях головних ефектів A і B.

Чим більша різниця в розмірах вибірок, тим більший вплив порядку членів рівняння на результати. У загальному випадку більш важливі ефекти повинні стояти у формулі раніше. Зокрема, спочатку потрібно вказувати коваріати, потім головні ефекти, потім парні взаємодії, далі – взаємодії трьох змінних тощо. З головних ефектів у першу чергу потрібно вказувати найбільш суттєві. Ось головна ідея: якщо дизайн дослідження не ортогональний (тобто чинники і / або коваріати корельовані), потрібно бути обережними при визначенні порядку ефектів.

При однофакторному дисперсійному аналізі існує інтерес у порівнянні середніх значень залежної змінної для двох і більше груп, заданих категоріальною групуючою змінною. Наступний приклад заснований на наборі запозичених з публікації зміст Westfall, Tobia, Rom, & Hochberg даних `cholesterol` з пакета `multcomp`. П'ятдесят пацієнтів проходили один з п'яти курсів лікування, спрямовані на зниження рівня холестерину (змінна `trt`). Три з цих курсів полягали у використанні одного і того ж препарату в кількості 20 мг один раз на день (`1time`), в кількості 10 мг двічі на день (`2times`) або 5 мг чотири рази на день (`4times`). Два інших курси лікування полягали в прийомі альтернативних препаратів (`drugD` і `drugE`). Який з курсів лікування привів до найбільш помітного зниження рівня холестерину (змінна `response`)? Проведений аналіз поданий нижче.

```
> library(multcomp)
> attach(cholesterol)
> table(trt)
trt
1time 2times 4times drugD drugE
10 10 10 10 10
> aggregate(response,by=list(trt),FUN=mean)
```

```
Group.1 x
1 1time 5.78197
2 2times 9.22497
3 4times 12.37478
4 drugD 15.36117
5 drugE 20.94752
> aggregate(response,by=list(trt),FUN=sd)
Group.1 x
1 1time 2.878113
2 2times 3.483054
3 4times 2.923119
4 drugD 3.454636
5 drugE 3.345003
> fit <- aov(response~trt)
> summary(fit)
Df Sum Sq Mean Sq F value Pr(>F)
trt 4 1351.4 337.8 32.43 9.82e-13
Residuals 45 468.8 10.4

trt ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> library(gplots)
> plotmeans(response~trt)
> detach(cholesterol)
```

За виведеними на екран результатами можна бачити, що кожен спосіб лікування був випробуваний на 10 пацієнтах. Порівняння середніх значень показує, що препарат E (drugE) дозволяє домогтися найбільш помітного зниження рівня холестерину, а протокол time1 найменш ефективний. Стандартні відхилення у всіх п'яти групах були приблизно однаковими, коливаючись від 2,88 до 3,48. F-тест для типу лікування (trt) був статистично значущим ($p < 0,0001$), що свідчить про нсоднакову сфсктивність цих п'яти способів лікування.

За допомогою функції `plotmeans()` з пакета `gplots` можна побудувати діаграму, що відображає середні значення та їх довірчі інтервали для кожної групи. Така діаграма з 95 % довірчими інтервалами зображена на рис. 5.1. На ній добре видно міжгрупові відмінності.

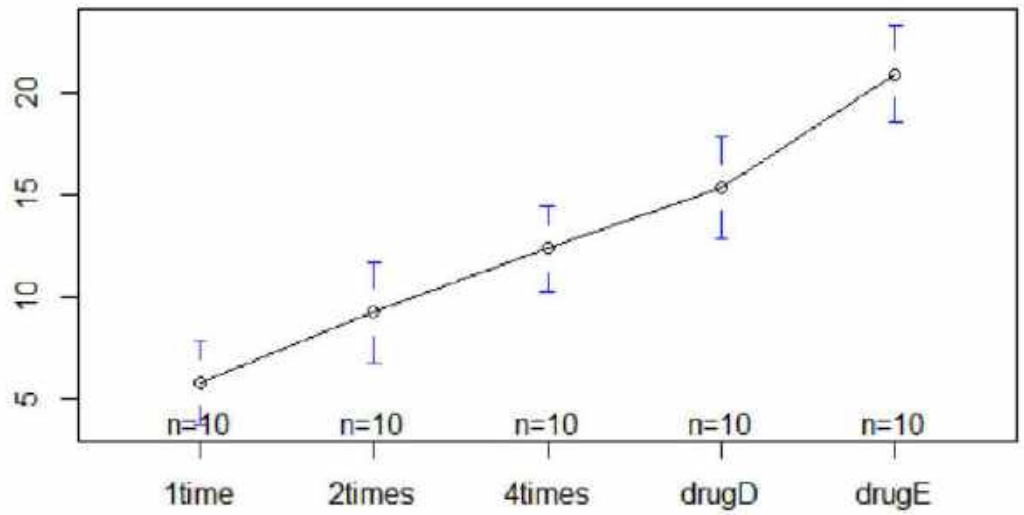


Рис. 5.1. Середні значення та 95 % довірчі інтервали курсів лікування

Результат F-тесту для способу лікування свідчить, що п'ять випробуваних методів лікування неоднаково ефективні, проте звідси не можна зрозуміти, які саме способи розрізняються між собою. Для відповіді на це питання можна використовувати множинні порівняння. Наприклад, функція `TukeyHSD()` дозволяє провести тест на попарні різниці між середніми значеннями для всіх груп. Функція `glht()` з пакета `multcomp` реалізує набір більш складних методів порівняння декількох середніх значень. Ці методи можна використовувати як для лінійних моделей, так і для узагальнених лінійних моделей. За допомогою наведеного нижче програмного коду можна провести тест чесних оцінок достовірності відмінностей Тьюкі (Tukey honest significant differences test) і графічно представити його результати (рис. 5.2).

```
> plotmeans(response~trt)
> detach(cholesterol)
> library(multcomp)
```

```
> par(mar=c(5,4,6,2))
> tuk <- glht(fit,linfct=mcp(trt="Tukey"))
> plot(cld(tuk,level=.05),col="lightgrey")
```

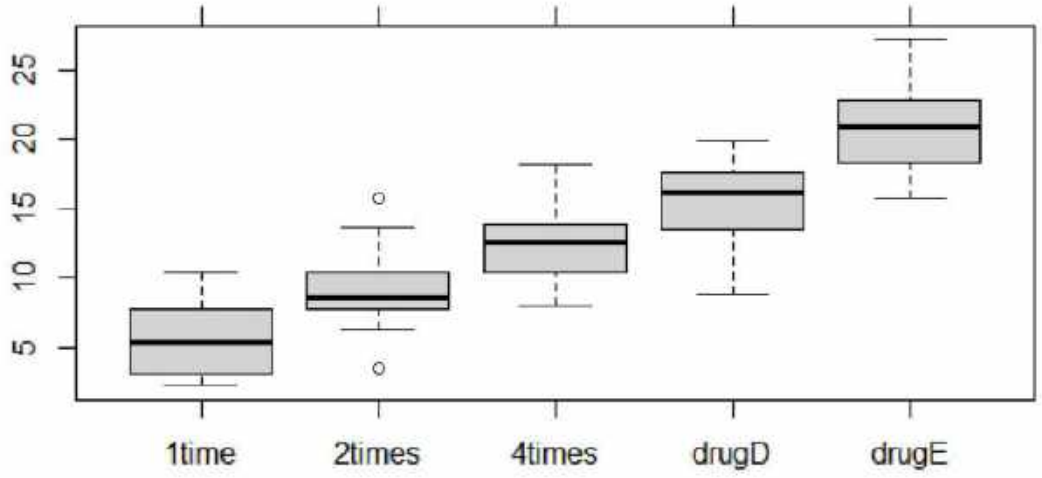


Рис. 5.2. Діаграма результатів тесту Тьюкі

У цьому програмному коді висота верхнього поля діаграми була збільшена за допомогою параметра `par`. Параметр `level` функції `cld()` дозволяє задати рівень значущості (у даному випадку 0,05 або 95% впевненості).

Групи (які подані у вигляді діаграм розмахів) позначені однаковими буквами, якщо їх середні значення статистично не розрізняються. Звідси можна бачити, що протоколи лікування `1time` і `2times` не розрізняються. Протоколи `2times` і `4times` також не розрізняються.

Із цих результатів видно, що краще приймати ліки, що знижують рівень холестерину, по 5 мг чотири рази на день, ніж по 20 мг один раз на день. Альтернативний препарат drugD не більш ефективний порівняно з чотириразовим прийомом першого препарату. Однак ліки drugE діють краще, ніж всі інші методи лікування. Методологія множинних порівнянь – це складна і швидкозмінна область досліджень.

5.2. Методика здійснення однофакторного коваріаційного аналізу в системі R

Однофакторний коваріаційний аналіз розширює однофакторний дисперсійний аналіз додаванням однієї або декількох кількісних коваріат. Наведений приклад заснований на наборі даних `litter` з пакета `multcomp`. Вагітні миші були розділені на чотири експериментальні групи. Мишам з кожної групи давали різні дози ліків (0, 5, 50 або 500). Середня вага новонароджених мишенят кожного посліду була залежною змінною, а час вагітності було включено в модель як коваріата. Проведений аналіз поданий у такому програмному коді:

```
> data(litter,package ="multcomp")
> attach(litter)
> table(dose)
dose
 0 5 50 500
 20 19 18 17
> aggregate(weight,by=list(dose),FUN=mean)
 Group.1 x
1 0 32.30850
2 5 29.30842
3 50 29.86611
4 500 29.64647
> fit<-aov(weight~gesttime+dose)
> summary(fit)
 Df Sum Sq Mean Sq F value Pr(>F)
gesttime 1 134.3 134.30 8.049 0.00597 **
dose 3 137.1 45.71 2.739 0.04988 *
Residuals 69 1151.3 16.69
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

За допомогою функції `table()` можна дізнатися, що миші з різних груп отримували різні дози препарату: 20 тварин не піддавали дії ліків, а 17 мишей отримали 500 одиниць препарату. На

підставі середніх значень по групах можна зазначити, що для мишей, які одержували ліки, у цілому характерна найбільша вага потомства (32,3). Результати F-тесту свідчать про те, що термін вагітності пов'язаний з вагою мишенят при народженні і кількість ліків впливали на вагу мишенят при постійних значеннях терміну вагітності. Середня вага мишенят з різних експериментальних груп є неоднаковою при постійних значеннях терміну вагітності.

Оскільки використовуємо коваріату, можна захотіти з'ясувати скориговані середні по групах – за вирахуванням ефекту коваріати. Для обчислення скоригованих середніх значень можна використовувати функцію `effect()` з пакета `effects`:

```
> library(effects)
> effect("dose",fit)

dose effect
dose
0 5 50 500
32.35367 28.87672 30.56614 29.33460
```

У цьому випадку скориговані середні значення подібні зі звичайними середніми, які обчислюються за допомогою функції `aggregate()`, але це не завжди буває так. У пакеті `effects` реалізований потужний метод обчислення та графічного представлення скоригованих середніх значень для досліджень зі складним дизайном. Більш детальна інформація наведена в інформаційній базі `CRAN`.

Як і в прикладі з однофакторним дисперсійним аналізом з попереднього питання, результати F-тесту для дози препарату, повідомлялось, що середня вага мишенят у різних експериментальних групах неоднакова, проте не зрозуміло, які саме пари середніх значень розрізняються. Знову потрібно використовувати множинні попарні порівняння середніх значень за допомогою пакета `multcomp`. Крім того, цей пакет можна використовувати для перевірки певних користувальницьких гіпотез про середні значення.

Припустимо, потрібно дізнатися, чи відрізняється група, яка не отримувала ліки, від тих груп, яким давали три різні дози препарату. Для перевірки цієї гіпотези можна використовувати наведений нижче програмний код:

```
> library(multcomp)
> contrast <- rbind("no drug vs.drug"=c(3,-1,-1,-1))
>
summary(glht(fit,linfct=mcp(dose=contrast)))

    Simultaneous Tests for General Linear Hypotheses

    Multiple Comparisons of Means: User-defined Contrasts


    Fit: aov(formula = weight ~ gesttime + dose)

    Linear Hypotheses:
    Estimate Std. Error t value
    no drug vs.drug == 0 8.284 3.209 2.581
    Pr(>|t|)
    no drug vs.drug == 0 0.012 *
    ---
    Signif. codes:
    0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Контраст `c(3,-1,-1,-1)` означає порівняння першої групи з усередненими по трьох інших групах значеннями. Гіпотеза перевіряється за допомогою t-статистики (її значення в цьому випадку дорівнює 2,581). Отже, можна стверджувати, що група, яка не отримувала ліків, характеризується великою вагою новонароджених мишенят, ніж всі інші експериментальні групи. Інші контрасти можна додавати за допомогою команди `rbind()` для отримання більш детальної інформації.

5.3. Багатофакторний дисперсійний аналіз у середовищі R

При двофакторному дисперсійному аналізі об'єкти розподіляють за групами, які задаються комбінацією двох факторів.

Розглянемо двофакторний дисперсійний аналіз для залежних змінних на прикладі набору даних `ToothGrowth`, що представлений з базовою версією програми. Шістдесят морських свинок були випадково розподілені по групах, які отримували три різних кількості аскорбінової кислоти (`dose`: 0,5, 1 або 2 мг) двома способами (`supp`: апельсиновий сік – `OJ` або вітамін С – `vc`) – по 10 свинок у кожній з шести груп. Залежна змінна – це довжина зубів (`len`). Нижче наведено програмний код для аналізу:

```
> attach(ToothGrowth)
> table(supp,dose)
dose
supp 0.5 1 2
OJ 10 10 10
VC 10 10 10
>
aggregate(len,by=list(supp,dose),FUN=mean)
Group.1 Group.2 x
1 OJ 0.5 13.23
2 VC 0.5 7.98
3 OJ 1.0 22.70
4 VC 1.0 16.77
5 OJ 2.0 26.06
6 VC 2.0 26.14
> aggregate(len,by=list(supp,dose),FUN=sd)
Group.1 Group.2 x
1 OJ 0.5 4.459709
2 VC 0.5 2.746634
3 OJ 1.0 3.910953
4 VC 1.0 2.515309
5 OJ 2.0 2.655058
6 VC 2.0 4.797731
```

```
> fit <- aov(len~supp*dose)
> summary(fit)
Df Sum Sq Mean Sq F value Pr(>F)
supp 1 205.4 205.4 12.317 0.000894
dose 1 2224.3 2224.3 133.415 < 2e-16
supp:dose 1 88.9 88.9 5.333 0.024631
Residuals 56 933.6 16.7

supp ***
dose ***
supp:dose *
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Результат застосування команди `table` вказує, що дизайн експерименту збалансований (однакове число об'єктів у всіх експериментальних групах), а функція `aggregate` дозволяє обчислити середні та стандартні відхилення для кожної групи. Результати дисперсійного аналізу, подані у вигляді таблиці за допомогою функції `summary`, вказують на те, що обидва головних – ефект (`supp` і `dose`) і взаємодія між ними – значимі.

Ці результати можна візуалізувати кількома способами. Взаємодію між факторами можна проілюструвати за допомогою функції `interaction.plot()`. Програмний код такий:

```
> interaction.plot(dose,supp,len,type="b",
+ col=c("red","blue"),pch=c(16,18))
```

Отримана діаграма подана на рис. 5.3. На ній зображена середня довжина зубів при всіх трьох дозах аскорбінової кислоти і двох способах її отримання.

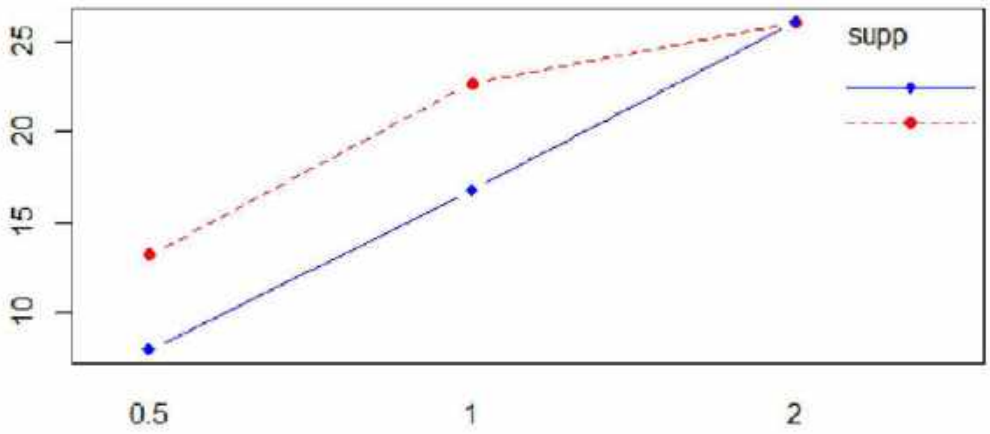


Рис. 5.3. Графічне відображення середніх значень, отримане за допомогою функції `interaction.plot()`

Також можна побудувати діаграму взаємодій за допомогою команди `plotmeans()` з пакета `gplots`. Наведений нижче програмний код

```
> interaction.plot(dose,supp,len,type="b",
+ col=c("red","blue"),pch=c(16,18))
> library(gplots)
>
plotmeans(len~interaction(supp,dose,sep=""),
+ connect=list(c(1,3,5),c(2,4,6)),
+ col=c("red","darkgreen"))
```

дозволяє отримати діаграму, подану на рис. 5.4.

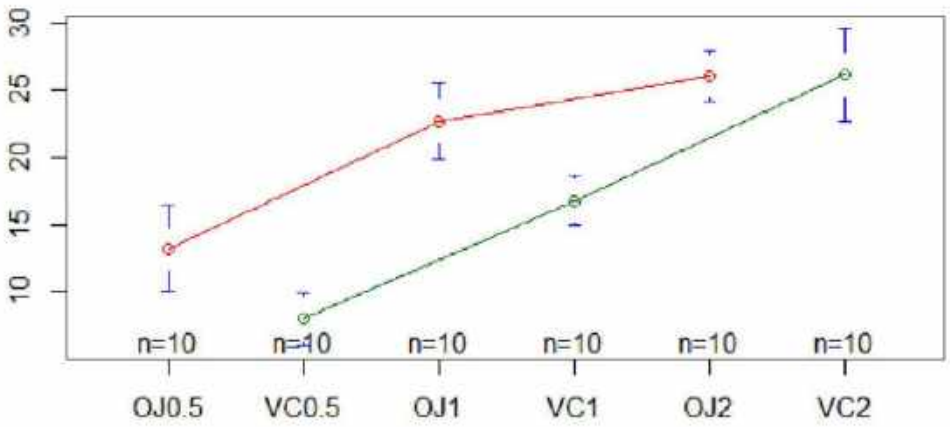


Рис. 5.4. Графічне відображення середніх значень з 95 % довірчими інтервалами, отримане за допомогою функції `plotmeans()`

На цій діаграмі відображені середні значення поряд з діапазонами точності (95 % довірчими інтервалами) і розмірами вибірок.

Можна досліджувати кілька залежних змінних одночасно за допомогою багатовимірного дисперсійного аналізу (multivariate analysis of variance, MANOVA). Розглянемо цей підхід на прикладі набору даних `uscereal` з пакета `MASS`. У цьому прикладі дізнаємося, як залежить вміст калорій (`calories`), жирів (`fat`) і вуглеводів (`sugars`) в американських сухих сніданках від того, на якій полиці в магазині вони знаходяться (`shelf`: 1 нижня полиця, 2 – середня, 3 – верхня). Кількість калорій, жирів і вуглеводів – це залежні змінні, а номер полиці – незалежна змінна з трьома рівнями (1,2 і 3). Аналіз поданий у наступному програмному коді:

```
> library(MASS)
> attach(UScereal)
> y <- cbind(calories,fat,sugars)
> aggregate(y,by=list(shelf),FUN=mean)
 Group.1 calories fat sugars
1 1 119.4774 0.6621338 6.295493
2 2 129.8162 1.3413488 12.507670
3 3 180.1466 1.9449071 10.856821
> cov(y)
calories fat sugars
calories 3895.24210 60.674383 180.380317
fat 60.67438 2.713399 3.995474
sugars 180.38032 3.995474 34.050018
> fit <- manova(y~shelf)
> summary(fit)
 Df Pillai approx F num Df den Df
shelf 1 0.19594 4.955 3 61
Residuals 63
Pr(>F)
shelf 0.00383 **
Residuals
---
Signif. codes:
```

```
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
> summary.aov(fit)
Response calories :
Df Sum Sq Mean Sq F value Pr(>F)
shelf 1 45313 45313 13.995 0.0003983
Residuals 63 203982 3238

shelf ***
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response fat :
Df Sum Sq Mean Sq F value Pr(>F)
shelf 1 18.421 18.4214 7.476 0.008108
Residuals 63 155.236 2.4641

shelf **
Residuals
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Response sugars :
Df Sum Sq Mean Sq F value Pr(>F)
shelf 1 183.34 183.34 5.787 0.01909 *
Residuals 63 1995.87 31.68
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

У цьому програмному коді для створення матриці трьох залежних змінних (калорії, жири, вуглеводи) використана функція `cbind()`. За допомогою функції `aggregate()` розраховуються середні значення для полиць, а функція `cov()` обчислює дисперсію і коваріацію зі злаків.

Функція `manova()` здійснює багатовимірний тест на міжгрупові відмінності. Статистично значуща величина F-статистики свідчить про те, що три групи злаків розрізняються за поживною цінністю.

Оскільки результати багатовимірного тесту значимі, можна використовувати функцію `summary.aov()` для виконання ряду однофакторних дисперсійних аналізів. У цьому випадку відмінності за кожним показником харчової цінності аналізуються окремо.

Питання для самоперевірки та контролю знань

1. Що розуміють під однофакторним дисперсійним аналізом? За яких умов він застосовується?

2. У яких умовах використовується однофакторний коваріаційний аналіз з однією коваріатою?

3. Яким чином можна інтерпретувати графік середніх значень та 95%-их довірчих інтервалів?

4. Для чого застосовується тест Тьюкі? Яка процедура його застосування та синтаксис програмного коду?

5. Охарактеризуйте особливості застосування однофакторного коваріаційного аналізу з однією коваріатою.

6. Яка процедура застосування двофакторного дисперсійного аналізу в середовищі R?

7. Для чого здійснюють графічне відображення середніх значень з 95 % довірчими інтервалами?

8. Що розуміють під множинним дисперсійним аналізом? У яких умовах він застосовується?

9. Охарактеризуйте дисперсійний аналіз з повторними змінами з одним внутрішньогруповим фактором і одним міжгруповим фактором.

10. Охарактеризуйте однофакторний дисперсійний аналіз для залежних змінних.

Тестові завдання для самоконтролю та перевірки знань

1. Для оцінки значущості оцінок параметрів регресії розраховують:

а) t-статистику Стьюдента;

б) F-критерій Фішера;
в) коефіцієнт детермінації;
г) коефіцієнт кореляції.

2. Як називається інтервал, в якому з наперед заданою ймовірністю знаходиться істинне значення параметра:
а) довірчий інтервал;
б) предикативний інтервал;
в) дескриптивний інтервал;
г) правильна відповідь відсутня.

3. F-критерій характеризує:
а) співвідношення факторної і залишкової дисперсій;
б) частку факторної дисперсії в загальній дисперсії результативної ознаки;
в) частку залишкової дисперсії в загальній дисперсії результативної ознаки;
г) частку залишкової дисперсії в загальній дисперсії факторної ознаки.

4. Команда, яка використовується в R для діагностики моделі на наявність гетероскедастичності на основі тесту Гольдфельда-Квандта:
а) qf;
б) mean;
в) predict;
г) gqtest.

5. Залишкова сума квадратів відхилень відображає вплив на відхилення y:
а) досліджуваного фактора x;
б) неврахованих у моделі факторів;
в) досліджуваного фактора x та інших факторів;
г) правильна відповідь відсутня.

6. Залишкова сума квадратів рівна нулю в тому разі, коли:
а) y пов'язаний з x функціонально;

б) значення у, розраховані за рівнянням регресії, рівні середньому значенню у;

в) уся загальна дисперсія обумовлена впливом інших факторів;

г) правильна відповідь відсутня.

7. Загальна сума квадратів відхилень збігається із залишковою, коли:

а) фактор х не впливає впливу на результат;

б) інші чинники не впливають на результат;

в) фактор х та інші фактори однаковою мірою впливають на результат;

г) правильна відповідь відсутня.

8. Вибірковою коваріацією двох змінних називається:

а) середня величина добутку відхилень цих змінних від своїх середніх;

б) сума добутків відхилень цих змінних від своїх середніх;

в) математичне сподівання добутку відхилень цих величин від своїх середніх значень;

г) правильна відповідь відсутня.

9. Як називається інтервал, в якому з наперед заданою ймовірністю знаходиться істинне значення параметра:

а) довірчий інтервал;

б) предикативний інтервал;

в) дескриптивний інтервал;

г) правильна відповідь відсутня.

10. Пояснена сума квадратів відхилень відображає вплив на відхилення у:

а) досліджуваного фактора х;

б) неврахованих в моделі факторів;

в) досліджуваного фактора х та інших факторів;

г) правильна відповідь відсутня.

ЧАСТИНА 2

Лабораторний практикум

ЛАБОРАТОРНА РОБОТА 1	Побудова, дослідження та аналіз лінійної багатофакторної економетричної моделі
ЛАБОРАТОРНА РОБОТА 2	Побудова та дослідження нелінійної багатофакторної економетричної моделі на прикладі неокласичної виробничої функції Кобба-Дугласа та її використання для аналізу та прогнозування процесу виробництва
ЛАБОРАТОРНА РОБОТА 3	Побудова та дослідження лінійної багатофакторної економетричної моделі з мільтиколіанерністю
ЛАБОРАТОРНА РОБОТА 4	Побудова та дослідження лінійної багатофакторної економетричної моделі з гетероскедастичністю
ЛАБОРАТОРНА РОБОТА 5	Побудова, дослідження лінійної багатофакторної економетричної моделі з автокореляцією залишків та моделі з інструментальними змінними

ЛАБОРАТОРНА РОБОТА 1
ПОБУДОВА, ДОСЛІДЖЕННЯ ТА АНАЛІЗ ЛІНІЙНОЇ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ

Мета: набуття практичних навичок побудови багатофакторної економетричної моделі, її дослідження та практичне використання.

Задачі роботи

1. Специфікація економетричної моделі.
2. Оцінювання параметрів моделі, складання рівняння регресії та економічна інтерпретація параметрів моделі.
3. Перевірка значимості оцінок параметрів та адекватності моделі в цілому.
4. Прогнозування.
5. Перевірка моделі щодо виконання передумов застосування МНК.
6. Формування економічних висновків на основі побудованої економетричної моделі.

Завдання роботи

На підставі даних варіанта для однієї з країн ОЕСР побудувати та дослідити двофакторну економетричну модель залежності рівня середньорічної заробітної плати (Y, одиниць національної валюти) від ВВП (X_1, млн одиниць національної валюти) та продуктивності праці (X_2, одиниць національної валюти на 1 зайнятого) для цього потрібно:

1. Побудувати діаграми розсіювання для результуючого показника та показників-факторів.
2. Розрахувати оцінки параметрів моделі методом найменших квадратів. Скласти рівняння регресії та надати економічну інтерпретацію оцінкам параметрів моделі.
3. Знайти довірчі інтервали параметрів моделі, β_1, β_2, , якщо $\alpha = 0{,}05$.

4. Перевірити на значущість оцінки параметрів моделі, якщо $\alpha = 0{,}05$.

5. Поясність значення коефіцієнта детермінації та скоригованого коефіцієнта детермінації.

6. Провести *F*-тест оціненої моделі на адекватність, якщо $\alpha = 0{,}05$.

7. Знайти точкову та інтервальну оцінку індивідуального прогнозного значення за умови, що показники-фактори будуть більшими від середнього їх значення на $(k+N)/100$, а $\alpha = 0{,}05$. Де k – номер групи $+N$ – номер студента за списком.

8. Провести діагностику моделі на наявність гетероскедастичності на основі тесту Гольдфельда-Квандта.

9. Провести діагностику автокореляції залишків за тестом Дарбіна-Уотсона та Бреуша-Годфрі, якщо $\alpha = 0{,}05$.

10. Провести діагностику багатофакторної економетричної моделі на наявність мультиколінеарності.

Інструкції до виконання завдань

Першим кроком до виконання лабораторної роботи є формування робочої папки. Створіть папку з короткою назвою, наприклад Lr-1. Далі сформуйте масив вихідних даних. Для цього відкрийте таблицю MS Excel та наберіть дані свого варіанта з короткими заголовками, використовуючи як десятковий розділювач (розділювача цілої та дробової частки) крапку, а не кому, оскільки система статистичного аналізу R сприймає саме цей розділювач. Не використовуйте функцію розділювача розрядів (рис. 1.1).

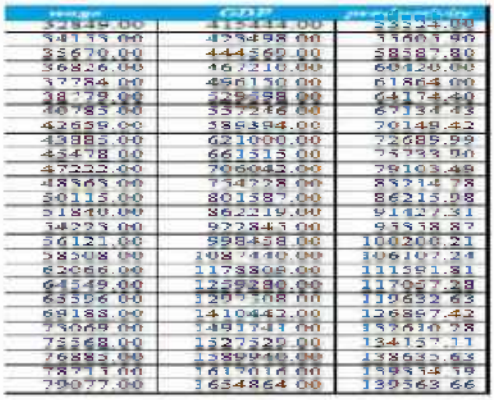
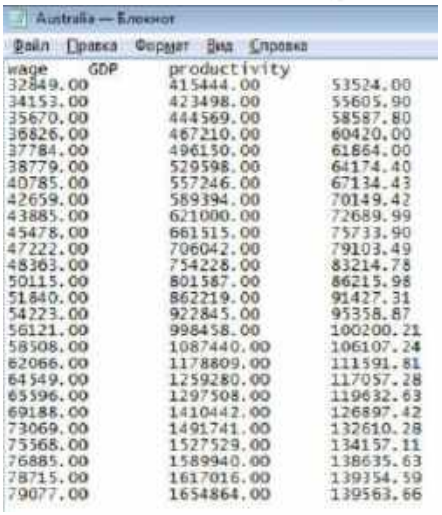


Рис. 1.1. Формування вихідного масиву інформації в MS Excel

Після формування вихідного масиву інформації в MS Excel скопіюйте дані з заголовками та вставте їх у файл програми *Блокнот* (рис. 1.2). Або збережіть файл MS Excel у форматі *.txt *(текстовый файл с разделителями табуляции)* (рис. 1.3).


Australia — Блокнот

Файл Правка Формат Вид Справка

wage	GDP	productivity	
32849.00		415444.00	53524.00
34153.00		423498.00	55605.90
35670.00		444569.00	58587.80
36826.00		467210.00	60420.00
37784.00		496150.00	61864.00
38779.00		529598.00	64174.40
40785.00		557246.00	67134.43
42659.00		589394.00	70149.42
43885.00		621000.00	72689.99
45478.00		661515.00	75733.90
47222.00		706042.00	79103.49
48363.00		754228.00	83214.78
50115.00		801587.00	86215.98
51840.00		862219.00	91427.31
54223.00		922845.00	95358.87
56121.00		998458.00	100200.21
58508.00		1087440.00	106107.24
62066.00		1178809.00	111591.81
64549.00		1259280.00	117057.28
65596.00		1297508.00	119632.63
69188.00		1410442.00	126897.42
73069.00		1491741.00	132610.28
75568.00		1527529.00	134157.11
76885.00		1589940.00	138635.63
78715.00		1617016.00	139354.59
79077.00		1654864.00	139563.66

Рис. 1.2. Формування вихідного масиву інформації в Блокноті

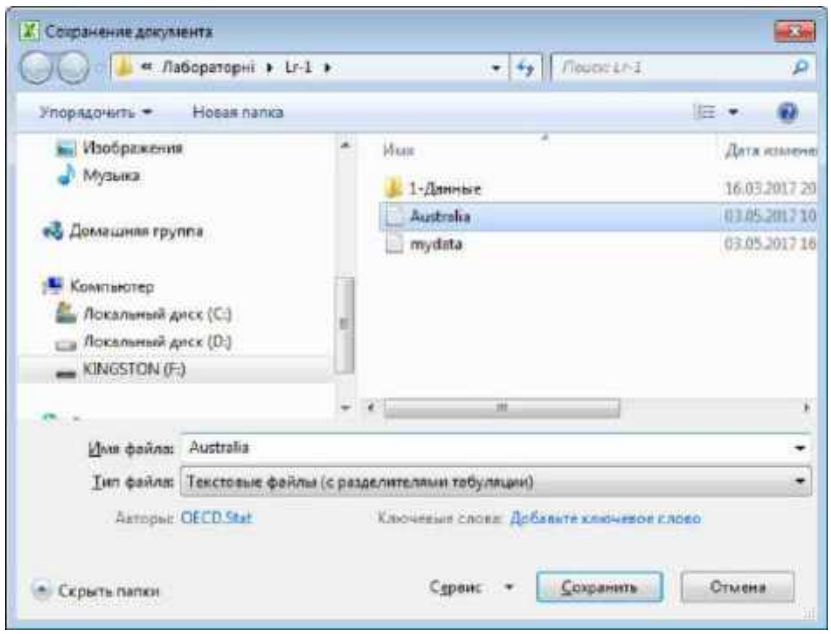


Рис. 1.3. Збереження файлу MS Excel у форматі *.txt

Після формування файлу з розширенням *txt відкрийте R та встановіть робочу папку. Для цього зайдіть у вкладку *Session* оберіть *Set Working directory* далі натисніть кнопку *Choose Directory* після цього вкажіть шлях до вашої папки (рис. 1.4).

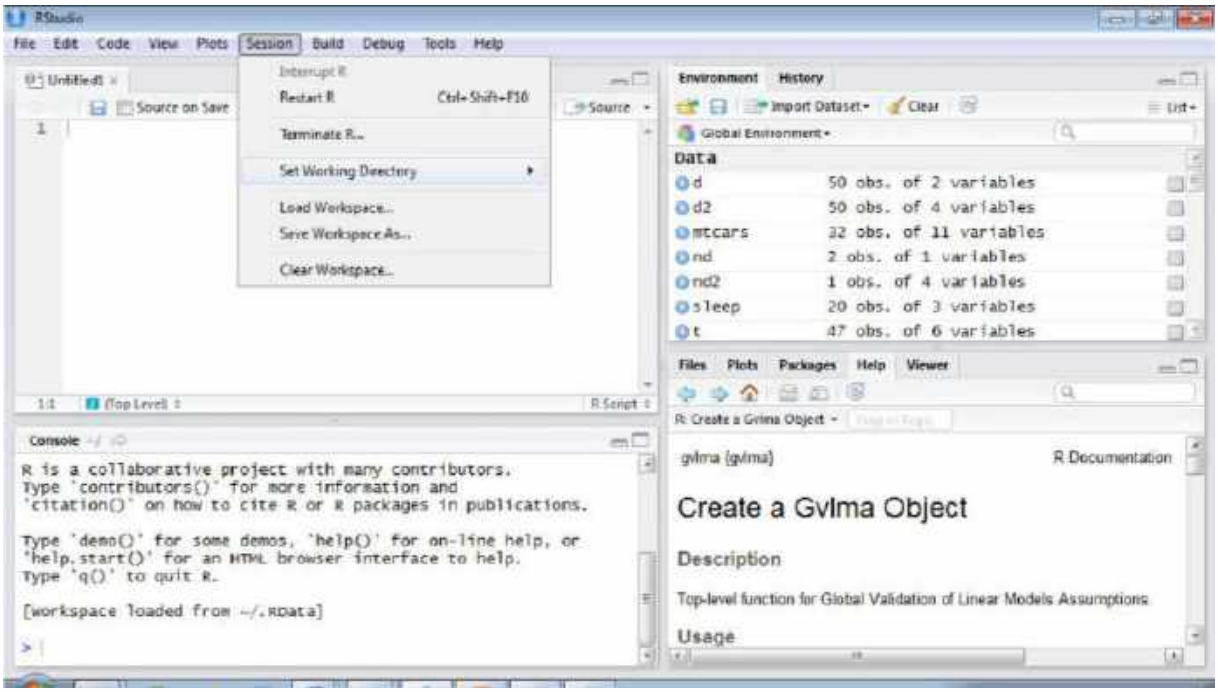


Рис. 1.4. Обрання робочої папки в R

Після обрання робочої папки ви повинні загрузити пакети, які будуть вам потрібні для виконання лабораторної роботи 1, для цього в робочому середовищі *R* наберіть такі команди:

```
library("dplyr") # маніпуляції з даними
library("psych") # описові статистики
library("lmtest") # тестування гіпотез в лінійних моделях
library("sjPlot") # графіки
library("sgof")#
library("ggplot2") # графіки
library("foreign") # завантаження даних в різних форматах
library("hexbin") # графіки
```

Якщо при копіюванні команд та вставці у робочу зону текст українською мовою відображається символами, які неможливо прочитати, необхідно у вкладці *File* натиснути кнопку *Reopen with encoding* та обрати кодування *UTF-8.* Розрахунки, які проводяться в інтегрованому середовищі *RStudio,* можна супроводжувати вашими власними коментарями, які повинні починатись із символу «#».

Після того, як введені дані команди, послідовно ставте курсор на кожну з них та натискайте клавіші *Ctrl+Enter*. На консолі ви побачите, як завантажуються пакети (рис. 1.5). Так ви активуєте всі пакети. Якщо на екрані з'являються червоні написи, це не означає, що в назвах пакетів є помилки. Ознакою помилки є тільки явне повідомлення *Error*. Загружати потрібні для проведення розрахунків бібліотеки з командами необхідно кожного разу перед початком роботи в інтегрованому середовищі.

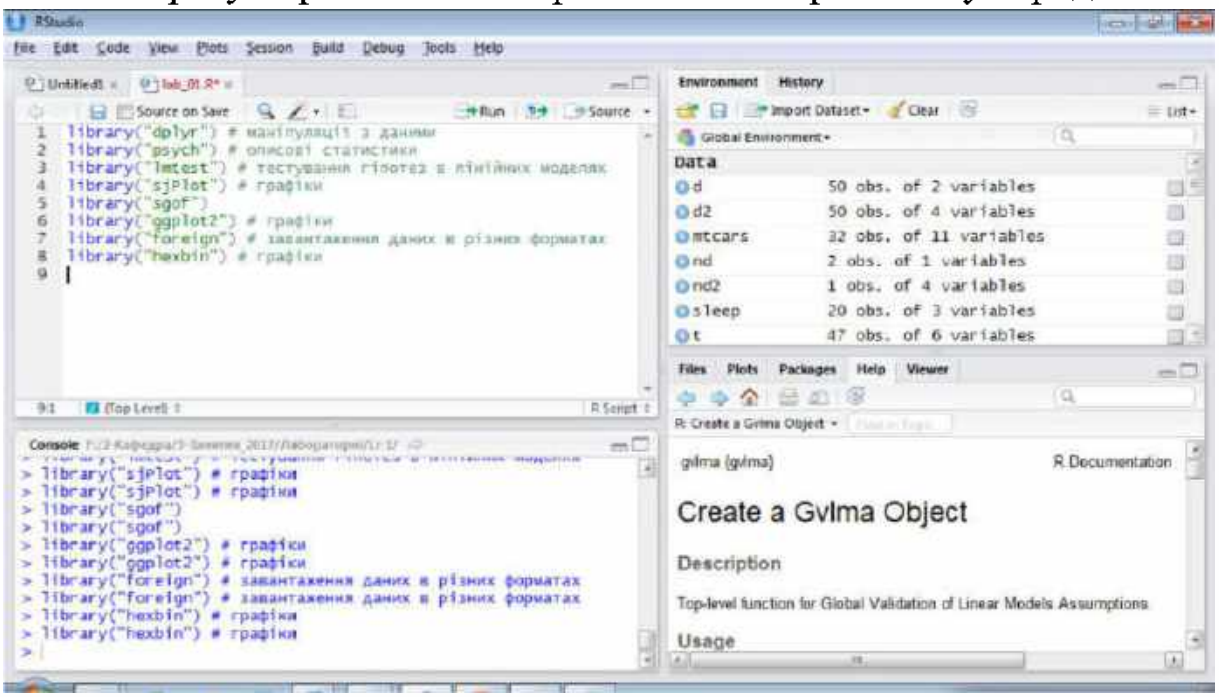


Рис. 1.5. Завантаження необхідних пакетів у R

Після того, як пакети активовано, вам необхідно завантажити підготовлений заздалегідь файл з даними (рис. 1.6). Для цього оберіть ім'я таблиці, в якій буде зберігатись масив інформації, та напишіть його в робочому середовищі, наприклад *t,* далі поставте знак менше і коротку рисочку «<-» (використовується замість «=» в R) і наберіть команду *read.csv*. Після цього відкрийте дужки, поставте лапки, помістіть курсор всередині них, натисніть клавішу *Tab* і оберіть файл з даними. Ця команда призначена для завантаження файлів у форматі **.csv,* а оскільки потрібно завантажити файл у форматі **.txt,* далі потрібно його описати. Через кому напишіть, який використовується розділювач, у файлі **txt* – табуляція, тому наберіть *sep="\t"*. Потім через кому напишіть, який використовується десятковий розділювач, у файлі

з даними розділювачем є крапка, тому пишіть *dec=“.”*. Останнє, що потрібно задати, чи використовуються заголовки, оскільки заголовки у файлі пишіть: *header= TRUE*. Приклад:

```
t<-read.csv(“Australia.txt”, sep=“\t”, dec=“.”, header= TRUE)
```

Далі натискайте клавіші *Ctrl+Enter*.

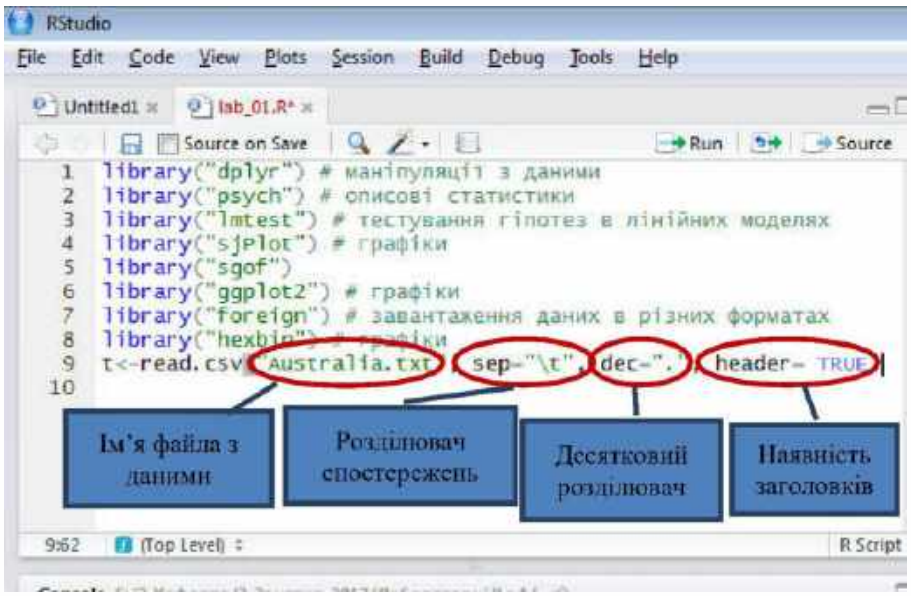


Рис. 1.6. Завантаження файлу з даними у форматі *txt в *R*

На набір вихідних даних можна подивитись, натиснувши на кнопку *t* у вікні глобального середовища. Таблиця з даними має виглядати як на рис. 1.7.

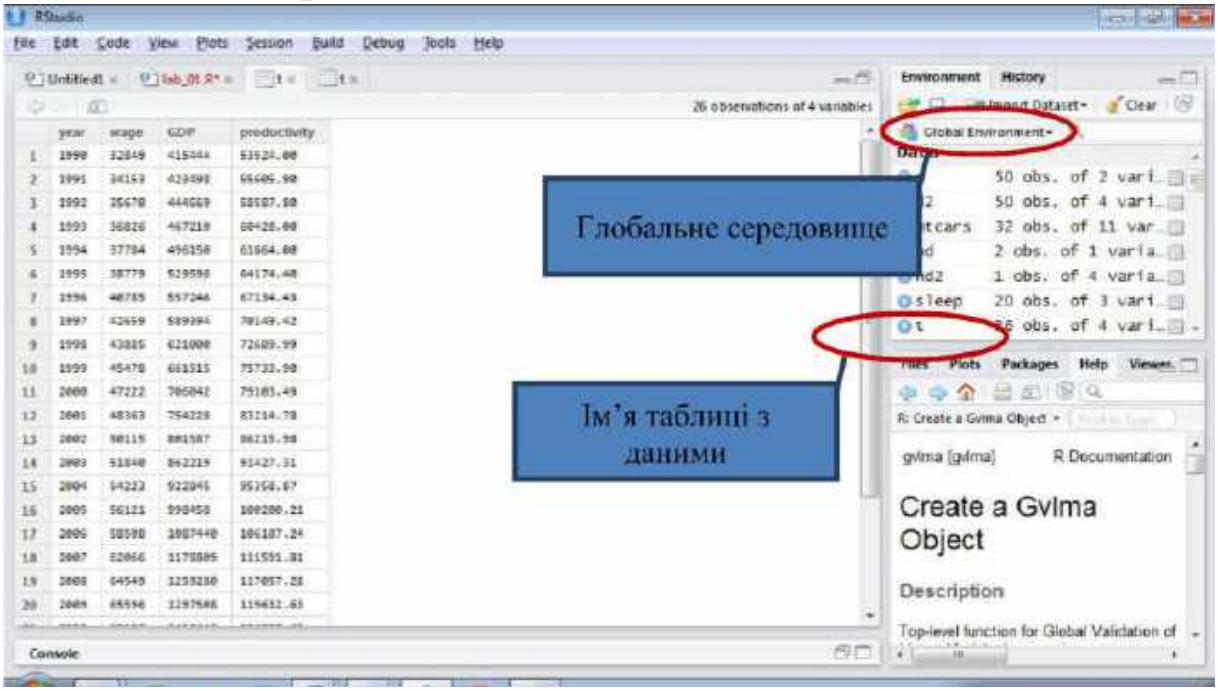


Рис. 1.7. Таблиця з даними в R

Після завантаження даних послідовно будуємо поля кореляції між залежною та факторними змінними. Для цього використовуємо функцію `qplot ()`, аргументами якої є вихідний масив інформації *data=t,* змінні, які відображаються на осі абсцис (вісь *X*) та ординат (вісь *Y*). Приклад:

```
qplot(data=t,productivity,wage)
```

Після введення команди необхідно натиснути клавіші *Ctrl+Enter*. Побудовані графіки з'являться у правому нижньому куті. Для того, щоб їх скопіювати, натисніть на кнопку *Export* і оберіть необхідний варіант, наприклад *Copy to Clipboard* (рис. 1.8).

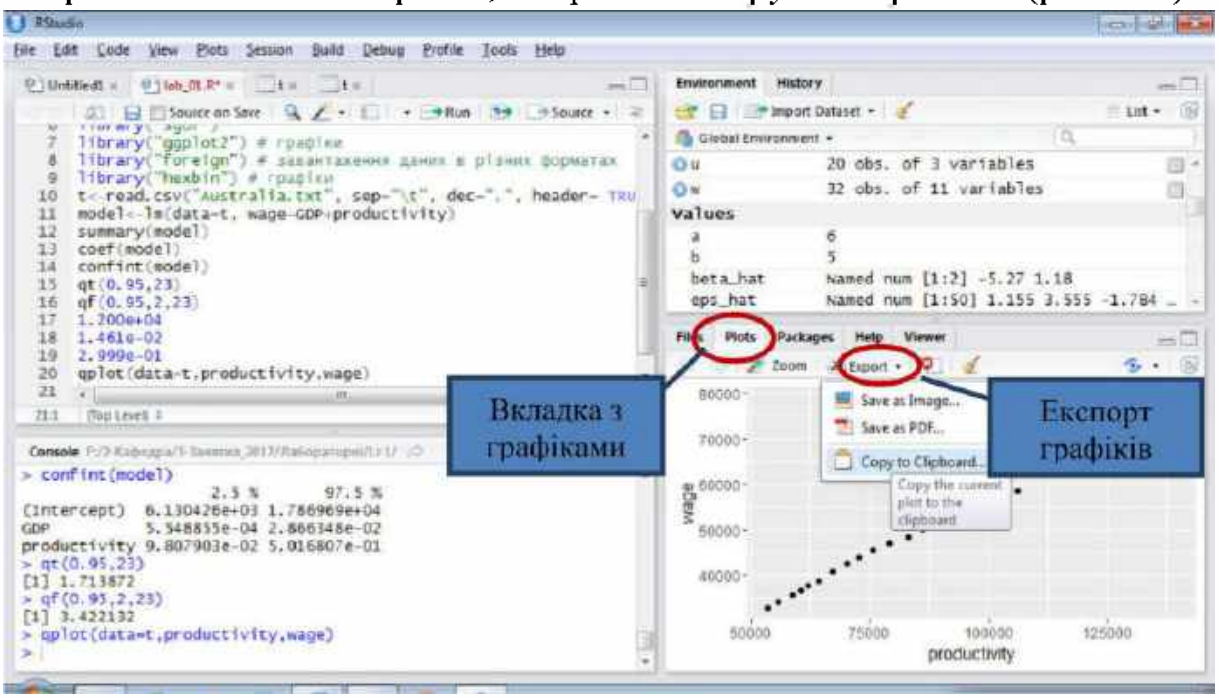


Рис. 1.8. Копіювання графіків в R

Графіки виглядають так.

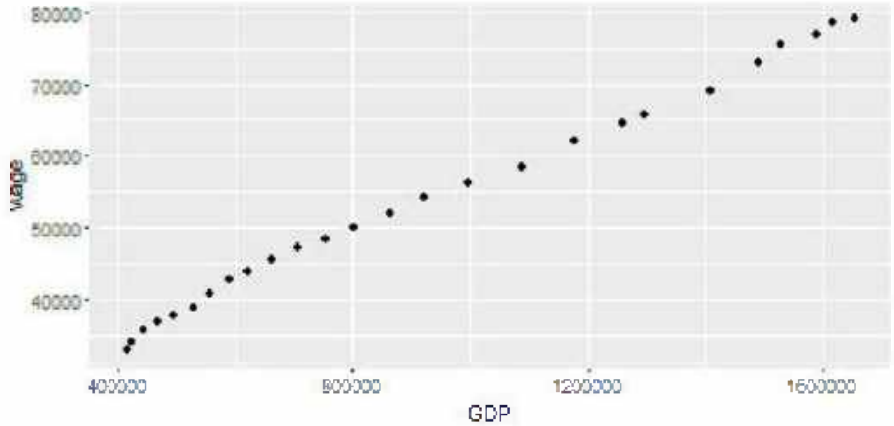


Рис. 1.9. Поле кореляції між середньорічною заробітною платою та ВВП

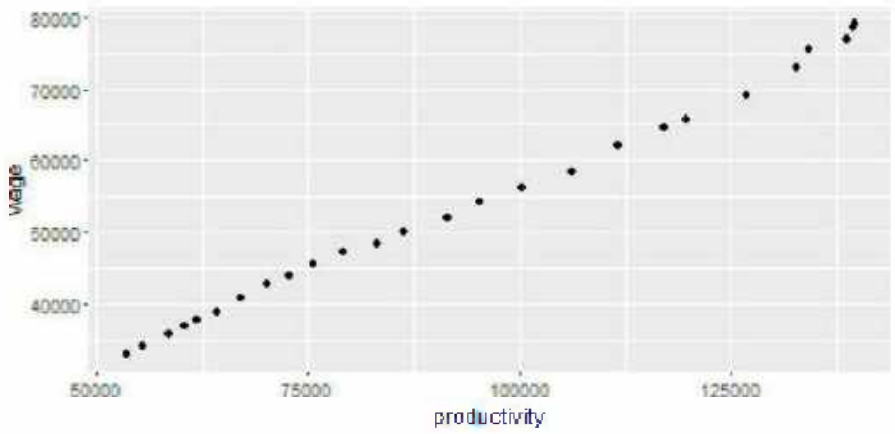


Рис. 1.10. Поле кореляції між середньорічною заробітною платою та продуктивністю праці

Аналогічно побудуйте поле кореляції для результуючого показника та ВВП. Після побудови графіків опишіть їх.

Приклад опису: *на основі візуального аналізу поля кореляції між середньорічною заробітною платою й продуктивністю праці та між середньорічною заробітною платою й ВВП можна простежити наявність лінійного зв'язку між результуючим показником і факторами, які на нього впливають.*

Далі будуємо економетричну модель. Для цього у робочому середовищі записуємо назву моделі, наприклад *model,* далі знак менше та коротке тире «<-» і функцію `lm()`. Аргументами цієї функції є вихідні дані для аналізу: *data=t,* далі через кому запишіть залежну змінну: *wage* потім *знак* «~» (тільда) і факторні змінні, які розділяються знаком «+» рис. 1.11. Приклад:

```
model<-lm(data=t, wage~GDP+productivity)
```

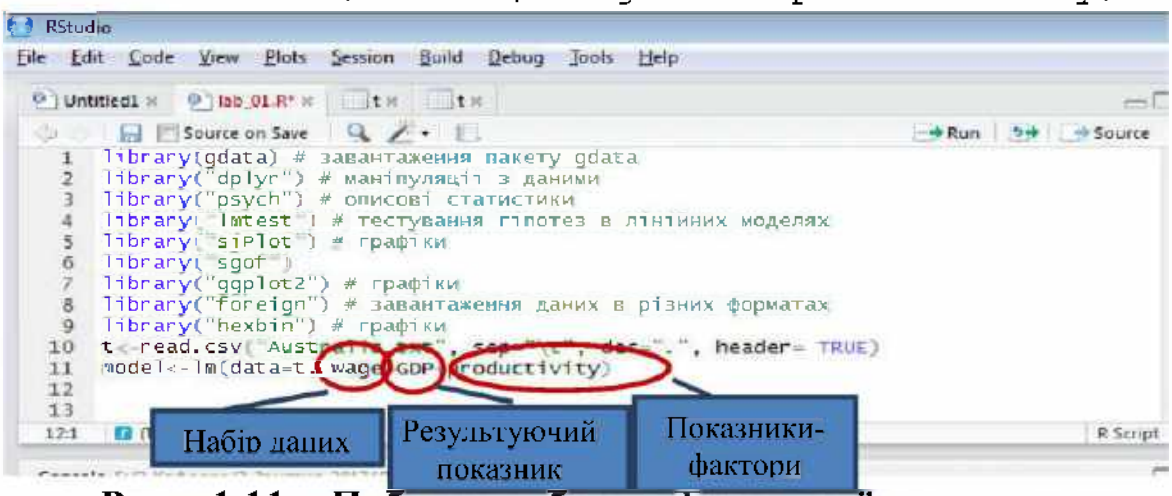


Рис. 1.11. Побудова багатофакторної економетричної моделі в *R*

Після введення команди необхідно натиснути клавіші *Ctrl+Enter*. Для отримання протоколу моделі використовується функція `summary()`. Приклад:

```
summary(model)
```

Потім необхідно натиснути клавіші *Ctrl+Enter* і протокол відобразиться на консолі у вигляді, який подано на рис. 1.12. Оцінки параметрів відображені у стовпчику *Estimate*, стандартні помилки – в стовпчику *Std. Error,* t-статистика – *в стовпчику t value.* У кожному рядку можна побачити зірочки, їх кількість означає рівень, з яким оцінки параметрів можна вважати значимими. Далі представлене стандартне відхилення вектора залишків на один ступінь свободи *(Residual standard error),* коефіцієнт детермінації (*Multiple R-squared*), скоригований коефіцієнт детермінації (*Adjusted R-squared*), критерій Фішера (*F-statistic*) і його ступені свободи.

Після отримання протоколу економетричної моделі виконайте завдання 2. Опишіть отримані результати.

Приклад опису економетричної моделі

В Австралії економетрична залежність між середньорічною заробітною платою ВВП та продуктивністю праці виглядає так:

$$\widehat{wage}=12000+0.0146\cdot GDP+0.2999\cdot productivity$$

Коли за всіх однакових умов ВВП в Австралії зросте на 100 млн австралійських доларів, середньорічна заробітна плата зросте на 1,5 австралійський долар. У разі, якщо за всіх однакових умов продуктивність праці зросте на 10 австралійських доларів на 1 зайнятого, середньорічна заробітна плата зросте на 3 австралійських долари.

Довірчі інтервали моделі можна знайти через функцію:

```
confint(model)
```

Після введення команди необхідно натиснути клавіші *Ctrl+Enter*. Опишіть інтервали.

Варто зазначити, що така функція дозволяє отримати довірчі інтервали за замовчуванням з рівнем значимості α = 0,05, у разі необхідності отримання інтервалів з іншим рівнем значимості,

через кому необхідно зазначити довічий рівень ймовірності. Приклад синтаксису функції з довічною ймовірністю 0,99 (α = 0,01, p =1–0,01):

```
confint(model, level = 0.99).
```

Приклад опису довірчих інтервалів

З ймовірністю 95 % істинне значення параметрів β_1, β_2 знаходиться в інтервалах:

[0.0005; 0.0287] і [0.0980; 0.1157].

Для того щоб виконати завдання 4, необхідно отримати критичне значення t-статистики Стьюдента, якщо α = 0,05. Для цього використовуємо функцію **qt** з аргументами: ймовірність та число ступенів свободи. Приклад:

```
qt(0.95,23)
```

Після введення команди необхідно натиснути клавіші *Ctrl+Enter.*

У консолі відобразиться результат розрахунків:

```
> qt(0.95,23)
[1] 1.713872
```

Після проведення розрахунків зробіть висновок про те, чи значимі оцінки параметрів.

Приклад висновку про значимість оцінок параметрів

Оскільки $t_{\beta_0} > t_{\text{кр}\cdot 0,05\,23}$, $t_{\beta_0} > t_{\text{кр}\cdot 0,05\,23}$, $t_{\beta_0} < t_{\text{кр}\cdot 0,05\,23}$, (4.229 > 1,7138, 2.150 > 1,7138, 3.074 > 1,7138) можна зробити висновок, про те, що з ймовірністю 95 % оцінки параметрів β_0, β_1 і β_2 є статистично значимими.

Виконайте завдання 5.

Приклад пояснення значення коефіцієнта детермінації та скоригованого коефіцієнта детермінації

Коефіцієнт детермінації показує, що рівняння регресії пояснює 99,67 % дисперсії середньорічної заробітної плати, а на частку інших факторів припадає лише 0,33 %. Скоригований коефіцієнт детермінації R=0,9965 показує високу частку поясненої дисперсії результативної ознаки з урахуванням ступенів свободи.

Для виконання завдання 6 отримайте критичне значення F-статистики з рівнем значимості α = 0,05. Для цього використайте

функцію *qf* з аргументами ймовірності та число ступенів свободи. Приклад команди:

```
qf(0.95,2,23)
```

Після введення команди необхідно натиснути клавіші *Ctrl+Enter* або *Run*. У консолі відобразиться результат розрахунків:

```
> qf(0.95,2,23)
[1] 3.422132
```

Приклад оцінки моделі на адекватність

Оскільки 3525>3,42→ *F*факт>*F*2 23, *то визнається статистична значимість (адекватність) рівняння регресії в цілому з ймовірністю 95 %.*

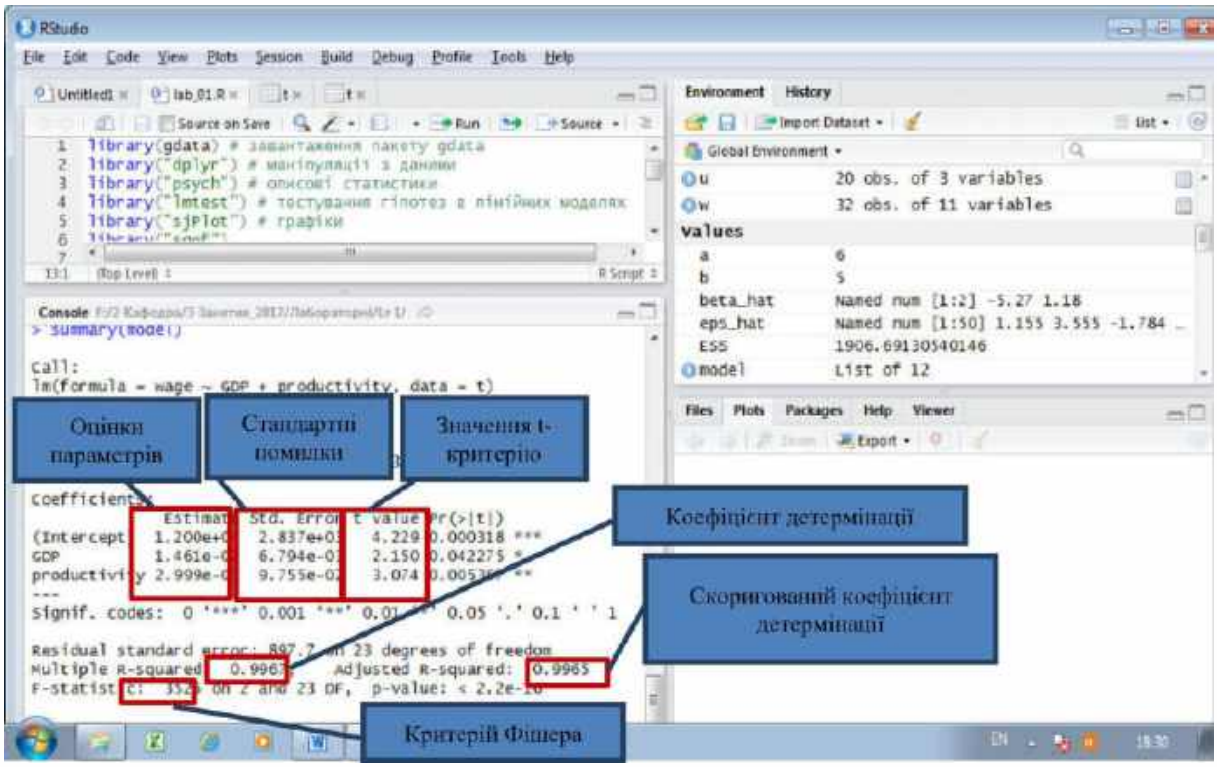


Рис. 1.12. Протокол економетричної моделі в R

Для виконання завдання 7 необхідно сформувати масив прогнозних значень ознак факторів. Для цього даємо ім'я цьому масиву: наприклад *newt*, далі набираємо знак менше та коротку риску «<-», потім команду `data.frame()`. У дужках необхідно записати, чому дорівнюють прогнозні значення ознак факторів (рис. 1.13). Оскільки за умовою лабораторної роботи прогнозні

значення розраховуються на підставі середніх значень, для їх розрахунку використовуємо функцію `mean()`. Аргументами цієї функції є назва таблиці з вихідними даними, яку ви дали раніше, та через знак долара без пробілу назва показника, для якого необхідно порахувати середнє значення. До середнього значення потрібно додати $(k+N)/100$. Приклад написання команди, якщо $(k+N)/100 = 0{,}5$:

```
newt <- data.frame(GDP=mean(t$GDP)+0.5, productivity=(mean(t$productivity)+0.5))
```

Натискаємо клавіші *Ctrl+Enter*.

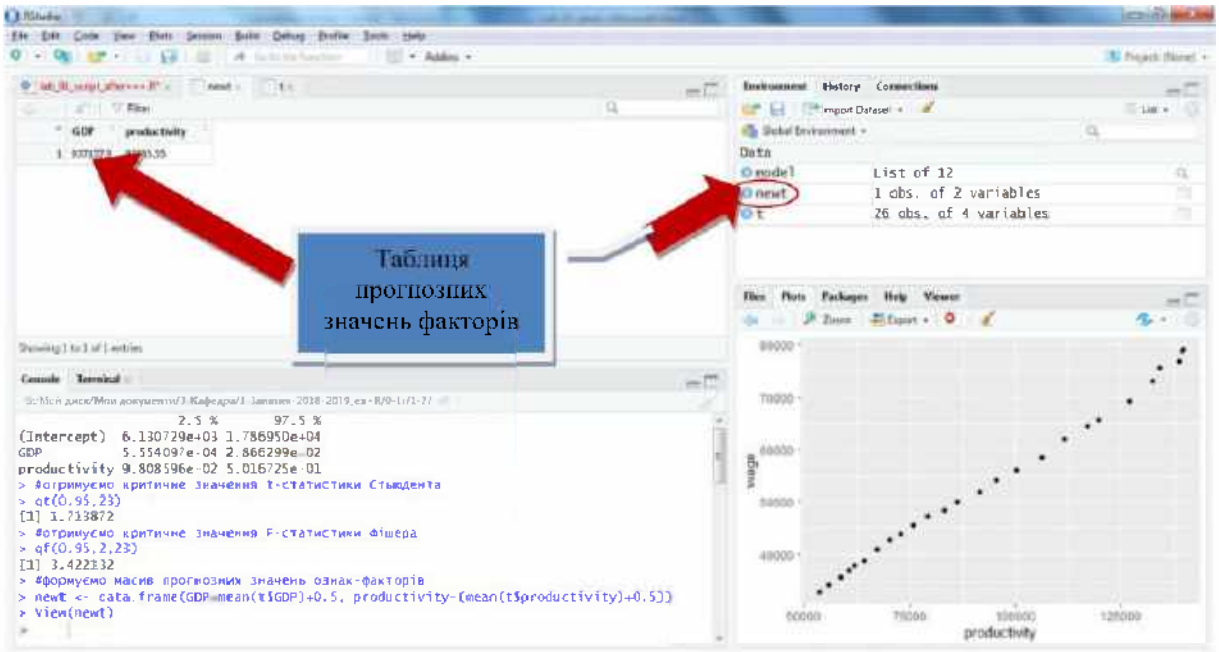


Рис. 1.13. Таблиця прогнозних значень факторів у R

Для знаходження точкового прогнозу використовуємо функцію *predict,* аргументами якої є назва моделі та масив прогнозних значень ознак-факторів. Приклад написання команди:

```
predict(model, newdata=newt)
```

Натисніть клавіші *Ctrl+Enter*. У консолі з'явиться шуканий точковий прогноз.

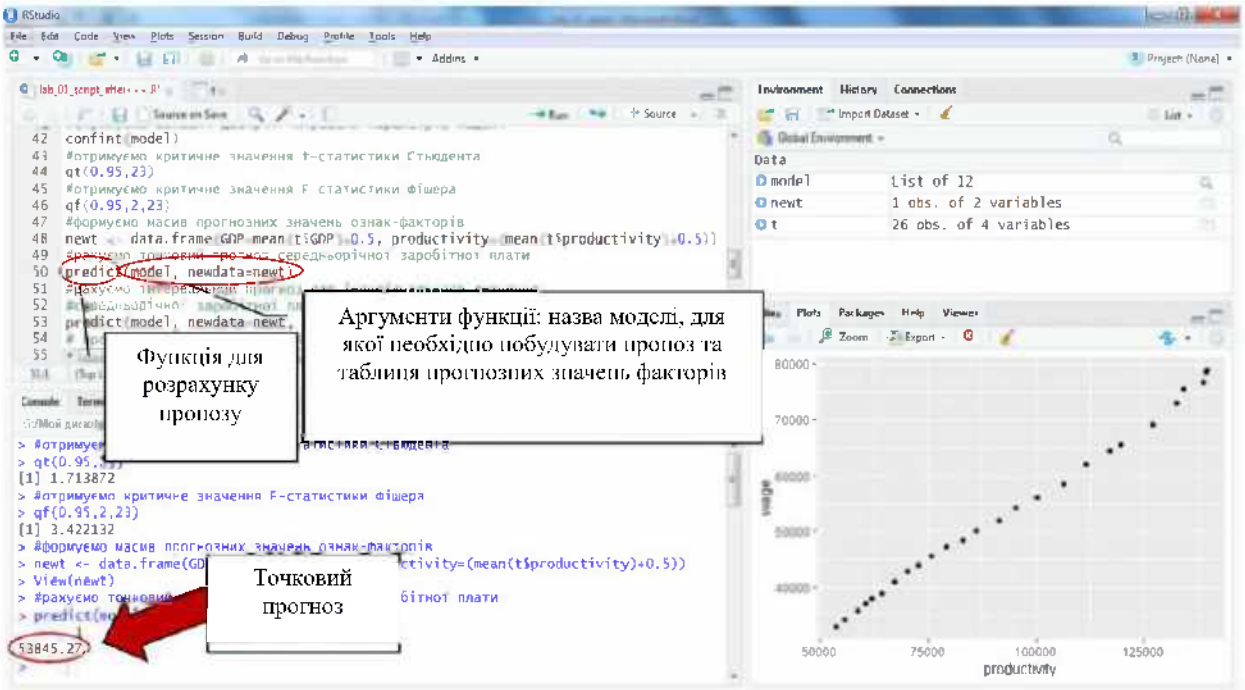


Рис. 1.14. Знаходження точкового прогнозу в R

Для знаходження інтервального прогнозу для індивідуального значення залежної змінної до функції `predict()`, крім зазначених вище аргументів, необхідно додати через кому ще один: *interval="prediction"*. Приклад написання:

```
predict(model, newdata=newt, interval="prediction")
```

Після написання команди потрібно натиснути комбінацію клавіш *Ctrl+Enter*. У консолі відобразиться шуканий інтервальний прогноз з рівнем значимості за замовчуванням 0,05. Опишіть його.

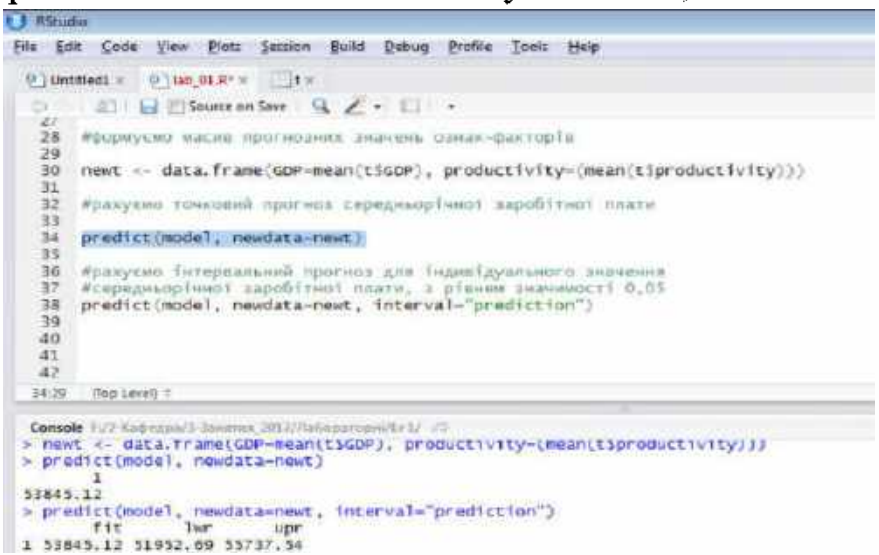


Рис. 1.15. Знаходження інтервального прогнозу залежної змінної в *R*

Приклад опису прогнозу

Якщо у прогнозному періоді в Австралії ВВП буде 937137.9 млн австралійських доларів, а продуктивність праці складе 93885.55 австралійських доларів на 1 зайнятого, середньорічна заробітна плата становитиме 53 845.12 австралійських доларів на рік. Окреме індивідуальне значення середньорічної заробітної плати в Австралії з ймовірністю 95 % міститиметься в інтервалі [51 952.69; 55 737.54].

Після знаходження прогнозного значення залежної змінної необхідно провести діагностику моделі на наявність гетероскедастичності на основі тесту Гольдфельда-Квандта. Для цього використовуємо функцію *gqtest,* аргументами якої є назва моделі, змінна, що зумовлює гетероскедастичність, наприклад ВВП: *order.by = ~GDP,* вихідний масив інформації: *data=t* і кількість спостережень, які необхідно вилучити *fraction = 0.25.* Приклад написання команди:

```
gqtest(model, order.by = ~GDP, data=t, fraction = 0.25)
```

Натисніть комбінацію клавіш *Ctrl+Enter.* У консолі відобразиться дисперсійне співвідношення і число ступенів свободи (рис. 1.16). Далі знайдіть критичне значення *F*-критерію Фішера для певного рівня ймовірності та ступенів свободи df1, df2. Приклад команди:

```
qf(0.95,7,6)
```

Порівняйте дисперсійне співвідношення та критичне значення *F*-критерію Фішера та зробіть висновок про відсутність або наявність гетероскедастичності.

Приклад висновку щодо наявності або відсутності гетероскедастичності в моделі

Порівняємо значення критерію R (BG) (21.328) з критичним (табличним) значенням F-критерію при $\gamma_1 = 7$ і $\gamma_2 = 6$ ступенях свободи і рівні значущості $\alpha = 0{,}05$ $F_{(\alpha = 0{,}05)\ 7\ 6} = 4.2067$. Оскільки $R^* > F\ (\alpha = 0{,}05)\ 7\ 6$, то вихідні дані з імовірністю 95 % мають гетероскедастичність.*

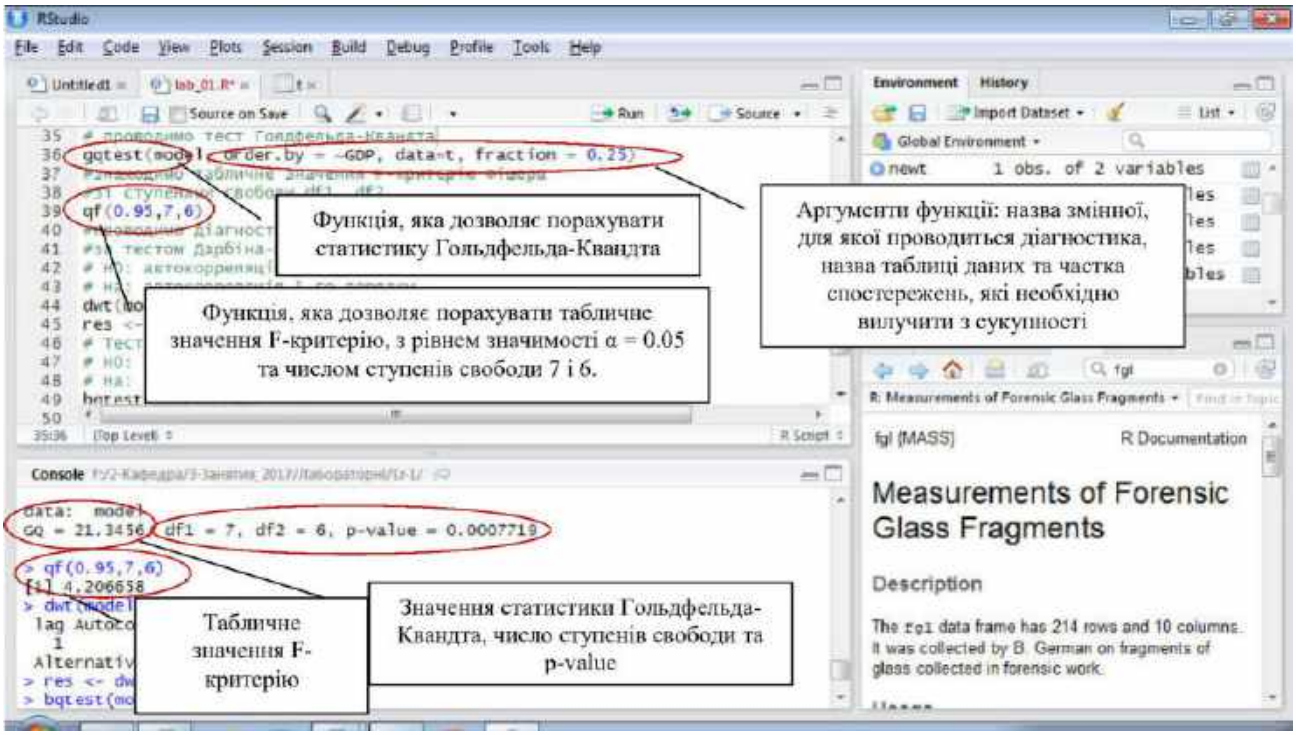


Рис. 1.16. Діагностика наявності гетероскедастичності в моделі

Проведіть таку ж процедуру для другої змінної.

Далі проведіть діагностику моделі на наявність автокореляції залишків за критерієм Дарбіна-Уотсона. Для цього використовується функція `dwt()`, аргументом якої є назва моделі. Функція дозволяє перевірити дві гіпотези: H_0 – в моделі відсутня автокореляція залишків та H_a – в моделі є автокореляція залишків 1-го порядку, з рівнем значимості за замовчуванням $\alpha = 0{,}05$. Для отримання результатів тестування необхідно набрати команду:

```
res <- dwt()
```

Приклад написання:

```
dwt(model)
res <- dwt(model)
```

Після введення кожної команди натисніть комбінацію клавіш *Ctrl+Enter*. В консолі відобразиться значення статистики Дарбіна-Уотсона та яка з гіпотез підтвердилась.

Проведіть тест Бреуша-Годфрі. Для його проведення необхідно використати функцію `bgtest()`. Аргументами цієї функції є назва модслі *model* та порядок автокорсляції *order = 1*. Приклад написання функції:

```
bgtest(model,order = 1)
```

Натисніть комбінацію клавіш *Ctrl+Enter.* У консолі відобразиться значення статистики Бреуша-Годфрі та число ступенів свободи. Знайдіть критичне значення χ^2 критерію для певної ймовірності та числа ступенів свободи, яке дорівнює порядку автокореляції. Приклад написання:

```
qchisq(0.95,1)
```

Натисніть *Ctrl+Enter.* Порівняйте значення статистики Бреуша-Годфрі з критичним значенням χ^2 критерію. Зробіть висновок про наявність або відсутність автокореляції в моделі за двома тестами.

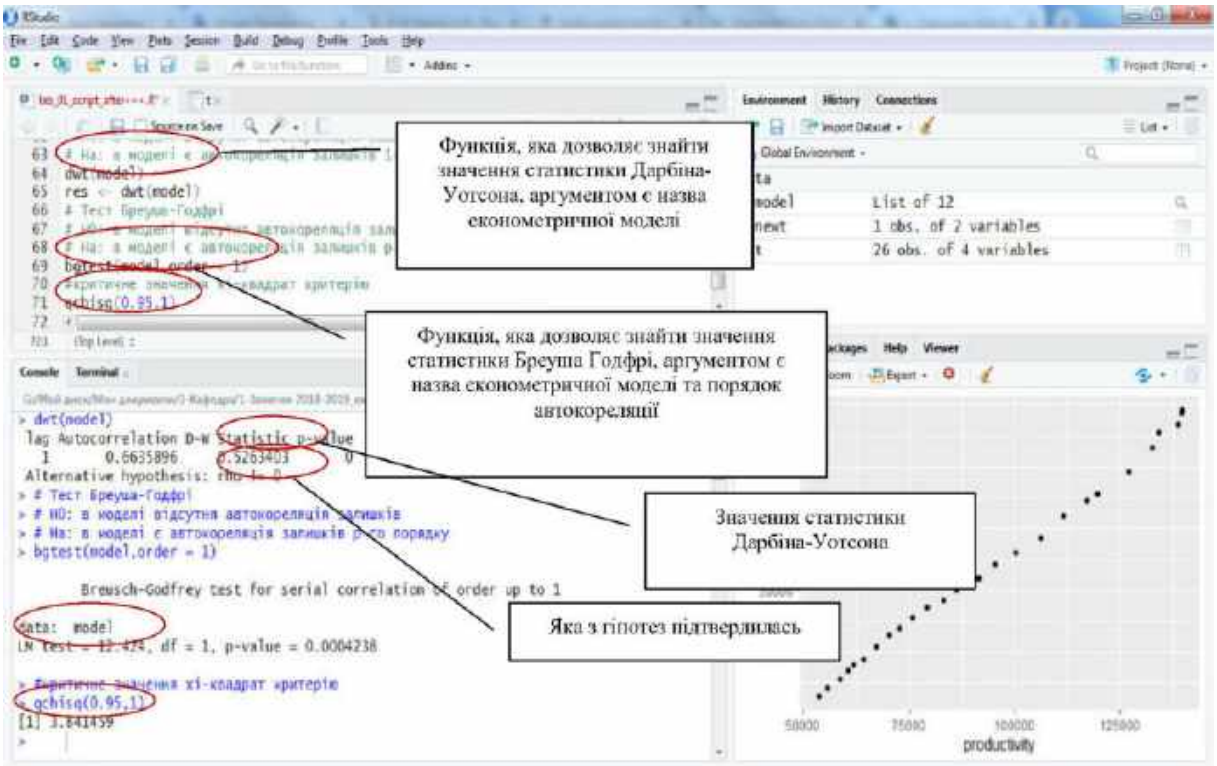


Рис. 1.17. Діагностика наявності автокореляції залишків у моделі

Приклад висновку про наявність автокореляції залишків у моделі

Оскільки підтвердилась альтернативна гіпотеза за тестом Дарбіна-Уотсона, можна зробити висновок про наявність у моделі автокореляції залишків.

Цей висновок підтверджує тест Бреуша-Годфрі. Оскільки $BG>\chi_1^2$ *(12.424>3.8415) в моделі присутня автокореляція залишків першого порядку з ймовірністю 95 %.*

Після проведення діагностики моделі на гетероскедастичність та автокореляцію залишків проведіть тестування моделі на мультиколінеарність. Для цього використайте функцію `vif()`. Аргументами цієї функції є назва моделі *model.* Така функція дозволяє знаходити дисперсійно-інфляційний VIF-фактор. Висновок про наявність мультиколінеарності в моделі варто роботи, якщо коефіцієнт є більшим за 10. Приклад написання:

```
vif(model)
```

Натисніть *Ctrl+Enter.* У консолі відобразиться значення дисперсійно-інфляційних VIF-факторів для незалежно змінних у моделі. Зробіть висновки про наявність або відсутність мультиколінеарності у моделі.

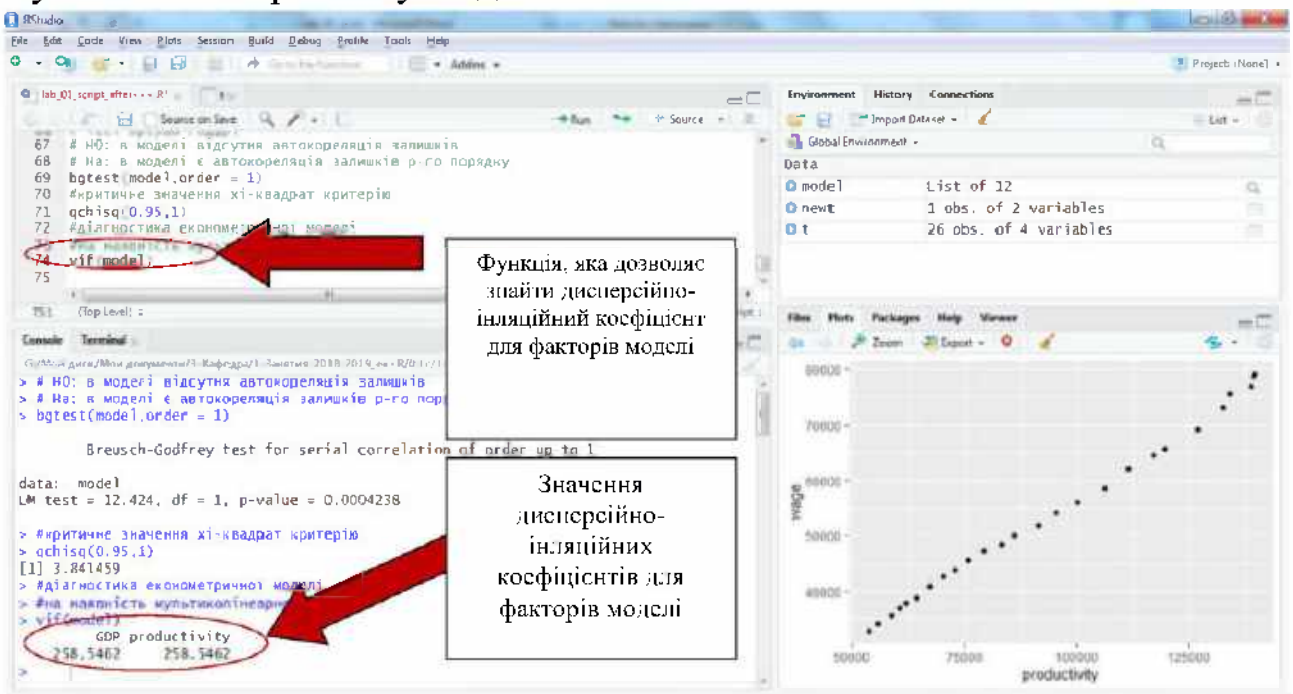


Рис. 1.18. Діагностика наявності мультиколінеарності у моделі

Приклад висновку про наявності мультиколінеарності у моделі

Оскільки значення дисперсійно-інфляційних коефіцієнтів для факторів моделі є більшим за 10, в моделі наявна мультиколінеарність.

Додаток А до лабораторної роботи 1

ВИХІДНІ ДАНІ ДЛЯ ВИКОНАННЯ ЛАБОРАТОРНОЇ РОБОТИ 1

Роки	Варіант 1			Варіант 2		
	Австралія			Австрія		
	Національна валюта: австралійський долар			Національна валюта: євро		
	Середньорічна заробітна плата (Y), од. н.в.	ВВП (X_1), млн од. н.в.	Продуктивність праці (X_2), од. н.в. на 1 зайнятого	Середньорічна заробітна плата (Y), од. н.в.	ВВП (X_1), млн од. н.в.	Продуктивність праці (X_2), од. н.в. на 1 зайнятого
	wage	*GDP*	*productivity*	*wage*	*GDP*	*productivity*
1996	40 785	557 246.0	67 134.4	25 595	182 132.8	50 557.5
1997	42 659	589 394.0	70 149.4	25 761	188 309.7	51 892.8
1998	43 885	621 000.0	72 690.0	26 669	195 828.2	53 423.7
1999	45 478	661 515.0	75 733.9	27 300	203 418.3	54 664.9
2000	47 222	706 042.0	79 103.5	27 930	213 196.1	56 754.7
2001	48 363	754 228.0	83 214.8	28 515	220 095.7	58 172.7
2002	50 115	801 587.0	86 216.0	29 249	226 302.7	59 870.5
2003	51 840	862 219.0	91 427.3	29 780	230 999.2	60 717.2
2004	54 223	922 845.0	95 358.9	30 871	241 505.0	63 084.1
2005	56 121	998 458.0	100 200.2	31 786	253 009.3	65 312.4
2006	58 508	1 087 440.0	106 107.2	32 913	266 478.0	67 610.7
2007	62 066	1 178 809.0	111 591.8	33 905	282 346.9	70 345.8
2008	64 549	1 259 280.0	117 057.3	35 237	291 930.4	71 371.0
2009	65 596	1 297 508.0	119 632.6	35 960	286 188.4	70 270.5
2010	69 188	1 410 442.0	126 897.4	36 416	294 627.5	71 822.4
2011	73 069	1 491 741.0	132 610.3	37 159	308 630.3	74 083.3
2012	75 568	1 527 529.0	134 157.1	38 258	317 117.0	75 329.7

2013	76 885	1 589 940.0	138 635.6	39 097	322 539.2	76 363.1
2014	78 715	1 617 016.0	139 354.6	39 987	330 417.6	77 505.3
2015	79 077	1 654 864.0	139 563.7	40 678	339 896.0	79 216.9
Роки	**Варіант 3**			**Варіант 4**		
	Бельгія			**Канада**		
	Національна валюта: євро			**Національна валюта: канадський долар**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	27 489	215 942.0	55 682.4	34 842	857 023.0	62 253.5
1997	28 296	225 969.0	57 874.0	36 656	903 902.0	64 387.3
1998	28 590	234 490.0	59 020.9	37 889	937 295.0	65 299.0
1999	30 304	244 226.0	60 635.1	38 987	1 004 456.0	68 188.3
2000	31 076	258 222.0	62 844.6	40 985	1 102 380.0	73 163.2
2001	31 885	265 788.0	63 811.6	41 667	1 140 505.0	75 003.6
2002	32 870	275 065.0	66 164.3	42 029	1 189 452.0	76 340.4
2003	33 523	282 637.0	68 041.4	42 745	1 250 315.0	78 541.5
2004	34 163	298 711.0	71 211.5	44 566	1 331 178.0	82 265.6
2005	34 868	311 481.0	73 200.1	47 250	1 417 028.0	86 259.1
2006	36 099	326 662.0	75 916.7	49 387	1 492 207.0	89 432.2
2007	36 973	344 713.0	78 811.4	51 387	1 573 532.0	92 352.7
2008	38 269	354 066.0	79 513.6	53 148	1 652 923.0	95 626.8
2009	38 439	348 781.0	78 457.1	53 824	1 567 365.0	92 275.7
2010	38 879	365 101.0	81 605.1	54 788	1 662 130.0	96 087.4
2011	40 405	379 106.0	83 590.1	56 809	1 769 921.0	100 722.2
2012	41 535	387 500.0	85 071.4	58 881	1 822 808.0	102 611.8

2013	42 301	391 712.0	86 278.3	60 821	1 897 531.0	105 402.9
2014	42 687	400 805.0	87 913.2	62 551	1 983 117.0	109 508.7
2015	42 862	410 351.0	89 183.5	63 430	1 986 193.0	108 619.9
Роки	**Варіант 5**			**Варіант 6**		
	Чехія			**Данія**		
	Національна валюта: чеська крона			**Національна валюта: датська Крона**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	125 567	1 812 622.0	353 200.5	238 818	1 088 024.0	412 442.8
1997	138 383	1 953 311.0	383 303.9	247 315	1 146 129.0	428 619.7
1998	149 282	2 142 587.0	427 868.2	255 549	1 185 988.0	437 311.2
1999	158 127	2 237 300.0	456 619.6	263 647	1 241 521.0	453 938.2
2000	169 073	2 372 630.0	488 261.7	270 971	1 326 912.0	481 637.7
2001	180 647	2 562 679.0	528 781.0	279 422	1 371 526.0	493 000.0
2002	197 175	2 674 634.0	548 443.3	291 130	1 410 271.0	506 562.9
2003	212 263	2 801 163.0	578 985.0	301 969	1 436 751.0	520 750.6
2004	229 607	3 057 660.0	633 201.9	312 556	1 506 001.0	548 834.2
2005	239 065	3 257 972.0	661 834.3	323 380	1 585 984.0	569 882.9
2006	253 040	3 507 131.0	702 976.0	336 330	1 682 260.0	591 096.3
2007	268 236	3 831 819.0	752 348.6	346 242	1 738 845.0	597 130.8
2008	281 069	4 015 346.0	771 576.7	361 205	1 801 470.0	611 289.4
2009	282 879	3 921 827.0	767 465.8	374 624	1 722 143.0	603 413.8
2010	290 934	3 953 651.0	781 780.2	387 992	1 810 926.0	649 543.0
2011	298 388	4 033 755.0	799 802.6	395 202	1 846 854.0	662 667.4
2012	304 662	4 059 912.0	801 621.8	403 673	1 895 002.0	684 858.0

2013	303 882	4 098 128.0	806 570.5	407 512	1 929 677.0	697 641.7
2014	307 539	4 313 789.0	844 356.4	417 647	1 977 255.0	707 679.0
2015	314 199	4 554 615.0	879 328.5	428 396	2 027 171.0	716 568.0
	Варіант 7			**Варіант 8**		
Роки	**Естонія**			**Фінляндія**		
	Національна валюта:євро			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	2 501	3 649.3	5 898.3	22 685	102 060.0	48 996.6
1997	3 001	4 494.8	7 261.3	23 070	110 738.0	51 398.5
1998	3 415	5 049.2	8 312.8	24 229	120 382.0	54 853.7
1999	3 686	5 378.7	9 268.8	25 187	126 923.0	56 427.8
2000	4 217	6 170.8	10 542.9	26 493	136 261.0	59 287.7
2001	4 627	6 976.4	11 852.5	27 490	144 437.0	61 947.6
2002	5 029	7 773.8	13 209.6	28 100	148 289.0	62 917.0
2003	5 630	8 708.9	14 488.2	28 925	151 569.0	64 218.7
2004	6 271	9 707.7	16 214.6	29 989	158 477.0	66 766.5
2005	6 971	11 262.3	18 393.4	30 994	164 387.0	68 193.4
2006	7 934	13 521.7	21 055.3	32 131	172 614.0	70 322.7
2007	9 878	16 246.4	25 250.9	33 241	186 584.0	74 428.2
2008	10 791	16 517.3	25 723.8	34 689	193 711.0	75 588.6
2009	10 367	14 145.9	24 533.2	35 610	181 029.0	72 382.6
2010	10 683	14 716.5	26 850.1	36 693	187 100.0	75 328.1
2011	10 764	16 667.6	28 540.5	38 005	196 869.0	78 262.4
2012	11 458	17 934.9	30 218.9	39 100	199 793.0	78 733.1

2013	12 026	18 890.1	31 436.4	39 629	203 338.0	80 702.5
2014	12 912	19 758.3	32 631.3	40 322	205 474.0	81 943.8
2015	13 599	20 251.7	32 511.9	40 893	209 511.0	83 841.3
	Варіант 9			**Варіант 10**		
Роки	**Франція**			**Німеччина**		
	Національна валюта: євро			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	23 550	1 258 950.0	52 848.2	26 518	1 926 320.0	50 734.0
1997	24 008	1 299 739.0	54 128.7	26 702	1 967 090.0	51 837.8
1998	24 429	1 358 776.0	55 603.2	27 099	2 018 230.0	52 548.5
1999	24 839	1 408 159.0	56 252.1	27 633	2 064 880.0	52 903.6
2000	25 523	1 485 303.0	57 859.2	27 990	2 116 480.0	53 022.0
2001	26 195	1 544 629.0	59 324.4	28 762	2 179 850.0	54 757.7
2002	27 134	1 594 259.0	60 940.3	29 283	2 209 290.0	55 747.9
2003	27 828	1 637 438.0	62 588.4	29 776	2 220 080.0	56 634.7
2004	28 890	1 710 760.0	65 306.2	30 062	2 270 620.0	57 722.2
2005	29 770	1 771 978.0	67 176.4	30 470	2 300 860.0	58 507.3
2006	30 753	1 853 267.0	69 496.6	30 820	2 393 250.0	60 382.2
2007	31 536	1 945 670.0	71 949.9	31 269	2 513 230.0	62 324.4
2008	32 340	1 995 850.0	73 463.3	31 997	2 561 740.0	62 701.7
2009	32 865	1 939 017. 0	72 181.7	31 992	2 460 280.0	60 165.3
2010	33 892	1 998 481.0	74 304.0	32 754	2 580 060.0	62 897.6
2011	34 447	2 059 284.0	75 968.7	34 006	2 703 120.0	65 014.8

2012	35 172	2 086 929.0	76 736.6	34 916	2 758 260.0	65 577.6
2013	35 634	2 115 256.0	77 575.7	35 703	2 826 240.0	66 770.0
2014	36 018	2 139 964.0	78 118.0	36 556	2 923 930.0	68 537.1
2015	36 491	2 181 064.0	79 245.1	37 613	3 032 820.0	70 437.3
	Варіант 11			**Варіант 12**		
Роки	**Греція**			**Угорщина**		
	Національна валюта: євро			**Національна валюта: форинт**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	10 744	103 036.6	24 932.3	705 333	7 103 898.0	1 770 348.8
1997	12 192	114 712.2	27 867.0	850 877	8 812 386.0	2 192 722.2
1998	12 956	125 262.6	29 149.3	975 599	10 433 144.0	2 556 188.3
1999	13 597	133 788.7	31 123.1	1 059 643	11 637 865.0	2 773 647.7
2000	13 902	141 247.3	32 750.8	1 228 823	13 321 531.0	3 144 100.3
2001	14 598	152 193.8	35 161.8	1 441 741	15 383 442.0	3 636 499.1
2002	16 275	163 460.8	36 862.9	1 606 947	17 421 576.0	4 121 576.0
2003	17 419	178 904.8	39 794.8	1 829 196	19 077 348.0	4 513 329.7
2004	18 325	193 715.8	42 074.7	2 019 402	21 023 630.0	5 021 840.8
2005	18 805	199 242.3	42 876.7	2 167 828	22 470 802.0	5 383 273.7
2006	19 644	217 861.6	46 046.5	2 283 324	24 153 022.0	5 761 398.7
2007	20 364	232 694.6	48 527.9	2 404 806	25 560 380.0	6 088 785.4
2008	20 917	241 990.4	49 829.6	2 571 881	27 071 868.0	6 578 050.5
2009	22 113	237 534.2	49 189.1	2 581 592	26 297 412.0	6 553 276.0
2010	21 382	226 031.4	48 035.8	2 660 179	27 085 900.0	6 823 932.3

2011	20 440	207 028.9	47 247.2	2 751 218	28 166 115.0	7 096 428.5
2012	19 600	191 203.9	46 575.8	2 830 649	28 660 518.0	7 209 398.8
2013	18 099	180 654.3	45 189.5	2 888 110	30 127 349.0	7 495 888.8
2014	17 716	177 940.6	44 493.0	2 889 151	32 400 148.0	7 690 000.4
2015	17 642	175 697.4	43 708.4	2 977 799	33 999 012.0	7 889 897.5
	Варіант 13			**Варіант 14**		
Роки	**Ісландія**			**Ірландія**		
	Національна валюта: ісландська крона			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	2 432 882	498 879.7	3 513 237.2	24 806	60 206.5	45 245.6
1997	2 644 437	538 598.0	3 792 943.7	26 744	69 367.3	49 373.1
1998	3 000 744	600 899.0	4 062 873.6	27 918	80 352.5	52 710.9
1999	3 196 977	646 053.0	4 214 305.3	29 328	92 716.4	57 124.8
2000	3 456 276	703 305.0	4 496 835.0	31 354	108 382.7	63 912.4
2001	3 707 420	793 628.0	4 991 371.1	33 707	121 940.6	69 737.1
2002	4 016 306	843 255.0	5 381 333.8	35 518	135 950.5	76 545.5
2003	3 951 784	867 125.0	5 519 573.5	38 021	145 556.6	80 460.3
2004	4 342 986	961 859.0	6 146 063.9	39 976	156 144.0	83 480.5
2005	4 737 822	1 051 241.0	6 505 204.2	42 213	170 215.8	86 743.0
2006	5 312 738	1 196 095.0	7 044 140.2	43 861	185 060.7	90 141.6
2007	5 735 451	1 364 052.0	7 684 800.0	46 406	197 293.4	92 075.4
2008	5 762 057	1 551 434.0	8 662 389.7	48 490	187 687.2	88 147.1
2009	5 671 372	1 593 337.0	9 484 148.8	48 835	169 704.3	86 482.4

2010	6 057 449	1 620 293.0	9 679 169.7	46 917	167 124.3	88 767.2
2011	6 562 716	1 701 585.0	10 164 784.9	47 426	173 070.2	92 430.3
2012	6 947 079	1 778 499.0	10 505 014.8	47 429	175 752.5	94 413.8
2013	7 357 839	1 891 239.0	10 813 259.0	46 664	180 209.3	94 436.1
2014	7 603 064	2 006 019.0	11 288 795.7	47 012	193 159.6	99 546.0
2015	8 059 248	2 213 543.0	12 049 771.4	47 366	255 815.1	128 588.2
	Варіант 15			**Варіант 16**		
Роки	**Італія**			**Японія**		
	Національна валюта: євро			**Національна валюта: ієна**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	18 709	1 043 085.8	47 317.9	4 563 719	511 934 800.0	7 608 905.9
1997	19 577	1 089 869.2	49 284.6	4 597 990	523 198 300.0	7 720 774.7
1998	20 042	1 135 499.5	50 836.3	4 580 708	512 438 600.0	7 649 593.2
1999	20 610	1 171 901.4	51 897.7	4 529 917	504 903 200.0	7 639 515.2
2000	21 291	1 239 266.3	53 831.3	4 506 104	509 860 000.0	7 771 071.5
2001	21 957	1 298 890.2	55 334.8	4 440 996	505 543 200.0	7 733 565.9
2002	22 442	1 345 794.2	56 386.3	4 302 687	499 147 000.0	7 725 896.6
2003	23 065	1 390 709.6	57 424.9	4 258 353	498 854 800.0	7 717 312.5
2004	24 103	1 448 362.7	59 445.4	4 236 132	503 725 300.0	7 748 427.9
2005	24 897	1 489 725.5	60 801.9	4 281 413	503 903 000.0	7 695 172.8
2006	25 721	1 548 473.4	61 979.1	4 221 088	506 687 000.0	7 700 176.3
2007	26 278	1 609 550.8	63 631.4	4 153 668	512 975 200.0	7 762 589.5
2008	27 124	1 632 150.8	64 386.7	4 175 655	501 209 300.0	7 620 868.8

2009	27 159	1 572 878.3	63 103.2	4 014 433	471 138 700.0	7 275 037.4
2010	27 809	1 604 514.5	64 787.8	4 020 812	482 676 900.0	7 485 219.6
2011	28 171	1 637 462.9	65 913.2	4 086 790	471 578 700.0	7 327 429.4
2012	28 065	1 613 265.0	65 143.5	4 004 867	475 331 700.0	7 384 596.4
2013	28 444	1 604 599.1	65 971.0	4 013 403	479 083 700.0	7 399 776.0
2014	28 647	1 621 827.2	66 633.8	4 033 913	486 938 800.0	7 474 806.6
2015	28 890	1 645 439.4	67 150.9	4 060 008	499 349 900.0	7 588 766.3

	Варіант 17			**Варіант 18**		
Роки	**Корея**			**Люксембург**		
	Національна валюта: вона			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	15 979 894	481 140 800.0	23 089 586.3	35 049	16 686.2	75 316.6
1997	16 648 113	530 347 100.0	25 015 192.7	36 332	17 465.2	76 248.4
1998	17 347 436	524 476 800.0	26 329 156.6	36 940	18 189.1	76 072.0
1999	17 966 286	576 872 800.0	28 452 419.2	38 385	20 838.3	83 165.9
2000	18 715 856	635 184 600.0	30 050 839.8	40 504	23 200.8	87 870.1
2001	19 998 824	688 164 900.0	31 923 036.6	41 761	23 525.7	84 290.1
2002	21 292 820	761 938 900.0	34 397 494.5	43 243	24 767.7	86 246.5
2003	22 782 060	810 915 300.0	36 665 956.8	43 938	25 877.3	88 478.2
2004	24 082 504	876 033 100.0	38 836 246.7	45 492	27 660.5	92 398.9
2005	25 428 004	919 797 300.0	40 242 791.9	47 130	29 733.5	96 649.3
2006	26 238 578	966 054 600.0	41 728 417.8	48 994	33 409.3	104 600.9
2007	27 393 540	1 043 257 800.0	44 521 450.8	51 418	36 766.1	110 249.1

2008	28 428 992	1 104 492 200.0	46 845 377.4	52 738	37 647.4	107 762.7
2009	29 111 324	1 151 707 800.0	48 997 375.1	53 911	36 268.2	102 737.2
2010	30 327 188	1 265 308 000.0	53 099 723.4	55 098	39 946.6	111 135.0
2011	31 821 592	1 332 681 000.0	54 968 838.0	55 939	42 856.4	115 813.1
2012	32 079 712	1 377 456 700.0	55 811 313.3	56 539	43 905.0	115 839.6
2013	32 859 444	1 429 445 400.0	57 025 899.0	58 489	46 352.6	120 098.9
2014	33 224 208	1 486 079 300.0	58 051 333.2	60 064	49 272.8	124 542.8
2015	33 911 008	1 558 591 600.0	60 093 058.8	60 381	51 216.2	126 272.7
	Варіант 19			**Варіант 20**		
Роки	**Мексика**			**Нідерланди**		
	Національна валюта: мексиканський песо			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	34 578	2 820 216.9	84 196.0	28 745	340 964.0	45 908.7
1997	42 612	3 544 596.6	100 058.6	29 473	364 961.0	47 669.9
1998	51 172	4 295 077.7	118 136.3	29 060	389 315.0	49 670.2
1999	61 303	5 130 759.9	139 517.8	30 150	414 838.0	51 449.6
2000	71 071	6 132 389.2	163 120.2	31 867	448 061.0	54 575.0
2001	78 834	6 487 465.5	172 152.2	33 419	476 696.0	56 946.1
2002	82 963	6 993 814.8	181 375.7	34 689	494 501.0	58 673.6
2003	88 668	7 696 035.0	197 955.5	35 738	506 671.0	60 454.7
2004	93 265	8 690 254.0	216 088.9	36 831	523 939.0	63 254.7
2005	98 849	9 424 602.0	232 876.5	37 502	545 609.0	65 428.6
2006	103 151	10 520 793.0	251 291.8	38 463	579 212.0	67 990.6

2007	109 206	11 399 472.0	267 798.8	39 796	613 280.0	69 921.3
2008	115 218	12 256 864.0	281 523.5	40 892	639 163.0	71 703.3
2009	119 059	12 072 542.0	280 345.5	42 043	617 540.0	69 873.3
2010	118 857	13 266 858.0	284 711.0	42 711	631 512.0	71 942.6
2011	125 333	14 527 337.0	309 806.9	43 343	642 929.0	72 614.5
2012	129 782	15 599 271.0	318 330.5	44 078	645 164.0	73 015.4
2013	134 108	16 078 960.0	326 170.2	45 131	652 748.0	74 753.6
2014	137 844	17 217 016.0	350 322.1	45 347	663 008.0	76 111.6
2015	141 129	18 194 758.0	361 991.9	46 384	676 531.0	76 948.5
	Варіант 21			**Варіант 22**		
Роки	**Нова Зеландія**			**Норвегія**		
	Національна валюта: новозеландський долар			**Національна валюта: норвезька крона**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	35 978	101 101.0	60 705.4	228 194	1 054 657.0	487 589.9
1997	37 467	104 814.0	62 854.6	238 720	1 141 324.0	512 724.2
1998	37 472	106 826.0	62 477.2	253 697	1 163 192.0	509 055.6
1999	37 341	113 223.0	64 189.9	267 068	1 265 701.0	548 872.9
2000	38 279	119 839.0	68 100.1	277 325	1 507 886.0	649 950.9
2001	39 815	128 712.0	71 268.0	290 603	1 564 585.0	672 072.6
2002	41 369	135 182.0	72 671.2	306 220	1 560 181.0	667 599.9
2003	43 089	144 499.0	75 585.9	318 926	1 619 613.0	701 434.8
2004	45 198	154 557.0	78 300.9	333 190	1 781 981.0	766 773.2
2005	46 845	162 935.0	79 145.0	346 392	1 988 942.0	844 920.1

2006	48 536	172 111.0	81 769.0	366 500	2 215 312.0	910 152.8
2007	52 235	186 846.0	87 845.6	387 077	2 349 861.0	926 966.9
2008	53 630	189 618.0	87 050.0	409 879	2 605 380.0	995 559.8
2009	55 709	194 252.0	91 027.2	424 222	2 429 698.0	932 705.6
2010	57 183	203 434.0	94 226.0	438 973	2 590 089.0	999 648.4
2011	58 897	213 241.0	97 503.9	458 672	2 791 973.0	1 061 586.7
2012	60 553	217 995.0	100 135.5	476 141	2 965 208.0	1 104 772.0
2013	64 792	232 530.0	103 854.4	495 891	3 071 134.0	1 132 006.6
2014	67 383	241 260.0	104 351.2	510 059	3 140 371.0	1 143 616.5
2015	69 404.9	248 956.0	105 464.8	524 466	3 117 433.0	1 132 376.7
	Варіант 23			**Варіант 24**		
Роки	**Польща**			**Португалія**		
	Національна валюта: злотий			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	15 111	431 222.0	29 866.6	10 339	94 351.4	20 489.5
1997	18 172	521 795.0	35 230.2	10 931	102 356.9	21 660.3
1998	20 720	606 069.0	40 385.8	11 342	111 385.2	22 927.6
1999	23 713	673 287.0	45 261.9	12 019	119 639.2	24 251.6
2000	26 557	747 032.0	51 460.5	12 637	128 466.3	25 479.9
2001	29 123	779 975.0	54 947.9	13 173	135 827.5	26 476.6
2002	29 966	810 617.3	58 884.2	13 626	142 631.4	27 695.8
2003	30 497	845 930.0	62 172.8	14 081	146 158.3	28 657.4
2004	30 964	933 061.8	67 808.2	14 434	152 371.6	30 088.1

2005	31 510	990 468.0	70 460.8	14 901	158 652.6	31 472.7
2006	32 229	1 069 824.0	73 761.1	15 093	166 248.7	32 849.9
2007	33 767	1 187 605.0	78 359.3	15 711	175 467.7	34 666.6
2008	37 164	1 286 069.0	81 749.1	16 097	178 872.6	35 210.3
2009	38 081	1 372 208.0	86 906.9	16 518	175 448.2	35 503.7
2010	40 644	1 445 297.0	94 031.8	16 752	179 929.8	36 936.5
2011	42 650	1 566 813.0	101 364.0	16 626	176 166.6	36 880.2
2012	43 675	1 629 392.0	105 292.6	16 201	168 398.0	36 756.5
2013	44 351	1 656 842.0	107 143.3	16 664	170 269.3	38 261.3
2014	45 139	1 719 704.0	109 319.4	16 428	173 079.1	38 351.3
2015	46 203	1 798 302.0	112 719.7	16 368	179 539.9	39 236.6
	Варіант 25			**Варіант 26**		
Роки	**Словаччина**			**Словенія**		
	Національна валюта: євро			**Національна валюта: євро**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	3 430	21 951.5	10 204.7	7 679	12 131.9	13 417.8
1997	4 069	24 415.6	11 468.6	8 773	13 827.2	15 576.8
1998	4 475	26 664.1	12 584.1	9 507	15 337.4	17 296.3
1999	4 812	28 547.6	13 823.2	10 288	17 210.4	19 114.2
2000	5 405	31 601.3	15 606.7	11 360	18 902.0	20 671.0
2001	5 702	34 310.6	16 847.7	12 718	21 146.8	22 994.5
2002	6 178	37 279.8	18 288.7	13 667	23 622.5	25 287.8
2003	6 730	41 404.3	20 094.7	14 684	25 667.5	27 561.3

2004	7 334	46 101.5	22 425.9	15 851	27 672.7	29 627.8
2005	8 024	50 415.1	24 134.7	16 834	29 226.6	31 443.6
2006	8 694	56 272.7	26 389.5	17 789	31 555.4	33 422.4
2007	9 469	63 053.9	28 964.1	18 921	35 152.6	36 024.3
2008	9 991	68 491.6	30 479.5	20 282	37 951.2	37 920.2
2009	10 334	64 023.1	29 059.7	20 645	36 166.2	36 795.4
2010	10 897	67 577.3	31 144.2	21 572	36 252.4	37 680.6
2011	11 218	70 627.2	31 982.4	21 893	36 896.3	39 001.4
2012	11 475	72 703.5	32 906.0	21 570	36 002.5	38 413.1
2013	11 716	74 169.9	33 832.7	21 705	35 917.1	38 753.2
2014	11 932	75 946.4	34 161.6	22 003	37 332.4	40 102.8
2015	12 305	78 685.6	34 707.6	22 192	38 570.0	40 966.8
	Варіант 27			**Варіант 28**		
Роки	**Іспанія**			**Швеція**		
	Національна валюта: євро			**Національна валюта: шведська крона**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	17 105	487 992.0	34 745.1	213 898	1 932 025.0	471 697.3
1997	17 595	518 049.0	35 555.6	224 646	2 019 261.0	499 421.5
1998	17 913	554 042.0	36 429.0	231 924	2 121 037.0	515 891.7
1999	18 227	594 316.0	37 376.2	241 393	2 237 854.0	533 114.3
2000	18 644	646 250.0	38 718.5	250 086	2 380 358.0	553 494.4
2001	19 293	699 528.0	40 557.5	256 870	2 478 130.0	564 352.9
2002	19 971	749 288.0	42 349.6	262 768	2 569 876.0	584 966.8

2003	20 548	803 472.0	43 949.3	270 342	2 677 446.0	612 982.4
2004	21 072	861 420.0	45 409.4	280 731	2 805 115.0	646 757.1
2005	21 912	930 566.0	47 036.3	289 360	2 907 352.0	668 556.6
2006	22 619	1 007 974.0	48 908.9	299 918	3 099 081.0	700 769.0
2007	23 676	1 080 807.0	50 778.1	314 038	3 297 053.0	728 727.1
2008	25 504	1 116 207.0	52 344.9	328 289	3 387 599.0	742 048.3
2009	26 995	1 079 034.0	53 987.3	337 842	3 288 509.0	738 194.5
2010	27 052	1 080 913.0	55 037.7	344 916	3 519 994.0	782 620.9
2011	27 316	1 070 413.0	56 005.9	355 060	3 656 577.0	796 015.5
2012	27 097	1 039 758.0	56 685.3	364 771	3 684 800.0	796 300.3
2013	27 603	1 025 634.0	57 420.2	371 189	3 769 909.0	806 950.0
2014	27 408	1 037 025.0	57 542.0	378 907	3 936 840.0	831 012.8
2015	27 479	1 075 639.0	58 224.9	389 385	4 181 103.0	869 306.4
	Варіант 29			**Варіант 30**		
Роки	**Швейцарія**			**Великобританія**		
	Національна валюта: швейцарський франк			**Національна валюта: фунт стерлінгів**		
	Середньорічна заробітна плата (*Y*), од. н.в.	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***	***wage***	***GDP***	***productivity***
1996	66 019	407 409.5	102 875.6	18 306	892 900.0	34 268.2
1997	67 528	415 948.4	105 103.1	19 181	938 855.0	35 398.1
1998	67 776	427 658.4	106 629.4	20 378	980 308.0	36 587.5
1999	68 988	435 463.9	107 716.6	21 275	1 021 205.0	37 589.9
2000	70 454	458 778.9	112 392.7	22 438	1 080 863.0	39 327.7
2001	73 505	470 213.9	113 301.9	23 649	1 120 575.0	40 437.2

2002	74 203	469 337.8	112 279.6	24 193	1 172 652.0	41 964.4
2003	74 940	474 015.0	113 806.5	25 115	1 242 449.0	44 022.6
2004	74 662	489 369.1	117 172.1	26 113	1 304 874.0	45 731.7
2005	76 580	507 463.4	120 622.8	26 601	1 379 457.0	47 809.4
2006	78 110	538 125.4	125 186.8	27 997	1 455 644.0	49 952.2
2007	80 272	573 080.3	129 962.4	29 374	1 530 890.0	52 108.8
2008	82 176	597 380.5	132 282.7	29 728	1 564 252.0	52 796.9
2009	83 380	587 060.7	129 319.7	30 509	1 519 459.0	52 117.9
2010	83 691	606 145.7	133 081.2	31 087	1 572 439.0	53 800.9
2011	84 128	618 324.8	132 638.1	31 426	1 628 274.0	55 431.5
2012	84 842	623 610.9	131 776.4	31 878	1 675 044.0	56 409.7
2013	85 563	634 776.0	132 258.1	32 177	1 739 563.0	57 904.8
2014	86 069	643 784.4	131 731.7	32 453	1 822 480.0	59 260.4
2015	85 422	645 555.7	130 083.5	33 102	1 872 714.0	59 844.5

	Варіант 31		
	США		
	Національна валюта: долар США		
Роки	**Середньорічна заробітна плата (*Y*), од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Продуктивність праці (X_2), од. н.в. на 1 зайнятого**
	wage	***GDP***	***productivity***
1996	32 268	8 100 201.0	62 717.4
1997	33 644	8 608 515.0	65 248.6
1998	35 306	9 089 168.0	67 938.6
1999	36 771	9 660 624.0	71 149.6
2000	38 963	10 284 779.0	73 898.2
2001	40 050	10 621 824.0	76 294.1

2002	40 898	10 977 514.0	79 084.7
2003	42 195	11 510 670.0	82 169.8
2004	44 061	12 274 928.0	86 706.3
2005	45 364	13 093 726.0	90 941.3
2006	47 339	13 855 888.0	94 464.7
2007	49 491	14 477 635.0	97 627.3
2008	50 800	14 718 582.0	99 690.3
2009	51 208	14 418 739.0	101 409.7
2010	52 384	14 964 372.0	105 840.5
2011	53 806	15 517 926.0	109 140.5
2012	55 266	16 155 254.8	111 608.0
2013	55 733	16 691 517.0	114 180.8
2014	57 417	17 393 103.0	117 115.0
2015	58 714	18 036 648.0	119 448.0

ЛАБОРАТОРНА РОБОТА 2
ПОБУДОВА ТА ДОСЛІДЖЕННЯ НЕЛІНІЙНОЇ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ НА ПРИКЛАДІ НЕОКЛАСИЧНОЇ ВИРОБНИЧОЇ ФУНКЦІЇ КОББА-ДУГЛАСА ТА ЇЇ ВИКОРИСТАННЯ ДЛЯ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ ПРОЦЕСУ ВИРОБНИЦТВА

Мета: набуття практичних навичок побудови нелінійної багатофакторної економетричної моделі на прикладі неокласичної виробничої функції Кобба-Дугласа та її використання для аналізу та прогнозування процесу виробництва.

Задачі роботи

1. Оцінювання параметрів неокласичної виробничої функції Кобба-Дугласа.
2. Дослідження побудованої моделі.
3. Аналіз виробництва на основі побудованої моделі.
4. Прогнозування на основі побудованої моделі.

Завдання роботи

На основі даних варіанта (додаток 1) щодо ВВП (Y, млн од. н.в.), вартості основного капіталу (K, млн од. н. в.) та чисельності зайнятих (L, тис. осіб) необхідно:

1. Побудувати діаграми розсіювання для логарифмів залежної змінної та показників-факторів.
2. Побудувати неокласичну виробничу функцію Кобба-Дугласа для однієї з країн ОЕСР:

$$Y = aK^{\alpha}L^{\beta},$$

де Y – випуск продукції;

F – вартість основного капіталу;

L – чиссльність зайнятих;

a, α, β – параметри моделі.

3. Оцінити якість побудованої виробничої функції на рівні значимості $\alpha = 0{,}05$.

4. Оцінити вплив виробничих ресурсів на річний випуск продукції.

5. Оцінити вплив зростання масштабів виробництва на темпи зростання випуску продукції та ефективність виробництва.

6. Для планового випуску продукції $Y = Y^*$ обчислити необхідну чисельність зайнятих у країні L^* на підставі припущення, що вартість основного капіталу залишиться на рівні останнього року вибірки. Плановий ВВП становитиме $100+(k+N)$ % від значення ВВП в останньому році вибірки, де k – номер групи + N – номер студента за списком.

7. Для планового випуску продукції $Y=Y^*$ обчислити необхідну вартість основного капіталу в країні K^* з огляду на припущення, що чисельність зайнятих у галузі залишиться на рівні останнього року вибірки. Плановий ВВП становить $100+(k+N)$ % від значення ВВП в останньому році вибірки, де k – номер групи + N – номер студента за списком.

8. Для прогнозних значень основного капіталу K_{pr} та кількості зайнятих країни L_{pr} обчислити середню та граничну продуктивність праці та основного капіталу. Прогнозне значення основного капіталу K_{pr} та кількості зайнятих країни L_{pr} визначається як $100+N$ % від їх значень в останньому році вибірки.

9. Для прогнозних значень основного капіталу K_{pr} і кількості зайнятих у країні L_{pr} розрахувати точковий прогноз ВВП.

Інструкції до виконання завдань

Спочатку створіть робочу папку з короткою назвою на диску D, наприклад Lr-2. Далі сформуйте масив вихідних даних у MS Excel, наберіть дані свого варіанта з десятковим розділювачем крапкою та без розділювача розрядів. Для змінних використайте короткі заголовки, наприклад *year, GDP, capital, labor*. Збережіть файл у форматі *.txt (текстовый файл с разделителями табуляции).

Після цього відкрийте *R* та встановіть робочу папку. Для цього зайдіть у вкладку *Session,* оберіть *Set Working directory* та натисніть кнопку *Choose Directory,* після цього вкажіть шлях до вашої папки.

Після обрання робочої папки активуйте пакети, потрібні для виконання лабораторної роботи 2, для цього в робочому середовищі *R* наберіть такі команди:

```
library("dplyr") # маніпуляції з даними
library("psych") # описові статистики
library("lmtest") # тестування гіпотез в лінійних моделях
library("sjPlot") # графіки
library("sgof")#
library("ggplot2") # графіки
library("foreign") # завантаження даних в різних форматах
library("hexbin") # графіки
```

Після того, як введені ці команди, послідовно поставте курсор на кожну з них та натисність клавіші *Ctrl+Enter.*

Після того, як пакети активовано, завантажте підготовлений файл з даними. Для цього оберіть ім'я таблиці, в якій буде зберігатись масив інформації, та напишіть його в робочому середовищі, наприклад *t2*, далі поставте знак менше і коротку рисочку «<-» і наберіть команду *read.csv.* Після цього відкрийте дужки, поставте лапки, помістіть курсор всередині них, натисніть клавішу *Tab* і оберіть файл з даними. Ця команда призначена для завантаження файлів у форматі ***.csv,** а оскільки потрібно завантажити файл у форматі **.txt,* опишіть його. Через кому напишіть, який використовується розділювач, у файлі **txt* – табуляція, тому наберіть *sep="\t".* Потім через кому напишіть, який використовується десятковий розділювач: *dec=".".* Останнє, що потрібно задати, чи використовуються заголовки: *header= TRUE.* Приклад:

```
t2<-read.csv("Australia.txt", sep="\t", dec=".", header= TRUE)
```

Далі натискайте клавіші *Ctrl+Enter.*

Для виконання завдання 1, а саме побудови діаграм розсіювання логарифмів залежної змінної та показників-факторів використайте функцію *qplot,* аргументами якої є вихідний масив інформації *data=t2,* ці змінні відображаються на осі абсцис та

ординат. Оскільки для побудови графіків необхідно використати логарифмовані дані, застосуйте функцію log(). Приклад написання команд:

```
qplot(data=t2,log(capital),log(GDP))
qplot(data=t2,log(labor),log(GDP))
```

Після введення команди потрібно натиснути клавіші *Ctrl+Enter*. Побудовані графіки з'являться у правому нижньому куті. Для того, щоб їх скопіювати, натисніть кнопку ***Export i*** оберіть необхідний варіант, наприклад *Copy to Clipboard.* Діаграми розсіювання будуть виглядати як на рис. 2.1 та рис. 2.2.

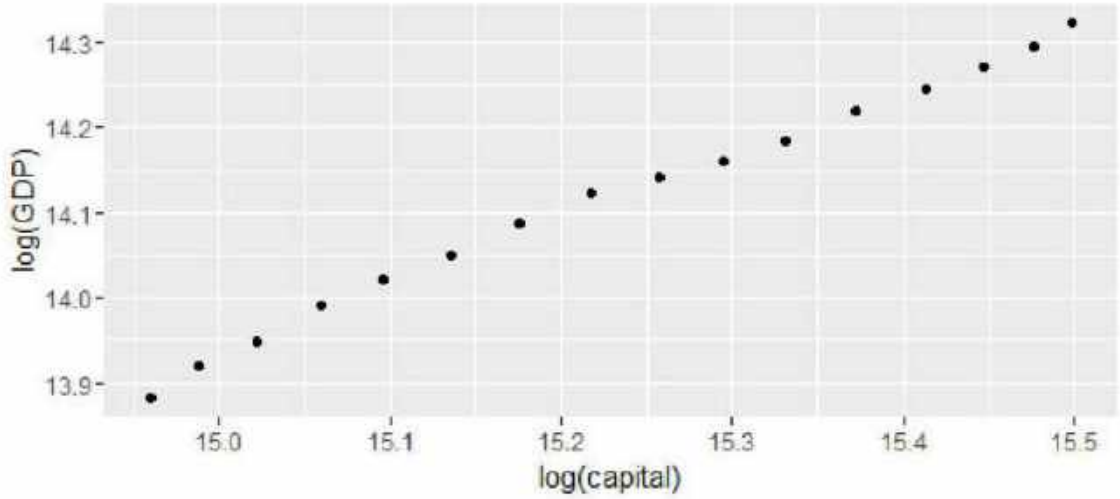


Рис. 2.1. Діаграма розсіювання між логарифмованими змінними ВВП та вартість основних фондів

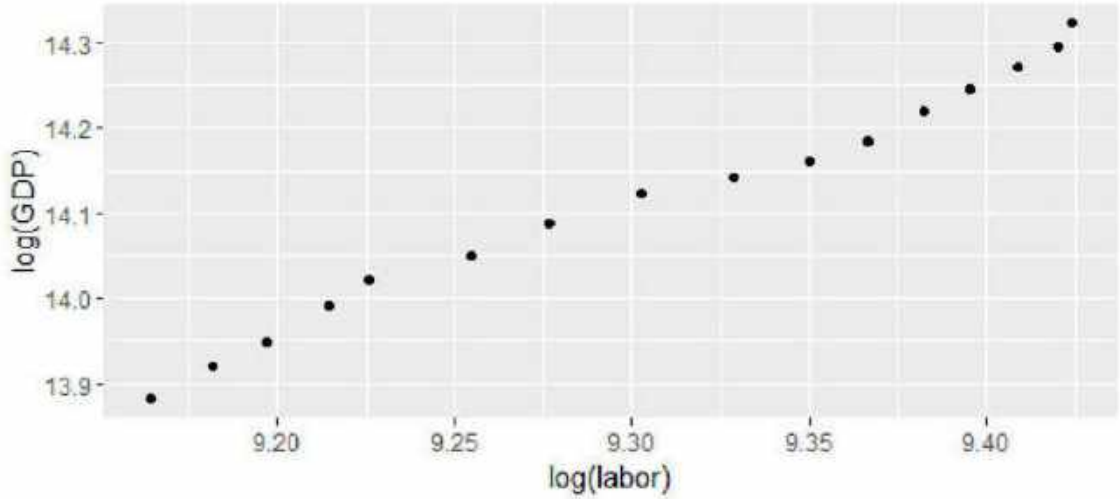


Рис. 2.2. Діаграма розсіювання між логарифмованими змінними ВВП та чисельність зайнятих

Після побудови графіків опишіть їх.

Далі будуємо економетричну модель. Для лінеаризації виробничої функції $Y = aK^{\alpha}L^{\beta}$ використовуємо метод логарифмування:

$$Ln(Y) = Ln(a) + \alpha\, Ln(K) + \beta\, Ln(L). \quad (2.1)$$

Виконуємо заміну змінних:

$$Y = Ln(Y),\ X_1 = Ln(K),\ X_2 = Ln(L). \quad (2.2)$$

Тепер можна будувати лінійну багатофакторну економетричну модель. Для цього у робочому середовищі записуємо назву моделі, наприклад *model,* далі знак менше та коротке тире «<-» і функцію `lm()`. Аргументами цієї функції є вихідні дані для аналізу: *data=t2,* логарифмована залежна змінна: *log(GDP)* потім *знак* «~» (тільда) і логарифмовані факторні змінні, які розділяються знаком «+». Приклад:

```
model<-lm(data=t2,
log(GDP)~log(capital)+log(labor))
```

Після введення команди необхідно натиснути клавіші *Ctrl+Enter.* Для отримання звіту побудованої моделі наберіть функцію `summary(model)` та натисніть клавіші *Ctrl+Enter,* звіт з'явиться в консолі (рис. 3).

Оцінки параметрів відображені у стовпчику *Estimate,* стандартні помилки – у стовпчику *Std. Error*, t-статистика – в стовпчику *t value*. У кожному рядку можна побачити зірочки, їх кількість означає рівень, з яким оцінки параметрів можна вважати значимими. Далі представлене стандартне відхилення вектора залишків на одну ступінь свободи *(Residual standard error),* коефіцієнт детермінації *(Multiple R-squared),* скоригований коефіцієнт детермінації *(Adjusted R-squared),* критерій Фішера *(F-statistic)* і його ступені свободи.

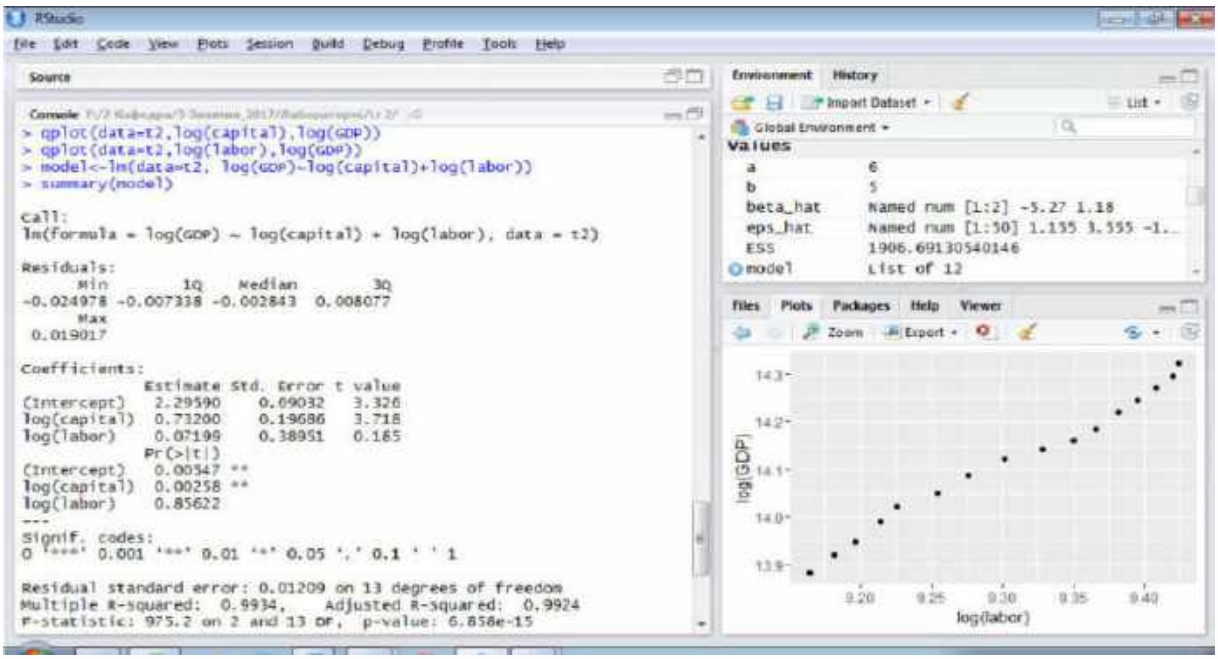


Рис. 2.3. Протокол лінеаризованої виробничої функції

Тепер можна записати лінеаризовану виробничу функцію:

$$\hat{Y} = 2{,}3 + 0{,}73 \cdot X_1 + 0{,}7 \cdot X_2.$$

Для отримання виробничої функції у нелінійній формі необхідно виконати зворотні перетворення. Параметри a, α, β визначаються на основі оцінок параметрів лінеаризованої функції за такими залежностями:

$$a = e^{\hat{\beta}_0},\ \alpha = \hat{\beta}_1,\ \beta = \hat{\beta}_2 \qquad (2.3)$$

Знайдіть значення експоненти оцінки ***a***, використавши функцію exp(). Приклад, написання: *exp(coef(model)[1])*. Натисніть клавіші *Ctrl+Enter*. У консолі отримаєте шукане значення:

```
> exp(coef(model)[1])
(Intercept)
9.933381
```

Тепер можна записати нелінійну форму виробничої функції:

$$\hat{Y} = 9{,}93 \cdot K^{0{,}73} \cdot L^{0{,}07}$$

Після отримання звіту з побудованої економетричної моделі виконайте завдання 3. Для отримання критичного значення t-статистики Стьюдента при рівні значимості $\alpha = 0{,}05$ використайте функцію `qt()` з аргументами: ймовірність та число ступенів свободи. Приклад: `qt(0.95,13)`. Після введення команди необхідно натиснути клавіші *Ctrl+Enter*. Для знаходження критичного значення

F-статистики з рівнем значимості $\alpha = 0{,}05$ використайте функцію `qf()` з аргументами ймовірність та число ступенів свободи. Приклад: `qf(0.95,2,13)`. Після введення команди треба натиснути клавіші *Ctrl+Enter*. Опишіть отримані результати.

Для виконання завдання 4, а саме оцінки впливу виробничих ресурсів на річний випуск продукції, необхідно визначити часткові коефіцієнти еластичності випуску продукції за виробничими ресурсами за формулами:

$$E_K = \alpha;\ E_L = \beta, \qquad (2.4)$$

де E_K – коефіцієнт еластичності випуску продукції за основним капіталом;

E_L – коефіцієнт еластичності випуску продукції за працею.

Для побудованої виробничої функції Австралії часткові коефіцієнти еластичності становлять $E_K = 0{,}73$; $E_L = 0{,}07$. Це означає, що зі зростанням вартості виробничих фондів на 1 % ВВП зросте на 0,73 %, а зі зростанням чисельності зайнятих на 1 % ВВП збільшиться на 0,07 %.

Для виконання завдання 5, а саме оцінки впливу зростання масштабів виробництва на темпи зростання випуску продукції та ефективність виробництва, необхідно знайти загальний коефіцієнт еластичності, який визначається за формулою:

$$E = E_K + E_L = \alpha + \beta. \qquad (2.5)$$

Залежно від цього значення за таблицею, яка наведена в додатку 2, потрібно зробити висновок про вплив зростання масштабів виробництва на темпи зростання випуску продукції і ефективність виробництва.

Для виробничої функції економіки Австралії загальний коефіцієнт еластичності становитиме: E = 0,73 + 0,07 = 0,8. Це означає, що темпи зростання випуску продукції нижчі за темпи зростання ресурсів. Отже, в Австралії відбувається зниження ефективності виробництва зі зростанням його масштабів, зростання витрат ресурсів на одиницю продукції.

Для виконання завдання 6, а саме: розрахунку необхідної чисельності зайнятих для планового випуску продукції $Y = Y^*$, нсобхідно використати таку формулу:

$$L^*=\left(\frac{Y^*}{aK^\alpha}\right)^{\frac{1}{\beta}}, \quad (2.6)$$

де Y^* – плановий випуск продукції;

L^* – необхідна чисельність зайнятих.

Для економіки Австралії за умови, що ВВП становитиме Y^*= 1 827 913 млн австралійських доларів, а обсяг виробничих фондів буде K = 5 383 272 млн австралійських доларів, необхідна чисельність зайнятих буде:

$$L^*=\left(\frac{1827913}{9{,}93\cdot 5\,383\,272^{0{,}73}}\right)^{\frac{1}{0{,}07}} = 49\,152{,}57 \text{ тис. осіб.}$$

Наведемо приклад виконаних розрахунків в R. У робочому середовищі вводимо команди для розрахунку планового ВВП та вартості основного капіталу. Для вилучення з таблиці вихідних даних значення ВВП в останньому році вибірки та його збільшення на 10 % (якщо, наприклад, студент групи 1 і 9-тий за списком) використайте функцію *t2[16,2]*1.1*. Пояснення: останній рік вибірки, це 16-те значення 2-го стовпчика таблиці вихідних даних *t2*. Аналогічні операції застосуйте для вартості основних фондів. Приклад введення команди:

```
#задаємо планове значення ВВП
y_plan<-t2[16,2]*1.1
#задаємо вартість основного капіталу
k<-t2[16,3]
```

Далі задайте, чому дорівнюють показники *a*, α, β:

```
#задаємо значення a
a<-exp(coef(model)[1])
#задаємо альфа
alfa<-coef(model)[2]
#задаємо бета
beta<-coef(model)[3]
```

Введіть команду для розрахунку необхідної чисельності зайнятих:

(y_plan/(a*(k^alfa)))^(1/beta)

У консолі отримаєте шукане значення:

```
> (y_plan/(a*(k^alfa)))^(1/beta)
(Intercept)
```

```
49152.57
```

Для виконання завдання 7, а саме знаходження планового випуску продукції $Y = Y^*$, необхідна вартість основного капіталу K^* розраховується за формулою:

$$K^* = \left(\frac{Y^*}{aL^{\beta}}\right)^{\frac{1}{\alpha}}, \tag{2.7}$$

де Y^* – плановий випуск продукції;

K^* – необхідна вартість основних фондів.

Для економіки Австралії за умови, що ВВП складе $Y^* = 1827913$ млн австралійських доларів, а чисельність зайнятих буде $L = 12\,385{,}2$ тис. осіб, необхідна вартість основного капіталу становитиме:

$$K^* = \left(\frac{1827913}{9{,}93 \cdot 12\,385{,}2^{\,0{,}07}}\right)^{\frac{1}{0{,}73}} = 6\,164\,859$$ млн австралійських доларів.

Приклад розрахунків в R:

У робочому середовищі задаємо чисельність зайнятих:

```
l<-t2[16,4]
```

Пояснення: 16-те значення 4-го стовпчика таблиці **t2**.

```
#знаходимо необхідну вартість основного капіталу в країні
(y_plan/(a*(l^beta)))^(1/alfa)
```

В консолі отримуємо шукане значення:

```
> (y_plan/(a*(l^beta)))^(1/alfa)
(Intercept)
 6164859
```

Для виконання завдання 8, а саме обчислення середньої та граничної продуктивності праці та основного капіталу, необхідно використати такі формули:

Середня та гранична продуктивність праці:

$$AP_L = \frac{Y}{L} = aK_{pr}{}^{\alpha} L_{pr}{}^{\beta-1}, \tag{2.8}$$

$$MP_L = \frac{dY}{dL} = \beta \cdot AP_L, \tag{2.9}$$

де AP_L – середня продуктивність праці;

MP_L – гранична продуктивність праці;

K_{pr} – прогнозне значення вартості основного капіталу;

L_{pr} – прогнозне значення чисельності зайнятих.

Середня та гранична фондовіддача:

$$AP_K = \frac{Y}{K} = aK_{pr}^{\alpha-1} L_{pr}^{\beta}, \quad (2.10)$$

$$MP_K = \frac{dY}{dK} = \alpha \cdot AP_K, \quad (2.11)$$

де AP_K – середня фондовіддача;

MP_K – гранична фондовіддача;

K_{pr} – прогнозне значення вартості основного капіталу;

L_{pr} – прогнозне значення чисельності зайнятих.

Розрахунок середнього та гранична продуктивність праці для прикладу в *R*:

```
#задаємо прогнозну вартість основного капіталу
k_predict<-t2[16,3]*1.09
#задаємо прогнозне значення чисельності зайнятих в країні
l_predict<-t2[16,4]*1.09
#знаходимо середню продуктивність праці
a*(k_predict^alfa)*(l_predict^(beta-1))
#задаємо значення середньої продуктивності праці
APl<-a*(k_predict^alfa)*(l_predict^(beta-1))
#знаходимо граничну продуктивність праці
beta*APl
```

У консолі отримуємо шукані значення:

```
> k_predict<-t2[16,3]*1.09
> l_predict<-t2[16,4]*1.09
> a*(k_predict^alfa)*(l_predict^(beta-1))
(Intercept)
 131.4071
>  APl<-a*(k_predict^alfa)*(l_predict^(beta-1))
> beta*APl
log(labor)
 9.460284
```

Рохрахунок середньої та граничної фондовіддачі для прикладу в *R*:

```
a*(k_predict^(alfa-1))*l_predict^beta
APk<-a*(k_predict^(alfa-1))*l_predict^beta
alfa*APk
```

У консолі отримуємо шукані значення:

```
> a*(k_predict^(alfa-1))*l_predict^beta
(Intercept)
0.302326
> APk<-a*(k_predict^(alfa-1))*l_predict^ beta
> alfa*APk
log(capital)
0.2213017
```

Отже, в Австралії середня продуктивність праці становитиме 131,4 млн австралійських доларів на 1 тис. зайнятих осіб, а гранична продуктивність відповідно – 9,5 австралійських доларів на 1 тис. зайнятих осіб. Середня фондовіддача становить 0,3 австралійських долара на 1 долар вартості основних фондів, гранична фондовіддача – 0,2 австралійських долара на 1 долар вартості основних фондів.

Рахуємо прогнозне значення ВВП. Приклад у R:

```
#формуємо масив прогнозних значень ознак факторів
newt2        <-data.frame(capital=k_predict, labor=l_predict)
#рахуємо точковий прогноз ВВП
exp(predict(model, newdata=newt2))
```

У консолі відобразится шукане значення точкового прогнозу:

```
>exp(predict(model, newdata=newt2))
1
1773979
```

Можна зробити висновок, що якщо у прогнозному періоді вартість основних фондів буде 5 867 766 млн австралійських доларів, а чисельність занятих становитиме 13499,87 тис. осіб, ВВП складе 1 773 979 млн австралійських доларів.

Додаток А до виконання лабораторної роботи 2

ВИХІДНІ ДАНІ ДЛЯ ВИКОНАННЯ ЛАБОРАТОРНОЇ РОБОТИ 2

	Варіант 1			**Варіант 2**		
	Австралія			**Австрія**		
	Національна валюта: австралійський долар			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	529 598	1 689 651.0	9 003.5	176 183	918 525.2	3 588.6
1996	557 246	1 757 819.0	9 127.6	182 133	962 377.7	3 602.5
1997	589 394	1 854 340.0	9 187.7	188 310	1 006 123.5	3 628.8
1998	621 000	1 968 247.0	9 281.5	195 828	1 045 945.2	3 665.6
1999	661 515	2 104 951.0	9 366.4	203 418	1 089 279.1	3 721.2
2000	706 042	2 270 417.0	9 549.4	213 196	1 139 906.8	3 756.5
2001	754 228	2 390 555.0	9 716.6	220 096	1 182 946.7	3 783.5
2002	801 587	2 566 012.0	9 865.5	226 303	1 223 436.1	3 779.9
2003	862 219	2 789 572.0	10 038.7	230 999	1 275 466.2	3 804.5
2004	922 845	3 033 028.0	10 155.4	241 505	1 336 553.6	3 828.3
2005	998 458	3 304 705.0	10 455.1	253 009	1 402 074.1	3 873.8
2006	1 087 440	3 593 938.0	10 683.7	266 478	1 477 034.0	3 941.4
2007	1 178 809	3 880 130.0	10 964.1	282 347	1 571 223.1	4 013.7
2008	1 259 280	4 096 045.0	11 258.6	291 930	1 661 552.0	4 090.3
2009	1 297 508	4 284 299.0	11 498.8	286 188	1 727 532.1	4 072.7

2010	1 410 442	4 498 069.0	11 687.2	294 628	1 799 345.8	4 102.2
2011	1 491 741	4 734 255.0	11 877.2	308 630	1 873 817.3	4 166.0
2012	1 527 529	5 017 142.0	12 029.5	317 117	1 950 255.3	4 209.7
2013	1 589 940	5 330 601.0	12 193.2	322 539	2 024 045.5	4 223.8
2014	1 617 016	5 618 422.0	12 335.8	330 418	2 096 255.3	4 263.2
2015	1 654 864	5 822 960.0	12 385.2	339 896	2 168 144.6	4 290.7
2016	1 820 350	6 405 256.0	12 434.7	349 493	2 384 959.0	4 347.9
	Варіант 3			**Варіант 4**		
	Бельгія			**Канада**		
	Національна валюта: євро			**Національна валюта: канадський долар**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	211 616	1 001 131.5	3 867.9	828 973	1 861 477.0	14 759.2
1996	215 942	1 030 980.9	3 878.1	857 023	1 939 921.0	14 912.5
1997	225 969	1 069 121.3	3 904.5	903 902	2 038 958.0	15 141.6
1998	234 490	1 109 733.5	3 973.0	937 295	2 143 191.0	15 374.8
1999	244 226	1 162 044.9	4 027.8	1 004 456	2 220 374.0	15 643.7
2000	258 222	1 215 646.1	4 108.9	1 102 380	2 340 688.0	15 898.9
2001	265 788	1 260 610.5	4 165.2	1 140 505	2 446 490.0	16 147.3
2002	275 065	1 291 965.8	4 157.3	1 189 452	2 529 293.0	16 615.1
2003	282 637	1 353 132.0	4 153.9	1 250 315	2 569 898.0	16 999.6
2004	298 711	1 419 427.3	4 194.7	1 331 178	2 701 431.0	17 202.7
2005	311 481	1 494 151.0	4 255.2	1 417 028	2 869 179.0	17 351.4
2006	326 662	1 594 818.6	4 302.9	1 492 207	3 096 806.0	17 562.9

2007	344 713	1 682 722.4	4 373.9	1 573 532	3 333 139.0	17 908.3
2008	354 066	1 796 608.6	4 452.9	1 652 923	3 653 192.0	18 185.0
2009	348 781	1 804 152.0	4 445.5	1 567 365	3 767 211.0	18 314.6
2010	365 101	1 871 206.3	4 474.0	1 662 131	3 872 819.0	18 517.0
2011	379 106	1 953 293.9	4 535.3	1 769 922	4 073 927.0	18 685.7
2012	387 500	2 022 685.3	4 555.0	1 822 808	4 298 790.0	18 875.2
2013	391 712	2 061 942.8	4 540.1	1 897 532	4 486 443.0	19 103.3
2014	400 805	2 094 044.8	4 559.1	1 983 117	4 760 749.0	19 189.3
2015	410 351	2 122 545.6	4 601.2	1 986 193	4 986 438.0	19 266.1
2016	421 974	2 334 800.2	4 660.6	2 184 812	5 485 081.8	19 343.1
	Варіант 5			**Варіант 6**		
	Чехія			**Данія**		
	Національна валюта: чеська крона			**Національна валюта: датська крона**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	1 580 115	12 087 629.0	5 105.5	1 036 483	5 834 574.0	2 612.0
1996	1 812 622	13 357 810.0	5 132.0	1 088 024	6 095 784.0	2 638.0
1997	1 953 311	15 174 843.0	5 096.0	1 146 129	6 380 047.0	2 674.0
1998	2 142 587	16 822 652.0	5 007.6	1 185 988	6 634 294.0	2 712.0
1999	2 237 300	17 945 395.0	4 899.7	1 241 521	6 944 124.0	2 735.0
2000	2 372 630	19 055 491.0	4 859.3	1 326 912	7 204 552.0	2 755.0
2001	2 562 679	20 135 397.0	4 846.4	1 371 526	7 612 999.0	2 782.0
2002	2 674 634	20 773 933.0	4 876.8	1 410 271	7 844 200.0	2 784.0
2003	2 801 163	21 456 267.0	4 838.1	1 436 751	8 133 480.0	2 759.0

2004	3 057 660	22 713 725.0	4 828.9	1 506 001	8 412 391.0	2 744.0
2005	3 257 972	23 722 687.0	4 922.6	1 585 984	8 786 900.0	2 783.0
2006	3 507 131	24 982 648.0	4 989.0	1 682 260	9 448 270.0	2 846.0
2007	3 831 819	27 188 673.0	5 093.1	1 738 845	9 963 088.0	2 912.0
2008	4 015 346	29 060 999.0	5 204.1	1 801 470	10 207 447.0	2 947.0
2009	3 921 827	29 429 585.0	5 110.1	1 722 143	9 670 373.0	2 854.0
2010	3 953 651	29 648 426.0	5 057.2	1 810 926	10 096 901.0	2 788.0
2011	4 033 755	30 161 561.0	5 043.4	1 846 854	10 463 368.0	2 787.0
2012	4 059 912	30 358 449.0	5 064.6	1 895 002	10 734 633.0	2 767.0
2013	4 098 128	30 608 105.0	5 080.9	1 929 677	10 912 724.0	2 766.0
2014	4 313 789	31 305 667.0	5 109.0	1 977 255	11 171 154.0	2 794.0
2015	4 554 615	32 511 002.0	5 179.7	2 027 171	11 434 131.0	2 829.0
2016	4 715 061	35 762 102.2	5 273.2	2 060 868	12 577 544.1	2 877.0
	Варіант 7			**Варіант 8**		
	Естонія			**Фінляндія**		
	Національна валюта: євро			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	2 779	17 221.9	633.6	98 556	541 991.0	2 053.6
1996	3 649	18 128.3	618.7	102 060	551 078.0	2 083.0
1997	4 495	19 082.4	619.0	110 738	575 517.0	2 154.5
1998	5 049	20 086.8	607.4	120 382	608 204.0	2 194.6
1999	5 379	21 144.0	580.3	126 923	641 142.0	2 249.3
2000	6 171	22 256.8	585.3	136 261	689 122.0	2 298.3

2001	6 976	25 839.8	588.6	144 437	732 824.0	2 331.6
2002	7 774	27 996.8	588.5	148 289	743 482.0	2 356.9
2003	8 709	31 229.4	601.1	151 569	762 092.0	2 360.2
2004	9 708	35 077.2	598.7	158 477	798 117.0	2 373.6
2005	11 262	40 380.4	612.3	164 387	850 882.0	2 410.6
2006	13 522	49 084.4	642.2	172 614	900 288.0	2 454.6
2007	16 246	59 386.4	643.4	186 584	973 815.0	2 506.9
2008	16 517	63 271.4	642.1	193 711	1 046 624.0	2 562.7
2009	14 146	63 723.0	576.6	181 029	1 049 368.0	2 501.0
2010	14 717	65 255.8	548.1	187 100	1 038 939.0	2 483.8
2011	16 668	69 258.1	584.0	196 869	1 087 629.0	2 515.5
2012	17 935	75 625.8	593.5	199 793	1 153 969.0	2 537.6
2013	18 890	85 489.5	600.9	203 338	1 183 627.0	2 519.6
2014	19 758	89 124.5	605.5	205 474	1 204 156.0	2 507.5
2015	20 252	98 037.0	622.9	209 511	1 219 382.0	2 498.9
2016	20 916	107 840.6	624.7	214 062	1 341 320.2	2 513.1
	Варіант 9			**Варіант 10**		
	Франція			**Німеччина**		
	Національна валюта: євро			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	1 224 967	5 180 151.0	23 686.0	1 898 880	9 866 399.0	2 612.0
1996	1 258 950	5 354 997.0	23 822.0	1 926 320	10 078 738.0	2 638.0
1997	1 299 739	5 515 687.0	24 012.0	1 967 090	10 317 720.0	2 674.0

1998	1 358 776	5 673 518.0	24 437.0	2 018 230	10 554 468.0	2 712.0
1999	1 408 159	5 881 342.0	25 033.0	2 064 880	10 756 454.0	2 735.0
2000	1 485 303	6 145 561.0	25 671.0	2 116 480	11 083 761.0	2 755.0
2001	1 544 629	6 430 781.0	26 037.0	2 179 850	11 325 928.0	2 782.0
2002	1 594 259	6 721 025.0	26 161.0	2 209 290	11 495 737.0	2 784.0
2003	1 637 438	7 074 489.0	26 162.0	2 220 080	11 656 475.0	2 759.0
2004	1 710 760	7 470 512.0	26 196.0	2 270 620	11 940 985.0	2 744.0
2005	1 771 978	7 925 283.0	26 378.0	2 300 860	12 195 808.0	2 783.0
2006	1 853 267	8 430 900.0	26 667.0	2 393 250	12 595 086.0	2 846.0
2007	1 945 670	8 994 591.0	27 042.0	2 513 230	13 388 182.0	2 912.0
2008	1 995 850	9 376 369.0	27 168.0	2 561 740	13 974 646.0	2 947.0
2009	1 939 017	9 554 351.0	26 863.0	2 460 280	14 280 852.0	2 854.0
2010	1 998 481	9 926 071.0	26 896.0	2 580 060	14 613 594.0	2 788.0
2011	2 059 284	10 341 016.0	27 107.0	2 703 120	15 188 039.0	2 787.0
2012	2 086 929	10 606 762.0	27 196.0	2 758 260	15 760 142.0	2 767.0
2013	2 115 256	10 779 808.0	27 267.0	2 826 240	16 291 222.0	2 766.0
2014	2 139 964	10 887 631.0	27 394.0	2 923 930	16 807 390.0	2 794.0
2015	2 181 064	10 934 645.0	27 523.0	3 032 820	17 280 324.0	2 829.0
2016	2 225 260	12 028 109.5	27 705.0	3 132 670	19 008 356.4	2 877.0

	Варіант 11			Варіант 12		
	Греція			Угорщина		
	Національна валюта: євро			Національна валюта: форинт		
	ВВП (*Y*), млн од. н.в.	Вартість основних фондів (*K*), млн од. н.в.	Чисельність зайнятих (*L*), тис. ос.	ВВП (*Y*), млн од. н.в.	Вартість основних фондів (*K*), млн од. н.в.	Чисельність зайнятих (*L*), тис. ос.
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	93 064	516 579.5	4 156.3	5 819 221	47 390 000.0	4 009.7
1996	103 037	568 385.4	4 132.7	7 103 898	60 297 000.0	4 012.7
1997	114 712	619 337.9	4 116.4	8 812 386	73 832 000.0	4 018.9
1998	125 263	679 344.3	4 297.3	10 433 144	84 402 000.0	4 081.5
1999	133 789	732 268.2	4 298.7	11 637 865	99 139 000.0	4 195.9
2000	141 247	788 540.5	4 312.8	13 321 531	110 832 000.0	4 237.0
2001	152 194	847 502.0	4 328.4	15 383 442	121 536 000.0	4 230.3
2002	163 461	907 843.3	4 434.3	17 421 576	128 293 000.0	4 226.9
2003	178 905	969 683.6	4 495.7	19 077 348	136 565 000.0	4 226.9
2004	193 716	1 030 199.9	4 604.1	21 023 630	144 569 000.0	4 186.4
2005	199 242	1 063 144.8	4 646.9	22 470 802	152 426 000.0	4 174.2
2006	217 862	1 165 336.3	4 731.3	24 153 022	166 105 000.0	4 192.2
2007	232 695	1 232 094.7	4 795.1	25 560 380	178 382 000.0	4 197.9
2008	241 990	1 311 157.3	4 856.4	27 071 868	191 024 000.0	4 115.5
2009	237 534	1 339 893.8	4 829.0	26 297 412	200 095 000.0	4 012.9
2010	226 031	1 347 474.0	4 705.5	27 085 900	205 966 000.0	3 969.3
2011	207 029	1 332 066.6	4 381.8	28 166 115	212 964 000.0	3 969.1
2012	191 204	1 318 348.3	4 105.2	28 660 518	220 125 000.0	3 975.4
2013	180 654	1 364 191.5	3 997.7	30 127 349	227 090 000.0	4 019.2

2014	177 941	1 356 909.4	3 999.3	32 400 148	234 473 000.0	4 213.3
2015	175 697	1 492 600.3	4 019.8	33 999 012	257 920 300.0	4 309.2
2016	175 888	1 641 860.3	4 071.6	35 005 439	283 712 330.0	4 402.8
	Варіант 13			**Варіант 14**		
	Ірландія			**Ізраїль**		
	Національна валюта: євро			**Національна валюта: шекелі**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	54 813	183 426.4	1 458.6	301 969	921 250.5	1 976.0
1996	60 206	197 468.9	1 507.4	350 918	1 057 447.2	2 023.9
1997	69 367	219 584.0	1 539.0	395 689	1 219 870.7	2 047.7
1998	80 353	254 525.3	1 634.4	440 630	1 384 208.9	2 086.8
1999	92 716	294 530.4	1 709.5	485 156	1 499 828.3	2 153.7
2000	108 383	342 656.4	1 766.2	539 546	1 621 458.5	2 234.2
2001	121 941	391 405.8	1 808.0	549 417	1 714 335.7	2 279.2
2002	135 951	434 795.7	1 851.3	572 970	1 914 818.7	2 294.6
2003	145 557	484 187.6	1 887.5	576 927	1 966 740.9	2 343.7
2004	156 144	547 891.2	1 941.0	606 706	2 133 087.3	2 416.6
2005	170 216	616 618.5	2 041.5	639 329	2 300 751.6	2 514.9
2006	185 061	691 299.2	2 134.1	685 647	2 428 635.3	2 601.9
2007	197 293	716 918.8	2 243.6	734 290	2 571 171.2	2 720.9
2008	187 687	676 939.0	2 278.3	774 758	2 728 929.7	2 821.2
2009	169 704	623 833.6	2 248.9	815 248	2 794 528.3	2 866.2
2010	167 124	614 397.3	2 198.7	874 010	2 932 810.0	2 967.7

2011	173 070	628 244.5	2 178.7	935 225	3 157 496.6	3 056.6
2012	175 754	662 604.6	2 159.1	993 440	3 391 852.2	3 399.6
2013	180 209	694 360.7	2 171.0	1 059 101	3 553 630.4	3 462.8
2014	193 160	756 428.6	2 156.1	1 104 746	3 723 440.2	3 476.7
2015	255 815	832 071.5	2 164.7	1 163 769	3 837 051.6	3 490.6
2016	265 835	915 278.6	2 173.4	1 280 146	4 220 756.8	3 504.6

	Варіант 15			**Варіант 16**		
	Італія			**Корея**		
	Національна валюта: євро			**Національна валюта: вона**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	984 983	4 485 659.1	21 911.0	428 927 100	1 317 484 400.0	12 618.2
1996	1 043 086	4 712 486.6	22 044.2	481 140 800	1 548 037 400.0	12 461.1
1997	1 089 869	4 942 444.9	22 113.8	530 347 100	1 828 989 000.0	12 600.7
1998	1 135 499	5 156 242.2	22 336.4	524 476 800	2 005 433 800.0	12 743.3
1999	1 171 901	5 333 394.3	22 581.0	576 872 800	2 132 332 600.0	12 933.3
2000	1 239 266	5 654 244.6	23 021.3	635 184 600	2 339 528 500.0	13 061.4
2001	1 298 890	5 913 462.7	23 473.3	688 164 900	2 566 672 300.0	13 161.3
2002	1 345 794	6 246 083.0	23 867.4	761 938 900	2 836 229 500.0	13 301.4
2003	1 390 710	6 522 752.9	24 217.9	810 915 300	3 168 989 600.0	13 399.7
2004	1 448 363	6 902 736.2	24 364.6	876 033 100	3 483 471 400.0	13 500.0
2005	1 489 725	7 313 471.8	24 501.3	919 797 300	3 748 165 700.0	13 606.4
2006	1 548 473	7 667 159.3	24 983.8	966 054 600	4 041 977 600.0	13 700.4
2007	1 609 551	8 103 196.1	25 294.9	1 043 257 800	4 546 468 500.0	13 801.9

2008	1 632 151	8 556 666.6	25 349.2	1 104 492 200	5 085 378 600.0	13 888.4
2009	1 572 878	8 806 282.6	24 925.5	1 151 707 800	5 463 126 500.0	13 988.0
2010	1 604 515	9 145 068.1	24 765.7	1 265 308 000	5 925 583 300.0	14 153.6
2011	1 637 463	9 600 327.5	24 842.7	1 332 681 000	6 392 111 500.0	14 284.1
2012	1 613 265	9 813 064.4	24 764.8	1 377 456 700	6 688 080 400.0	14 399.4
2013	1 604 599	9 884 452.8	24 322.8	1 429 445 400	6 949 137 500.0	14 504.6
2014	1 621 827	9 955 364.7	24 339.4	1 486 079 300	7 262 100 300.0	14 562.6
2015	1 645 439	9 990 479.7	24 503.6	1 558 591 600	7 988 310 330.0	14 620.9
2016	1 672 438	10 989 527.6	24 814.4	1 714 450 760	8 787 141 363.0	14 679.4

	Варіант 17			**Варіант 18**		
	Латвія			**Люксембург**		
	Національна валюта: євро			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	4 059	62 812.4	930.2	15 777	62 078.7	216.1
1996	4 678	67 079.5	935.4	16 714	64 665.4	221.7
1997	5 394	72 838.2	977.0	17 498	67 359.8	229.2
1998	6 022	80 167.4	973.4	18 184	70 166.4	239.3
1999	6 272	81 627.2	955.9	20 871	73 090. 0	250.8
2000	6 850	79 953.7	923.7	23 079	76 135.4	264.1
2001	7 460	76 931.2	938.7	23 772	82 590.3	279.1
2002	8 397	83 677.9	952.3	25 095	87 410.2	287.2
2003	9 553	91 806.2	958.3	26 188	91 299.1	292.5
2004	11 049	99 542.1	960.6	27 936	96 694.0	299.4

2005	13 597	117 837.9	969.2	30 031	102 451.6	307.7
2006	17 102	160 757.8	1 025.0	33 808	108 072.4	319.5
2007	22 592	204 097.4	1 064.1	37 179	115 675.5	333.6
2008	24 351	222 452.9	1 055.1	38 129	123 843.6	349.5
2009	18 827	168 104.9	903.7	36 977	129 620.7	353.1
2010	17 938	166 866.5	843.5	40 178	135 779.6	359.6
2011	20 303	187 979.6	856.2	43 165	143 916.7	370.2
2012	21 886	206 897.7	868.6	44 112	153 083.6	379.1
2013	22 787	209 410.4	888.6	46 551	160 810.8	386.1
2014	23 631	217 675.8	876.6	49 971	169 105.9	396.0
2015	24 368	239 443.4	887.9	52 340	177 868.2	406.3
2016	25 021	263 387.7	887.4	54 195	195 655.0	418.5
	Варіант 19			**Варіант 20**		
	Нідерланди			**Польща**		
	Національна валюта: євро			**Національна валюта: злотий**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисель-ність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	325 341	1 747 169.8	7 276.0	344 683	2 289 489.7	17 205.0
1996	340 964	1 782 826.3	7 427.0	431 222	2 360 298.6	17 199.8
1997	364 961	1 819 210.6	7 656.0	521 795	2 433 297.6	17 225.3
1998	389 315	1 856 337.3	7 838.0	606 069	2 508 554.2	17 284.5
1999	414 838	1 894 221.7	8 063.0	673 287	2 586 138.3	17 255.5
2000	448 061	1 932 879.3	8 210.0	747 032	2 666 122.0	17 405.0
2001	476 696	2 075 076.2	8 371.0	779 975	2 871 797.0	17 457.3

2002	494 501	2 212 392.1	8 428.0	810 617	3 031 765.0	17 277.0
2003	506 671	2 324 297.5	8 381.0	845 930	3 149 840.0	17 014.0
2004	523 939	2 409 969.3	8 283.0	933 062	3 247 961.0	17 085.0
2005	545 609	2 497 922.6	8 339.0	990 468	3 332 499.0	17 218.0
2006	579 212	2 619 145.8	8 519.0	1 069 824	3 454 636.0	16 992.0
2007	613 280	2 732 427.6	8 771.0	1 187 605	3 705 334.0	16 909.0
2008	639 163	2 873 843.5	8 914.0	1 286 069	3 936 314.0	17 058.0
2009	617 540	2 993 863.7	8 838.0	1 372 208	4 010 376.0	17 126.2
2010	631 512	3 075 692.3	8 778.0	1 445 297	4 149 452.0	17 194.7
2011	642 929	3 087 542.6	8 854.0	1 566 813	4 360 529.0	17 263.5
2012	645 164	3 104 342.9	8 836.0	1 629 392	4 533 206.0	17 332.6
2013	652 748	3 127 529.0	8 732.0	1 656 842	4 731 307.0	17 401.9
2014	663 008	3 125 940.0	8 711.0	1 719 704	4 882 665.0	17 471.5
2015	676 531	3 124 323.0	8 792.0	1 799 321	5 370 931.5	17 541.4
2016	696 871	3 436 755.3	8 887.0	1 851 171	5 908 024.7	17 611.6
	Варіант 21			**Варіант 22**		
	Словаччина			**Словенія**		
	Національна валюта: євро			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисель-ність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	19 686	290 057.4	2 516.9	10 522	92 297.3	922.5
1996	21 951	295 976.9	2 519.4	12 132	94 180.9	904. 2
1997	24 416	302 017.3	2 521.9	13 827	96 103.0	887.7
1998	26 664	308 180.9	2 544.8	15 337	98 064.2	886.7

1999	28 548	314 470.3	2 573.0	17 210	100 065.6	900.4
2000	31 601	320 888.0	2 608.2	18 902	102 107.7	914.4
2001	34 311	327 436.8	2 652.5	21 147	111 503.6	919.6
2002	37 280	334 119.2	2 628.2	23 622	119 398.4	934.2
2003	41 404	340 937.9	2 634.3	25 668	128 712.3	931.3
2004	46 102	347 895.8	2 658.6	27 673	142 029.5	934.0
2005	50 415	361 816.9	2 645.7	29 227	153 413.8	929.5
2006	56 273	385 292.7	2 656.3	31 555	163 037.4	944.1
2007	63 054	401 909.9	2 666.9	35 153	176 483.4	975.8
2008	68 492	439 560.7	2 677.6	37 951	190 734.1	1 000.8
2009	64 023	455 534.5	2 688.3	36 166	195 686.4	982.9
2010	67 577	467 092.3	2 699.0	36 252	202 864.0	962.1
2011	70 627	484 883.9	2 709.8	36 896	211 214.3	946.0
2012	72 704	496 030.0	2 720.7	36 002	217 351.0	937.2
2013	74 170	511 458.0	2 731.6	35 917	222 492.9	926.8
2014	75 946	529 979.9	2 742.5	37 332	226 758.7	930.9
2015	78 686	547 099.7	2 753.5	38 570	232 302.4	941.5
2016	80 958	601 809.6	2 764.5	39 769	255 532.6	960.6
	Варіант 23			**Варіант 24**		
	Великобританія			**Литва**		
	Національна валюта: фунт стерлінгів			**Національна валюта: євро**		
	ВВП (*Y*), млн од. н.в.	**Вартість основних фондів (*K*), млн од. н.в.**	**Чисельність зайнятих (*L*), тис. ос.**	**ВВП (*Y*), млн од. н.в.**	**Вартість основних фондів (*K*), млн од. н.в.**	**Чиссль-ність зайнятих (*L*), тис. ос.**
	GDP	***capital***	***labor***	***GDP***	***capital***	***labor***
1995	836 646	2 760 219.0	25 814.3	7 765	73 538.0	1 483.2

1996	892 900	2 872 556.0	26 056.3	9 714	75 038.8	1 497.0
1997	938 855	3 072 416.0	26 522.8	11 724	76 570.2	1 506.2
1998	980 308	3 224 589.0	26 793.5	13 022	78 132.9	1 494.4
1999	1 021 205	3 420 128.0	27 167.1	12 712	79 727.4	1 461.5
2000	1 080 863	3 625 052.0	27 483.3	13 368	81 354.5	1 399.2
2001	1 120 575	3 851 111.0	27 711.4	14 195	82 618.7	1 346.0
2002	1 172 652	4 106 059.0	27 943.9	15 202	84 793.2	1 394.7
2003	1 242 449	4 337 869.0	28 223.1	16 668	89 608.2	1 425.7
2004	1 304 874	4 607 744.0	28 533.5	18 238	97 582.3	1 410.1
2005	1 379 457	5 011 174.0	28 853.2	21 002	108 438.6	1 420.7
2006	1 455 644	5 293 719.0	29 140.5	24 079	124 517.1	1 416.3
2007	1 530 890	5 666 291.0	29 378.8	29 041	144 422.7	1 445.1
2008	1 564 252	6 108 343.0	29 627.6	32 696	151 355.2	1 426.1
2009	1 519 459	6 159 608.0	29 154.3	26 935	145 170.8	1 316.5
2010	1 572 439	6 246 735.0	29 226.9	28 028	145 447.1	1 246.8
2011	1 628 274	6 518 197.0	29 374.6	31 275	154 171.6	1 252.7
2012	1 675 044	6 754 865.0	29 694.4	33 348	163 272.7	1 274.9
2013	1 739 563	7 053 639.0	30 041.7	35 002	172 084.9	1 291.9
2014	1 822 480	7 273 528.0	30 753.6	36 590	179 435.3	1 317.8
2015	1 872 714	7 539 802.0	31 293.1	37 331	197 378.9	1 334.7
2016	1 940 198	8 293 782.2	31 738.0	38 637	217 116.7	1 361.2

Додаток Б до лабораторної роботи 2

ОЦІНКА ВПЛИВУ ЗРОСТАННЯ МАСШТАБІВ ВИРОБНИЦТВА НА ТЕМПИ ЗРОСТАННЯ ВИПУСКУ ПРОДУКЦІЇ І ЕФЕКТИВНІСТЬ ВИРОБНИЦТВА

№ з/п	***E***	**Висновок про вплив зростання масштабів виробництва на темпи зростання випуску продукції і ефективність виробництва**
1	$E > 1$	темпи зростання випуску продукції вищі за темпи зростання виробничих ресурсів. Маємо зростання ефективності виробництва при зростанні масштабів виробництва, економію виробничих ресурсів
2	$E < 1$	темпи зростання випуску продукції нижчі за темпи зростання ресурсів. Маємо падіння ефективності виробництва при зростанні масштабів виробництва, зростання витрат ресурсів на одиницю продукції
3	$E = 1$	постійна ефективність виробництва, тобто темпи зростання випуску продукції дорівнюють темпу зростання виробничих ресурсів

ЛАБОРАТОРНА РОБОТА 3
ПОБУДОВА ТА ДОСЛІДЖЕННЯ ЛІНІЙНОЇ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ З МУЛЬТИКОЛІНЕАРНІСТЮ

Мета: набуття практичних навичок аналізу динаміки та структури й побудови багатофакторної економетричної моделі з мультиколінеарністю, її дослідження та практичне використання.

Задачі роботи

1. Специфікація економетричної моделі.
2. Побудова економетричної моделі МНК.
3. Дослідження економетричної моделі на мультиколінеарність.
4. Побудова кореляційної матриці.
5. Знаходження оптимального регуляційного параметра λ з застосуванням узагальненої перехресної перевірки на основі зовнішнього критерію крос-валідації GCV (*generalized cross-validation*).
6. Побудова гребеневої або рідж-регресії (ridge regression).
7. Формування економічних висновків на основі побудованої економетричної моделі.

Завдання роботи

На основі з даних інтерактивної статистичної бази країн ОЕСР (https://stats.oecd.org/) для однієї з країн проаналізувати динаміку та структуру надходжень персонального прибуткового податку. Побудувати економетричну модель залежності рівня надходжень персонального прибуткового податку (*Personal Income Tax*) (Y, млн од. національної валюти) від ВВП (X_1, млн одиниць національної валюти), заробітної плати (X_2, млн одиниць національної валюти), доходів від операцій з капіталом (X_3, млн одиниць національної валюти) для цього потрібно:

1. Знайти дані, побудувати та проаналізувати графік частки надходжень персонального прибуткового податку в досліджуваній країні.

2. Знайти дані, побудувати та проаналізувати структуру надходжень персонального прибуткового податку у розрізі складових у досліджуваній країні.

3. Побудувати діаграми розсіювання для результуючого показника та показників-факторів та обрати найкращу специфікацію економетричної моделі.

2. Побудувати економетричну модель залежності надходжень персонального прибуткового податку від визначених факторів.

3. Провести діагностику економетричної моделі на наявність мультиколінеарності.

4. Побудувати та проаналізувати кореляційну матрицю ознак-факторів.

5. Знайти оптимальний регуляційний параметр λ із застосуванням узагальненої перехресної перевірки на основі зовнішнього критерію крос-валідації GCV (generalized cross-validation). Побувати графік залежності значення параметра регуляризації λ від критерію крос-валідації (GCV).

6. Побудувати гребеневу або рідж-регресію (ridge regression) та сформулювати економічні висновки.

Приклад виконання роботи (досліджувана країна – Україна, часовий інтервал 2006–2015)

У податковій системі України податок на доходи фізичних осіб (далі – ПДФО) є одним із основних бюджетоформуючих. Надходження цього податку є стабільним і регулярним джерелом наповнення бюджету, має важливе фіскальне значення і широку базу оподаткування, а також здійснює безпосередній вплив на розвиток економіки держави. Саме тому дослідження податкових надходжень минулих років та виявлення факторів впливу на них є вкрай важливим та актуальним сьогодні.

Аналізуючи динаміку частки надходжень з ПДФО, варто зазначити, що протягом досліджуваного періоду вона була нестабільною і складала від 18 % податкових находжень у 2006 р.

до 20 % у III кв 2015 р. Починаючи з 2006 р. надходження зростали, що було обумовлено податковою реформою ПДФО (рис. 3.1). У 2007 р. ставку в розмірі 13 % об'єкта оподаткування було підвищено до 15 %. Спадна динаміка надходжень була характерною для періоду економічної кризи, у 2009 р., порівняно з 2008 р., частка ПДФО в податкових надходження зменшилася. З 2010 року зростання надходжень ПДФО відновилось. Другим періодом, в якому спостерігалось зменшення частки ПДФО в надходженнях, стали 2014–2015 рр. через політико-економічні фактори, характерні для цього періоду.

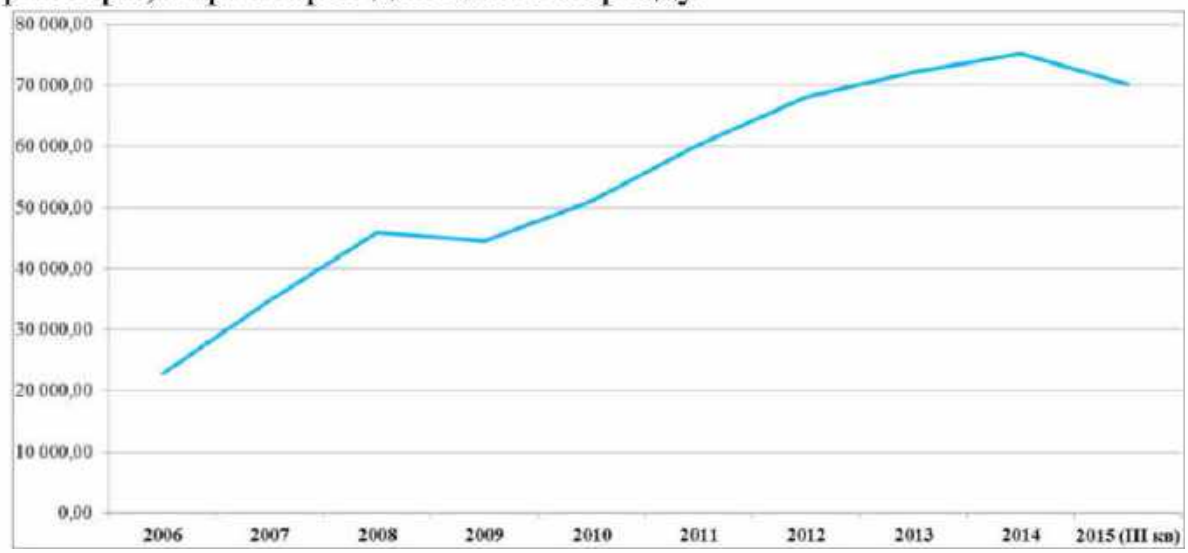


Рис. 3.1. Надходження з податку на доходи фізичних осіб (млн грн) в 2006–2015 рр.

Джерело: складено авторами за даними Державної казначейської служби України.

Переважну частку в надходженнях складають податок на доходи фізичних осіб із заробітної плати (рис. 3.2). Частка надходжень податку з інших доходів була незначною в 2006–2013 рр., а з 2014 почала зростати. Основними причинами цього стало розширення бази ПДФО на відсотки за банківськими депозитами. У зв'язку з цим доцільно включати показники, що характеризують базу оподаткування, у економетричну модель як ознаки-фактори.

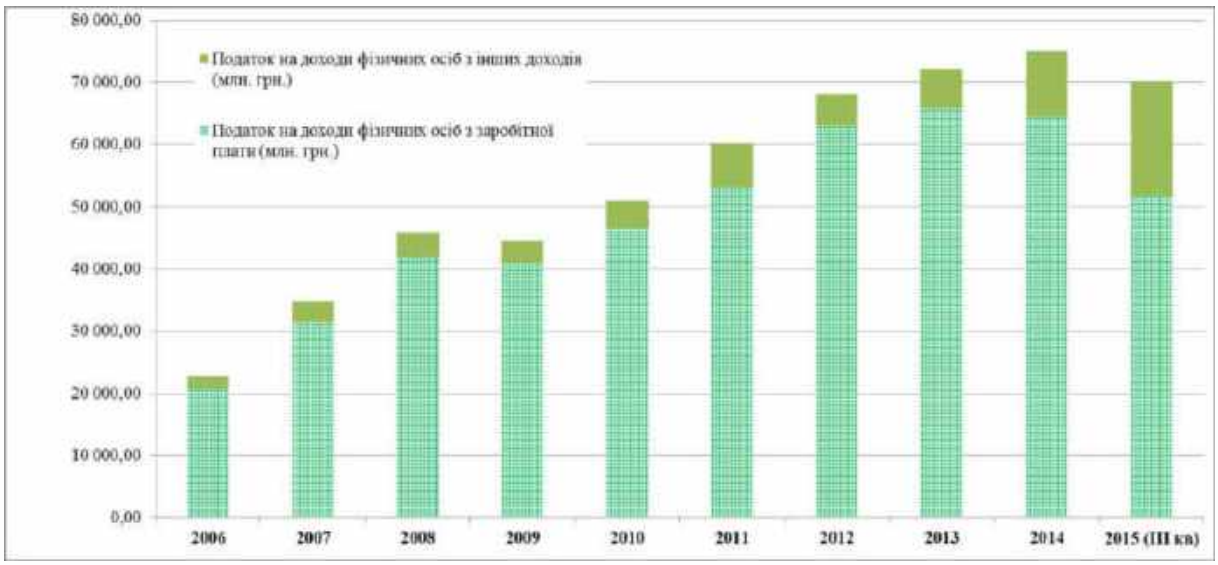


Рис. 3.2. Структура надходжень ПДФО в Україні в 2006–2015 рр.

Джерело: *складено авторами за даними Державної казначейської служби України.*

На підставі цього на надходження ПДФО впливають показники, що відображають базу оподаткування, а саме ВВП, обсяг заробітної плати в економіці, обсяг депозитів фізичних осіб. Крім того, зважаючи на значну проблему, зумовлену виплатою заробітної плати «в конвертах», додатково було проаналізовано вплив на надходження ПДФО кінцевих витрат населення на придбання товарів та послуг як ознак-факторів, у модель було включено кінцеві споживчі витрати населення як показник що відображає потенційну фазу оподаткування ПДФО, легальні та нелегальні доходи фізичних осіб. Проте з такими факторами в моделі знаки при змінних ВВП та обсяг заробітної плати в економіці вийшли від'ємними, у зв'язку з цим змінну «витрати населення на придбання товарів та послуг» вирішено було вилучити.

Внаслідок проведених попередніх розрахунків було прийнято рішення зупинитися на логарифмічній специфікації моделі. За результуючий показник відібрано надходження з податку на доходи фізичних осіб (Y) за період 2006–2015 рр. (з квартальною дискретністю), факторами, що на нього впливають – показники, що відображають базу оподаткування:

ВВП (X_1), заробітна плата (X_2) та обсяг депозитів фізичних осіб (X_3).

Бібліотеки в R, які будуть вам необхідні для виконання завдання

```
library("dplyr") # маніпуляції з даними
library("psych") # описові статистики
library("lmtest") # тести для лінійних моделей
library("glmnet") # LASSO + ridge
library("ggplot2") # графіки
library("car") # vif
```

Для завантаження вихідних даних введемо таку команду:

```
t3<-read.csv("data3.txt", sep="\t", dec=".", header= TRUE)
```

Поля кореляції між результуючим показником та показниками факторами побудовані за допомогою таких команд:

```
qplot(data=t3,log(PIT),log(GDP))
qplot(data=t3,log(PIT), log(earnes))
qplot(data=t3,log(PIT), log(deposits))
```

Відповідні графіки подані на рис. 3.3–3.5.

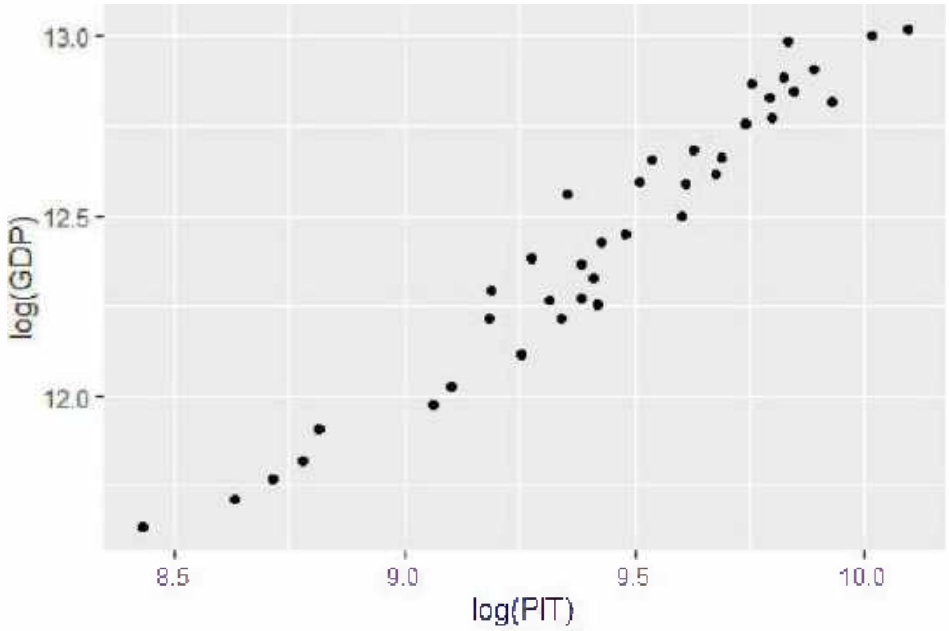


Рис. 3.3. Поле кореляції між надходженнями ПДФО в Україні та ВВП (логарифмовані дані)

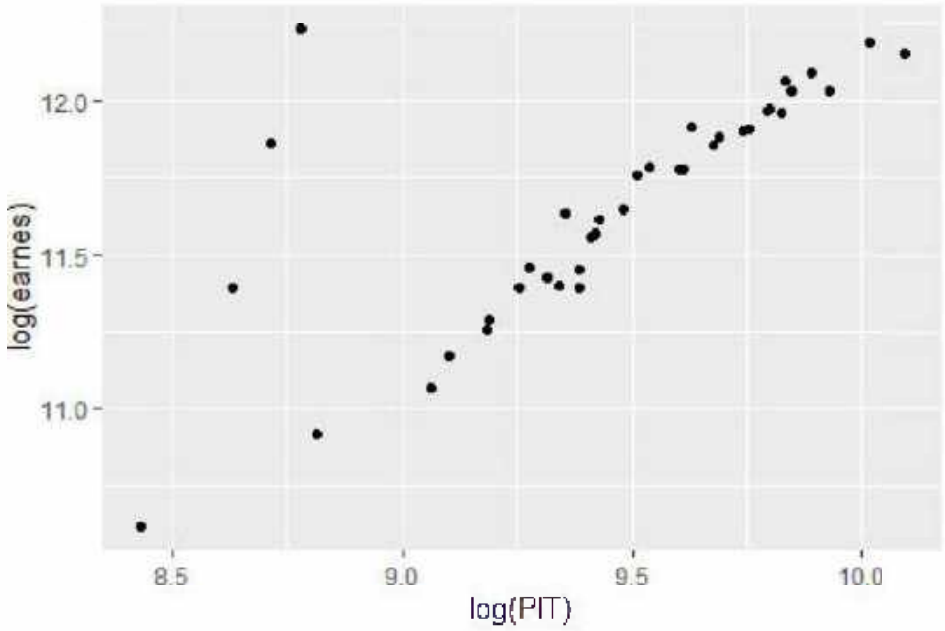


Рис. 3.4. Поле кореляції між надходженнями ПДФО в Україні та заробітною платою (логарифмовані дані)

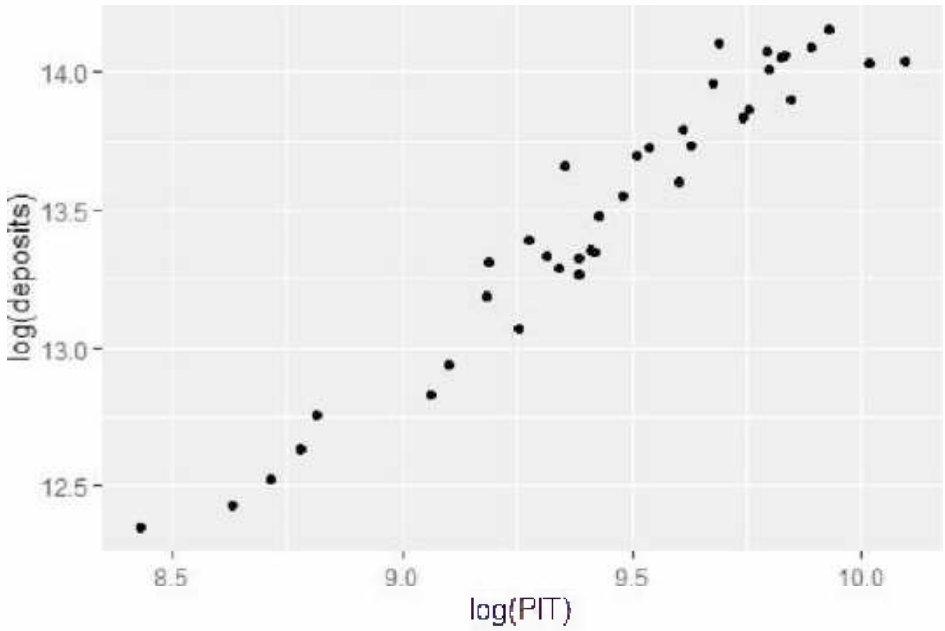


Рис. 3.5. Поле кореляції між надходженнями ПДФО в Україні та обсягом депозитів (логарифмовані дані)

За даними, поданими на графіках, можна побачити наявність візуального зв'язку з результуючим показником та відібраними для дослідження факторами.

Будуємо логарифмічну модель та викликаємо її протокол за допомогою функцій `lm ()` та `summary()`:

```
model <- lm(data=t3, log(PIT)~log(GDP)+
log(earnes)+log(deposits))
```

Протокол економетричної моделі викликаємо через функцію `summary()`. Приклад виконання:

```
summary(model)
```

Результати моделювання подано на рис. 3.6.

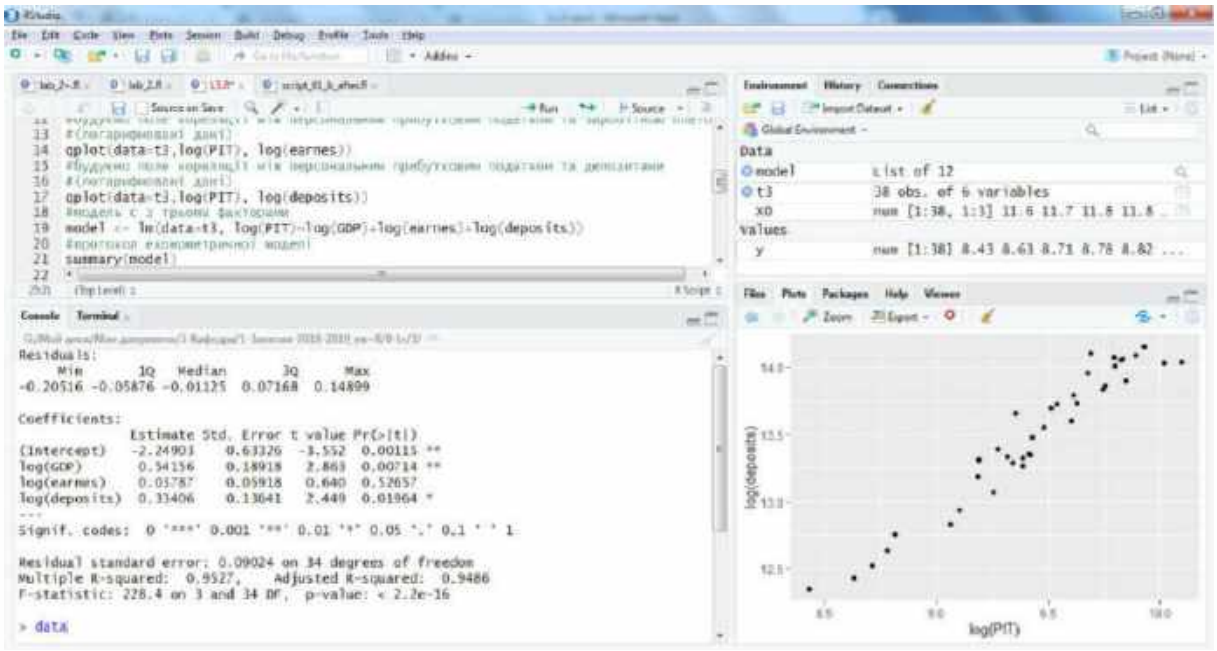


Рис. 3.6. Протокол економетричної моделі залежності надходжень ПДФО, ВВП, обсягу заробітної плати та депозитів, побудованої за допомогою МНК

Економетрична модель між надходженнями ПДФО, ВВП, обсягом заробітної плати та депозитів виглядає так:

$$\widehat{logPIT} = -2.2493 + 0.5416 \cdot \log(GDP) + 0.0379 \cdot \log(earnes) + 0.3341 \cdot \log(deposits)$$

Оскільки ми очікуємо, що ВВП, обсяг заробітної плати та депозитів будуть залежними між собою, побудуємо кореляційну матрицю та проаналізуємо коефіцієнти кореляції між показниками-факторами. Для цього використаємо функцію `cor()` та проведемо ряд маніпуляцій.

```
#вектор залежної змінної
y <- log(t3$PIT)
```

```
# матриця ознак-факторів (без стовпчика одиниць)
X0 <- model.matrix(data=t3, log(PIT)~0+log(GDP)+log(earnes)+log(deposits))
# кореляційна матриця ознак-факторів
cor(X0)
```

Кореляційна матриця, складена за результуючим показником та ознаками-факторами, представлена на рис. 3.7.

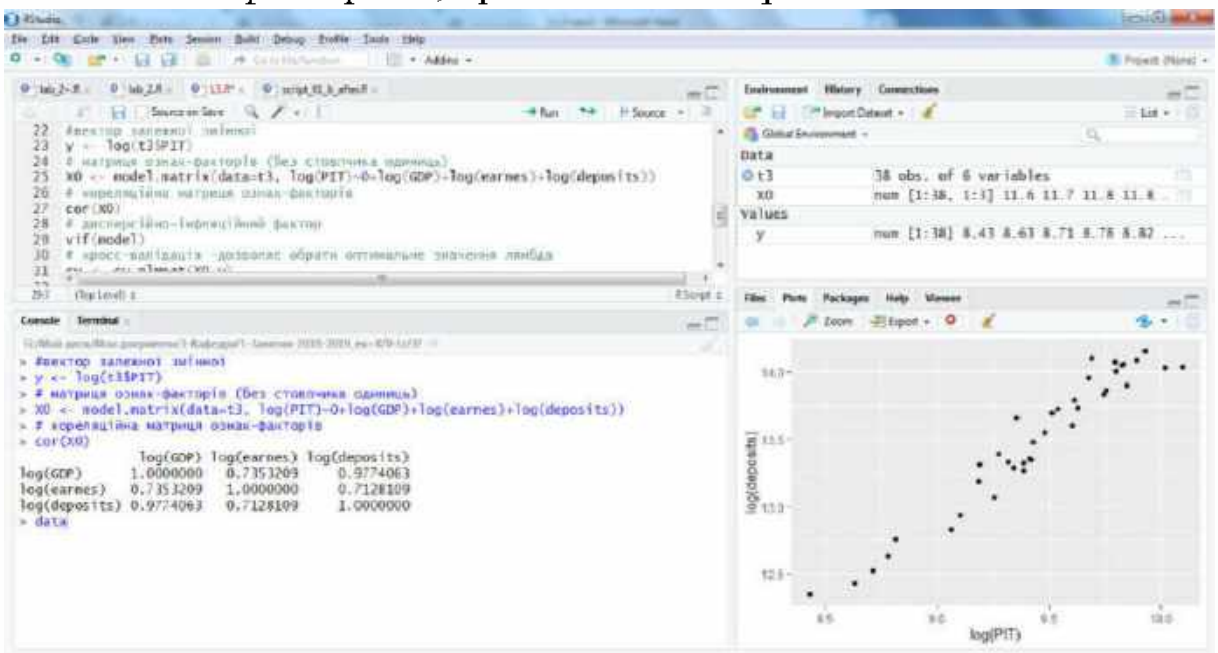


Рис. 3.7. Кореляційна матриця між ВВП, обсягом заробітної плати та депозитів

Аналізуючи коефіцієнти кореляції між результуючим показником та ознаками-факторами, що подані в табл. 1, можна зробити висновок про щільний зв'язок між усіма факторами, оскільки вони є більшими за 0,7.

Для того, щоб упевнитись в тому, що в моделі наявна мультиколінеарність, проведено її діагностику за допомогою розрахунку дисперсійно-інфляційних коефіцієнтів через використання функції `vif()`. Приклад виконання:

```
vif(model)
```

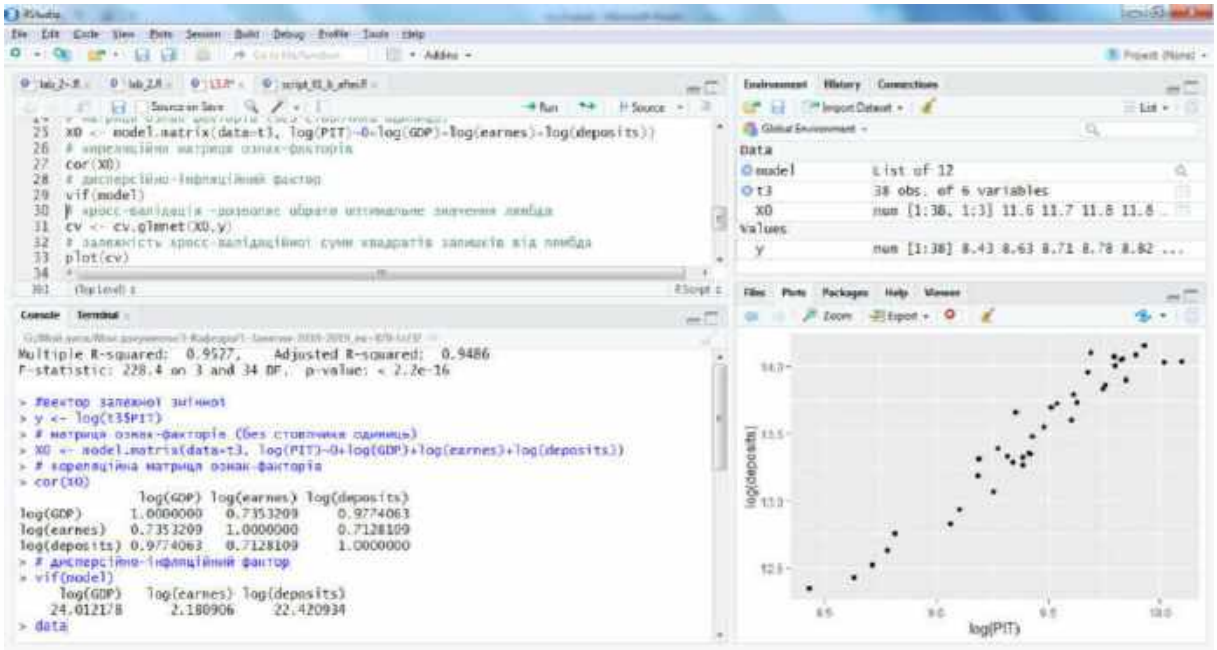


Рис. 3.8. Розрахунок дисперсійно-інфляційних коефіцієнтів

Оскільки дисперсійно-інфляційні коефіцієнти для ВВП та обсягу депозитів є більшими за 10, можна стверджувати, що в моделі наявна мультиколінеарність.

На підставі того, що обсяг депозитів фізичних осіб корелює з ВВП, вирішено було побудувати рідж-регресію. Це один з методів регуляризації, який виконує певні „коригування" відхилень від нормального розподілу залишків для забезпечення надійності і ефективності оцінок параметрів регресії. Цей метод передбачає введення штрафу за невиправдано великі значення оцінок, причому його величина пропорційна величині параметра λ, за допомогою якого можна налаштовувати алгоритм підгонки моделі й отримати більш стійке рішення. Метод рідж-регресії (ridge regression) реалізує класичну L2-регуляризацію Тихонова.

Для знаходження оптимального регуляційного параметра λ доцільно використовувати узагальнену перехресну перевірку на основі зовнішнього критерію крос-валідації GCV (generalized cross-validation).

Підготовчі етапи до побудови рідж-регресії (ridge regression) в R були зроблені раніше, а саме: завантаження вектора залежної змінної – надходжень персонального прибуткового податку та матрицю ознак-факторів: ВВП (X_1), заробітної плати (X_2), обсягу депозитів (X_3) без стовпчика одиниць.

Наступним нашим завданням є обрання оптимального значення лямбда з використанням методу крос-валідації. Для цього використовуємо функцію `cv.glmnet()`. Процедура проводиться так: вихідні дані випадковим способом r разів розбивається на k блоків (folds) майже рівної довжини. При реалізації кожної ітерації один з блоків по черзі стає перевірочною послідовністю, а об'єднання всіх інших блоків – навчальною вибіркою. За замовчуванням кількість блоків дорівнює 10. У разі невеликої кількості спостережень кількість блоків можна зменшити.

Приклад виконання:

```
# крос-валідація - дозволяє обрати оптимальне значення лямбда
cv <- cv.glmnet(X0,y,alpha=1, nfolds = 5)
```

Будуємо графік залежності крос-валідаційної суми квадратів залишків від лямбда за домопогою функції plot() (рис. 3.9). Приклад виконання

```
plot(cv)
```

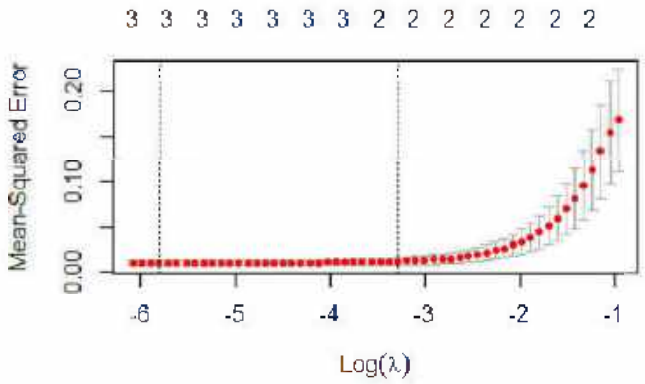


Рис. 3.9. Залежність крос-валідаційної суми квадратів залишків від лямбда

Далі знаходимо оптимальне значення лямбда за допомогою команди cv$lambda.min. Приклад виконання:

```
# лямбда, яка мінімізує крос-валідаційну суму квадратів залишків
cv$lambda.min
```

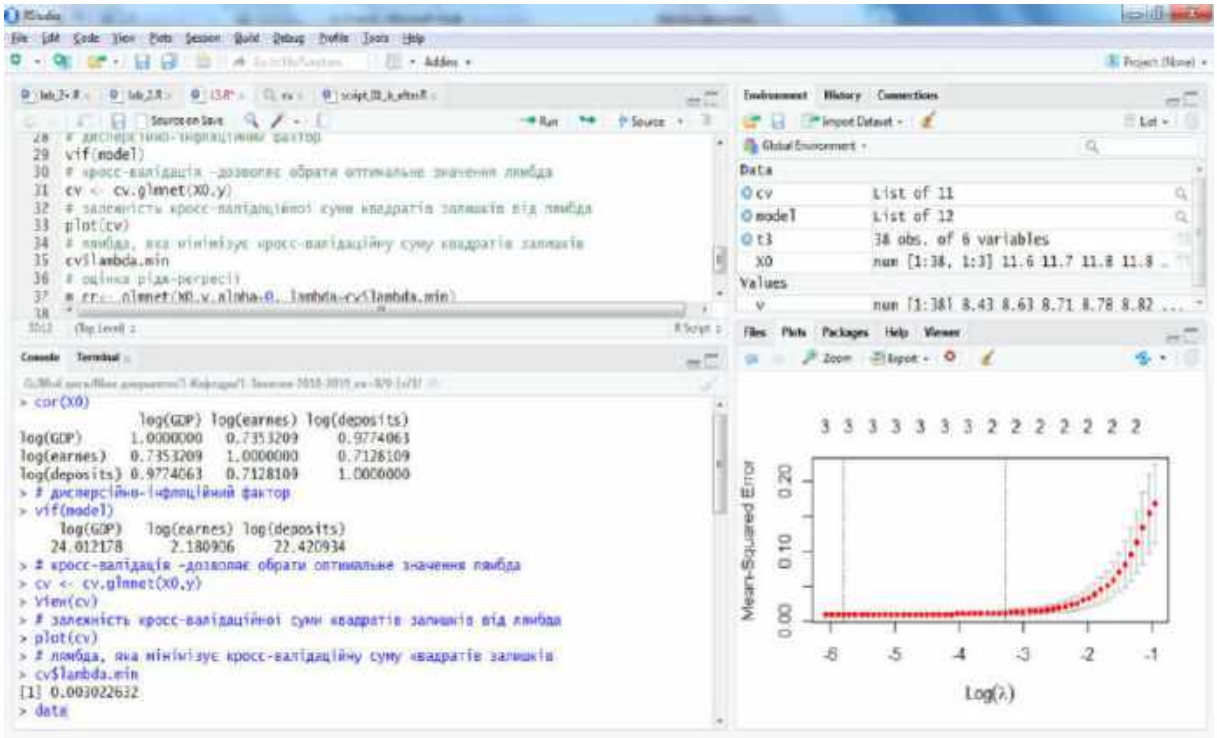


Рис. 3.10. Знаходження λ, що мінімізує крос-валідаційну суму квадратів залишків

Відповідно до проведених розрахунків лямбда, яке необхідно використати для побудови рідж-регресії, становить $\lambda = 0.003022632$.

Дані будуємо модель з використанням методу рідж-регресії за допомогою функції `glmnet()` та знаходимо оцінки параметрів моделі за допомогою функції `coef ()`. Приклад виконання:

```
    # оцінка рідж-регресії
    m_rr <- glmnet(X0,y,alpha=0, lambda=
cv$lambda.min)
    # коефіцієнти моделі, які відповідають
мінімальній лямбда
    coef(m_rr,s="cv$lambda.min")
```

Результати моделювання подані на рис. 11.

Рівняння регресії виглядає так:

$$\widehat{logPIT} = -2.2075 + 0.528 \cdot \log(GDP) + 0.044 \cdot \log(earnes) + 0.3382 \cdot \cdot \log(deposits)$$

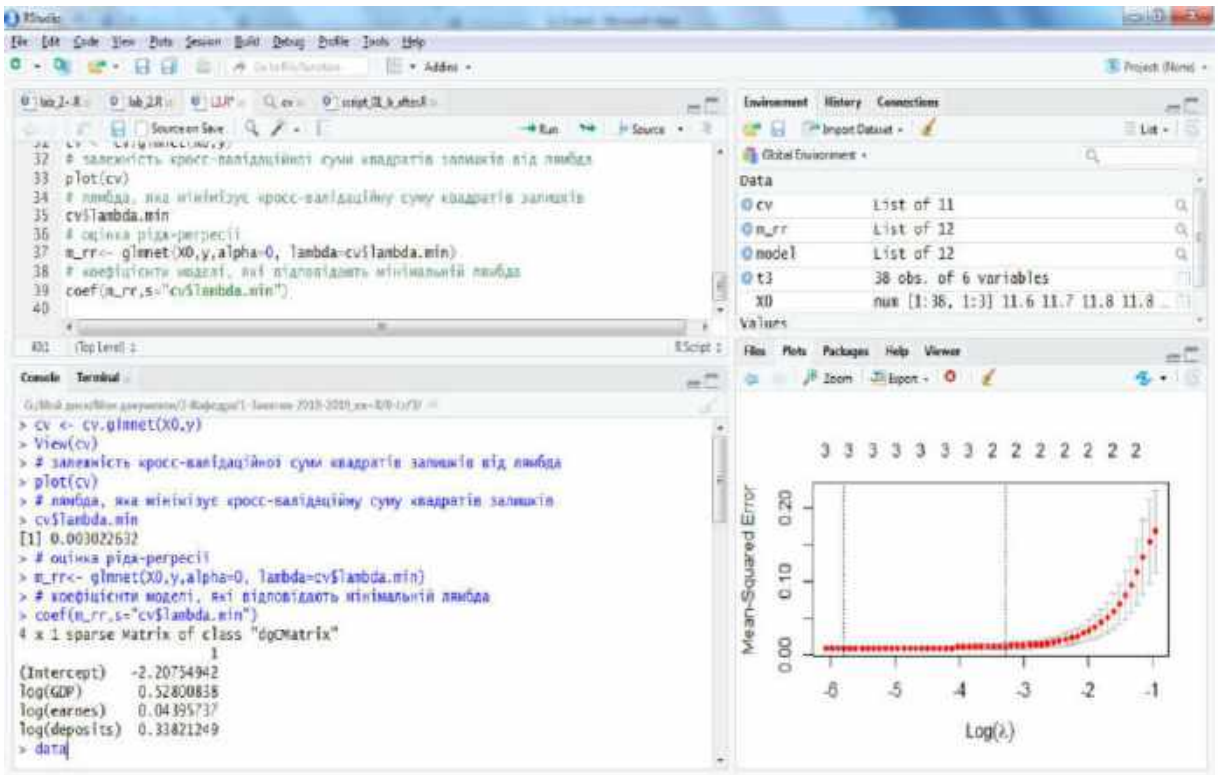


Рис. 3.11. Побудова рідж-регресії та знаходження оцінок параметрів

За результатами моделювання можна зробити такі висновки. Зростання ВВП на 1 % зумовлює зростання надходжень ПДФО на 0,528 %, а зростання заробітної плати на 1 % зумовлює зростання надходжень ПДФО на 0,053 %, зростання депозитів на 1 % призводить до зростання відповідних надходжень 0,34 %. Результати вийшли суперечливими і ми представляємо їх для публічного обговорення. Адже надходження ПДФО більшою мірою обумовлюються заробітною платою, аніж обсягом депозитів.

Крім того, варто відмітити високий рівень щільності зв'язку між надходженнями ПДФО і ВВП, з цього можна зробити висновок про те, що певною мірою рівень економічного зростання зумовлює зростання надходжень з податку.

Додаток А до лабораторної роботи 3

ВИХІДНІ ДАНІ ДО ЛАБОРАТОРНОЇ РОБОТИ 3

	Варіант 1					**Варіант 2**			
	Австралія					**Австрія**			
	Національна валюта: австралійський долар					**Національна валюта: євро**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
year	***PIT***	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	59893	529598	346949	16947	**1995**	15127	176183	99228	2467
1996	65455	557246	369926	17832	**1996**	16471	182133	99111	2550
1997	69212	589394	389499	18861	**1997**	17986	188310	100119	2636
1998	81052	621000	404855	19872	**1998**	18913	195828	103741	2742
1999	87496	661515	423513	21168	**1999**	19821	203418	106769	2848
2000	80991	706042	448544	22593	**2000**	19890	213196	109509	2985
2001	89864	754228	467550	24135	**2001**	21866	220096	110707	3081
2002	94368	801587	491834	25651	**2002**	21870	226303	115544	3168
2003	102524	862219	518096	27591	**2003**	22361	230999	119081	3234
2004	112060	922845	548450	29531	**2004**	22781	241505	121669	3381
2005	118239	998458	584451	31951	**2005**	22763	253009	126256	3542
2006	119073	1087440	622833	34798	**2006**	24006	266478	132963	3731
2007	127587	1178809	677278	37722	**2007**	25741	282347	139877	3901
2008	127294	1259280	723374	40297	**2008**	28008	291930	146808	3679

2009	124649	1297508	750832	41520	**2009**	26152	286188	151279	3894
2010	138163	1410442	806842	45134	**2010**	27098	294628	153766	4359
2011	152884	1491741	865793	47736	**2011**	28546	308630	157859	4892
2012	162410	1527529	903832	39426	**2012**	30289	317117	164276	3976
2013	169972	1589940	933881	50878	**2013**	31474	322539	169676	4505
2014	183023	1617016	969795	65292	**2014**	33428	330418	174621	4602
2015	191747	1654864	994945	52956	**2015**	35841	339896	180275	4277
2016	198534	1820350	1018440	52266	**2016**	32250	349493	188549	5034
	Варіант 3					**Варіант 4**			
	Бельгія					**Канада**			
	Національна валюта: євро					**Національна валюта: канадський долар**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	29444	211616	115452	2963	**1995**	108812	828973	499238	37707
1996	29316	215942	117744	3023	**1996**	113922	857023	517349	41769
1997	31186	225969	121767	3164	**1997**	123207	903902	552787	39225
1998	32416	234490	123391	3283	**1998**	127682	937295	580260	45914
1999	32890	244226	132808	3419	**1999**	136815	1004456	607555	50055
2000	35260	258222	136751	3615	**2000**	143652	1102380	649287	53181
2001	37071	265788	138469	3721	**2001**	145930	1140505	670595	56378
2002	38253	275065	144659	3851	**2002**	138699	1189452	696031	59569
2003	38392	282637	148600	3957	**2003**	140824	1250315	724283	63252

2004	39955	298711	154363	4179	**2004**	151331	1331178	764170	69394
2005	39030	311481	161312	6842	**2005**	164801	1417028	817061	72938
2006	39324	326662	167780	5460	**2006**	173895	1492207	864382	75207
2007	40967	344713	175004	5046	**2007**	190025	1573532	917086	82148
2008	43583	354066	182893	5209	**2008**	195133	1652923	963160	86136
2009	41997	348781	184513	6231	**2009**	184090	1567365	982304	86945
2010	43956	365101	190291	5032	**2010**	180894	1662131	1010872	85112
2011	46072	379106	196221	3778	**2011**	197077	1769922	1057756	98107
2012	47651	387500	203050	5187	**2012**	207452	1822808	1107530	112500
2013	50471	391712	209720	4314	**2013**	214054	1897532	1157897	116188
2014	51875	400805	212768	4783	**2014**	226521	1983117	1198477	121249
2015	51879	410351	212853	4247	**2015**	240295	1986193	1226902	123900
2016	51565	421974	215240	4633	**2016**	241500	2184812	1241799	121834
	Варіант 5					**Варіант 6**			
	Чехія					**Данія**			
	Національна валюта: чеська крона					**Національна валюта: датська крона**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	70361	1580115	547515	7901	**1995**	264058	1036483	639863	21766
1996	80206	1812622	642322	9063	**1996**	274959	1088024	667884	22849
1997	87356	1953311	710332	9767	**1997**	287748	1146129	699410	24069
1998	93632	2142587	769312	10713	**1998**	293196	1185988	720674	24906

1999	92962	2237300	817130	11187	**1999**	309838	1241521	745680	26072
2000	99668	2372630	868080	11863	**2000**	327624	1326912	762105	27865
2001	105406	2562679	921397	12813	**2001**	342502	1371526	792013	28802
2002	114957	2674634	1006425	13373	**2002**	347917	1410271	820930	29616
2003	125399	2801163	1082424	14006	**2003**	353775	1436751	853795	30172
2004	136317	3057660	1174639	14392	**2004**	361551	1506001	894185	31626
2005	137937	3257972	1236950	15254	**2005**	380555	1585984	923045	33306
2006	139350	3507131	1315707	17055	**2006**	402033	1682260	968271	35327
2007	153374	3831819	1394204	21907	**2007**	425942	1738845	994432	36516
2008	142269	4015346	1470512	21463	**2008**	434457	1801470	1051080	37831
2009	136023	3921827	1495226	22319	**2009**	433506	1722143	1087737	36165
2010	131706	3953651	1535505	20654	**2010**	421020	1810926	1118629	38029
2011	142752	4033755	1559357	16996	**2011**	428312	1846854	1138861	33917
2012	144831	4059912	1603016	18772	**2012**	441324	1895002	1157320	35148
2013	149831	4098128	1614499	21282	**2013**	484032	1929677	1167414	39591
2014	153763	4313789	1655551	19660	**2014**	521766	1977255	1203424	41378
2015	164678	4554615	1706258	25949	**2015**	516169	2027171	1241183	57421
2016	183103	4715061	1798575	24267	**2016**	510009	2060868	1304972	43278

	Варіант 7					Варіант 8			
	Естонія					Фінляндія			
	Національна валюта: євро					Національна валюта: євро			
	ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.		ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.
	PIT	*GDP*	*Earnes*	*Capital income*		*PIT*	*GDP*	*Earnes*	*Capital income*
1995	234	2779	1357	19	**1995**	13639	98556	54111	1568
1996	278	3649	1715	26	**1996**	15256	102060	56282	1670
1997	335	4495	2043	31	**1997**	15325	110738	57120	1790
1998	398	5049	2288	35	**1998**	16195	120382	60266	2094
1999	417	5379	2424	38	**1999**	16511	126923	64408	2637
2000	421	6171	2816	43	**2000**	19118	136261	68598	3055
2001	454	6976	3066	49	**2001**	19627	144437	71604	3475
2002	499	7774	3309	54	**2002**	20058	148289	73329	3674
2003	564	8709	3730	61	**2003**	19905	151569	75195	4328
2004	608	9708	4154	68	**2004**	20177	158477	77775	4998
2005	622	11262	4592	79	**2005**	21211	164387	81219	4691
2006	746	13522	5459	95	**2006**	22038	172614	85078	4777
2007	936	16246	6801	114	**2007**	23396	186584	88920	5251
2008	1011	16517	7507	116	**2008**	24593	193711	93765	4938
2009	789	14146	7161	99	**2009**	23011	181029	95371	4818
2010	776	14717	7347	103	**2010**	22545	187100	98035	5629
2011	846	16668	7500	117	**2011**	24181	196869	101945	5960

2012	931	17935	8055	113	**2012**	24989	199793	105182	5299
2013	1031	18890	8370	151	**2013**	25998	203338	106045	5328
2014	1134	19758	8802	110	**2014**	27523	205474	107568	5594
2015	1182	20252	9253	170	**2015**	27824	209511	109423	5574
2016	1259	20916	9925	146	**2016**	28136	214062	110906	6194
	Варіант 9					**Варіант 10**			
	Франція					**Німеччина**			
	Національна валюта: євро					**Національна валюта: євро**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	58633	1224967	585590	39199	**1995**	189069	1898880	1025305	34180
1996	63184	1258950	601508	40286	**1996**	170046	1926320	1048803	34674
1997	69291	1299739	617140	41592	**1997**	167013	1967090	1062863	35408
1998	100554	1358776	635890	43481	**1998**	179105	2018230	1087576	36328
1999	107326	1408159	657190	45061	**1999**	187124	2064880	1094682	37168
2000	115199	1485303	682683	47530	**2000**	194173	2116480	1106495	38097
2001	117055	1544629	708397	49428	**2001**	207025	2179850	1141360	39237
2002	116122	1594259	740670	51016	**2002**	190745	2209290	1163334	39767
2003	119774	1637438	748081	52398	**2003**	183565	2220080	1185866	39961
2004	122512	1710760	783087	54744	**2004**	176104	2270620	1203453	40871
2005	136179	1771978	812922	56703	**2005**	179645	2300860	1247212	41415
2006	139274	1853267	844880	59305	**2006**	199996	2393250	1276502	43079

2007	140375	1945670	873861	62261	**2007**	218848	2513230	1300463	45238
2008	146349	1995850	902634	63867	**2008**	240294	2561740	1333565	51805
2009	138173	1939017	925167	62049	**2009**	223610	2460280	1334036	44285
2010	143124	1998481	957656	63951	**2010**	219387	2580060	1338169	46441
2011	152777	2059284	973780	65897	**2011**	238671	2703120	1400550	49794
2012	167519	2086929	1002749	77186	**2012**	258564	2758260	1443114	45155
2013	178369	2115256	1020338	66983	**2013**	270481	2826240	1488612	49889
2014	182329	2139964	1059781	70860	**2014**	283271	2923930	1534614	47122
2015	186909	2181064	1073762	68093	**2015**	299764	3032820	1585098	50677
2016	190956	2225260	1088655	61181	**2016**	314800	3132670	1652268	56388
	Варіант 11					**Варіант 12**			
	Греція					**Угорщина**			
	Національна валюта: євро					**Національна валюта: форинт**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	3112	93064	39719	1396	**1995**	383744	5819221	2337382	52373
1996	3464	103037	43690	1546	**1996**	490666	7103898	2852094	63935
1997	4285	114712	50035	1721	**1997**	560156	8812386	3428137	79311
1998	5804	125263	56225	1879	**1998**	656749	10433144	3972217	93898
1999	6463	133789	60397	2007	**1999**	779782	11637865	4452470	104741
2000	6802	141247	64114	2119	**2000**	954698	13321531	5150058	119894
2001	6612	152194	68764	2283	**2001**	1134447	15383442	6072056	138451

2002	7119	163461	77428	2452	**2002**	1318857	17421576	6775722	156794
2003	7496	178905	84059	2684	**2003**	1342535	19077348	7839146	171696
2004	8229	193716	89965	2843	**2004**	1385726	21023630	8738435	189213
2005	9043	199242	92838	3515	**2005**	1476032	22470802	9528813	202237
2006	9677	217862	97745	3732	**2006**	1613706	24153022	10123946	217377
2007	10767	232695	101461	3398	**2007**	1838563	25560380	10546075	209001
2008	11226	241990	104548	3679	**2008**	2067806	27071868	11166802	243647
2009	10563	237534	111467	4264	**2009**	1928107	26297412	11165277	247216
2010	8918	226031	107534	3641	**2010**	1768830	27085900	11815801	243773
2011	9823	207029	100897	3204	**2011**	1422112	28166115	12317860	253495
2012	13242	191204	95845	2264	**2012**	1622030	28660518	12846108	254049
2013	10698	180654	87196	2383	**2013**	1610329	30127349	13213161	271146
2014	10478	177941	85980	2117	**2014**	1705112	32400148	13578032	286744
2015	10047	175697	84061	2096	**2015**	1814908	33999012	13408098	305991
2016	10226	175888	83214	2229	**2016**	1853431	35005439	14199560	315049
	Варіант 13					**Варіант 14**			
	Ірландія					**Ізраїль**			
	Національна валюта: євро					**Національна валюта: шекелі**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	5507	54813	32788	548	**1995**	28399	301969	148438	4530
1996	6086	60206	35163	602	**1996**	31875	350918	176951	5264

1997	6902	69367	38749	694	**1997**	37239	395689	198158	5935
1998	7651	80353	45898	804	**1998**	40443	440630	223026	6609
1999	8546	92716	50123	927	**1999**	45268	485156	247527	7277
2000	9876	108383	55452	1084	**2000**	53698	539546	271243	8093
2001	9939	121941	61222	1219	**2001**	56820	549417	284794	8241
2002	9751	135951	66054	1360	**2002**	53420	572970	291003	8595
2003	10082	145557	72259	1456	**2003**	47756	576927	288886	8654
2004	11737	156144	78018	1049	**2004**	45208	606706	293483	9101
2005	12586	170216	86194	1502	**2005**	46380	639329	305668	9590
2006	13750	185061	93846	4107	**2006**	50453	685647	331085	10285
2007	15029	197293	108469	3205	**2007**	55205	734290	351991	11014
2008	14642	187687	114279	1573	**2008**	52488	774758	369719	11621
2009	13602	169704	112623	1261	**2009**	46964	815248	382675	12229
2010	13247	167124	106804	994	**2010**	47563	874010	400050	13110
2011	14166	173070	98444	1327	**2011**	50980	935225	423701	11256
2012	15230	175754	97854	1234	**2012**	52791	993440	493094	14294
2013	15836	180209	98047	1366	**2013**	56279	1059101	514507	17584
2014	17187	193160	99811	1221	**2014**	59490	1104746	530494	19689
2015	18498	255815	103041	1594	**2015**	68150	1163769	553814	21586
2016	19322	265835	106699	2658	**2016**	72941	1280146	580967	13656

	Варіант 15					Варіант 16			
	Італія					Корея			
	Національна валюта: євро					Національна валюта: вона			
	ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.		ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	97869	984983	401117	15760	**1995**	13088538	428927100	286148335	10723178
1996	104051	1043086	422923	16689	**1996**	14512420	481140800	326473389	12028520
1997	113663	1089869	444713	17438	**1997**	14585930	530347100	347411633	13258678
1998	113354	1135499	464597	18168	**1998**	18157719	524476800	357967095	13111920
1999	125667	1171901	481501	18750	**1999**	16821000	576872800	374240759	14421820
2000	124233	1239266	501911	19828	**2000**	18569000	635184600	404388935	15879615
2001	133212	1298890	522120	20782	**2001**	19679000	688164900	437794159	17204123
2002	136044	1345794	538388	21533	**2002**	18830000	761938900	477051434	19048473
2003	140192	1390710	556967	22251	**2003**	20347000	810915300	511683716	20272883
2004	144974	1448363	585891	14147	**2004**	22274000	876033100	555139444	21900828
2005	148022	1489725	604588	13308	**2005**	23118000	919797300	587971118	22994933
2006	159043	1548473	627836	18822	**2006**	26672000	966054600	610806223	24151365
2007	172122	1609551	640528	21353	**2007**	31984000	1043257800	642343945	26081445
2008	182565	1632151	671426	20249	**2008**	31585000	1104492200	670884251	27612305
2009	175452	1572878	668229	20053	**2009**	31310000	1151707800	690779783	28792695
2010	180221	1604515	683611	21903	**2010**	33935000	1265308000	731550611	31632700
2011	180492	1637463	694621	25473	**2011**	39910000	1332681000	778930512	33317025

2012	187293	1613265	708773	23868	**2012**	43730000	1377456700	798701530	34436418
2013	185991	1604599	718412	25512	**2013**	46654000	1429445400	832094109	35736135
2014	182149	1621827	730316	20616	**2014**	51410000	1486079300	859125309	37151983
2015	185094	1645439	737547	21673	**2015**	55744000	1558591600	892330013	38964790
2016	184493	1672438	751862	22111	**2016**	62028000	1714450760	940947846	42861269
	Варіант 17					**Варіант 18**			
	Латвія					**Люксембург**			
	Національна валюта: євро					**Національна валюта: євро**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	197	4059	2472	12	**1995**	1214	15777	5924	142
1996	229	4678	2486	14	**1996**	1286	16714	6098	150
1997	278	5394	2800	16	**1997**	1330	17498	6271	157
1998	321	6022	3034	18	**1998**	1307	18184	6395	164
1999	342	6272	3184	19	**1999**	1428	20871	6866	188
2000	376	6850	3359	21	**2000**	1573	23079	7503	208
2001	409	7460	3508	22	**2001**	1589	23772	7874	214
2002	458	8397	3613	25	**2002**	1528	25095	8350	226
2003	530	9553	4069	29	**2003**	1690	26188	8538	236
2004	629	11049	4728	34	**2004**	1825	27936	9048	251
2005	734	13597	5975	41	**2005**	2162	30031	9552	270
2006	958	17102	7447	20	**2006**	2546	33808	10037	304

2007	1287	22592	10412	81	**2007**	2678	37179	10863	335
2008	1459	24351	12389	104	**2008**	3036	38129	11245	343
2009	1010	18827	10373	53	**2009**	2917	36977	12350	333
2010	1116	17938	9322	16	**2010**	3157	40178	12717	362
2011	1137	20303	9294	45	**2011**	3546	43165	13212	388
2012	1258	21886	10032	52	**2012**	3699	44112	14079	397
2013	1320	22787	10370	73	**2013**	4062	46551	14724	419
2014	1397	23631	11001	76	**2014**	4329	49971	15541	450
2015	1445	24368	11905	136	**2015**	4717	52340	16933	421
2016	1587	25021	12592	83	**2016**	4986	54195	17087	491
	Варіант 19					**Варіант 20**			
	Нідерланди					**Польща**			
	Національна валюта: євро					**Національна валюта: злотий**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	23530	325341	206787	6995	**1995**	28364	344683	202812	1103
1996	23026	340964	212907	7331	**1996**	33615	431222	259087	1380
1997	21583	364961	224145	7847	**1997**	39178	521795	312421	1670
1998	22108	389315	224986	8370	**1998**	45672	606069	356823	1939
1999	23294	414838	237899	8919	**1999**	32736	673287	411450	2155
2000	25057	448061	258419	9633	**2000**	32415	747032	459229	2391
2001	27695	476696	275844	10249	**2001**	32979	779975	505608	2496

2002	31856	494501	291356	10632	**2002**	33870	810617	515798	2594
2003	31187	506671	301343	10893	**2003**	35058	845930	516622	2707
2004	29471	523939	312812	11265	**2004**	37075	933062	527128	2986
2005	33820	545609	319271	11731	**2005**	42377	990468	540740	1871
2006	37318	579212	330667	12453	**2006**	49228	1069824	545886	3269
2007	41276	613280	347872	13186	**2007**	61403	1187605	569278	3271
2008	42080	639163	361292	13742	**2008**	67940	1286069	632176	3538
2009	47974	617540	374183	13277	**2009**	62078	1372208	658012	4337
2010	48771	631512	374160	13578	**2010**	62818	1445297	695951	6832
2011	47646	642929	378009	14834	**2011**	67768	1566813	734467	5115
2012	44943	645164	390286	15064	**2012**	72516	1629392	757321	5514
2013	44592	652748	402899	13398	**2013**	74124	1656842	769959	6701
2014	46361	663008	402758	20460	**2014**	78617	1719704	787726	5176
2015	52167	676531	409814	10719	**2015**	83861	1799321	794033	5529
2016	50162	696871	415677	10759	**2016**	89943	1851171	823040	5264
	Варіант 21					**Варіант 22**			
	Словаччина					**Словенія**			
	Національна валюта: євро					**Національна валюта: євро**			
	ППП (Y), млн од. н.в.	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**		**ППП (Y), млн од. н.в.**	**ВВП (X_1), млн од. н.в.**	**Заробітна плата (X_2), млн од. н.в.**	**Доходи від капіталу, (X_3), млн од. н.в.**
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	695	19686	7482	20	**1995**	604	10522	6147	88
1996	855	21951	8606	22	**1996**	707	12132	7162	102

1997	1037	24416	10185	24	**1997**	798	13827	8381	116
1998	1138	26664	11257	27	**1998**	859	15337	9271	129
1999	1208	28548	12265	29	**1999**	942	17210	9801	145
2000	1055	31601	13980	32	**2000**	1039	18902	10940	159
2001	1193	34311	14988	34	**2001**	1187	21147	12377	178
2002	1205	37280	16128	37	**2002**	1320	23622	13266	198
2003	1304	41404	17647	41	**2003**	1437	25668	14115	216
2004	1394	46102	19442	40	**2004**	1556	27673	15973	232
2005	1496	50415	21226	34	**2005**	1583	29227	17084	246
2006	1627	56273	23107	66	**2006**	1782	31555	18173	265
2007	1820	63054	25124	66	**2007**	1916	35153	19579	229
2008	2095	68492	26871	89	**2008**	2178	37951	21121	334
2009	1794	64023	27795	46	**2009**	2063	36166	21508	278
2010	1790	67577	29495	70	**2010**	2017	36252	22485	299
2011	2000	70627	30069	70	**2011**	2053	36896	22302	316
2012	2123	72704	31063	73	**2012**	2040	36002	21836	345
2013	2144	74170	31813	60	**2013**	1847	35917	21871	428
2014	2275	75946	32479	63	**2014**	1896	37332	22376	448
2015	2464	78686	33846	66	**2015**	1974	38570	22523	449
2016	2682	80958	35136	61	**2016**	2116	39769	22842	341

	Варіант 23					Варіант 24			
	Великобританія					Литва			
	Національна валюта: фунт стерлінгів					Національна валюта: євро			
	ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.		ППП (Y), млн од. н.в.	ВВП (X_1), млн од. н.в.	Заробітна плата (X_2), млн од. н.в.	Доходи від капіталу, (X_3), млн од. н.в.
	PIT	***GDP***	***Earnes***	***Capital income***		***PIT***	***GDP***	***Earnes***	***Capital income***
1995	71121	836646	495366	16733	**1995**	487	7765	4496	43
1996	71206	892900	514237	17858	**1996**	601	9714	5792	53
1997	72525	938855	542088	18777	**1997**	838	11724	6885	64
1998	85903	980308	575266	19606	**1998**	995	13022	7861	72
1999	93080	1021205	606282	20424	**1999**	1047	12712	8028	70
2000	101925	1080863	652424	21617	**2000**	1020	13368	7740	74
2001	107661	1120575	690105	22412	**2001**	1021	14195	8183	78
2002	108525	1172652	711431	20181	**2002**	1033	15202	8541	84
2003	110203	1242449	745259	22863	**2003**	1073	16668	9109	92
2004	117369	1304874	780459	25281	**2004**	1206	18238	9859	23
2005	127803	1379457	803932	29991	**2005**	1389	21002	10923	69
2006	136611	1455644	860164	31449	**2006**	1567	24079	12410	68
2007	147309	1530890	908450	38885	**2007**	1810	29041	13674	212
2008	149228	1564252	931575	34529	**2008**	2056	32696	14991	447
2009	138023	1519459	960518	31679	**2009**	1062	26935	13874	168

2010	144502	1572439	983435	29847	**2010**	972	28028	13792	117
2011	149353	1628274	1002068	29963	**2011**	1064	31275	14087	108
2012	146474	1675044	1026407	29302	**2012**	1135	33348	14389	156
2013	152668	1739563	1060442	36991	**2013**	1226	35002	14890	193
2014	156187	1822480	1080848	37927	**2014**	1293	36590	15611	239
2015	163291	1872714	1105731	40252	**2015**	1390	37331	16304	228
2016	169582	1940198	1145318	29771	**2016**	1479	38637	17278	263

ЛАБОРАТОРНА РОБОТА 4
ПОБУДОВА ТА ДОСЛІДЖЕННЯ ЛІНІЙНОЇ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ З ГЕТЕРОСКЕДАСТИЧНІСТЮ

***Мета*:** набуття практичних навичок побудови багатофакторної економетричної моделі з гетероскедастичністю, її дослідження та практичне використання.

Задачі роботи

1. Специфікація економетричної моделі.
2. Побудова економетричної моделі МНК.
3. Дослідження економетричної моделі на наявність гетероскедастичності.
4. Побудова економетричних моделей узагальненим методом найменших квадратів на основі гіпотези про взаємозв'язок між залишками та значенням ознаки фактора та обрання найкращої моделі.
5. Формування економічних висновків на основі побудованої економетричної моделі.

Завдання роботи

На підставі статистичних даних Державної служби статистики України у розрізі регіонів для одного року з часового проміжку 2002–2017 побудувати економетричну модель залежності рівня заощаджень населення (*saving*) (Y, млн грн) від наявного доходу (*income*) (X_1, млн грн) та витрат (*expenditure*) (X_2, млн грн) для цього потрібно:

1. Побудувати діаграми розсіювання для результуючого показника та показників-факторів.
2. Побудувати економетричну модель залежності заощаджень населення від визначених факторів.
3. Провести діагностику економетричної моделі на наявність гетероскедастичності.

4. Побудувати економетричну модель узагальненим методом найменших квадратів на основі гіпотези про те, що дисперсія залишків пропорційна до зміни наявного доходу.

5. Побудувати економетричну модель узагальненим методом найменших квадратів на основі гіпотези про те, що дисперсія залишків пропорційна до зміни витрат.

6. Обрати найкращу економетричну модель на основі критерію найменшого стандартного відхилення вектора залишків моделі та сформулювати економічні висновки.

Приклад виконання

Бібліотеки в R, які будуть вам необхідні для виконання завдання

```
library("gdata") # завантаження пакета gdata
library("dplyr") # маніпуляції з даними
library("psych") # описові статистики
library("lmtest") # тестування гіпотез в лінійних моделях
library("sjPlot") # графіки
library("sgof")# завантаження пакета sgof
library("ggplot2") # графіки
library("foreign") # завантаження даних в різних форматах
library("hexbin") # графіки
library("car") # тести
library("MASS") # узагальнений МНК
```

Після того, як пакети активовано завантажуємо підготовлений заздалегідь файл з даними (рис. 4.1). Для цього обираємо ім'я таблиці, в якій буде зберігатись масив інформації, напишемо його в робочому середовищі – t, далі поставимо знак менше і коротку рисочку «<-» та набираємо команду read.csv. Після цього відкриваємо дужки, ставимо лапки, поміщуємо курсор всередині них, натискаємо клавішу Tab і обираємо файл з даними. Ця команда призначена для завантаження файлів у форматі *.csv, а оскільки потрібно завантажити файл у форматі *.txt, далі

потрібно його описати. Тому далі через кому напишемо, який використовується розділювач, у файлі *txt – табуляція, тому набираємо sep="\t". Потім через кому напишемо, який використовується десятковий розділювач, у файлі з даними розділювачем є крапка, тому пишемо dec=".". Останнє, що потрібно задати, чи використовуються заголовки, оскільки заголовки у файлі є, пишемо: header=TRUE. Отримуємо таку команду:

```
t<-read.csv("2012-saving.txt", sep="\t", dec=".", header= TRUE)
```

Після того, як введена ця команда, ставимо на неї курсор та натискаємо клавіші Ctrl+Enter для їх завантаження.

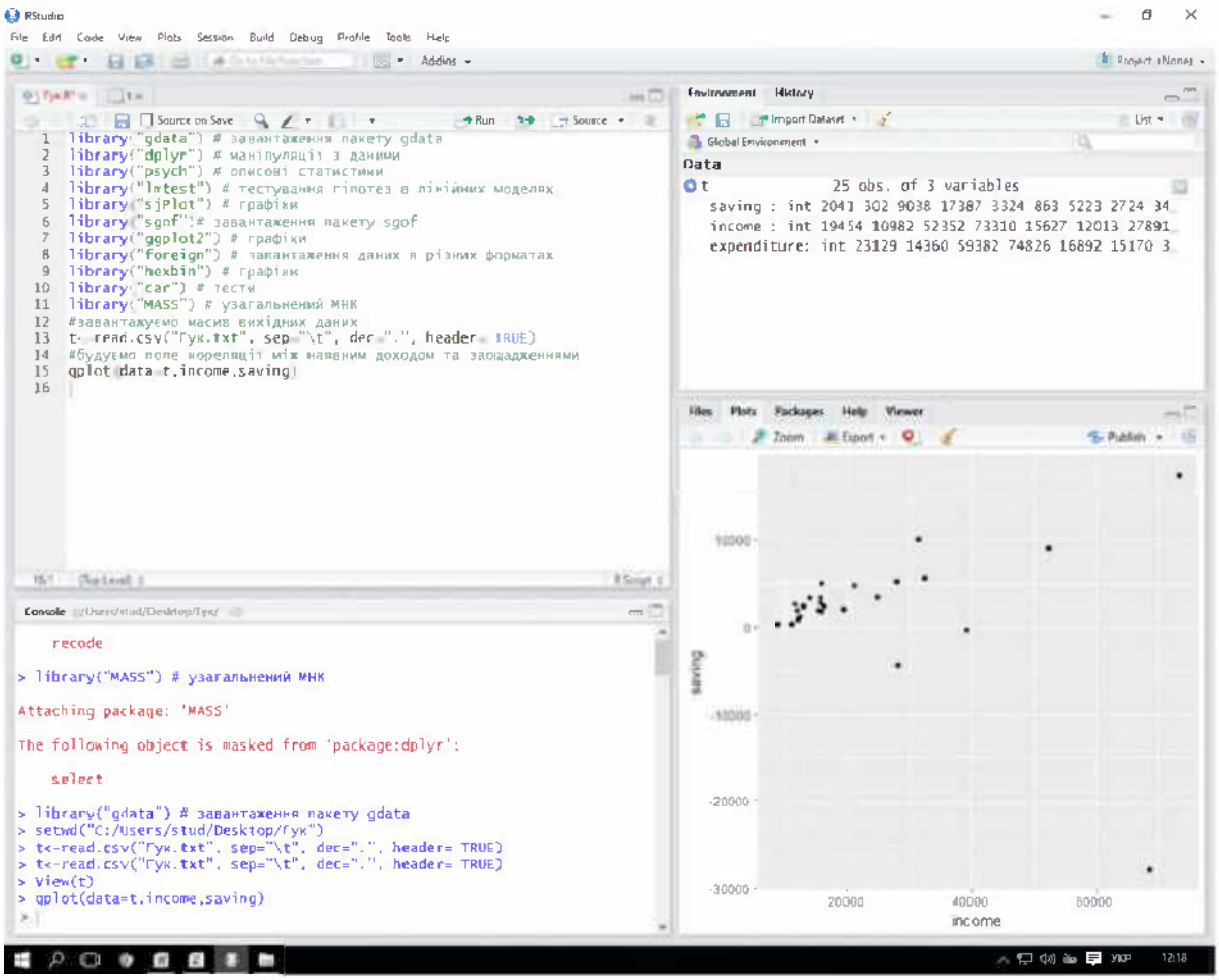


Рис. 4.1. Завантаження файлу з даними у форматі *txt в R

На набір вихідних даних можна подивитись, натиснувши на кнопку t у вікні глобального середовища (рис. 4.2).

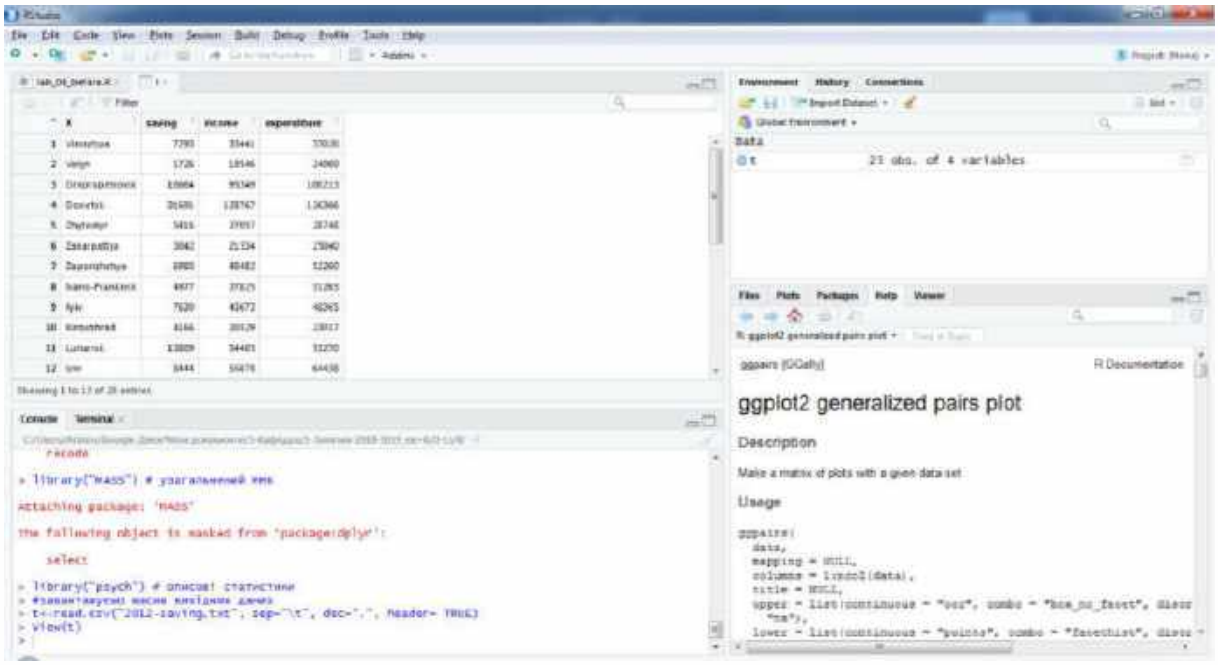


Рис. 4.2. Завантажені в R вихідні дані

Побудуємо діаграми розсіювання для результуючого показника та показників-факторів (рис. 4.3, рис. 4.4), використовуючи команди:

```
qplot(data=t,income,saving)
qplot(data=t,expenditure,saving)
```

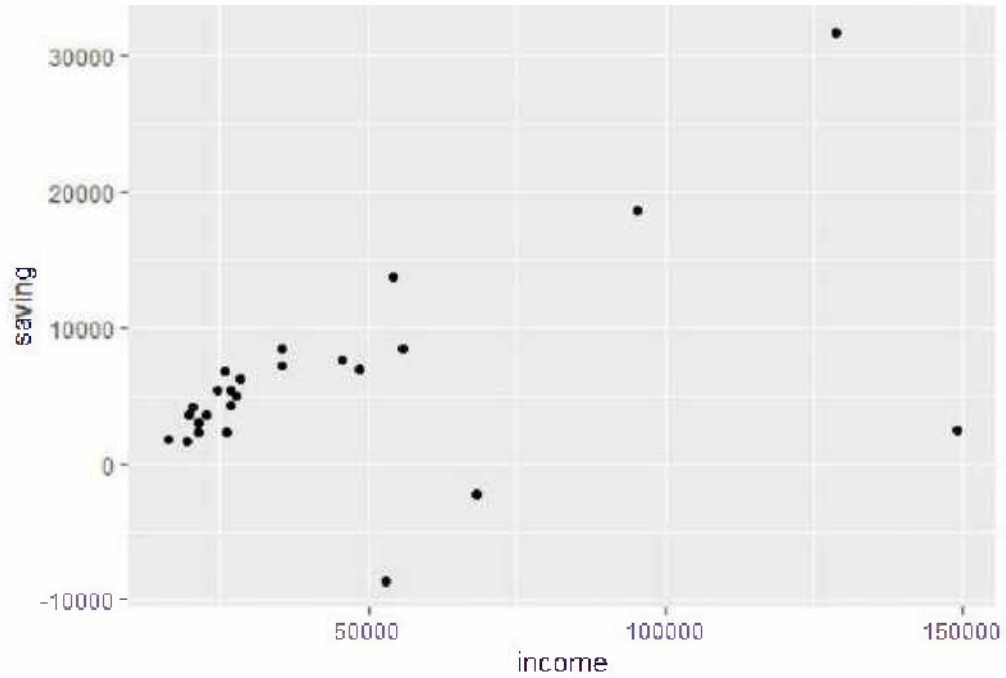


Рис. 4.3. Поле кореляції між наявним доходом та заощадженнями

На основі візуального аналізу поля кореляції між змінними можна дійти висновку, що спостерігається розкидання точок. Це означає, що виражена залежність між параметрами.

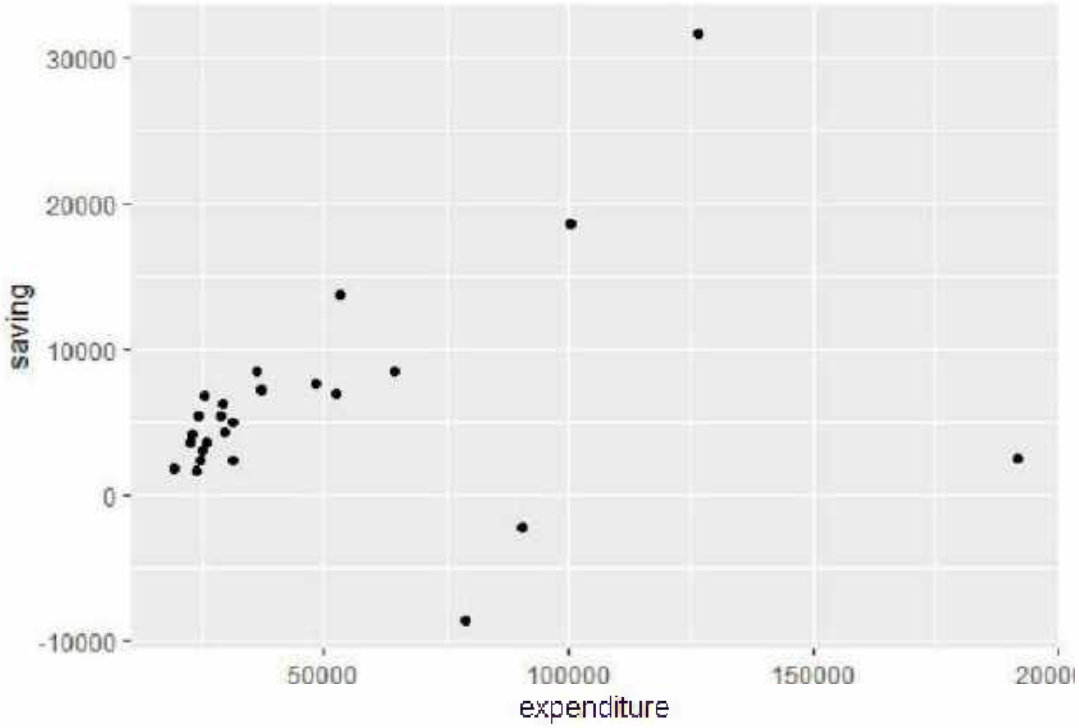


Рис. 4.3. Поле кореляції між витратами та заощадженнями

За отриманими діаграмами розсіювання можна зробити висновок, що щільність поля кореляції між результативною ознакою (заощадження) та факторною ознакою (наявні доходи) майже така, як у полі кореляції між заощадженнями та витратами. Це свідчить про те, що фактори-ознаки (витрати та наявні доходи) майже однаково впливають на результативну ознаку.

Будуємо економетричну модель МНК та викликаємо її протокол за допомогою функцій lm () та summary():

model<-lm(data=t,saving~income+expenditure)

summary(model)

Результати моделювання подано на рис. 4.4.

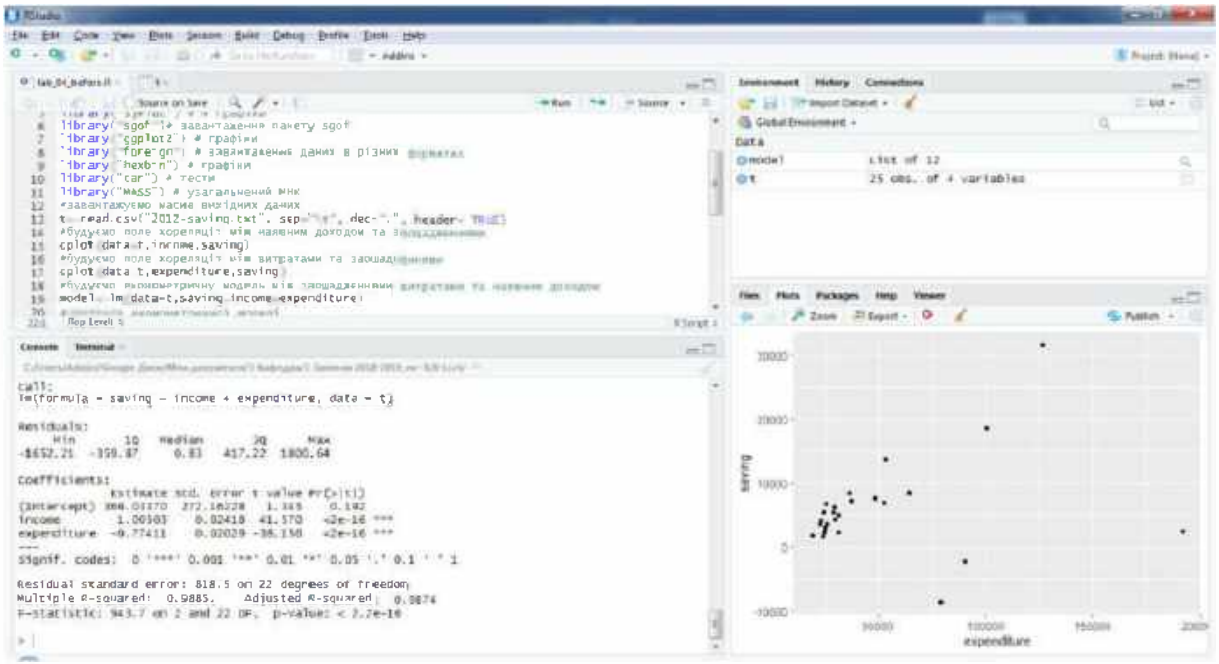


Рис. 4.5. Протокол економетричної моделі, побудованої класичним МНК

Економетрична залежність між заощадженнями, витратами і наявним доходом виглядає так:

$\widehat{saving}$= 336,03 + 1,005 · *income* - 0.7141· *expenditure*

Проводимо діагностику економетричної моделі на наявність гетероскедастичності, використовуючи функцію `gqtest()`.

```
gqtest(model, order.by = ~income, data=t, fraction = 0.25)
```

```
gqtest(model, order.by = ~expenditure, data=t, fraction = 0.25)
```

Після чого знаходимо критичне значення критерію Фішера з числом ступенів свободи df1 = 7, df2 =6, тобто виконуємо команду:

```
qf(0.90, 7, 6)
```

У консолі отримуємо результат розрахунків для гіпотези, що квадрат залишків пропорційний змінній:

```
Goldfeld-Quandt test

data: model
GQ = 3.5781, df1 = 7, df2 = 6, p-value = 0.0707
```

```
    alternative hypothesis: variance increases
from segment 1 to 2

    > #знаходимо критичне значення критерію
Фішера з числом ступенів свободи df1, df2
    > qf(0.90,7,6)
    [1] 3.014457
```

У консолі отримуємо результат розрахунків для гіпотези, що квадрат залишків пропорційний змінній:

```
    Goldfeld-Quandt test

    data: model
    GQ = 3.5781, df1 = 7, df2 = 6, p-value =
0.0707
    alternative hypothesis: variance increases
from segment 1 to 2
```

Порівняємо значення критерію R^*_{income} /$R^*_{expediture}$ (BG) = 3,5781 з критичним (табличним) значення F-критерію при γ1 = 7 і γ2 = 6 ступенях свободи і рівні значущості α = 0,05 F (α = 0,05) 7 6 =3,014457. Оскільки R^*_{income} /$R^*_{expediture}$> F(α = 0,05) 7 6, то вихідні дані з імовірністю 90 % мають гетероскедастичність.

Будуємо економетричну модель узагальненим методом найменших квадратів на основі з гіпотези про те, що дисперсія залишків пропорційна до зміни наявного доходу та викликаємо протокол, використовуючи функцію `rlm ()` та `summary ()`.

```
    model2<-rlm(data=t,
saving~income+expenditure,weight=1/incom
```

гіпотеза

Результати моделювання подані на рис. 4.6.

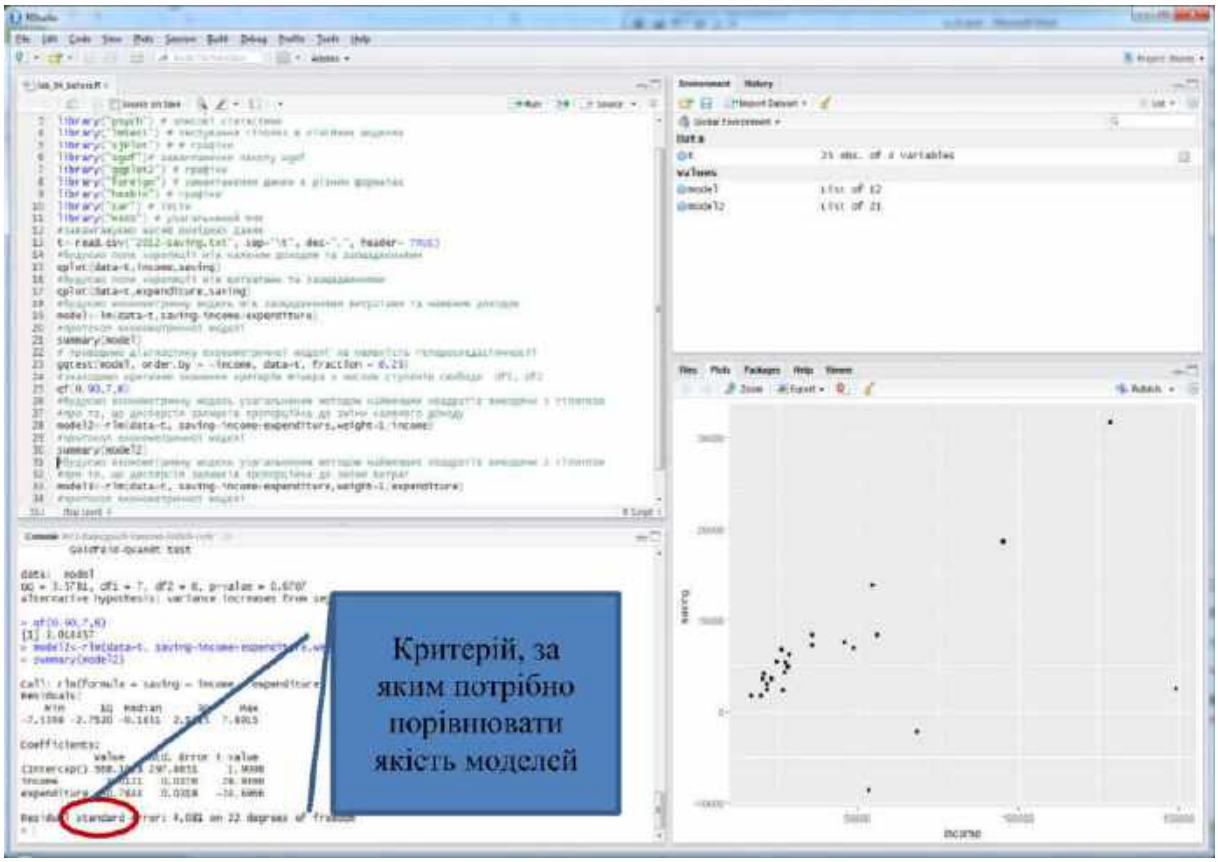


Рис. 4.6. Протокол економетричної моделі, побудованої узагальненим методом найменших квадратів на основі гіпотези про те, що дисперсія залишків пропорційна до зміни наявного доходу

Далі будуємо нову економетричну модель узагальненим методом найменших квадратів на основі з гіпотези про те, що дисперсія залишків пропорційна до зміни витрат (рис. 4.7):

```
model3<-rlm(data=t, saving~income+-
expenditure,weight=1/expenditure)
```

Порівнюємо дві моделі, побудовані на основі узагальненого МНК на критерії найменшого стандартного відхилення вектора залишків моделі. Можна дійти висновку, що більш якісною є модель, при якій була поставлена гіпотеза, що дисперсія залишків пропорційна до зміни витрат. Рівняння регресії виглядає так:

$$\widehat{saving} = 574.8041 + 1.0070 \cdot income - 0.7799 \cdot expenditure$$

З цієї економетричної моделі можна зробити такі висновки: коли за всіх однакових умов наявні доходи збільшаться на 1 млн грн, то заощадження зростуть на 1.0070 млн грн. У разі якщо за

всіх однакових умов витрати збільшаться на 1 млн грн, то заощадження зменшаться на 0.7799 млн грн.

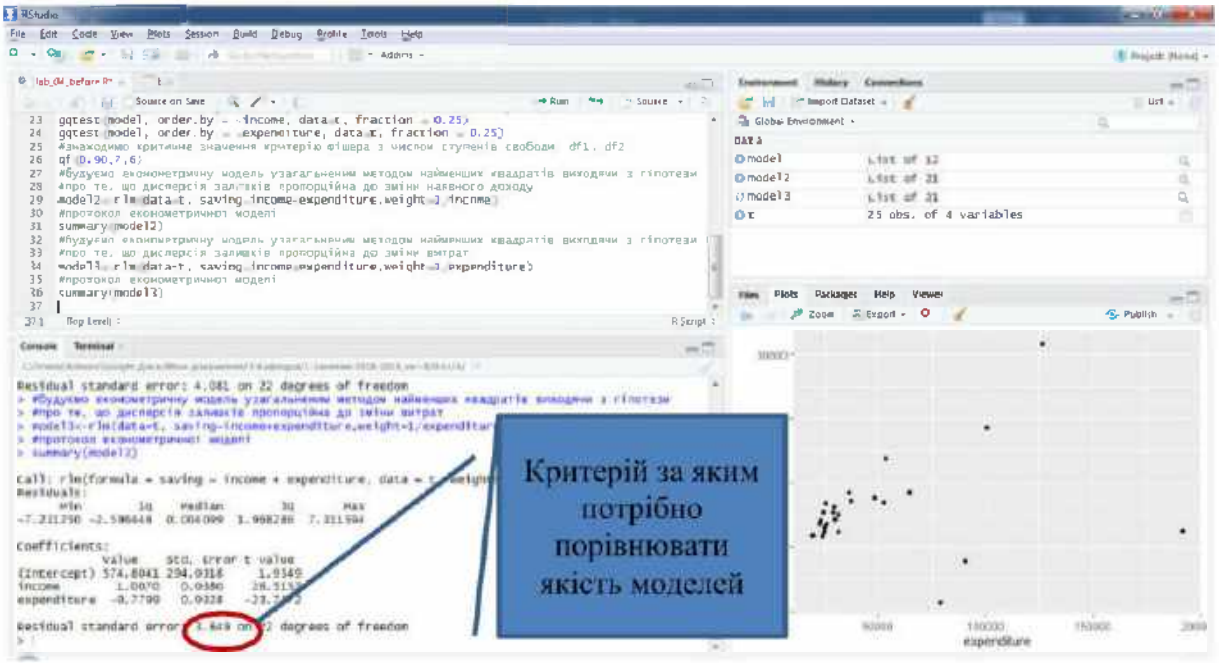


Рис. 4.7. Протокол економетричної моделі, побудованої узагальненим методом найменших квадратів на основі гіпотези про те, що дисперсія залишків пропорційна до зміни витрат

Додаток А до лабораторної роботи 4

ВИХІДНІ ДАНІ ДО ВИКОНАННЯ ЛАБОРАТОРНОЇ РОБОТИ 4*

Варіант 1				Варіант 2				Варіант 3			
2002				2003				2004			
	saving	income	expenditure		saving	income	expenditure		saving	income	expenditure
Vinnytsya	1444	4769	4370	**Vinnytsya**	1642	5604	5411	**Vinnytsya**	2142	7016	6605
Volyn	334	2742	3111	**Volyn**	355	3128	3688	**Volyn**	532	3904	4409
Dnipropetrovsk	922	11233	13835	**Dnipropetrovsk**	1254	13022	16077	**Dnipropetrovsk**	2663	17124	19578
Donetsk	4884	14852	14688	**Donetsk**	5275	17270	17572	**Donetsk**	8733	23407	20764
Zhytomyr	1058	3510	3423	**Zhytomyr**	1147	4156	4179	**Zhytomyr**	1679	5390	5188
Zakarpattya	-16	2601	3376	**Zakarpattya**	-76	3355	4413	**Zakarpattya**	-179	4305	5630
Zaporizhzhya	358	6071	7599	**Zaporizhzhya**	607	7166	8774	**Zaporizhzhya**	1117	9414	10690
Ivano-Frankivsk	729	3302	3462	**Ivano-Frankivsk**	1183	4120	4006	**Ivano-Frankivsk**	1806	5276	4739
Kyiv	1035	5659	6032	**Kyiv**	1598	6343	6546	**Kyiv**	2271	7908	7658
Kirovohrad	248	2894	3474	**Kirovohrad**	525	3379	3810	**Kirovohrad**	870	4243	4502
Luhansk	1635	6603	6905	**Luhansk**	2194	7845	7908	**Luhansk**	3773	10167	9112
Lviv	444	7177	8777	**Lviv**	1137	8860	10329	**Lviv**	1812	11400	12540
Mykolayiv	672	3483	3886	**Mykolayiv**	1007	4218	4471	**Mykolayiv**	1298	5378	5456
Odesa	-766	6076	8748	**Odesa**	-164	7717	10337	**Odesa**	39	10228	13428
Poltava	913	4934	5366	**Poltava**	1176	5696	6158	**Poltava**	2109	7407	7151
Rivne	186	2996	3686	**Rivne**	420	3481	4181	**Rivne**	745	4403	4930
Sumy	556	3463	3948	**Sumy**	967	4164	4421	**Sumy**	1569	5188	4962
Ternopil	393	2455	2726	**Ternopil**	799	3159	3184	**Ternopil**	1256	3958	3686
Kharkiv	-1032	8493	12285	**Kharkiv**	-1200	9981	14595	**Kharkiv**	-986	12829	17510
Kherson	277	2800	3252	**Kherson**	381	3273	3828	**Kherson**	764	4338	4716
Khmelnytskiy	183	3754	4510	**Khmelnytskiy**	149	4330	5285	**Khmelnytskiy**	764	5612	6295
Cherkasy	775	3461	3677	**Cherkasy**	945	4086	4369	**Cherkasy**	1348	5300	5396
Chernivtsi	-277	1867	2678	**Chernivtsi**	10	2513	3188	**Chernivtsi**	344	3294	3801

Chernihiv	809	3419	3489	**Chernihiv**	964	3960	4086	**Chernihiv**	1555	5063	4748
city of Kyiv	-4495	11139	22148	**city of Kyiv**	-5877	14577	28646	**city of Kyiv**	-6402	20224	36996
Варіант 4				**Варіант 5**				**Варіант 6**			
2005				**2006**				**2007**			
	saving	**income**	**expenditure**		**saving**	**income**	**expenditure**		**saving**	**income**	**expenditure**
Vinnytsya	2887	9908	9306	**Vinnytsya**	2748	11825	12040	**Vinnytsya**	2832	15052	16432
Volyn	738	5478	6070	**Volyn**	618	6638	7796	**Volyn**	655	8321	10288
Dnipropetrovsk	4633	24287	26711	**Dnipropetrovsk**	6100	29652	32594	**Dnipropetrovsk**	8331	39015	42624
Donetsk	11913	32945	28840	**Donetsk**	12121	40752	38826	**Donetsk**	13713	53011	53344
Zhytomyr	2269	7449	7149	**Zhytomyr**	2359	8940	9113	**Zhytomyr**	2887	11624	12245
Zakarpattya	-34	6012	7538	**Zakarpattya**	307	7166	8827	**Zakarpattya**	885	9105	11218
Zaporizhzhya	1758	13223	14602	**Zaporizhzhya**	2902	16118	17246	**Zaporizhzhya**	4260	20747	22164
Ivano-Frankivsk	2561	7404	6559	**Ivano-Frankivsk**	2481	8966	8751	**Ivano-Frankivsk**	2568	11537	12211
Kyiv	3441	11168	10383	**Kyiv**	3155	13978	14269	**Kyiv**	3084	17876	20061
Kirovohrad	1310	5989	6141	**Kirovohrad**	1422	7073	7569	**Kirovohrad**	1659	8962	9943
Luhansk	5431	14308	12429	**Luhansk**	6239	17591	15803	**Luhansk**	7540	22786	21427
Lviv	2691	16020	17132	**Lviv**	3215	19025	20589	**Lviv**	4262	24420	27099
Mykolayiv	1954	7516	7347	**Mykolayiv**	1728	8897	9518	**Mykolayiv**	1783	11324	12989
Odesa	-148	14173	18706	**Odesa**	-1351	16807	23808	**Odesa**	-3107	21031	32293
Poltava	3077	10404	9754	**Poltava**	3301	12481	12272	**Poltava**	3883	15913	16400
Rivne	1207	6192	6614	**Rivne**	1626	7557	8030	**Rivne**	2169	9555	10386
Sumy	2260	7376	6854	**Sumy**	2883	9092	8398	**Sumy**	4112	11892	10853
Ternopil	1923	5616	5046	**Ternopil**	2095	6967	6613	**Ternopil**	2560	8886	8807
Kharkiv	-976	18042	23902	**Kharkiv**	-185	22056	28373	**Kharkiv**	248	28961	37587
Kherson	1253	6115	6411	**Kherson**	1118	7283	8163	**Kherson**	1061	9315	11009
Khmelnytskiy	1382	7916	8486	**Khmelnytskiy**	1662	9494	10305	**Khmelnytskiy**	2301	12156	13250
Cherkasy	1851	7467	7541	**Cherkasy**	1576	9070	10013	**Cherkasy**	1451	11609	13726
Chernivtsi	589	4620	5187	**Chernivtsi**	614	5514	6396	**Chernivtsi**	713	7006	8465
Chernihiv	2271	7044	6379	**Chernihiv**	2292	8355	8141	**Chernihiv**	3037	10681	10523

city of Kyiv	-10622	28629	53970	city of Kyiv	-15717	36708	73024	city of Kyiv	-24000	49762	101435
Варіант 7				**Варіант 8**				**Варіант 9**			
2008				**2009**				**2010**			
	saving	**income**	**expenditure**		**saving**	**income**	**expenditure**		**saving**	**income**	**expenditure**
Vinnytsya	2041	19454	23129	**Vinnytsya**	2795	20181	24018	**Vinnytsya**	6591	26364	27011
Volyn	302	10982	14360	**Volyn**	611	11366	14566	**Volyn**	2136	14840	17001
Dnipropetrovsk	9038	52352	59382	**Dnipropetrovsk**	13300	55139	58838	**Dnipropetrovsk**	18953	69394	69969
Donetsk	17387	73310	74826	**Donetsk**	20821	76284	75775	**Donetsk**	29476	94860	88747
Zhytomyr	3324	15627	16892	**Zhytomyr**	3203	15712	17452	**Zhytomyr**	5730	20231	20394
Zakarpattya	863	12013	15170	**Zakarpattya**	543	12148	15949	**Zakarpattya**	2364	15825	18477
Zaporizhzhya	5223	27891	30405	**Zaporizhzhya**	5224	28433	31795	**Zaporizhzhya**	8340	36530	37439
Ivano-Frankivsk	2724	15563	17419	**Ivano-Frankivsk**	2662	16146	18361	**Ivano-Frankivsk**	4625	20716	21879
Kyiv	3404	24763	28262	**Kyiv**	4973	25915	29385	**Kyiv**	8545	33558	34187
Kirovohrad	1934	11873	13518	**Kirovohrad**	2056	11990	14093	**Kirovohrad**	4056	15632	16157
Luhansk	10009	31430	29852	**Luhansk**	10314	32670	31602	**Luhansk**	14060	41080	37278
Lviv	5631	32513	36589	**Lviv**	6707	33706	37106	**Lviv**	9821	42064	45017
Mykolayiv	1847	15148	18103	**Mykolayiv**	1972	15689	18751	**Mykolayiv**	4466	20161	21568
Odesa	-4440	28133	43807	**Odesa**	-4434	29666	46856	**Odesa**	-2595	38896	55519
Poltava	4730	21135	22294	**Poltava**	5603	22195	22636	**Poltava**	8143	26873	26319
Rivne	2345	12810	14520	**Rivne**	2224	13029	15234	**Rivne**	4052	16855	18310
Sumy	4880	15698	14939	**Sumy**	4966	15819	15352	**Sumy**	6737	19692	18181
Ternopil	2645	11599	12354	**Ternopil**	2347	11603	13261	**Ternopil**	3928	15020	15659
Kharkiv	-347	39232	52318	**Kharkiv**	194	40617	54325	**Kharkiv**	2391	50962	64711
Kherson	1208	12410	14998	**Kherson**	1292	12542	15439	**Kherson**	3048	16233	17930
Khmelnytskiy	2399	16067	18197	**Khmelnytskiy**	2694	16371	18832	**Khmelnytskiy**	5255	20995	21732
Cherkasy	1751	15702	18673	**Cherkasy**	2238	16269	19113	**Cherkasy**	3981	20347	22213
Chernivtsi	344	8856	11870	**Chernivtsi**	436	9054	12183	**Chernivtsi**	1982	12211	14132
Chernihiv	3359	13942	14549	**Chernihiv**	4089	14675	14828	**Chernihiv**	5422	18354	17757
city of Kyiv	-27761	68713	137542	**city of Kyiv**	-15307	76250	137823	**city of Kyiv**	-987	103347	144890

Варіант 10				Варіант 11				Варіант 12			
2011				2012				2013			
	saving	income	expenditure		saving	income	expenditure		saving	income	expenditure
Vinnytsya	6290	31265	32700	**Vinnytsya**	7239	35441	37026	**Vinnytsya**	5090	37323	41067
Volyn	1530	17289	21054	**Volyn**	1672	19546	24069	**Volyn**	-95	20609	27002
Dnipropetrovsk	15423	80888	86445	**Dnipropetrovsk**	18610	95349	100213	**Dnipropetrovsk**	15467	99995	109127
Donetsk	27328	109968	108271	**Donetsk**	31637	128767	126366	**Donetsk**	26408	135362	139958
Zhytomyr	4578	23884	25491	**Zhytomyr**	5362	27057	28748	**Zhytomyr**	3121	27405	31826
Zakarpattya	2170	18738	22276	**Zakarpattya**	2988	21534	25040	**Zakarpattya**	1961	22512	27141
Zaporizhzhya	5624	42550	46648	**Zaporizhzhya**	6931	48482	52260	**Zaporizhzhya**	6616	50545	56055
Ivano-Frankivsk	4003	24471	27221	**Ivano-Frankivsk**	4923	27825	31263	**Ivano-Frankivsk**	1301	29003	36009
Kyiv	6487	38703	42503	**Kyiv**	7576	45673	48365	**Kyiv**	2374	47216	56520
Kirovohrad	3502	18151	19941	**Kirovohrad**	4112	20529	23017	**Kirovohrad**	2993	21485	24702
Luhansk	12231	47648	46388	**Luhansk**	13755	54405	53270	**Luhansk**	10284	57527	61201
Lviv	7720	48833	55882	**Lviv**	8390	55878	64438	**Lviv**	7687	58762	68075
Mykolayiv	3597	23664	26203	**Mykolayiv**	4267	26900	29640	**Mykolayiv**	2513	27948	32612
Odesa	-7303	45704	68738	**Odesa**	-8595	53154	79024	**Odesa**	-7937	61265	86222
Poltava	7516	31010	31783	**Poltava**	8378	35380	36457	**Poltava**	7967	37118	39017
Rivne	3182	19984	22962	**Rivne**	3577	22950	25980	**Rivne**	3718	24507	28093
Sumy	5879	22667	22468	**Sumy**	6739	25920	25676	**Sumy**	5914	26812	27555
Ternopil	2970	17697	19742	**Ternopil**	3587	20025	22515	**Ternopil**	2478	20424	23867
Kharkiv	-2204	59886	79520	**Kharkiv**	-2221	68226	90573	**Kharkiv**	-6530	71530	97863
Kherson	2213	19169	21883	**Kherson**	2389	21551	24832	**Kherson**	2185	23362	27304
Khmelnytskiy	4806	24800	26530	**Khmelnytskiy**	6198	28437	29260	**Khmelnytskiy**	5639	29865	31131
Cherkasy	2183	23379	27463	**Cherkasy**	2310	26376	31174	**Cherkasy**	1160	27353	33864
Chernivtsi	1286	14470	17396	**Chernivtsi**	1717	16409	19295	**Chernivtsi**	1841	17646	20567
Chernihiv	4238	21283	22039	**Chernihiv**	5401	24664	24253	**Chernihiv**	3852	25306	26541
city of Kyiv	-841	119500	164898	**city of Kyiv**	2469	149145	191967	**city of Kyiv**	7311	159534	211436

Варіант 12				Варіант 13				Варіант 14			
2014				2015				2016			
	saving	income	expenditure		saving	income	expenditure		saving	income	expenditure
Vinnytsya	563	37812	48855	**Vinnytsya**	1445	47609	59478	**Vinnytsya**	323	55761	71565
Volyn	-3923	20985	31909	**Volyn**	-6341	26049	40405	**Volyn**	-4957	31267	45749
Dnipropetrovsk	9324	105223	127486	**Dnipropetrovsk**	11652	127830	154424	**Dnipropetrovsk**	9119	143861	179697
Donetsk	33983	113343	108762	**Donetsk**	42598	91388	74873	**Donetsk**	32467	89036	85268
Zhytomyr	-188	27831	37002	**Zhytomyr**	-816	34799	45869	**Zhytomyr**	-1208	41026	54892
Zakarpattya	-2069	21840	32057	**Zakarpattya**	-2674	28282	39856	**Zakarpattya**	-2924	33812	47061
Zaporizhzhya	3690	53449	64637	**Zaporizhzhya**	3957	63841	77780	**Zaporizhzhya**	4514	75910	92181
Ivano-Frankivsk	-4844	28139	42692	**Ivano-Frankivsk**	-4266	36689	51418	**Ivano-Frankivsk**	-4570	43807	60988
Kyiv	-3562	49133	66904	**Kyiv**	-11465	58767	87615	**Kyiv**	-10417	69552	100922
Kirovohrad	38	21605	28863	**Kirovohrad**	-890	26750	36240	**Kirovohrad**	-2263	31746	44138
Luhansk	19801	44124	36432	**Luhansk**	21215	34594	22942	**Luhansk**	13420	30348	27847
Lviv	2276	59887	77102	**Lviv**	-594	74919	98334	**Lviv**	-2106	89517	118391
Mykolayiv	-1151	27362	37524	**Mykolayiv**	-1191	34075	45466	**Mykolayiv**	-1807	40363	54197
Odesa	-18171	58096	98609	**Odesa**	-17101	77509	118280	**Odesa**	-22280	93463	140752
Poltava	4500	38078	45428	**Poltava**	4065	46203	56545	**Poltava**	5037	54362	66889
Rivne	579	25266	32735	**Rivne**	168	31021	40141	**Rivne**	210	36374	47146
Sumy	2755	27007	32620	**Sumy**	4828	34189	39483	**Sumy**	4650	40014	47901
Ternopil	-1383	19718	28275	**Ternopil**	-738	25670	34589	**Ternopil**	-1186	29954	41463
Kharkiv	-18370	71841	114267	**Kharkiv**	-22772	87736	139652	**Kharkiv**	-25550	103509	161225
Kherson	-3042	22183	33119	**Kherson**	-3330	29695	41563	**Kherson**	-4894	34913	49162
Khmelnytskiy	1973	29585	36880	**Khmelnytskiy**	3933	38015	44720	**Khmelnytskiy**	3049	44362	54318
Cherkasy	-2248	27329	38942	**Cherkasy**	-2072	33642	46780	**Cherkasy**	-2429	39992	55925
Chernivtsi	-1699	16798	24640	**Chernivtsi**	-1874	21773	30190	**Chernivtsi**	-2081	25780	35738
Chernihiv	761	24509	31237	**Chernihiv**	571	29871	38209	**Chernihiv**	-608	34534	46324
city of Kyiv	11187	180513	229011	**city of Kyiv**	12765	221683	276091	**city of Kyiv**	29082	269030	309001

Варіант 15			
2017			
	saving	income	expenditure
Vinnytsya	8542	71930	85875
Volyn	-578	40043	53782
Dnipropetrovsk	19987	185230	225791
Donetsk	38599	106733	102741
Zhytomyr	-1239	52753	71365
Zakarpattya	-5314	42652	61882
Zaporizhzhya	6791	93942	115968
Ivano-Frankivsk	3847	55947	69627
Kyiv	-29796	88379	147551
Kirovohrad	1282	40580	53232
Luhansk	13856	35814	35486
Lviv	-9664	113883	161920
Mykolayiv	2649	51964	64909
Odesa	-24738	119505	178378
Poltava	4117	69116	88651
Rivne	4651	46846	57180
Sumy	6660	50410	60627
Ternopil	2936	38224	49260
Kharkiv	-27277	130484	203127
Kherson	-5173	43834	62317
Khmelnytskiy	501	55848	73019
Cherkasy	-3419	51304	72818
Chernivtsi	-6100	32861	49642
Chernihiv	-426	43636	58681
city of Kyiv	29944	346360	417615

*У разі якщо кількість осіб у групі більша за 15, 16-й за списком студент обирає 1-й варіант.

ЛАБОРАТОРНА РОБОТА 5
ПОБУДОВА, ДОСЛІДЖЕННЯ ЛІНІЙНОЇ БАГАТОФАКТОРНОЇ ЕКОНОМЕТРИЧНОЇ МОДЕЛІ З АВТОКОРЕЛЯЦІЄЮ ЗАЛИШКІВ ТА МОДЕЛІ З ІНСТРУМЕНТАЛЬНИМИ ЗМІННИМИ

***Мета*:** набуття практичних навичок побудови багатофакторної економетричної моделі з автокореляцією залишків, її дослідження, практичне використання та моделі з інструментальними змінними.

Задачі частини 1

1. Побудова економетричної моделі МНК.
2. Дослідження економетричної моделі на наявність автокореляції залишків.
3. Побудова економетричної моделі з використанням методу усунення автокореляції залишків: додання як пояснюючу змінну значення залежної змінної попереднього періоду.
4. Формування економічних висновків на основі побудованої економетричної моделі.

Задачі частини 2

1. Побудова економетричної моделі МНК.
2. Побудова економетричної моделі з інструментальними змінними.
3. Формування економічних висновків на основі побудованої економетричної моделі.

Завдання частини 1

На підставі з інформації Державної служби статистики України (http://www.ukrstat.gov.ua/) для одного з регіонів (залежно від варіанта, додаток Б до лабораторної роботи) побудувати економетричну модель цінової еластичності попиту на бензин марки А-95/дизельного палива (залежно від варіанта) зі

щомісячною дискретністю на основі даних за 2017–2021 рр. Як залежна змінна необхідно обрати обсяг споживання палива на роздрібному ринку (Y_t, л), а змінними факторами ціна палива *(*X_{1_t}*, грн/л)* та обсяг споживання палива на роздрібному ринку в попередньому періоді (Y_{t-1}, л) для цього потрібно:

1. Побудувати економетричну модель залежності споживання палива на роздрібному ринку (Y_t, л) від його ціни *(*X_{1_t}*, грн/л).*

2. Провести діагностику економетричної моделі на наявність автокореляції, зробити висновки за тестом Дарбіна-Уотсона та Бреуша-Годфрі.

3. Побудувати економетричну модель з використанням методу усунення автокореляції залишків: додання як пояснюючу змінну значення залежної змінної попереднього періоду (лагової змінної), оцінити значимість оцінок параметрів та адекватність моделі в цілому.

4. Провести діагностику нової економетричної моделі на наявність автокореляції, зробити висновки за тестом Дарбіна-Уотсона та Бреуша-Годфрі.

5. Провести діагностику економетричної моделі на наявність гетероскедастичності, зробити висновки за тестом Гольдфельда-Квандта.

6. Провести тест Рамсея на пропущені змінні.

7. Розрахувати коефіцієнт цінової еластичності попиту на бензин марки А-95.

Завдання до частини 2

На підставі інформації Державної служби статистики України (http://www.ukrstat.gov.ua/) для одного з регіонів (залежно від варіанта, які наведно в додатку Б до лабораторної роботи) побудувати економетричну модель цінової еластичності попиту на дизельне паливо зі щомісячною дискретністю на основі даних за 2017–2021 рр. Як залежну змінну, необхідно обрати обсяг споживання палива на роздрібному ринку (Y_t, л), а змінною-фактором ціну палива (X_{1_t}, грн/л) за умови, що на неї впливає курс долара США до гривні:

1. Побудувати економетричну модель з використанням методу інструментальних змінних, а саме врахувати вплив на ціну дизельного палива курсу долара США щодо гривні, оцінити значимість оцінок параметрів.

2. Провести діагностику економетричної моделі на наявність автокореляції, зробити висновки за тестом Бреуша-Годфрі.

3. Провести діагностику економетричної моделі на наявність гетероскедастичності, зробити висновки за тестом Гольдфельда-Квандта.

4. Провести тест Рамсея на пропущені змінні та зробити висновок.

5. Розрахувати коефіцієнт цінової еластичності попиту на дизельне паливо та зробити висновки.

Приклад виконання (на основі даних 2011–2015 рр.)

У період тривалих податкових трансформацій, які мають на меті підвищення фіскальної ефективності податків та забезпечення збалансованості бюджету, надзвичайно актуальними є питання, пов'язані з реалізацією потенціалу екологічних податків, зокрема акцизів на нафтопродукти та скраплений газ.

Зважаючи на заявлені євроінтеграційні наміри, Україна не може залишитися осторонь процесів трансформації податкових систем країн ЄС. В Україні зазначені процеси, розпочаті з розроблення й прийняття Податкового кодексу України, мають також торкнутися і екологічного оподаткування, розвиток якого повинен відбуватись у напрямі підвищення ефективності реалізації фіскального та регулюючого потенціалу. Необхідне поступове підвищення ставок енергетичних податків, адже їх рівень за всіма складовими в Україні значно нижчий відповідного рівня у країнах ЄС.

Проте під час прийняття рішень щодо реалізації фіскального потенціалу енергетичних податків має бути враховано вплив зазначених заходів на рівень купівельної спроможності населення та конкурентоспроможність підприємств, виробництво яких пов'язане зі значним обсягом споживання енергоресурсів, а також зміщення попиту в „тіньовий сектор" економіки.

Нині рівень купівельної спроможності населення значно нижчий за європейський. Обсяг дизпалива та бензину, який можна придбати на середньомісячну заробітну плату в Україні, у 7,5 і 6,7 раза менший порівняно з Нідерландами (країною з найвищими ставками акцизів на нафтопродукти) (рис. 5.1). Найбільший обсяг нафтопродуктів можна купити на середньомісячну заробітну плату у Люксембурзі – 3 434,5 л бензину і 3 772,6 л дизельного палива, що менше ніж в Україні в 11,6 раза. При тому, що в цій країні ціни на нафтопродукти є одними з найнижчих в Європі (четверте місце з кінця), а середньомісячна заробітна плата найвища серед країн ЄС і 2014 р. становила 4 590 євро (у 15 разів вище ніж в Україні).

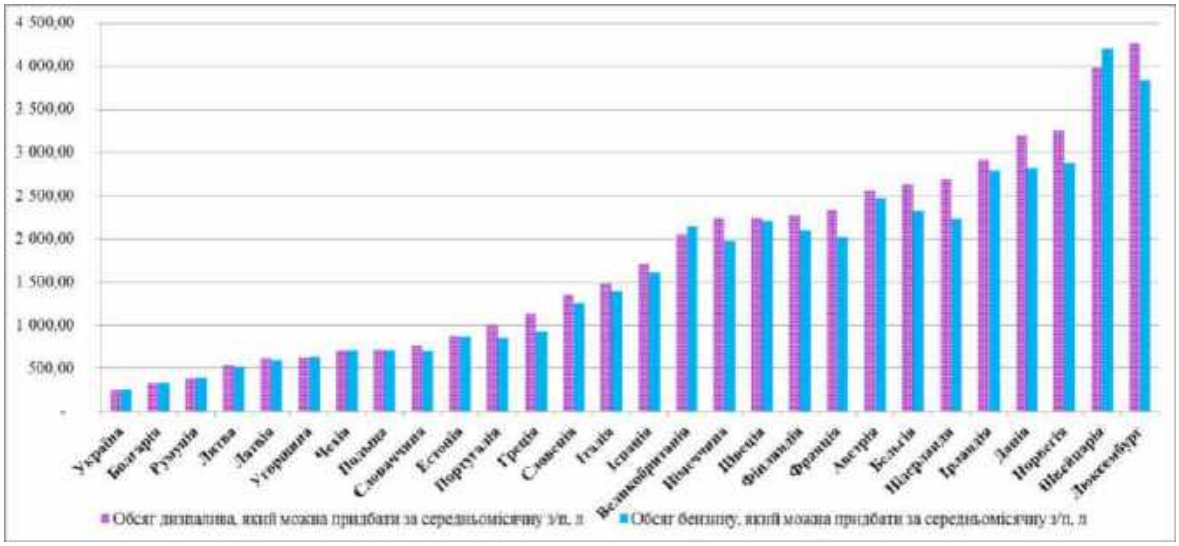


Рис. 5.1. Обсяг бензину та дизельного палива, які можна придбати в Україні та обраних країнах ЄС на середньомісячну заробітну плату 2014 р.

Джерело: розраховано за даними Oil Bulletin Prices History, OECP, Державної служби статистики України та Євростату.

Дослідивши рівень купівельної спроможності в Україні, порівняно з європейськими країнами, варто зазначити його значне зниження протягом 2014–2015 рр. Цей факт, а також втрата частини територій та девальвація гривні стали основними причинами значного падіння обсягів споживання нафтопродуктів на 40 %. Для з'ясування того, наскільки зростання ціни нафтопродуктів впливає на попит і пропозицію, доцільно провести аналіз їх цінової еластичності.

Аналізуючи динаміку середньозважених цін на бензин та дизпаливо та їх споживання (рис. 5.2), варто зазначити, що в Україні найважливішим фактором, який впливає на ціни нафтопродуктів, є курс національної валюти. Значні зміни в динаміці цін на нафтопродукти нерівномірно вплинули на обсяги їх реалізації. Наявні розриви в коливаннях попиту і ціни є аргументом на користь гіпотези щодо низької цінової еластичності попиту в середньостроковому періоді.

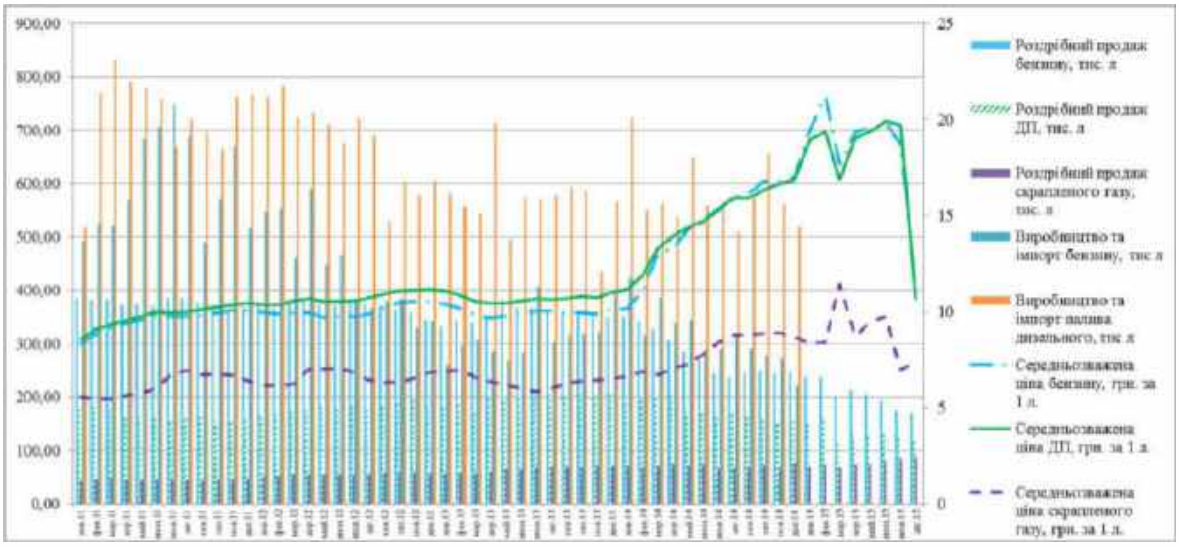


Рис. 5.2. Динаміка реальних середньозважених цін, обсягів виробництва, імпорту і реалізації бензину, дизельного палива та скрапленого газу на роздрібному ринку, скоригованих на сезонність у 2011–2015 рр.

Джерело: розраховано за даними щомісячних експрес-випусків Державної служби статистики України [106] та запитів у зазначену службу в програмному пакеті статистичного аналізу Statistica 7, використовуючи опцію *X-11 Census Method II*.

Для оцінки цінової еластичності попиту і пропозиції нафтопродуктів побудуємо набір економетричних моделей для кожного його виду палива. Вихідні дані для аналізу наведені на рис. 2. Оскільки в Україні щомісячні дані щодо обсягу споживання не формуються, за показник, що відображає споживання, обрано обсяг реалізації палива на роздрібному ринку. Зважаючи на те, що ціна й обсяги споживання, виробництва та імпорту нафто-продуктів і скрапленою газу значно обумовлені сезонною

компонентою, проведено коригування вихідних даних на сезонність, а також враховано вплив інфляційної компоненти на ціну.

Для специфікації моделей було проаналізовано досвід іноземних фахівців і визначено, що більшість з них використовують такий вигляд:

$$Ln(\hat{Y}_t) = \hat{\beta}_0 + \hat{\beta}_1 \cdot Ln(X_t) + \hat{\beta}_2 \cdot Ln(Y_{t-1}), \qquad (5.1)$$

де $\hat{Y}_t$ – розрахунковий обсяг споживання (виробництва та імпорту) палива в поточному періоді скоригований на сезонність;

X_t – ціна палива в поточному періоді скоригована на сезонність та інфляційну компоненту;

Y_{t-1} – обсяг споживання (виробництва та імпорту) палива в попередньому періоді скоригований на сезонність.

На основі інформаційних критеріїв Акаіке та Шварца було підтверджено доцільність такої специфікації для моделей цінової еластичності пропозиції. Для моделей цінової еластичності попиту було обрано лінійну специфікацію. Використовуючи аналітичний метод специфікації, зазначимо, що на підставі того, що вихідні дані є часовими рядами, в моделі може бути наявна автокореляція залишків, яка також може бути зумовлена інерцією у зміні економічних показників, ефектом павутини та проведеним сезонним корегуванням даних. Без урахування цього факту при специфікації моделі оцінки параметрів можуть бути неефективними, а також неможливо буде провести перевірку значимості оцінок параметрів за критерієм Стьюдента та адекватності моделі в цілому за критерієм Фішера. У зв'язку з цим було побудовано два типи моделей: на основі узагальненого методу найменших квадратів і класичного методу найменших квадратів з включенням додатково у модель як незалежну змінну значення залежної змінної попереднього періоду. Це дозволило дотриматись усіх передумов методу найменших квадратів та отримати достовірні оцінки параметрів економетричної моделі.

У табл. 5.1 побудовані рівняння регресії цінової еластичності попиту та пропозиції на бензин і дизельне паливо. Були проведені всі

необхідні етапи економетричного моделювання. Крім специфікації моделі, оцінено невідомі параметри економетричних моделей, перевірено їх значимість та адекватність моделей у цілому, проведено діагностику моделей на автокореляцію залишків та гетероскедастичність. Основні результати економетричного моделювання представлено в додатку А до лабораторної роботи. Оцінки параметрів моделі є статистично значимими, а рівняння регресій адекватні в цілому, значення коефіцієнтів детермінації – високі. Попит на бензин на 97,99 % обумовлюється ціною та попитом минулого періоду, а попит на дизпаливо – на 85,3 % (табл. 5.1). Значення коефіцієнтів детермінації за моделями цінової еластичності пропозиції нафтопродуктів є дещо меншими, пропозиція бензину обумовлюється ціною та пропозицію минулого періоду на 77,56 %, а дизпалива – на 53,92 %.

Коефіцієнти еластичності в економетричних моделях цінової еластичності попиту дозволяють зробити висновки, що попит на бензин та дизпаливо в Україні є нееластичним у середньостроковій перспективі. Еластичність попиту на бензин становила – 0,21, на дизпаливо – 0,12, це свідчить про те, що зростання ціни на 1 % зумовлює зменшення попиту на бензин на 0,21 %, а на дизпаливо – на 0,12 %.

Порівнюючи отримані показники еластичності попиту з результатами розрахунків закордонних фахівців, варто зазначити, що в Україні еластичність попиту на бензин перебуває на рівні розвинених країн, а попиту на дизпаливо є дещо нижчою. Так, у роботах[1,2,3,4] цінова еластичність попиту на паливо в середньостроковій перспективі становила від –0,2 до –0,3.

---

[1] Meta-analysis of the price elasticity of gasoline demand. A SUR approach / M. Brons, P. Nijkamp, E. Pels, P. A. Rietveld. *Energy Economics*. 2008. № 30. P. 2105–2122.

[2] Dahl C., Sterner T. A survey of econometric gasoline demand elasticities. *International Journal of Energy Systems*. 1991. № 11. P. 53–76.

[3] Espey M. Gasoline demand revisited: an international meta-analysis of elasticities. *Energy Economics*. 1998. № 20. P. 273–295.

[4] Graham D., Glaister S. The demand for automobile fuel: a survey of elasticities. *Journal of Transport Economics and Policy*. 2002. № 36. P. 1–26.

Таблиця 5.1

Економетричні моделі цінової еластичності попиту і пропозиції нафтопродуктів*

Вид палива	Вид економетричної моделі
Модель цінової еластичності попиту	
Бензин	$Cons_{petr_t} = 201{,}22 - 7{,}45 \cdot \Pr ice_{petr} + 0{,}66 \cdot Cons_{petr_{t-1}}$
	Коефіцієнт детермінації: $R^2 = 0{,}9799$
Дизпаливо	$Cons_{gasoil_t} = 64{,}06 - 1{,}68 \cdot \Pr ice_{gasoil} + 0{,}74 \cdot Cons_{gasoil_{t-1}}$
	Коефіцієнт детермінації: $R^2 = 0{,}853$
Модель цінової еластичності пропозиції	
Бензин	$Ln(\Pr_I_{petr_t}) = 22{,}22 - 0{,}56 \cdot Ln(\Pr ice_{petr}) + 0{,}8 \cdot Ln(\Pr_I_{petr_{t-1}})$
	Косфіцієнт детермінації: $R^2 = 0{,}7756$
Дизпаливо	$Ln(\Pr_I_{gasoil_t}) = 4{,}51 - 0{,}14 \cdot Ln(\Pr ice_{gasoil}) + 0{,}38 \cdot Ln(\Pr_I_{gasoil_{t-1}})$
	Косфіцієнт детермінації: $R^2 = 0{,}5392$

Джерело: розрахунки автора в системі статистичного аналізу R.

*Позначення: $Cons_{petr_t}$, $Cons_{gasoil_t}$, $\Pr_I_{petr_t}$, $\Pr_I_{gasoil_t}$ – обсяги споживання, виробництва й імпорту бензину та дизельного палива в поточному періоді, скориговані на сезонність;

$\Pr ice_{petr}$, $\Pr ice_{gasoil}$ – ціни бензину та дизельного палива в поточному періоді, скориговані на сезонність та інфляційну складову;

$Cons_{petr_{t-1}}$, $Cons_{gasoil_{t-1}}$, $\Pr_I_{petr_{t-1}}$, $\Pr_I_{gasoil_{t-1}}$ – обсяги споживання, виробництва та імпорту бензину та дизельного палива в попередньому періоді, скориговані на сезонність.

За результатами моделювання було з'ясовано, що попит на дизпаливо є менш еластичним, ніж попит на бензин. Вважаємо, що це обумовлюється тим, що дизпаливо більшою мірою використовується в комерційному секторі, а бензин – домогосподарствами. І зростання цін на дизпаливо меншою мірою зменшує його споживання, якщо у домогосподарств є вибір: їхати чи ні, то у комерційних підприємств вибір полягає у тому: закривати бізнес чи продовжувати працювати. Отже, підприємства реагують на зростання витрат на паливо відповідним зростанням витрат на продукцію, перекладаючи певну частину витрат на паливо на споживачів. Крім того, вплив зростання цін на бензин у

секторі домогосподарств певною мірою зумовив ефект заміщення приватного транспорту громадським.

Варто також зважати, що можливість зміни споживчої поведінки обумовлюється рівнем доходів населення. Отримуючи більші доходи, споживачі палива мають можливість змінити споживчу поведінку: придбати менш енергоємні автомобілі, скоротити відстань між роботою і житлом та ін.

У зазначених вище працях зарубіжних фахівців розрахована довгострокова цінова еластичність, яка є трохи більшою: від –0,6 до –0,8. Така різниця пояснюється тим, що під впливом високих цін на світлі нафтопродукти в довгостроковій перспективі споживачі можуть змінювати свою поведінку, купуючи менш енергоємні автомобілі, а також змінюючи місце проживання. Така різниця в еластичності пов'язана з триступеневою природою реакції споживачів та виробників на зміни в економічних стимулах. Короткостроковий ефект підвищення податку (ціни) на споживання енергоносіїв проявляється у тому, що споживачі стають менше користуватися автомобілем. У середньостроковому періоді споживачі купують більш економічні автомобілі, а в довгостроковому – винаходять електромобілі[5].

Проте цінова еластичність значно відрізняється від нуля, що означає, що зростання ціни може істотно знизити потребу в енергоресурсах. Отже, екологічні податки можуть мати значний вплив на зниження попиту на нафтопродукти, особливо в довгостроковій перспективі.

На основі проведеного економетричного моделювання цінової еластичності пропозиції нафтопродуктів можна зробити висновки, що пропозиція бензину і дизпалива в Україні нееластична в середньостроковому періоді. Коефіцієнти еластичності становили для бензину –0,56, для дизпалива –0,14. Варто зазначити, що отримані показники еластичності значно відрізняються від показників розвинених країн світу, в яких пропозиція нафтопродуктів є еластичною. Ці результати також

---

5 Чаплигін О. В. Податкове стимулювання енергоефективних інвестицій в Україні / дис… канд. екон. наук : 08.00.08. Ірпінь, 2016. 198 с.

підтверджують висновки щодо того, що зростання ціни зміщує пропозицію палива у тіньовий сектор, зумовлює появу на ринку контрабандної та фальсифікованої продукції.

Результати економетричного моделювання цінової еластичності попиту та пропозиції бензину і дизельного палива.

Таблиця 5.2

Основні результати економетричного моделювання цінової еластичності попиту та пропозиції нафтопродуктів і скрапленого газу

Показник	Значення показника для моделі цінової еластичності				
	попиту на бензин	попиту на дизельне паливо	пропозиції бензину	пропозиції дизельного палива	попиту на скраплений газ
Коефіцієнт детермінації, R^2	0,9799	0,853	0,7756	0,5392	0,8585
Скоригований коефіцієнт детермінації, $\overline{R}^2$	0,9791	0,8473	0,7654	0,5183	0,8506
Критерій Фішера, F	1268		76,03	25,74	109,2
t_{β_0}	4,101	1,937	2,559	8,326	0,542
t_{β_1}	–4,444	–2,477	–3,433	–1,457	1,458
t_{β_2}	7,564	5,033	8,505	5,723	6,245
Коефіцієнт еластичності, E	–0,29	-0,12	–0,56	–0,14	0,2891
Статистика Дарбіна-Уотсона, d	1,7	-	2,4893	1,837	2,3113
Статистка Бреуша-Годфрі, BG	1,7874	2,3497	6,1984	0,6671	4,1551
Статистика Гольдфельда-Квандта, GQ	1,4952	3,6959	0,662	0,8096	3,9937
Тест Рамсея, $RESET$	0,8433	0,8905	0,8997	4,6295	0,7372

Джерело: розраховано авторами у системі статистичного аналізу R.

Модель цінової еластичності попиту на бензин

Створіть файл в Excel з вихідними даними для дослідження, одразу занесіть у нього дані щодо значення залежної змінної попереднього періоду. Дані щодо обсягів реалізації палива потрібно скопіювати з файлу realization.xlsx, а дані щодо середньої за місяць ціни – вибрати з файлів sctp_2017_u.xlsx, sctp_2018_u.xlsx, sctp_2019_u.xlsx, sctp_2020_u.xlsx, sctp_2021_u.xlsx. Після цього дані занесіть у блокнот та сформуйте файл з розширенням *txt. Далі відкрийте R та завантажте файл lab_5_before.R і встановіть робочу папку. Для цього зайдіть у вкладку **Session,** оберіть **Set Working directory,** далі натисніть кнопку **Choose Directory,** після цього вкажіть шлях до вашої папки (рис. 5.3).

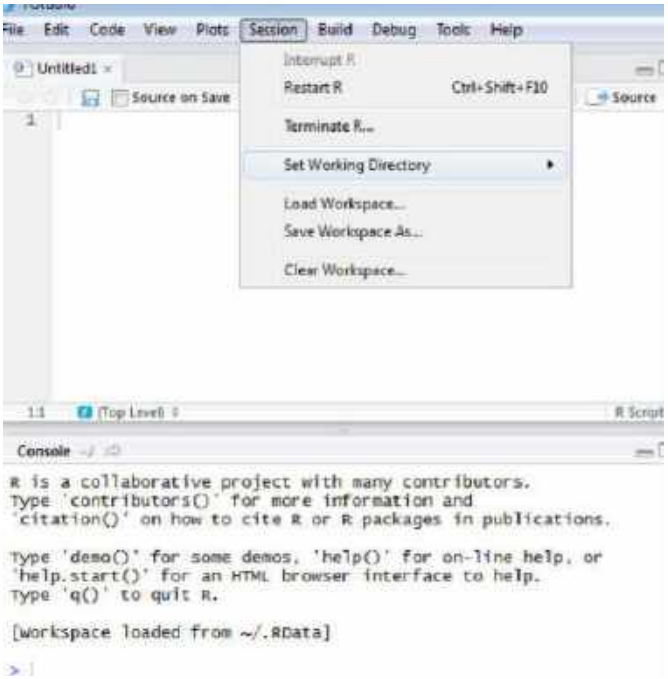


Рис. 5.3. Обрання робочої папки в R

Після обрання робочої папки ви повинні завантажити пакети, які будуть вам потрібні для виконання лабораторної роботи. Для цього в робочому середовищі ***R,*** наберіть такі команди:

```
library("dplyr") # маніпуляції з даними
library("psych") # описові статистики
library("lmtest") # тестування гіпотез в лінійних моделях
library("ggplot2") # графіки
```

```
library("foreign") # завантаження даних в різних форматах
library("car") # тести
library("AER") # інструментальні змінні
library("ivpack") # додаткові опції для інструментальних змінних
```

Якщо у разі копіювання команд та вставки у робочу зону текст українською мовою відображається символами, які неможливо прочитати, необхідно у вкладці **File** натиснути кнопку **Reopen with encoding** та обрати кодування **UTF–8.**

Загружати необхідні для проведення розрахунків бібліотеки з командами потрібно щоразу перед початком роботи в інтегрованому середовищі.

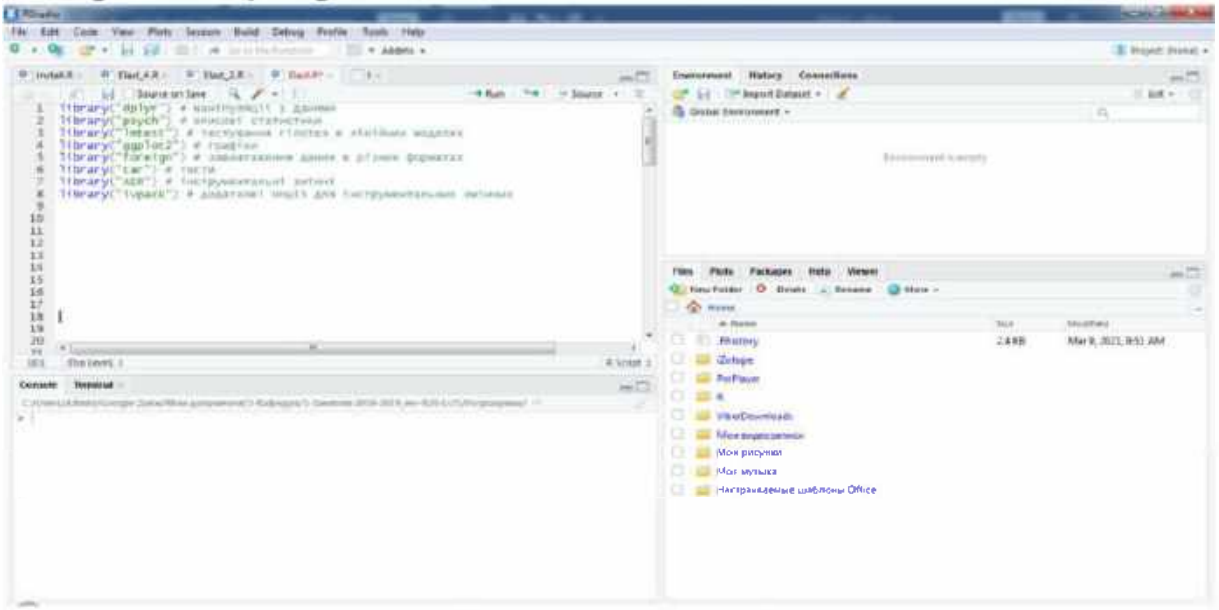


Рис. 5.4. Завантаження необхідних пакетів у R

Після того, як пакети активовано, вам необхідно завантажити підготовлений заздалегідь файл з даними (рис. 4). Для цього оберіть ім'я таблиці, в якій буде зберігатись масив інформації, та напишіть його в робочому середовищі, наприклад ***t***, далі поставте знак менше і коротку рисочку «<-» (використовується замість «=» у R) і наберіть команду *read.csv***.** Приклад:

```
t<-read.csv("elast_lag.txt", sep="\t", dec=".", header= TRUE)
```

Далі натискайте клавіші *Ctrl+Enter.*

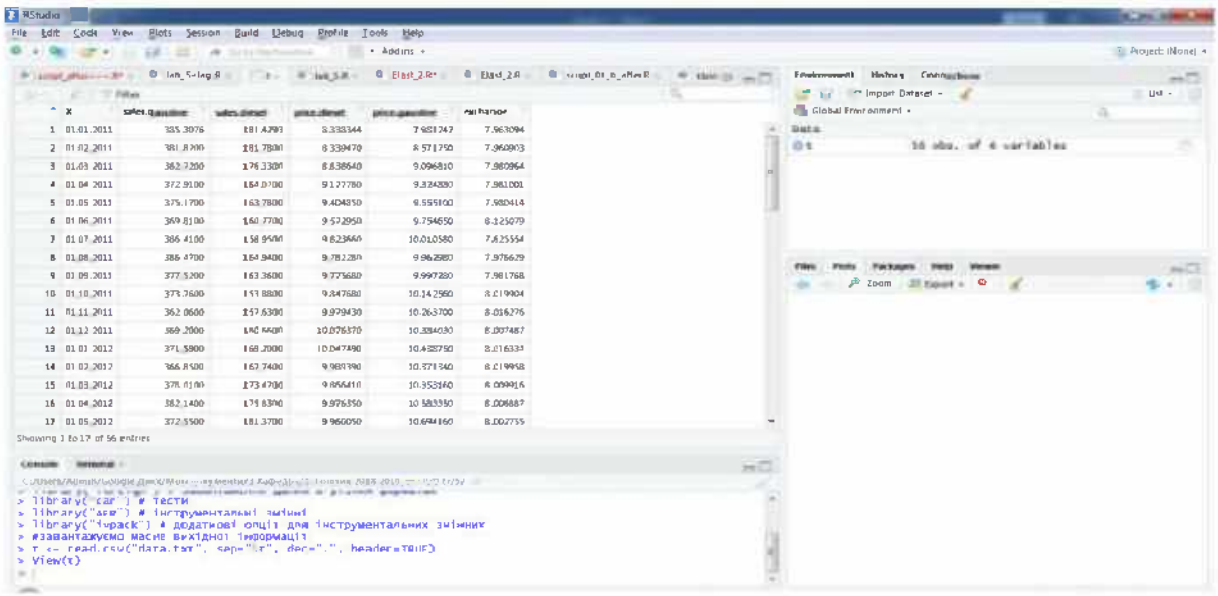


Рис. 5.5. Таблиця з даними в R

Після цього формуємо лагові змінні за допомогою функції with(). Приклад виконання:

```
t$sales.gasoline.lag <- with(t, c(NA, sales.gasoline[-length(sales.gasoline)]))
t$sales.diesel.lag <- with(t, c(NA, sales.diesel[-length(sales.diesel)]))
```

Фрейм видозміниться за допомогою функції head(t), дивимося на початок нового фрейму:

```
    X    sales.gasoline    sales.diesel    price.diesel
price.gasoline exchange sales.gasoline.lag
1  01.01.2011  385.3076  181.4293  8.338344  7.981242
7.963094 NA
2  01.02.2011  381.8200  181.7800  8.339470  8.571250
7.960903 385.3076
3  01.03.2011  382.7200  176.3300  8.838640  9.096810
7.980964 381.8200
4  01.04.2011  372.9100  164.0700  9.177780  9.324880
7.981001 382.7200
5  01.05.2011  375.1700  163.7800  9.404850  9.555100
7.980414 372.9100
6  01.06.2011  369.8100  160.7700  9.572950  9.754650
8.125079 375.1700
     sales.diesel.lag
1  NA
2  181.4293
```

Для січня 2011 р. значення лагової змінної не визначені через відсутність значень

```
3 181.7800
4 176.3300
5 164.0700
6 163.7800
```

Дані за період 01.01.2011 для лагових змінних відсутні, тому видаляємо пропущені значення за допомогою функції `na.omit(t)`. Приклад виконання:

```
t <- na.omit(t)
```

Після цього побудуйте лінійну однофакторну економетричну модель цінової еластичності попиту на бензин марки А-95 МНК, використовуючи команду `lm()`, та отримайте протокол економетричної моделі за допомогою функції `summary()`. Приклад виконання:

```
model_1<- lm(data=t, sales.gasoline~price.gasoline)
summary(model_1)
```

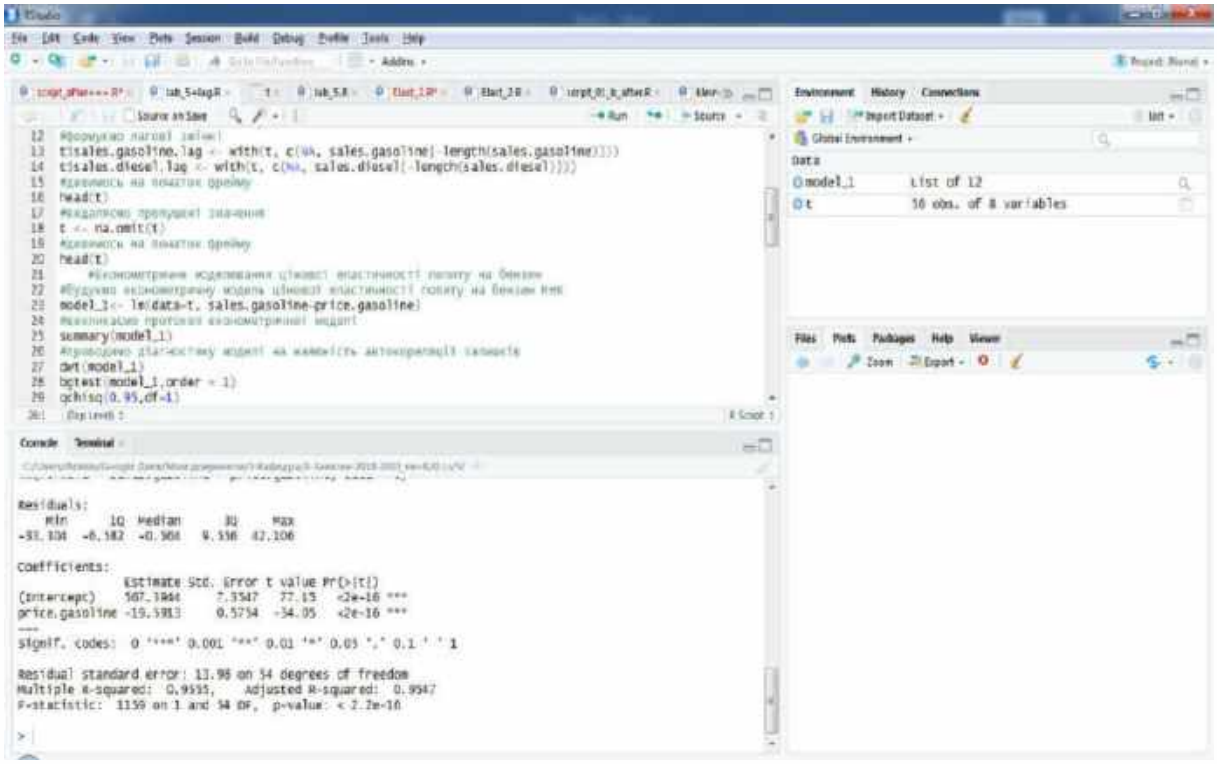


Рис. 5.6. Протокол лінійної однофакторної економетричної моделі цінової еластичності попиту на бензин марки А-95

Після цього перевірте модель на наявність автокореляції залишків першого порядку за двома тестами Дарбіна-Уотсона та Бреуша-Годфрі за допомогою функцій `dwt()` та `bgtest()` і знайдіть критичне значення χ^2 при рівні значимості 5 %. Приклад виконання:

```
dwt(model_1)
bgtest(model_1,order = 1)
qchisq(0.95,df=1)
```

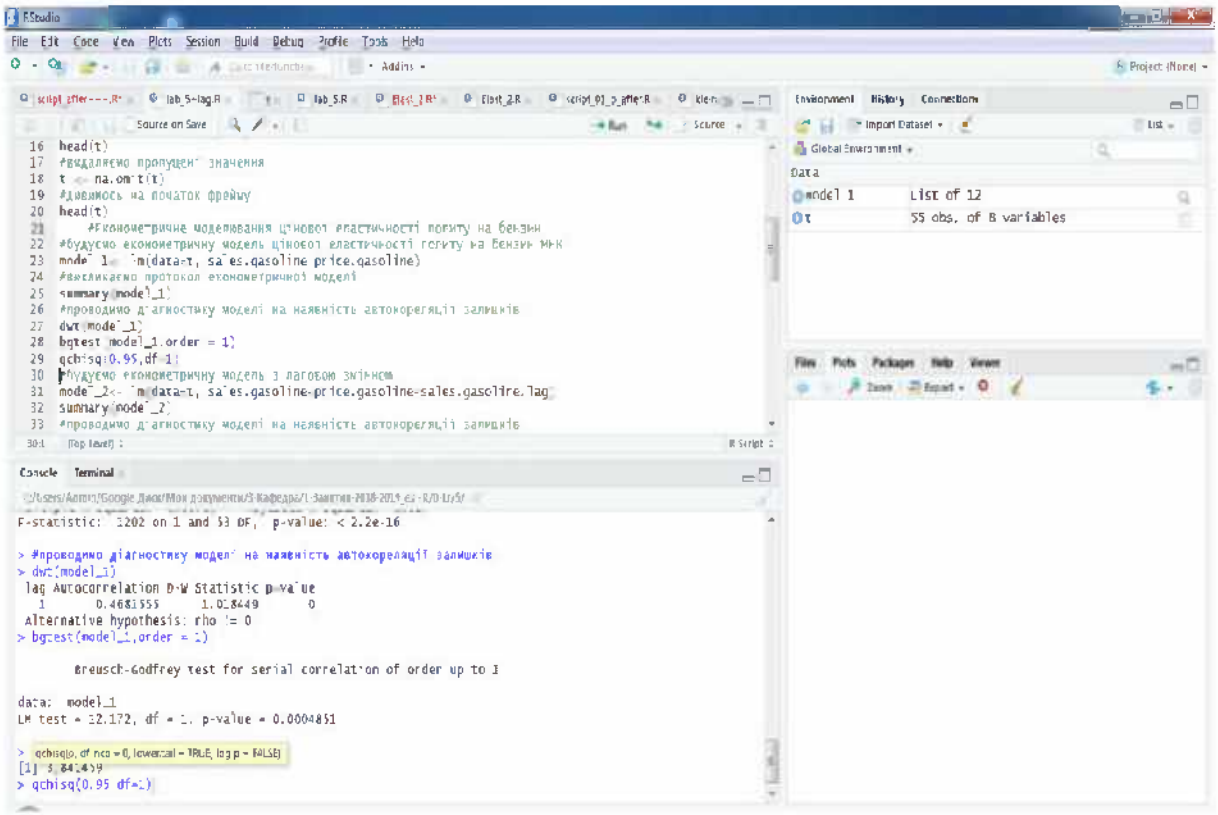
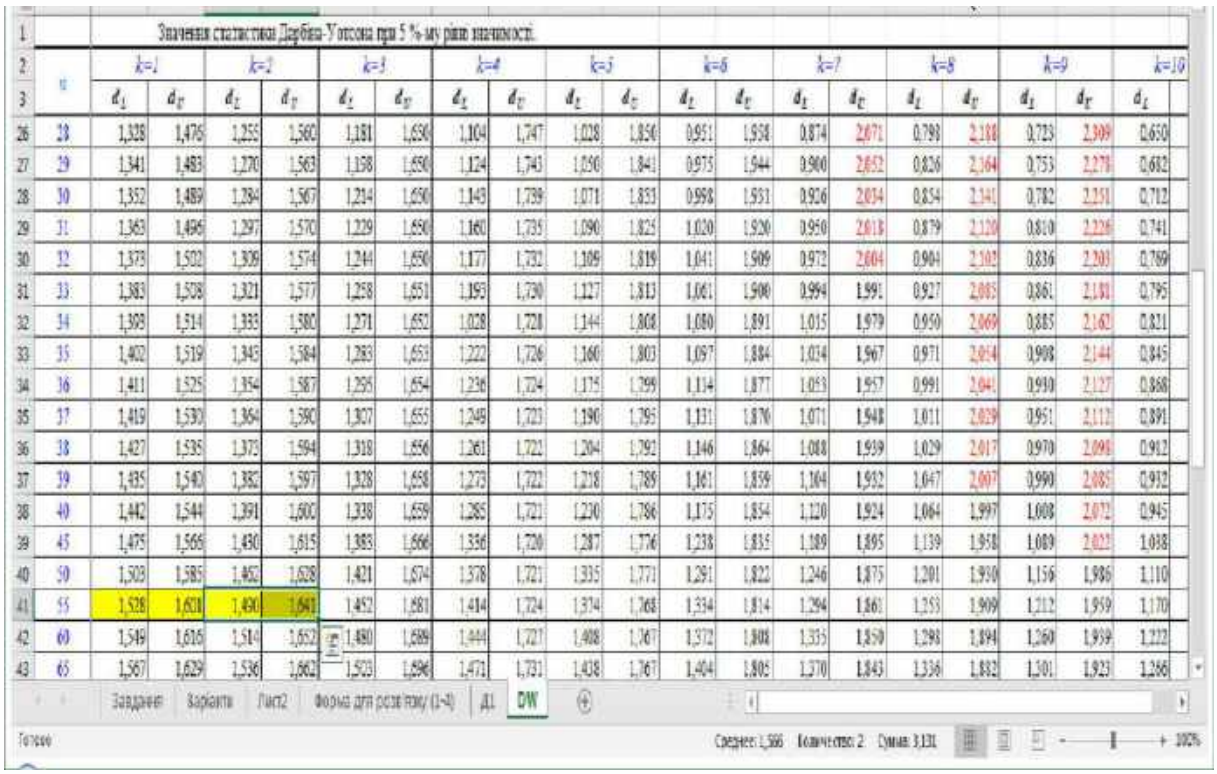


Рис. 5.7. Діагностика наявності автокореляції залишків у моделі

Проінтерпретуємо результати розрахунків. За таблицею значень статистики Дарбіна-Уотсона, за 5 %-им рівнем значимості, одним фактором ($k = 1$) та кількістю спостережень 55 ($n = 55$) знаходимо критичні значення d_L= 1,528; d_U = 1,601.



Значення статистики Дарбіна-Уотсона при 5 %-му рівні значимості.

n	k=1		k=2		k=3		k=4		k=5		k=6		k=7		k=8		k=9		k=10
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L
26	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,856	0,951	1,958	0,874	2,071	0,798	2,188	0,723	2,309	0,650
27	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,682
28	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833	0,998	1,933	0,926	2,034	0,854	2,141	0,782	2,251	0,712
29	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741
30	1,373	1,502	1,308	1,574	1,244	1,650	1,177	1,732	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769
31	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,795
32	1,393	1,514	1,333	1,580	1,271	1,652	1,028	1,728	1,144	1,808	1,080	1,891	1,015	1,979	0,950	2,069	0,885	2,162	0,821
33	1,402	1,519	1,345	1,584	1,283	1,653	1,222	1,726	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,998	2,144	0,845
34	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799	1,114	1,877	1,053	1,957	0,991	2,041	0,930	2,127	0,868
35	1,419	1,530	1,364	1,590	1,307	1,655	1,248	1,723	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891
36	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912
37	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932
38	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,945
39	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022	1,038
40	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110
41	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170
42	1,549	1,616	1,514	1,652	1,480	1,688	1,444	1,727	1,408	1,767	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222
43	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266

Завдання | Варіанти | Лист2 | Форма для розв'язку (1-4) | Д1 | DW

Готово | Среднее: 1,566 | Количество: 2 | Сумма: 3,131 | 100%

Рис. 5.8. Таблиця значень статистики Дарбіна-Уотсона за 5 %-им рівнем значимості

Оскільки значення статистики Дарбіна-Уотсона ставить $d = 0.947052$ і перебуває в інтервалі $0 < d < dL$ $(1.0184 < 1,528)$, можна зробити висновок про наявність позитивної автокореляції залишків з ймовірністю 95 %. Цей висновок підтверджується і тестом Бреуша-Годфрі, оскільки BG > χ^2 (12.172 > 3,841) в моделі наявна автокореляція залишків першого порядку з ймовірністю 95 %.

Для усунення автокореляції залишків побудуємо економетричну модель з використанням методу усунення автокореляції залишків: додання як пояснюючу змінну значення залежної змінної попереднього періоду (лагової змінної).

Приклад виконання:

```
model_2<- lm(data=t, sales.gasoline~price.-
gasoline+sales.gasoline.lag))
summary(model_2)
```

Протокол економетричної моделі подано на рис. 5.9

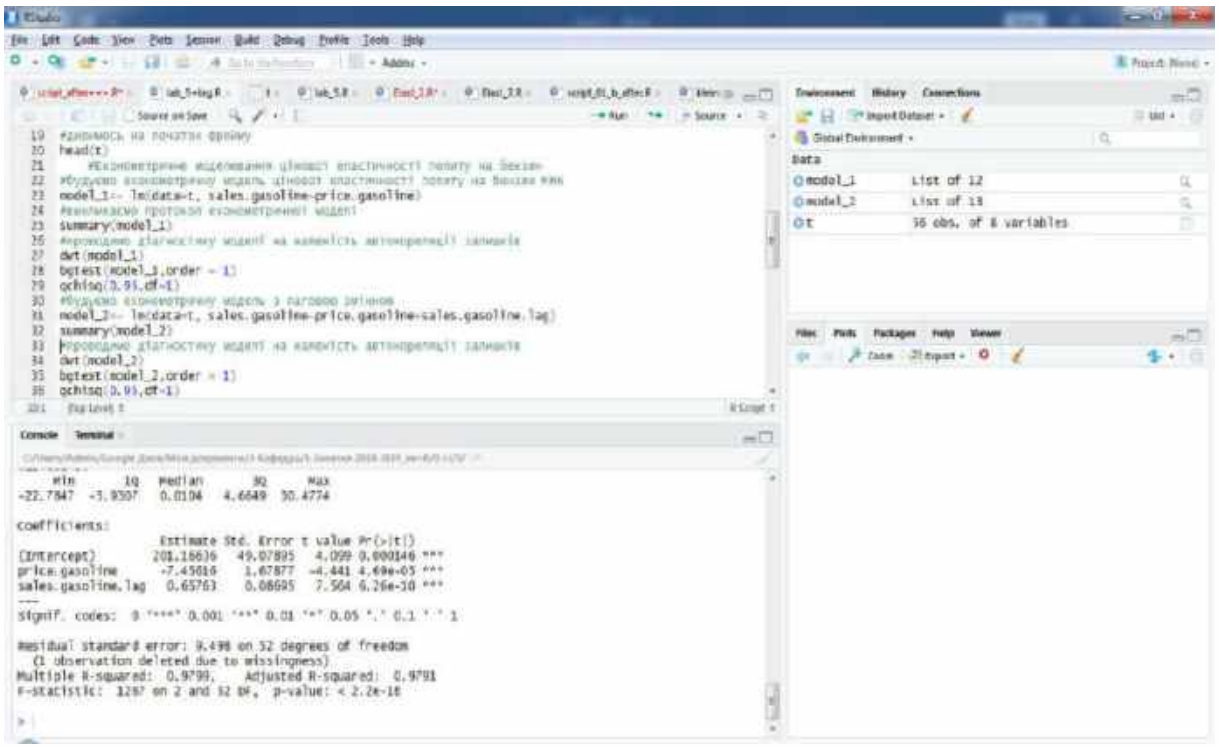


Рис. 5.9. Протокол економетричної моделі з доданою як пояснюючою змінною значення залежної змінної попереднього періоду

Проведіть перевірку нової моделі на наявність автокореляції залишків. У робочій зоні наберіть команди:

```
dwt(model_2)
bgtest(model_2,order = 1)
```

У консолі отримуємо результат:

```
lag Autocorrelation D-W Statistic p-value
 1 0.1415549 1.700208 0.164
 Alternative hypothesis: rho != 0
> bgtest(model_2,order = 1)

    Breusch-Godfrey test for serial correlation
of order up to 1

    data: model_2
    LM test = 1.7874, df = 1, p-value = 0.1812

    > qchisq(0.95,df=1)
    [1] 3.841459
```

На основі результатів діагностики моделі за тестом Дарбіна-Уотсона зроблено висновок про відсутність автокореляції залишків з ймовірністю 95 %, оскільки d потрапило в інтервал: $d_U < d < 4 - d_U$ (1,601 < 1.7 < 2,399). Тестування автокореляції першого порядку за тестом Бреуша-Годфрі також не виявило автокореляції залишків з ймовірністю 95,0 %, оскільки $BG < \chi^2_{cr}$ (1.7874 < 3,841).

Оскільки модель не містить автокореляції залишків можна провести інші необхідні процедури для перевірки можливості використання цієї моделі для формування економічних висновків. Знаходимо критичні значення t і F критеріїв:

```
> qt(0.95,df=52)
[1] 1.674689
> qf(0.95,df1=2,df2=52)
[1] 3.175141
```

Перевірка значимості оцінок параметрів за критерієм Стьюдента дозволила зробити висновок про значимість оцінки за ознаками-факторами з ймовірністю 95 %. Коефіцієнт детермінації показує, що рівняння регресії пояснює 98 % дисперсії результативної ознаки, а на частку інших факторів припадає 2 %. F-тест оцінки моделі дозволяє зробити висновок, що побудоване рівняння регресії є статистично значимим у цілому з ймовірністю 95 %.

Результати діагностики на наявність гетероскедастичності:

```
Goldfeld-Quandt test

data: model_2
GQ = 1.4952, df1 = 19, df2 = 19, p-value =
0.1942
alternative hypothesis: variance increases
from segment 1 to 2

> qf(0.95,df1=19,df2=19)
[1] 2.168252
```

Діагностика гетероскедастичності за тестом Гольдфельда-Квандта дозволила зробити висновок про її відсутність з ймовірністю 95 %, оскільки $GQ < F_{19\,19}$.

Перевірка наявності пропущених змінних у моделі за тестом Рамсея:

```
resettest(model_2)
qf(0.95,df1=2,df2=50)
```

Результати розрахунків у консолі:

```
RESET test

data: model_2
RESET = 0.8433, df1 = 2, df2 = 50, p-value =
0.4363

> qf(0.95,df1=2,df2=50)
[1] 3.18261
```

За тестом Рамсея в моделі відсутні пропущені змінні з ймовірністю 95 %, оскільки $F_{obs} < F_{cr}$.

Проведемо розрахунок коефіцієнта еластичності у моделі, введемо таку функцію:

```
elast<-coef(model_2)[2]*mean(t$price.-
gasoline)/mean(t$sales.gasoline)
```

У глобальному середовищі отримуємо, що коефіцієнт еластичності становить –0,286.

Аналіз коефіцієнта еластичності в моделі дозволяє зробити висновок про те, що у разі зростання ціни бензину на 1 % обсяг його споживання знизиться на 0,29 %. Порівнюємо дві економетричні моделі за допомогою функції `mtable ()`. Приклад виконання:

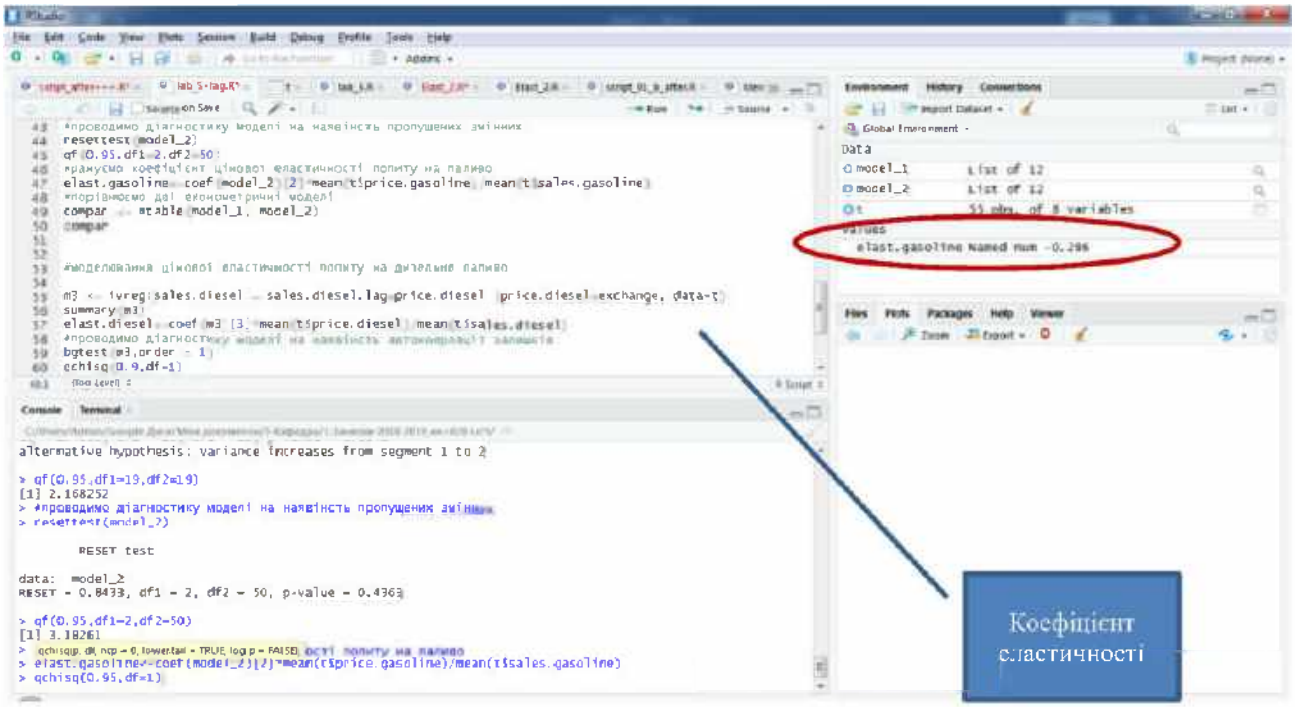


Рис. 5.10. Розрахунок коефіцієнта цінової еластичності безнину

```
compar <- mtable(model_1, model_2)
compar
```

У консолі отримуємо результат:

```
Calls:
model_1: lm(formula = sales.gasoline ~
price.gasoline, data = t)
model_2: lm(formula = sales.gasoline ~
price.gasoline + sales.gasoline.lag,
data = t)

=============================================
model_1 model_2
---------------------------------------------
(Intercept) 570.365*** 201.166***
(7.337) (49.079)
price.gasoline -19.792*** -7.45
(0.571) (1.679)
sales.gasoline.lag 0.658***
(0.087)
---------------------------------------------
```

середньоквадратичні помилки оцінок параметрів є заниженими

```
R-squared 0.958 0.980
N 55 55
================================================
Significance: *** = p < 0.001;
** = p < 0.01; * = p < 0.05
```

Аналізуючи дані, подані вище, можна побачити різницю між оцінками параметрів та стандартними помилками в обох моделях. У першій моделі середньоквадратичні помилки оцінок параметрів значно занижені. Крім того, у другій моделі коефіцієнт детермінації значно збільшився.

Модель цінової еластичності попиту на дизпаливо

При оцінці параметрів МНК та діагностиці моделі на гетероскедастичність з'ясовано її наявність. У зв'язку з цим було вирішено провести економетричне моделювання, використовуючи метод інструментальних змінних. На першому кроці була побудована модель залежності ціни дизельного палива від курсу гривні до долара США. На другому кроці побудована модель, у якій замість ціни дизельного палива було використано її розрахункове значення, отримане на першому кроці. У результаті отримали оцінки інструментальних змінних і модель набула вигляду:

$$Cons_{diesel_t} = \beta_0^{OLS} + \beta_1^{OLS} \cdot Pr\,i\,ce_{diesel} + \beta_2^{OLS} \cdot Cons_{diesel_{t-1}},$$

де $Cons_{diesel_t}$ – обсяг споживання дизельного палива в поточному періоді, скоригований на сезонність;

$\mathrm{Pr\,i\,ce_{diesel}}$ – ціна дизельного палива в поточному періоді, скоригована на сезонність та інфляційну складову;

$\mathrm{Cons_{diesel_{t-1}}}$ – обсяг споживання дизельного палива в попередньому періоді, скоригований на сезонність;

β_0^{OLS}, β_1^{OLS}, β_2^{OLS} – інструментальні змінні.

Економетричне моделювання, використовуючи метод інструментальних змінних, проводимо за допомогою функції `ivreg ()`.

Приклад виконання:

```
    m3 <- ivreg(sales.diesel ~ sales.diesel.-
t.1+price.diesel          |price.diesel+exchange,
data=t)
    summary(m3)
```

Результати моделювання подані на рис. 5.11.

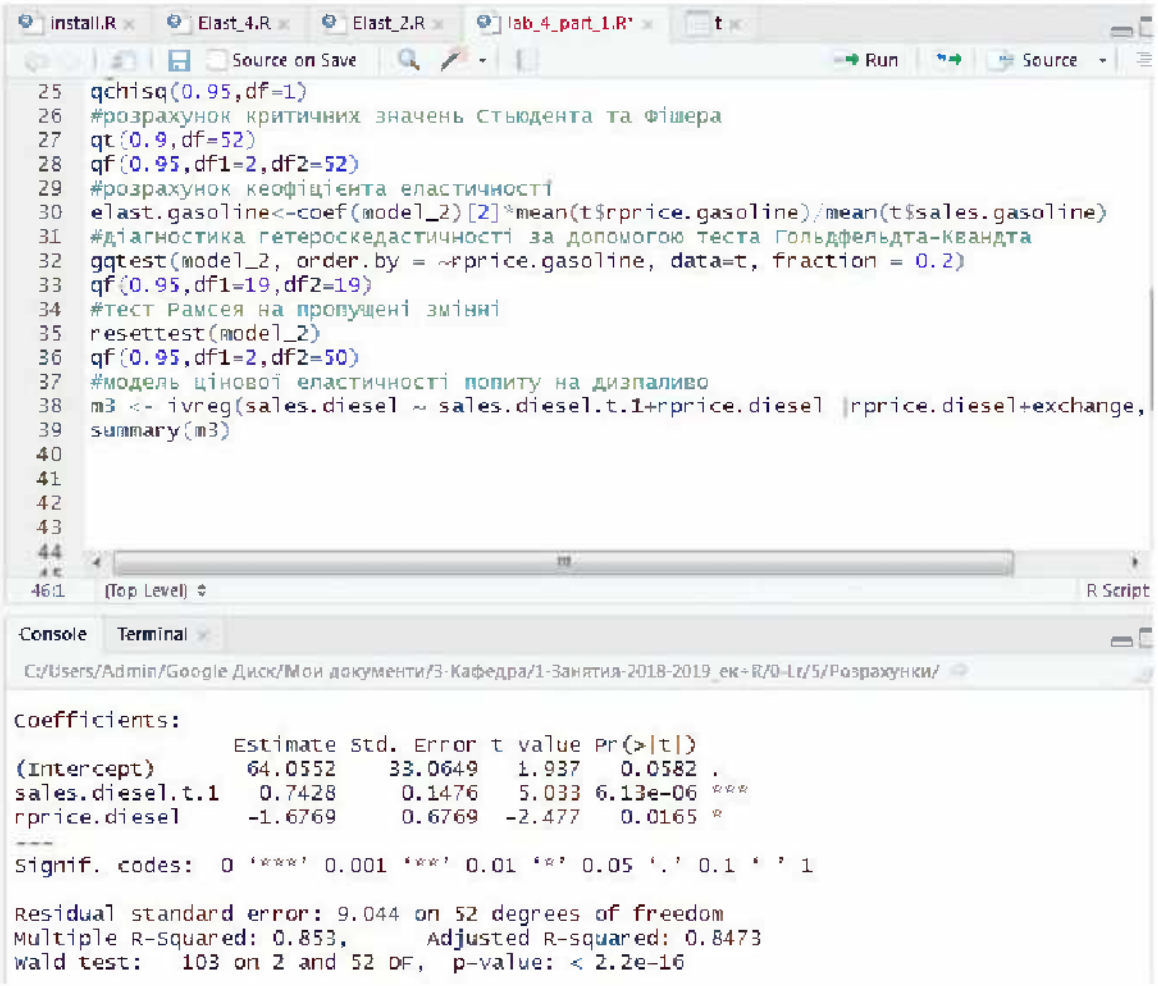


Рис. 5.11. Протокол економетричної моделі з використанням методу інструментальних змінних

Перевірка значимості оцінок параметрів за критерієм Стьюдента дозволила зробити висновок про значимість оцінки за ознаками-факторами з ймовірністю 90 %. Коефіцієнт детермінації показує, що рівняння регресії пояснює 85,3 % дисперсії результативної ознаки, а на частку інших факторів припадає 14,7 %. Тестування автокореляції першого порядку за тестом Бреуша-Годфрі не виявило автокореляції з ймовірністю 99,0 %, оскільки $BG < \chi^2_{\omega}$. Діагностика гетероскедастичності за тестом Гольдфельда Квандта дозволила зробити висновок про її

відсутність з ймовірністю 99 %, оскільки $GQ < F_{19\,19}$. За тестом Рамсея в моделі відсутні пропущені змінні з ймовірністю 95 %, оскільки $F_{obs} < F_{cr}$. Розрахунок коефіцієнта еластичності проводимо з використанням такої формули:

```
    elast.gasoline<-
coef(model_2)[2]*mean(t$price.gasoline)/mean(t$
sales.gasoline)
```

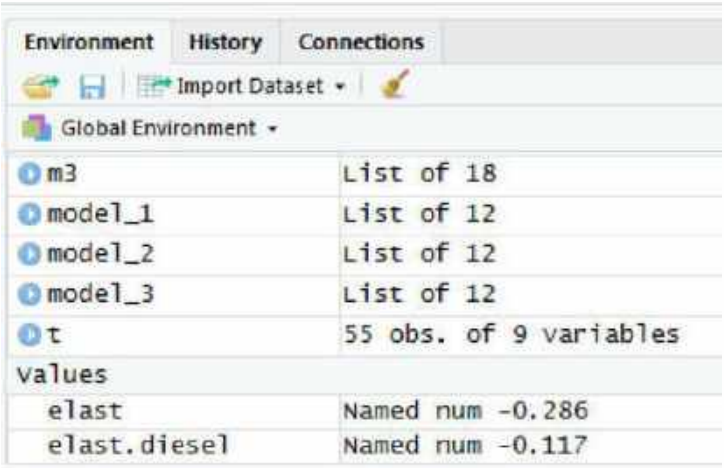


Рис. 5.12. Розрахунок коефіцієнта цінової еластичності дизельного палива

Аналіз коефіцієнта еластичності в моделі дозволяє зробити висновок про те, що під час зростання ціни дизельного палива, залежної від курсу гривні до долара США, на 1 % обсяг його споживання знизиться на 0,12 %.

Модель цінової еластичності пропозиції бензину

Перевірка значимості оцінок параметрів за критерієм Стьюдента дозволила зробити висновок про значимість оцінки при ознаках-факторах з ймовірністю 95 %. Аналіз коефіцієнта еластичності в моделі дозволяє зробити висновок про те, що при зростанні ціни бензину на 1 % обсяг його виробництва та імпорту знизиться на 0,56 %.

Коефіцієнт детермінації показує, що рівняння регресії пояснює 77,56 % дисперсії результативної ознаки, а на частку інших факторів припадає 22,44 %. *F*-тест оцінки моделі дозволяє

зробити висновок, що побудоване рівняння регресії є статистично значимим у цілому з ймовірністю 95 %.

На основі результатів діагностики моделі за тестом Дарбіна-Уотсона зроблено висновок про відсутність автокореляції залишків першого порядку з ймовірністю 99,95 %, оскільки d потрапило в інтервал: $d_U < d < 4 - d_U$. Тестування автокореляції другого порядку за тестом Бреуша-Годфрі також не виявило автокореляції з ймовірністю 99,0 %, оскільки $BG < \chi^2_{cr}$. Діагностика гетероскедастичності за тестом Гольдфельда-Квандта дозволила зробити висновок про її відсутність з ймовірністю 99,9 %, оскільки $GQ < F_{19\,19}$. За тестом Рамсея в моделі відсутні пропущені змінні з ймовірністю 95,99 %, оскільки $F_{obs} < F_{cr}$.

Модель цінової еластичності пропозиції дизпалива

Перевірка значимості оцінок параметрів за критерієм Стьюдента дозволила зробити висновок про значимість оцінки при ознаках-факторах з ймовірністю 95 %. Аналіз коефіцієнта еластичності в моделі дозволяє зробити висновок про те, що при зростанні ціни дизпалива на 1 % обсяг його пропозиції знизиться на 0,14 %.

Коефіцієнт детермінації показує, що рівняння регресії пояснює 53,92 % дисперсії результативної ознаки, а на частку інших факторів припадає 46,08 %. F-тест оцінки моделі дозволяє зробити висновок, що побудоване рівняння регресії є статистично значимим у цілому з ймовірністю 95 %.

На основі результатів діагностики моделі за тестом Дарбіна-Уотсона зроблено висновок про відсутність автокореляції залишків першого порядку з ймовірністю 99,95 %, оскільки d потрапило в інтервал: $d_U < d < 4 - d_U$. Тестування автокореляції другого порядку за тестом Бреуша-Годфрі також не виявило автокореляції з ймовірністю 99,0 %, оскільки $BG < \chi^2_{cr}$. Діагностика гетероскедастичності за тестом Гольдфельда-Квандта дозволила зробити висновок про її відсутність з ймовірністю 99,9 %, оскільки $GQ < F_{19\,19}$. За тестом Рамсея в моделі відсутні пропущені змінні з ймовірністю 95,99 %, оскільки $F_{obs} < F_{cr}$.

Додаток А до лабораторної роботи 5

ВИХІДНІ ДАНІ ДО ВИКОНАННЯ ЛАБОРАТОРНОЇ РОБОТИ 5

	Варіант 1					Варіант 2				
	Україна					Вінницька				
	sales gasoline	**sales diesel**	**price gasoline**	**prise diesel**	**exchange**	**sales gasoline**	**sales diesel**	**price gasoline**	**prise diesel**	**exchange**
01.01.2017	70648513.51	95241647.06	23.97	21.50	27.15	1742027.03	3612000.00	23.34	21.15	27.15
01.02.2017	70617588.93	95250142.78	24.53	22.12	27.03	1742609.57	3610805.75	23.87	21.71	27.03
01.03.2017	71465000.00	98107647.06	24.58	22.11	27.00	1597972.97	3632470.59	23.88	21.22	27.00
01.04.2017	87319729.73	113258941.18	24.86	22.35	26.86	2150675.68	4529058.82	24.17	21.43	26.86
01.05.2017	85532702.70	109411294.12	25.08	22.67	26.42	2010810.81	4073647.06	24.49	21.85	26.42
01.06.2017	89339729.73	106229882.35	24.85	22.46	26.11	2150945.95	3858705.88	24.27	21.75	26.11
01.07.2017	92651351.35	113126235.29	24.62	22.07	25.97	2209054.05	4213764.71	24.07	21.31	25.97
01.08.2017	102466081.08	123855882.35	24.65	21.91	25.64	2334729.73	4507647.06	24.38	21.14	25.64
01.09.2017	91413378.38	112228352.94	25.16	22.28	26.11	2206216.22	3952588.24	25.02	21.86	26.11
01.10.2017	86105135.14	123685647.06	26.54	23.97	26.65	2009459.46	4425411.76	26.51	24.22	26.65
01.11.2017	79696081.08	111064705.88	27.47	25.14	26.71	1781891.89	3771294.12	27.40	25.37	26.71
01.12.2017	75184729.73	98545882.35	28.14	25.80	27.52	1580000.00	3038117.65	27.85	25.93	27.52
01.01.2018	65162027.03	86606941.18	29.43	27.42	28.43	1478783.78	2878705.88	29.44	26.75	28.43
01.02.2018	57032027.03	76927764.71	30.05	27.78	27.17	1208378.38	2574117.65	29.70	27.60	27.17
01.03.2018	60161351.35	83757411.76	29.33	26.93	26.34	1470540.54	3367764.71	28.74	26.78	26.34
01.04.2018	84532297.30	108865411.76	29.15	26.95	26.15	2207702.70	4683058.82	28.59	26.37	26.15
01.05.2018	76429864.86	109395176.47	29.34	27.28	26.18	1947702.70	4530117.65	28.81	26.77	26.18
01.06.2018	78916351.35	110454235.29	29.40	27.15	26.20	1967567.57	4411411.76	28.78	26.88	26.20
01.07.2018	83565135.14	118191882.35	29.26	27.04	26.40	2072567.57	4774000.00	28.65	26.72	26.40
01.08.2018	88052972.97	123448470.59	29.54	27.44	27.48	2256891.89	5199647.06	29.11	27.01	27.48
01.09.2018	73945405.41	110833529.41	31.17	29.38	28.19	1991081.08	5054705.88	30.94	28.67	28.19
01.10.2018	76698108.11	124376117.65	32.92	31.96	28.13	1907702.70	5697529.41	32.65	31.56	28.13

01.11.2018	74775540.54	111731764.71	32.28	31.85	27.93	1833918.92	5050823.53	31.79	31.69	27.93
01.12.2018	68408918.92	95910470.59	30.43	29.80	27.79	1596216.22	3906588.24	29.79	29.91	27.79
01.01.2019	63408378.38	85364470.59	28.69	28.61	27.88	1505405.41	3383058.82	28.51	28.58	27.88
01.02.2019	60688378.38	80852588.24	28.38	28.24	27.16	1378918.92	3077058.82	28.03	28.35	27.16
01.03.2019	69644729.73	88309647.06	28.33	28.18	26.86	1501891.89	3205294.12	28.03	28.33	26.86
01.04.2019	74597567.57	92807882.35	28.85	28.33	26.81	1615405.41	3446588.24	28.53	28.67	26.81
01.05.2019	77578918.92	93166823.53	30.03	29.57	26.38	1793918.92	3391411.76	30.03	29.94	26.38
01.06.2019	85029864.86	104628470.59	29.99	29.34	26.50	1920000.00	3640588.24	29.83	29.78	26.50
01.07.2019	86678378.38	113036941.18	29.08	28.39	25.75	1909864.86	3977882.35	28.65	28.45	25.75
01.08.2019	91350675.68	117220705.88	28.77	28.00	25.25	2157297.30	4337882.35	28.32	28.17	25.25
01.09.2019	82192027.03	114221411.76	28.70	27.95	24.77	1918918.92	4394941.18	28.22	28.10	24.77
01.10.2019	82867837.84	118962470.59	28.68	27.99	24.81	1877702.70	5075764.71	28.21	28.12	24.81
01.11.2019	80615945.95	109808823.53	28.62	27.87	24.37	1686081.08	3826941.18	28.18	27.99	24.37
01.12.2019	78140540.54	104240588.24	28.09	27.45	23.61	1571891.89	3605058.82	27.78	27.39	23.61
01.01.2020	72864054.05	107507058.82	28.09	27.45	24.12	1650000.00	4346000.00	27.78	27.39	24.12
01.02.2020	76712432.43	111926705.88	26.89	26.31	24.60	1525675.68	4246352.94	26.52	26.37	24.60
01.03.2020	76615135.14	110627647.06	26.51	25.81	26.41	1654054.05	4385411.76	26.12	25.78	26.41
01.04.2020	65236486.49	87638000.00	26.22	25.49	27.22	1423918.92	3687529.41	25.72	25.39	27.22
01.05.2020	82161891.89	100570705.88	23.91	23.31	26.81	1718243.24	4376588.24	23.25	23.16	26.81
01.06.2020	98023513.51	116153294.12	21.88	21.23	26.71	2112297.30	4498705.88	21.22	21.27	26.71
01.07.2020	106139864.86	134240588.24	21.91	21.24	27.31	2362702.70	5357647.06	21.22	21.22	27.31
01.08.2020	111049324.32	137177882.35	22.62	21.85	27.52	2441756.76	5561058.82	21.81	21.91	27.52
01.09.2020	101652837.84	134267294.12	23.22	22.45	27.98	1995945.95	5408352.94	22.04	22.20	27.98
01.10.2020	101617027.03	140613882.35	23.36	22.60	28.32	2067567.57	6838470.59	22.12	22.31	28.32
01.11.2020	89856486.49	125787529.41	23.54	22.65	28.31	1958783.78	6113176.47	22.63	22.51	28.31
01.12.2020	90389729.73	126021529.41	24.27	23.33	28.17	1970405.41	5267058.82	23.60	23.40	28.17
01.01.2021	87768918.92	110732941.18	25.66	24.84	28.22	2049189.19	4440235.29	24.62	24.62	28.22

	Варіант 3					Варіант 4				
	Волинська					Дніпропетровська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	999594.5946	4833294.118	23.62	21.1	27.15	6303783.8	4120941.18	23.62	21.07	27.15
01.02.2017	999262.6414	4832428.06	24.21	21.73	27.03	6306306.3	4121947.47	24.1	21.66	27.03
01.03.2017	952297.2973	4919411.765	24.13	21.71	27.00	6337837.8	4736117.65	24.18	21.67	27.00
01.04.2017	1201621.622	5348823.529	24.56	21.98	26.86	7639189.2	4882000	24.79	22.19	26.86
01.05.2017	1163918.919	5139411.765	24.85	22.43	26.42	7479729.7	4707294.12	24.96	22.49	26.42
01.06.2017	1241621.622	5116117.647	24.67	22.12	26.11	7760270.3	4498117.65	24.78	22.31	26.11
01.07.2017	1317837.838	4940941.176	24.4	21.62	25.97	7869324.3	4588000	24.47	21.81	25.97
01.08.2017	1527837.838	6104823.529	24.38	21.42	25.64	8548513.5	4920117.65	24.51	21.58	25.64
01.09.2017	1248108.108	5598352.941	24.87	21.75	26.11	7917027	5035529.41	25.09	21.88	26.11
01.10.2017	1162567.568	6451647.059	26.35	23.69	26.65	7645540.5	5951176.47	26.44	23.57	26.65
01.11.2017	1052972.973	5510588.235	27.23	25.05	26.71	7112027	5416941.18	27.34	24.96	26.71
01.12.2017	1000270.27	5114823.529	27.82	25.45	27.52	6728378.4	4580000	28.07	25.54	27.52
01.01.2018	888513.5135	4357882.353	29.18	26.69	28.43	5749459.5	3981058.82	29.19	26.4	28.43
01.02.2018	739459.4595	3727764.706	29.77	27.45	27.17	5327162.2	3749058.82	29.76	27.56	27.17
01.03.2018	775135.1351	3958823.529	29.14	26.57	26.34	5346351.4	4002352.94	29.09	28.21	26.34
01.04.2018	1049459.459	4854000	28.79	26.57	26.15	6939054.1	4691647.06	28.95	27.72	26.15
01.05.2018	985810.8108	5240470.588	28.96	26.65	26.18	6968513.5	5070470.59	29.05	27.82	26.18
01.06.2018	1001216.216	5256941.176	29.08	26.78	26.20	7292973	5165294.12	29.21	27.65	26.20
01.07.2018	1076216.216	5478588.235	28.99	26.67	26.40	7614864.9	5588235.29	29.15	27.46	26.40
01.08.2018	1168783.784	5853529.412	29.34	26.82	27.48	7850540.5	5581176.47	29.38	27.15	27.48
01.09.2018	984459.4595	5327529.412	31.03	28.61	28.19	7296216.2	5568235.29	31.12	27.34	28.19
01.10.2018	934594.5946	5812117.647	32.85	31.15	28.13	6965405.4	5901058.82	32.95	28.38	28.13
01.11.2018	882972.973	4962352. 941	32.09	31.86	27.93	6878648.6	5428588.24	32.24	30.37	27.93
01.12.2018	798783.7838	4292705.882	30.38	30.14	27.79	6321756.8	4608941.18	30.43	31.37	27.79
01.01.2019	1073513.514	4698470.588	28.34	28.16	27.88	5484864.9	3833764.71	28.72	28.44	27.88

01.02.2019	722162.1622	3670588.235	28.18	27.84	27.16	5544594.6	3878705.88	28.5	28.22	27.16
01.03.2019	780675.6757	3916823.529	28.13	27.8	26.86	6243783.8	4156823.53	28.48	28.21	26.86
01.04.2019	878108.1081	4019529.412	28.77	28.04	26.81	6549459.5	4275764.71	28.95	28.35	26.81
01.05.2019	932027.027	3992000	29.84	29.26	26.38	6615135.1	4237411.76	30.51	29.87	26.38
01.06.2019	1040540.541	4536941.176	29.69	28.65	26.50	6972837.8	4536705.88	30.44	29.74	26.50
01.07.2019	1097567.568	4845529.412	28.85	27.71	25.75	7313918.9	5076235.29	29.24	28.51	25.75
01.08.2019	1176486.486	4950941.176	28.49	27.35	25.25	7352027	5055411.76	28.83	28.02	25.25
01.09.2019	1005270.27	5477647.059	28.3	27.36	24.77	6794189.2	5233882.35	28.79	27.93	24.77
01.10.2019	1016486.486	5461764.706	28.35	27.52	24.81	7134054.1	5517647.06	28.75	27.93	24.81
01.11.2019	992837.8378	4982823.529	28.31	27.31	24.37	6967297.3	5229647.06	28.72	27.92	24.37
01.12.2019	1015270.27	5055411.765	27.7	26.76	23.61	6624324.3	4741647.06	28.28	27.72	23.61
01.01.2020	988108.1081	5907411.765	27.7	26.76	24.12	5549324.3	4754823.53	28.28	27.72	24.12
01.02.2020	987972.973	6450470.588	26.1	25.27	24.60	6675270.3	6302941.18	26.85	26.29	24.60
01.03.2020	969729.7297	6200941.176	25.86	24.83	26.41	6305540.5	5986470.59	26.55	25.98	26.41
01.04.2020	882702.7027	4840352.941	25.63	24.55	27.22	6295270.3	5402000	26.35	25.87	27.22
01.05.2020	1061891.892	5375529.412	23.05	22.36	26.81	8238513.5	6236117.65	24.06	23.61	26.81
01.06.2020	1205945.946	5643647.059	20.85	19.52	26.71	9637702.7	7288705.88	21.9	21.19	26.71
01.07.2020	1387972.973	6418235.294	21.03	19.69	27.31	10305405	8573764.71	22.01	21.22	27.31
01.08.2020	1473648.649	6351764.706	21.9	20.61	27.52	10267162	8388941.18	22.4	21.42	27.52
01.09.2020	1440810.811	6534823.529	22.65	21.41	27.98	9960810.8	8908941.18	22.76	21.76	27.98
01.10.2020	1374459.459	7091764.706	22.67	21.45	28.32	10178108	9182941.18	22.81	21.87	28.32
01.11.2020	1151081.081	6077411.765	22.88	21.3	28.31	9711621.6	8598352.94	23.25	22.26	28.31
01.12.2020	1156891.892	6256705.882	27.7	26.76	28.17	10452703	9593058.82	28.28	27.72	28.17
01.01.2021	1163918.919	5877882.353	25.1	23.71	28.22	8516486.5	7245411.76	25.02	23.96	28.22

	Варіант 5					Варіант 6				
	Донецька					Житомирська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	1423513.514	1132470.588	24.37	21.76	27.15	1540000	3566588.24	23.80	21.31	27.15
01.02.2017	1423477.859	1132760.683	25.04	22.00	27.03	1539607.5	3566349.95	24.34	21.77	27.03
01.03.2017	1413513.514	1225647.059	25.52	22.52	27.00	1382567.6	3930117.65	24.30	21.82	27.00
01.04.2017	1725675.676	1342470.588	25.57	22.67	26.86	2055000	4724117.65	24.73	22.23	26.86
01.05.2017	1589189.189	1246000	25.67	22.74	26.42	1852432.4	4338470.59	24.90	22.40	26.42
01.06.2017	1894054.054	1437411.765	25.38	22.80	26.11	1956621.6	4328588.24	24.67	22.18	26.11
01.07.2017	2002972.973	1503058.824	25.28	22.68	25.97	2125135.1	4612823.53	24.50	21.78	25.97
01.08.2017	2185675.676	1713176.471	25.18	22.71	25.64	2348513.5	5035294.12	24.48	21.59	25.64
01.09.2017	1907027.027	1460117.647	25.63	23.12	26.11	2015405.4	4646941.18	24.91	21.89	26.11
01.10.2017	1759189.189	1505647.059	26.86	24.43	26.65	1834594.6	5226000	26.34	23.65	26.65
01.11.2017	1639324.324	1432705.882	27.93	25.55	26.71	1601756.8	4660588.24	27.28	24.80	26.71
01.12.2017	1671081.081	1457176.471	28.66	26.40	27.52	1406216.2	4147411.76	27.97	25.50	27.52
01.01.2018	1399864.865	1252823.529	28.66	27.58	28.43	1343243.2	3488941.18	29.77	26.79	28.43
01.02.2018	1276081.081	1274470.588	29.91	28.12	27.17	1020945.9	2985411.76	30.32	27.84	27.17
01.03.2018	1391891.892	1377058.824	30.38	27.45	26.34	1142567.6	3367411.76	29.59	27.10	26.34
01.04.2018	1673783.784	1481411.765	29.84	27.12	26.15	1746081.1	4137882.35	29.46	26.89	26.15
01.05.2018	1688918.919	1550941.176	29.77	27.43	26.18	1546756.8	4254470.59	29.76	26.97	26.18
01.06.2018	1855405.405	1604352.941	29.52	27.61	26.20	1602702.7	4276352.94	29.68	27.17	26.20
01.07.2018	1946891.892	1750117.647	29.18	27.34	26.40	1746216.2	4531294.12	29.50	26.90	26.40
01.08.2018	1998378.378	1821411.765	29.18	27.65	27.48	1910270.3	4483058.82	29.87	27.06	27.48
01.09.2018	1591081.081	1579529.412	29.30	29.35	28.19	1696351.4	3918941.18	31.51	28.56	28.19
01.10.2018	1529864.865	1694588.235	30.47	31.84	28.13	1655405.4	4439411.76	33.27	31.44	28.13
01.11.2018	1540945.946	1648470.588	32.65	32.24	27.93	1477973	3831647.06	32.46	31.57	27.93
01.12.2018	1467297.297	1381529.412	32.76	30.61	27.79	1296621.6	3185882.35	30.67	29.92	27.79
01.01.2019	1317837.838	1273764.706	31.20	30.69	27.88	1314864.9	3499529.41	28.94	28.78	27.88

01.02.2019	1294324.324	1278117.647	29.09	29.02	27.16	1144459.5	3233294.12	28.87	28.58	27.16
01.03.2019	1486351.351	1555176.471	28.31	28.37	26.86	1316891.9	3590235.29	28.72	28.50	26.86
01.04.2019	1594054.054	1591764.706	28.15	28.13	26.81	1448918.9	3829529.41	29.35	28.70	26.81
01.05.2019	1659054.054	1621294.118	28.40	28.15	26.38	1593108.1	3881764.71	30.48	29.88	26.38
01.06.2019	1798108.108	1680470.588	29.44	29.33	26.50	1673108.1	4001294.12	30.50	29.71	26.50
01.07.2019	1920135.135	1974823.529	29.32	28.95	25.75	1715945.9	4570705.88	29.56	28.78	25.75
01.08.2019	2004729.73	1948470.588	28.66	28.27	25.25	1842567.6	4621647.06	29.18	28.35	25.25
01.09.2019	1771486.486	1623529.412	28.33	27.99	24.77	1583918.9	4419529.41	29.03	28.28	24.77
01.10.2019	1834729.73	1576235.294	28.35	28.00	24.81	1546891.9	4548823.53	29.01	28.28	24.81
01.11.2019	1801891.892	1509882.353	28.47	28.05	24.37	1412027	4053647.06	28.99	28.21	24.37
01.12.2019	1734324.324	1436000	28.38	27.87	23.61	1415810.8	3885764.71	28.59	27.61	23.61
01.01.2020	1496756.757	1521647.059	27.99	27.39	24.12	1456891.9	4035058.82	28.59	27.61	24.12
01.02.2020	1684459.459	1705647.059	27.11	26.47	24.60	1366756.8	4488117.65	27.29	27.03	24.60
01.03.2020	1767432.432	1740470.588	26.53	25.94	26.41	1402432.4	4270000	26.93	26.49	26.41
01.04.2020	1657972.973	1523058.824	26.26	25.82	27.22	1273378.4	3419058.82	26.87	26.44	27.22
01.05.2020	2026216.216	1839882.353	24.56	23.83	26.81	1488513.5	4052000	24.59	24.46	26.81
01.06.2020	2358918.919	2012117.647	22.17	21.56	26.71	1737432.4	4892352.94	22.74	23.26	26.71
01.07.2020	2650405.405	2338470.588	22.15	21.43	27.31	1824189.2	5852117.65	22.77	22.57	27.31
01.08.2020	2768108.108	2396235.294	22.55	21.65	27.52	2088108.1	5922823.53	23.65	23.48	27.52
01.09.2020	2469189.189	2260588.235	23.06	22.54	27.98	1828918.9	5690941.18	24.19	24.08	27.98
01.10.2020	2485945.946	2126705.882	23.45	22.65	28.32	1857837.8	5822117.65	24.44	24.32	28.32
01.11.2020	2431486.486	1966352.941	23.62	22.78	28.31	1447702.7	5170941.18	24.64	24.38	28.31
01.12.2020	2618783.784	2071529.412	28.38	27.87	28.17	1536486.5	5152941.18	28.59	27.61	28.17
01.01. 2021	2088243.243	1633764.706	25.60	24.80	28.22	1829594.6	4556941.18	26.48	26.30	28.22

	Варіант 7					Варіант 8				
	Закарпатська					Запорізька				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	3870540.541	7759764.706	23.94	21.49	27.15	3060540.5	3163058.82	23.76	21.34	27.15
01.02.2017	3872174.061	7761836.442	24.43	21.9	27.03	3060135.1	3161942.21	24.32	21.89	27.03
01.03.2017	3694054.054	7001176.471	24.45	22.03	27.00	3060135.1	3405411.76	24.38	21.94	27.00
01.04.2017	4277567.568	8714470.588	24.69	22.21	26.86	3832837.8	3824941.18	24.91	22.32	26.86
01.05.2017	4331351.351	8496352.941	24.99	22.52	26.42	4053783.8	3872588.24	25.09	22.81	26.42
01.06.2017	4254189.189	8009647.059	24.67	22.55	26.11	4442973	4159529.41	24.78	22.58	26.11
01.07.2017	4460540.541	8346823.529	24.39	22.25	25.97	4721081.1	4546470.59	24.51	22.21	25.97
01.08.2017	4542297.297	9024117.647	24.39	22	25.64	5499459.5	4819411.76	24.57	21.87	25.64
01.09.2017	4480810.811	8917411.765	24.89	22.35	26.11	4229324.3	4213529.41	25.03	22.16	26.11
01.10.2017	4219324.324	8636705.882	26.1	23.61	26.65	3860135.1	4425058.82	26.49	23.8	26.65
01.11.2017	3969729.73	7749647.059	27.14	24.79	26.71	3567432.4	3856588.24	27.3	24.93	26.71
01.12.2017	3904459.459	7453411.765	27.83	25.48	27.52	3568918.9	3599647.06	27.95	25.68	27.52
01.01.2018	3781216.216	7409882.353	28.97	26.55	28.43	2880675.7	3031882.35	29.05	27.81	28.43
01.02.2018	3079729.73	6644000	29.71	27.59	27.17	2673378.4	2949764.71	29.8	28.34	27.17
01.03.2018	3420675.676	7449294.118	29.03	26.93	26.34	3054324.3	3283411.76	29.14	27.56	26.34
01.04.2018	4403378.378	9461529.412	28.76	26.76	26.15	3515810.8	3707529.41	28.99	27.35	26.15
01.05.2018	4058513.514	8846352.941	28.91	26.64	26.18	3649459.5	4153294.12	28.95	27.61	26.18
01.06.2018	3931216.216	8634235.294	29.07	26.89	26.20	4062973	4485058.82	29.07	27.8	26.20
01.07.2018	4293513.514	9208352.941	28.99	26.81	26.40	4661756.8	5180588.24	29.01	27.6	26.40
01.08.2018	4265540.541	9419529.412	29.16	26.93	27.48	4739864.9	5200352.94	29.25	28.01	27.48
01.09.2018	3789864.865	8749176.471	30.68	28.54	28.19	3612567.6	4438000	30.96	29.46	28.19
01.10.2018	3333108.108	7489294.118	32.7	31.03	28.13	3501216.2	4459411.76	32.78	31.97	28.13
01.11.2018	3087162.162	6732117.647	31.84	31.97	27.93	3496891.9	3873176.47	32.23	32.05	27.93
01.12.2018	3180405.405	6681764.706	30.19	30.43	27.79	3257973	3453176.47	30.39	30.44	27.79
01.01.2019	2962027.027	6205882.353	28.47	28.37	27.88	2582432.4	2493058.82	28.69	28.59	27.88

01.02.2019	2501621.622	5410588.235	28.05	27.83	27.16	2586351.4	2694117.65	28.46	28.33	27.16
01.03.2019	2682297.297	5810000	28.01	27.74	26.86	3243648.6	3200470.59	28.44	28.36	26.86
01.04.2019	3027972.973	6231764.706	28.47	27.96	26.81	3509459.5	3357294.12	28.87	28.48	26.81
01.05.2019	3071891.892	6287294.118	29.75	29.25	26.38	3562027	3437294.12	30.33	30.01	26.38
01.06.2019	3237297.297	6596823.529	29.78	28.89	26.50	4227027	4076823.53	30.39	29.91	26.50
01.07.2019	3617702.703	6872117.647	29.09	28.12	25.75	4639189.2	4523058.82	29.04	28.66	25.75
01.08.2019	3757972.973	7398000	28.69	27.61	25.25	4431891.9	4749176.47	28.67	28.18	25.25
01.09.2019	2791756.757	6352823.529	28.49	27.61	24.77	3762027	4268588.24	28.62	28.12	24.77
01.10.2019	2671486.486	6300705.882	28.35	27.59	24.81	3718378.4	4120352.94	28.61	28.1	24.81
01.11.2019	2467027.027	5628000	28.4	27.61	24.37	3641621.6	3885411.76	28.6	28.07	24.37
01.12.2019	2614324.324	5821294.118	27.82	27.02	23.61	3484864.9	3665411.76	28.25	27.87	23.61
01.01.2020	2690135.135	6372470.588	27.82	27.02	24.12	2895540.5	3152705.88	28.25	27.87	24.12
01.02.2020	2531486.486	6442352.941	26.8	25.88	24.60	3118243.2	3597529.41	26.95	26.59	24.60
01.03.2020	2262162.162	5511764.706	26.29	25.37	26.41	3362973	3854705.88	26.61	26.24	26.41
01.04.2020	1717297.297	3864117.647	25.77	24.93	27.22	2959864.9	3014705.88	26.45	26.21	27.22
01.05.2020	2325270.27	4923764.706	23.05	22.57	26.81	3329324.3	3153058.82	23.79	23.45	26.81
01.06.2020	2655000	6089058.824	21.37	20.32	26.71	4697567.6	4527294.12	21.85	21.47	26.71
01.07.2020	2956216.216	7057294.118	21.57	20.57	27.31	5551216.2	5570352.94	21.85	21.4	27.31
01.08.2020	3327702.703	7228823.529	22.22	21.39	27.52	5753243.2	5536470.59	22.31	21.76	27.52
01.09.2020	2952027.027	6286588.235	22.56	22.01	27.98	4779459.5	5018235.29	22.75	22.22	27.98
01.10.2020	2744189.189	6541411.765	22.65	22.08	28.32	4544189.2	4605647.06	22.96	22.49	28.32
01.11.2020	2490135.135	6183411.765	22.95	21.98	28.31	4205540.5	4172000	23.05	22.02	28.31
01.12.2020	2532297.297	6347529.412	27.82	27.02	28.17	4503108.1	4392588.24	28.25	27.87	28.17
01.01.2021	2863783.784	6490000	25.06	24.13	28.22	3588918.9	3212000	26.06	24.99	28.22

	Варіант 9					Варіант 10				
	Івано-Франківська					Київська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	2314459.459	4740588.235	24.82	21.81	27.15	6427297.3	7352588.24	24.64	21.48	27.15
01.02.2017	2314515.881	4740498.669	25.48	22.46	27.03	6428890.1	7351009.02	25.1	22.15	27.03
01.03.2017	2034459.459	4607764.706	25.53	22.52	27.00	6621756.8	8056705.88	25.06	21.86	27.00
01.04.2017	2307162.162	4793058.824	25.82	22.79	26.86	8790270.3	10022000	25.2	21.91	26.86
01.05.2017	2275135.135	4723882.353	25.97	23.2	26.42	7468918.9	9300588.24	25.48	22.27	26.42
01.06.2017	2380810.811	4569058.824	25.78	23.03	26.11	7986756.8	9450823.53	25.34	22	26.11
01.07.2017	2456756.757	4740000	25.53	22.59	25.97	8222837.8	9861647.06	24.94	21.41	25.97
01.08.2017	2950540.541	5619529.412	25.53	22.42	25.64	8868648.6	10063529.4	25.06	21.41	25.64
01.09.2017	2548378.378	4751882.353	26.03	22.73	26.11	7567837.8	8812000	25.43	21.89	26.11
01.10.2017	2248378.378	5506705.882	27.35	24.43	26.65	7187973	9702941.18	26.71	23.92	26.65
01.11.2017	2043648.649	5125411.765	28.37	25.62	26.71	6382027	8827058.82	27.53	24.84	26.71
01.12.2017	2058648.649	4687647.059	28.94	26.25	27.52	5883513.5	8114705.88	28.44	25.91	27.52
01.01.2018	2125000	4435176.471	30.56	27.09	28.43	5132297.3	6630117.65	29.78	27.09	28.43
01.02.2018	1591351.351	3529411.765	31.13	27.32	27.17	4560000	5823647.06	30.2	27.63	27.17
01.03.2018	1740270.27	3796705.882	30.34	26.54	26.34	4544729.7	5996470.59	29.36	27.63	26.34
01.04.2018	2503648.649	5142823.529	30.13	26.24	26.15	6845135.1	7691411.76	28.84	27.63	26.15
01.05.2018	1982162.162	4580000	30.46	26.81	26.18	6477702.7	8953764.71	29.54	27.63	26.18
01.06.2018	1947162.162	4412941.176	30.43	27.15	26.20	6771216.2	8870705.88	29.75	27.63	26.20
01.07.2018	2139324.324	4793411.765	30.25	27.01	26.40	6940945.9	9082588.24	29.56	27.63	26.40
01.08.2018	2398108.108	5318000	30.77	27.08	27.48	7297027	9525176.47	29.86	27.63	27.48
01.09.2018	1972297.297	4694705.882	32.36	28.76	28.19	6323783.8	8627411.76	31.43	27.63	28.19
01.10.2018	1794324.324	5125411.765	33.95	31.09	28.13	7395135.1	11622941.2	32.93	27.63	28.13
01.11.2018	1705405.405	4495764.706	32.96	31.75	27.93	7172162.2	10723647.1	32.3	27.63	27.93
01.12.2018	1675675.676	4031882.353	30.77	29.49	27.79	5991756.8	8651176.47	30.4	27.63	27.79
01.01.2019	1883918.919	3977176.471	29.05	28.66	27.88	6362567.6	7486705.88	28.83	28.64	27.88

01.02.2019	1586216.216	3531294.118	29	28.55	27.16	5650135.1	7604705.88	28.35	28.35	27.16
01.03.2019	1616351.351	3465882.353	29	28.53	26.86	6504459.5	8400352.94	28.26	28.32	26.86
01.04.2019	1712432.432	3750470.588	29.95	28.81	26.81	7197162.2	8856470.59	29.17	28.51	26.81
01.05.2019	1779729.73	3669647.059	30.97	29.86	26.38	7814594.6	9058588.24	30.08	29.46	26.38
01.06.2019	1881891.892	3948823.529	30.9	29.75	26.50	9353513.5	11725882.4	30.06	29.22	26.50
01.07.2019	2036216.216	4357764.706	29.89	28.67	25.75	8952027	12318470.6	29.25	28.21	25.75
01.08.2019	2295810.811	4872117.647	29.66	28.34	25.25	9899729.7	13080941.2	29.02	27.96	25.25
01.09.2019	1894459.459	4509882.353	29.63	28.28	24.77	8922027	12608470.6	28.97	27.9	24.77
01.10.2019	1867162.162	4725647.059	29.62	28.35	24.81	8860540.5	13035294.1	28.94	28.05	24.81
01.11.2019	1755945.946	4174235.294	29.55	28.25	24.37	8436486.5	12261882.4	28.82	27.92	24.37
01.12.2019	1850540.541	4260823.529	28.7	27.69	23.61	7835540.5	11477058.8	28.2	27.32	23.61
01.01.2020	1927162.162	4447764.706	28.7	27.69	24.12	7663918.9	11732588.2	28.2	27.32	24.12
01.02.2020	1853108.108	4321882.353	27.59	26.63	24.60	8127432.4	12837058.8	27.03	26.52	24.60
01.03.2020	1695135.135	3829058.824	27.36	26.4	26.41	8771891.9	13828352.9	26.74	25.98	26.41
01.04.2020	1314729.73	2898470.588	27.17	26.05	27.22	7841756.8	11329411.8	26.4	25.6	27.22
01.05.2020	1799189.189	3548000	25.16	24.24	26.81	10025000	12876941.2	24.31	23.59	26.81
01.06.2020	2081216.216	4174705.882	23.48	22.39	26.71	11505135	14531294.1	22.63	22.39	26.71
01.07.2020	2435945.946	4975882.353	23.46	22.46	27.31	11973108	15457411.8	22.76	22.37	27.31
01.08.2020	2809189.189	5387647.059	24.23	23.39	27.52	12426351	15613882.4	23.85	23.18	27.52
01.09.2020	2180405.405	4768000	24.86	23.75	27.98	11305946	16057647.1	24.65	23.92	27.98
01.10.2020	2091486.486	4843647.059	25.07	23.86	28.32	12189324	17320941.2	25.01	24.14	28.32
01.11.2020	1734324.324	4382823.529	25.28	24.07	28.31	10524054	15878470.6	24.91	24.01	28.31
01.12.2020	1755135.135	4489058.824	28.7	27.69	28.17	9997162.2	15664117.6	28.48	27.79	28.17
01.01.2021	2608108.108	5007882.353	27.39	26.37	28.22	8958918.9	12370235.3	26.81	25.44	28.22

	Варіант 11					Варіант 12				
	Кіровоградська					Луганська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	922027.027	1282352.941	23.83	21.49	27.15	332027.03	343058.824	24.66	21.99	27.15
01.02.2017	921946.1641	1281902.378	24.68	22.17	27.03	331918.95	343087.989	25.34	22.64	27.03
01.03.2017	913648.6486	1506235.294	24.68	22.17	27.00	291756.76	409647.059	25.34	22.64	27.00
01.04.2017	1260270.27	1818352.941	25.11	22.17	26.86	387297.3	397764.706	25.34	22.64	26.86
01.05.2017	1247432.432	1741411.765	25.36	22.17	26.42	378378.38	383882.353	25.34	22.64	26.42
01.06.2017	1275945.946	1679176.471	25.08	22.17	26.11	425270.27	389764.706	25.34	22.64	26.11
01.07.2017	1327567.568	1792823.529	24.87	22.17	25.97	431621.62	403764.706	25.34	22.64	25.97
01.08.2017	1480540.541	1883882.353	24.76	22.17	25.64	477297.3	439294.118	25.34	22.64	25.64
01.09.2017	1306351.351	1969529.412	25.45	22.17	26.11	434594.59	408000	25.34	22.64	26.11
01.10.2017	1219459.459	2372588.235	26.62	22.17	26.65	426486.49	467176.471	25.34	22.64	26.65
01.11.2017	1109729.73	1920235.294	27.67	22.17	26.71	380405.41	435529.412	25.34	22.64	26.71
01.12.2017	1028648.649	1529058.824	24.68	22.17	27.52	394864.86	404705.882	25.34	22.64	27.52
01.01.2018	892027.027	1271176.471	29.84	27.34	28.43	332702.7	354941.176	29.43	27.12	28.43
01.02.2018	772027.027	1225529.412	30.28	28.28	27.17	292567.57	318705.882	30.19	28.02	27.17
01.03.2018	791081.0811	1267176.471	30.28	28.28	26.34	317837.84	335058.824	30.19	28.02	26.34
01.04.2018	1205945.946	1799411.765	30.28	28.28	26.15	387297.3	356823.529	30.19	28.02	26.15
01.05.2018	1114324.324	1892117.647	30.28	28.28	26.18	391621.62	367882.353	30.19	28.02	26.18
01.06.2018	1116351.351	2060117.647	30.28	28.28	26.20	400270.27	373294.118	30.19	28.02	26.20
01.07.2018	1242432.432	2153176.471	30.28	28.28	26.40	422162.16	408117.647	30.19	28.02	26.40
01.08.2018	1315000	2130941.176	30.28	28.28	27.48	449594.59	422823.529	30.19	28.02	27.48
01.09.2018	701621.6216	1475647.059	30.28	28.28	28.19	377432.43	380117.647	30.19	28.02	28.19
01.10.2018	1100405.405	3028117.647	30.28	28.28	28.13	385405.41	468823.529	30.19	28.02	28.13
01.11.2018	1040675.676	2323294.118	30.28	28.28	27.93	392297.3	452470.588	30.19	28.02	27.93
01.12.2018	896486.4865	1600117.647	30.28	28.28	27.79	363918.92	380823.529	30.19	28.02	27.79
01.01.2019	808513.5135	1227647.059	28.87	28.32	27.88	331486.49	356705.882	28.96	29.17	27.88

01.02.2019	771891.8919	1285058.824	28.7	28.36	27.16	322972.97	351529.412	28.6	28.82	27.16
01.03.2019	910945.9459	1393529.412	28.7	28.36	26.86	381621.62	378941.176	28.6	28.82	26.86
01.04.2019	980270.2703	1437411.765	28.7	28.36	26.81	420810.81	406588.235	28.6	28.82	26.81
01.05.2019	1063918.919	1410941.176	28.7	28.36	26.38	444864.86	414588.235	28.6	28.82	26.38
01.06.2019	1143783.784	1639764.706	28.7	28.36	26.50	456351.35	410470.588	28.6	28.82	26.50
01.07.2019	1201216.216	2087294.118	28.7	28.36	25.75	453513.51	383529.412	28.6	28.82	25.75
01.08.2019	1286621.622	1982000	28.7	28.36	25.25	467027.03	379176.471	28.6	28.82	25.25
01.09.2019	1162972.973	2360235.294	28.7	28.36	24.77	457297.3	365529.412	28.6	28.82	24.77
01.10.2019	1184189.189	2462705.882	28.7	28.36	24.81	485540.54	404588.235	28.6	28.82	24.81
01.11.2019	1147567.568	2165411.765	28.7	28.36	24.37	485810.81	396705.882	28.6	28.82	24.37
01.12.2019	1075000	1858117.647	28.7	28.36	23.61	476621.62	387176.471	28.6	28.82	23.61
01.01.2020	981756.7568	1686823.529	28.48	27.79	24.12	404594.59	344352.941	27.37	26.96	24.12
01.02.2020	989594.5946	1758823.529	27.44	26.79	24.60	445810.81	395176.471	26.43	25.93	24.60
01.03.2020	1066216.216	1758117.647	27.07	26.26	26.41	491081.08	400117.647	26.18	25.65	26.41
01.04.2020	968513.5135	1446705.882	26.97	26.15	27.22	447567.57	293411.765	25.95	25.52	27.22
01.05.2020	1229189.189	1650705.882	24.84	24.01	26.81	550810.81	334000	23.14	22.56	26.81
01.06.2020	1407567.568	1993058.824	22.83	21.89	26.71	638378.38	408352.941	21.77	21.06	26.71
01.07.2020	1566486.486	2333294.118	22.63	21.71	27.31	716891.89	472000	21.59	21.17	27.31
01.08.2020	1639324.324	2329294.118	23.1	22.24	27.52	726756.76	492352.941	22.66	22.07	27.52
01.09.2020	1429729.73	2313764.706	23.55	22.67	27.98	694864.86	467764.706	23.59	22.51	27.98
01.10.2020	1465540.541	2521529.412	23.74	22.8	28.32	702837.84	442117.647	23.69	22.94	28.32
01.11.2020	1282297.297	2259764.706	23.96	23.04	28.31	670135.14	444235.294	23.82	22.93	28.31
01.12.2020	1243243.243	2141411.765	28.7	28.36	28.17	740270.27	530000	28.6	28.82	28.17
01.01.2021	1276351.351	1756117.647	26	25.03	28.22	590810.81	374470.588	26.32	25.32	28.22

	Варіант 13					Варіант 14				
	Львівська					Миколаївська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	5277567.568	11333882.35	24.15	21.62	27.15	1529594.6	1658117.65	23.7	21.6	27.15
01.02.2017	5276574.749	11331135.4	24.78	22.25	27.03	1529250.8	1658119.66	24.58	22.33	27.03
01.03.2017	5202702.703	11523764.71	24.78	22.25	27.00	1570540.5	1746000	24.58	22.33	27.00
01.04.2017	6113378.378	13006823.53	24.78	22.25	26.86	2005810.8	2004117.65	24.58	22.33	26.86
01.05.2017	6129324.324	12577176.47	24.78	22.25	26.42	2057702.7	1938941.18	24.58	22.33	26.42
01.06.2017	6241081.081	12192588.24	24.78	22.25	26.11	2196621.6	2007058.82	24.58	22.33	26.11
01.07.2017	6159324.324	12132470.59	24.78	22.25	25.97	2542027	2925411.76	24.58	22.33	25.97
01.08.2017	6633243.243	12552705.88	24.78	22.25	25.64	2728108.1	2837176.47	24.58	22.33	25.64
01.09.2017	6343243.243	11885294.12	24.78	22.25	26.11	2195540.5	2441411.76	24.58	22.33	26.11
01.10.2017	5966891.892	12596470.59	24.78	22.25	26.65	2083378.4	2814470.59	24.58	22.33	26.65
01.11.2017	5497027.027	11305176.47	24.78	22.25	26.71	1959864.9	2373647.06	24.58	22.33	26.71
01.12.2017	5177297.297	10107176.47	24.78	22.25	27.52	1817027	2110235.29	24.58	22.33	27.52
01.01.2018	4875945.946	9473764.706	29.63	27.45	28.43	1515675.7	1785411.76	29.48	26.66	28.43
01.02.2018	3972432.432	7975764.706	30.33	28.07	27.17	1393513.5	1670352.94	29.93	27.94	27.17
01.03.2018	4397567.568	8822705.882	30.33	28.07	26.34	1406216.2	1784117.65	29.93	27.94	26.34
01.04.2018	6505270.27	11877647.06	30.33	28.07	26.15	2008108.1	2265764.71	29.93	27.94	26.15
01.05.2018	5614864.865	11944117.65	30.33	28.07	26.18	1868648.6	2175176.47	29.93	27.94	26.18
01.06.2018	5488243.243	11459882.35	30.33	28.07	26.20	1955810.8	2562470.59	29.93	27.94	26.20
01.07.2018	5729459.459	11788941.18	30.33	28.07	26.40	2243378.4	3301647.06	29.93	27.94	26.40
01.08.2018	6006756.757	12942352.94	30.33	28.07	27.48	2345945.9	3101176.47	29.93	27.94	27.48
01.09.2018	5386216.216	12252000	30.33	28.07	28.19	1837162.2	2684352.94	29.93	27.94	28.19
01.10.2018	5067162.162	12541058.82	30.33	28.07	28.13	1780675.7	2959058.82	29.93	27.94	28.13
01.11.2018	4826486.486	11320941.18	30.33	28.07	27.93	1674459.5	2661882.35	29.93	27.94	27.93
01.12.2018	4491621.622	10043882.35	30.33	28.07	27.79	1522432.4	2389764.71	29.93	27.94	27.79
01.01. 2019	4379864.865	8919058.824	29	28.95	27.88	1299324.3	1991882.35	28.6	28.81	27.88

01.02.2019	4255675.676	8313058.824	28.88	28.8	27.16	1267837.8	2008117.65	28.31	28.39	27.16
01.03.2019	4383918.919	8277882.353	28.88	28.8	26.86	1522027	2240941.18	28.31	28.39	26.86
01.04.2019	4794054.054	8745647.059	28.88	28.8	26.81	1637162.2	2277529.41	28.31	28.39	26.81
01.05.2019	4955945.946	8816117.647	28.88	28.8	26.38	1747432.4	2336235.29	28.31	28.39	26.38
01.06.2019	5367297.297	10623294.12	28.88	28.8	26.50	1905000	2554705.88	28.31	28.39	26.50
01.07.2019	5523378.378	11213058.82	28.88	28.8	25.75	2168378.4	3067647.06	28.31	28.39	25.75
01.08.2019	5952702.703	11690000	28.88	28.8	25.25	2147837.8	3140588.24	28.31	28.39	25.25
01.09.2019	5363243.243	11982470.59	28.88	28.8	24.77	1834864.9	3115882.35	28.31	28.39	24.77
01.10.2019	5322567.568	12036941.18	28.88	28.8	24.81	1844054.1	3244235.29	28.31	28.39	24.81
01.11.2019	5112702.703	11123882.35	28.88	28.8	24.37	1765000	3185294.12	28.31	28.39	24.37
01.12.2019	5077432.432	11116117.65	28.88	28.8	23.61	1762162.2	2841764.71	28.31	28.39	23.61
01.01.2020	4781621.622	12976941.18	28.72	28.28	24.12	1605000	2668705.88	28.29	27.85	24.12
01.02.2020	4791216.216	11174470.59	27.24	27.02	24.60	1617432.4	2863411.76	27.02	26.48	24.60
01.03.2020	4497702.703	10108352.94	26.76	26.42	26.41	1721216.2	2786941.18	26.68	26.1	26.41
01.04.2020	3690810.811	7338705.882	26.69	26.24	27.22	1554054.1	2338941.18	26.23	25.98	27.22
01.05.2020	4653918.919	7875529.412	24.44	24.17	26.81	1898513.5	2467882.35	24.09	24.13	26.81
01.06.2020	5465945.946	9350588.235	22.38	22.03	26.71	2368378.4	2864000	22.15	21.41	26.71
01.07.2020	6032567.568	10738588.24	22.43	22.13	27.31	2821216.2	4070941.18	22.24	21.79	27.31
01.08.2020	6263513.514	11220352.94	22.76	22.19	27.52	2975405.4	3737764.71	23.14	22.49	27.52
01.09.2020	5539594.595	10767294.12	22.93	22.42	27.98	2375135.1	3824235.29	23.63	22.97	27.98
01.10.2020	5342972.973	11129411.76	23.02	22.58	28.32	2403918.9	3895529.41	23.75	23.21	28.32
01.11.2020	4650000	9886823.529	23.23	22.85	28.31	2231351.4	3694470.59	24.02	23.35	28.31
01.12.2020	4624054.054	9878941.176	28.88	28.8	28.17	2329864.9	3722235.29	28.31	28.39	28.17
01.01.2021	5393243.243	9671294.118	25.24	24.76	28.22	1974729.7	2633647.06	26.3	25.61	28.22

	Варіант 15					Варіант 16				
	Одеська					Полтавська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	6658108.108	4732941.176	23.88	21.68	27.15	1750810.8	2651764.71	23.3	21.08	27.15
01.02.2017	6660955.598	4731416.513	24.36	22.27	27.03	1751626.8	2651729.12	23.77	21.84	27.03
01.03.2017	7460270.27	5275529.412	24.36	22.27	27.00	1753378.4	2805529.41	23.77	21.84	27.00
01.04.2017	9035405.405	6007764.706	24.36	22.27	26.86	2385000	3378705.88	23.77	21.84	26.86
01.05.2017	9034459.459	5928588.235	24.36	22.27	26.42	2357027	3251176.47	23.77	21.84	26.42
01.06.2017	9666486.486	6175058.824	24.36	22.27	26.11	2423783.8	3110705.88	23.77	21.84	26.11
01.07.2017	10615405.41	7597647.059	24.36	22.27	25.97	2490675.7	3329764.71	23.77	21.84	25.97
01.08.2017	11738648.65	8221764.706	24.36	22.27	25.64	3070000	4558352.94	23.77	21.84	25.64
01.09.2017	9779729.73	6650588.235	24.36	22.27	26.11	2507973	3464705.88	23.77	21.84	26.11
01.10.2017	9087972.973	6977411.765	24.36	22.27	26.65	2212567.6	4284235.29	23.77	21.84	26.65
01.11.2017	8768108.108	6422941.176	24.36	22.27	26.71	2177567.6	3990470.59	23.77	21.84	26.71
01.12.2017	8446621.622	5727176.471	24.36	22.27	27.52	1874729.7	3149411.76	23.77	21.84	27.52
01.01.2018	6476216.216	4505058.824	28.86	26.44	28.43	1645270.3	2735647.06	28.17	26.88	28.43
01.02.2018	5947702.703	4220588.235	29.6	27.28	27.17	1380270.3	2318235.29	28.94	27.9	27.17
01.03.2018	6369189.189	4654117.647	29.6	27.28	26.34	1391486.5	2435411.76	28.94	27.9	26.34
01.04.2018	9210270.27	6340235.294	29.6	27.28	26.15	2117432.4	3125058.82	28.94	27.9	26.15
01.05.2018	8245540.541	5935176.471	29.6	27.28	26.18	2092432.4	3518705.88	28.94	27.9	26.18
01.06.2018	8922702.703	6876352.941	29.6	27.28	26.20	2192973	3527764.71	28.94	27.9	26.20
01.07.2018	9876351.351	7920352.941	29.6	27.28	26.40	2350675.7	3722000	28.94	27.9	26.40
01.08.2018	10437837.84	8091411.765	29.6	27.28	27.48	2482162.2	3702352.94	28.94	27.9	27.48
01.09.2018	7730270.27	5796588.235	29.6	27.28	28.19	2069864.9	3981411.76	28.94	27.9	28.19
01.10.2018	7942432.432	6115529.412	29.6	27.28	28.13	2006486.5	4755294.12	28.94	27.9	28.13
01.11.2018	7979729.73	5923882.353	29.6	27.28	27.93	1888378.4	4029294.12	28.94	27.9	27.93
01.12.2018	7410000	5292941.176	29.6	27.28	27.79	1602702.7	3225176.47	28.94	27.9	27.79
01.01.2019	6458513.514	4449647.059	28.39	28.39	27.88	1452567.6	2416470.59	28.22	28.43	27.88

01.02.2019	6640405.405	4647411.765	28.01	27.99	27.16	1403378.4	2341647.06	27.91	27.95	27.16
01.03.2019	7570810.811	5068705.882	28.01	27.99	26.86	1637432.4	2636352.94	27.91	27.95	26.86
01.04.2019	7741351.351	5106588.235	28.01	27.99	26.81	1874054.1	2850470.59	27.91	27.95	26.81
01.05.2019	7913108.108	5181647.059	28.01	27.99	26.38	2045810.8	2771058.82	27.91	27.95	26.38
01.06.2019	8983378.378	6165294.118	28.01	27.99	26.50	2299459.5	3378352.94	27.91	27.95	26.50
01.07.2019	9171756.757	6656823.529	28.01	27.99	25.75	2362702.7	3566352.94	27.91	27.95	25.75
01.08.2019	9615945.946	6949176.471	28.01	27.99	25.25	2509864.9	3553294.12	27.91	27.95	25.25
01.09.2019	8190675.676	5766235.294	28.01	27.99	24.77	2211756.8	3712588.24	27.91	27.95	24.77
01.10.2019	8256216.216	5692000	28.01	27.99	24.81	2252297.3	4073294.12	27.91	27.95	24.81
01.11.2019	8268513.514	6421647.059	28.01	27.99	24.37	2192702.7	3740352.94	27.91	27.95	24.37
01.12.2019	8145810.811	5495294.118	28.01	27.99	23.61	1985405.4	3226941.18	27.91	27.95	23.61
01.01.2020	7326756.757	5847764.706	27.56	26.99	24.12	1880135.1	3141176.47	27.73	27.22	24.12
01.02.2020	7356351.351	5637176.471	26.64	26.98	24.60	1837567.6	3378705.88	26.67	26.45	24.60
01.03.2020	6897027.027	5430823.529	25.99	25.67	26.41	2039594.6	3682000	26.4	25.93	26.41
01.04.2020	5697702.703	4258352.941	25.79	25.01	27.22	1800135.1	2951058.82	26.24	25.64	27.22
01.05.2020	6644594.595	5489411.765	23.11	22.89	26.81	2412432.4	3420235.29	24.26	23.91	26.81
01.06.2020	8737567.568	5730823.529	20.97	20.83	26.71	2765000	4154117.65	21.62	21.58	26.71
01.07.2020	9766891.892	6911411.765	21.05	20.7	27.31	3023513.5	4726470.59	21.4	21.49	27.31
01.08.2020	10296351.35	7201058.824	22.03	21.64	27.52	3186486.5	4728823.53	22.55	22.18	27.52
01.09.2020	8740270.27	5985647.059	23.22	22.7	27.98	2938378.4	4987411.76	23.13	22.53	27.98
01.10.2020	8206486.486	5713647.059	23.25	22.83	28.32	3061216.2	5598941.18	23.23	22.65	28.32
01.11.2020	6247297.297	4323176.471	23.12	22.52	28.31	2769729.7	4928941.18	23.43	22.59	28.31
01.12.2020	6754864.865	4570941.176	28.01	27.99	28.17	2845135.1	4833294.12	27.91	27.95	28.17
01.01.2021	7512837.838	5098823.529	25.27	25.02	28.22	2531756.8	3638000	25	24.3	28.22

	Варіант 17					Варіант 18				
	Рівненська					Сумська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	1658243.243	4063882.353	23.68	21.19	27.15	1157297.3	1770588.24	22.93	20.87	27.15
01.02.2017	1657323.604	4062146.563	24.03	21.75	27.03	1156951.4	1770461.61	23.41	21.38	27.03
01.03.2017	1430270.27	4098705.882	24.03	21.75	27.00	1092162.2	1809411.76	23.41	21.38	27.00
01.04.2017	1977702.703	4847176.471	24.03	21.75	26.86	1354459.5	2150588.24	23.41	21.38	26.86
01.05.2017	1875405.405	4404941.176	24.03	21.75	26.42	1392702.7	1931529.41	23.41	21.38	26.42
01.06.2017	1946216.216	4105764.706	24.03	21.75	26.11	1439864.9	1790000	23.41	21.38	26.11
01.07.2017	2012567.568	4270941.176	24.03	21.75	25.97	1493513.5	1922352.94	23.41	21.38	25.97
01.08.2017	2111621.622	4650235.294	24.03	21.75	25.64	1636486.5	2050235.29	23.41	21.38	25.64
01.09.2017	1937837.838	4162588.235	24.03	21.75	26.11	1440405.4	2054117.65	23.41	21.38	26.11
01.10.2017	1771216.216	4824117.647	24.03	21.75	26.65	1323918.9	2580705.88	23.41	21.38	26.65
01.11.2017	1590810.811	4288000	24.03	21.75	26.71	1218918.9	2489058.82	23.41	21.38	26.71
01.12.2017	1483918.919	3829176.471	24.03	21.75	27.52	1121486.5	1758352.94	23.41	21.38	27.52
01.01.2018	1472567.568	3514235.294	28.86	26.02	28.43	1046351.4	1745529.41	28.41	26.99	28.43
01.02.2018	1134864.865	3129764.706	29.42	26.91	27.17	886351.35	1588117.65	29	27.97	27.17
01.03.2018	1249594.595	3391882.353	29.42	26.91	26.34	905135.14	1689529.41	29	27.97	26.34
01.04.2018	1774054.054	4393529.412	29.42	26.91	26.15	1234324.3	2072941.18	29	27.97	26.15
01.05.2018	1580945.946	4400235.294	29.42	26.91	26.18	1254594.6	2253764.71	29	27.97	26.18
01.06.2018	1597297.297	4317647.059	29.42	26.91	26.20	1295540.5	2187647.06	29	27.97	26.20
01.07.2018	1741351.351	4517529.412	29.42	26.91	26.40	1244594.6	2117882.35	29	27.97	26.40
01.08.2018	1878783.784	4717764. 706	29.42	26.91	27.48	1494729.7	2877529.41	29	27.97	27.48
01.09.2018	1585540.541	4362823.529	29.42	26.91	28.19	1274324.3	2618235.29	29	27.97	28.19
01.10.2018	1490135.135	4837882.353	29.42	26.91	28.13	1238648.6	3315176.47	29	27.97	28.13
01.11.2018	1384054.054	4248352.941	29.42	26.91	27.93	1231891.9	2838941.18	29	27.97	27.93
01.12.2018	1259594.595	3624000	29.42	26.91	27.79	1035810.8	1882941.18	29	27.97	27.79
01.01.2019	1241756.757	3216588.235	28.51	28.79	27.88	990270.27	1674235.29	28.12	27.63	27.88

01.02.2019	1121081.081	3079294.118	28.22	28.48	27.16	905405.41	1707764.71	27.97	27.21	27.16
01.03.2019	1258783.784	3321411.765	28.22	28.48	26.86	1056351.4	1895058.82	27.97	27.21	26.86
01.04.2019	1466756.757	3527058.824	28.22	28.48	26.81	1202162.2	2132823.53	27.97	27.21	26.81
01.05.2019	1726216.216	3576352.941	28.22	28.48	26.38	1285675.7	2025764.71	27.97	27.21	26.38
01.06.2019	1501081.081	3422941.176	28.22	28.48	26.50	1391486.5	2019294.12	27.97	27.21	26.50
01.07.2019	1666621.622	3772352.941	28.22	28.48	25.75	1443783.8	2276235.29	27.97	27.21	25.75
01.08.2019	1868648.649	3945882.353	28.22	28.48	25.25	1511081.1	2316588.24	27.97	27.21	25.25
01.09.2019	1608783.784	4162000	28.22	28.48	24.77	1373513.5	2612823.53	27.97	27.21	24.77
01.10.2019	1635945.946	4732941.176	28.22	28.48	24.81	1406756.8	3129764.71	27.97	27.21	24.81
01.11.2019	1503513.514	4379176.471	28.22	28.48	24.37	1345000	2332705.88	27.97	27.21	24.37
01.12.2019	1482972.973	4204823.529	28.22	28.48	23.61	1278378.4	2023882.35	27.97	27.21	23.61
01.01.2020	1804189.189	4219058.824	27.67	27.65	24.12	1173108.1	1939529.41	27.6	26.57	24.12
01.02.2020	1650135.135	4056352.941	26.58	25.66	24.60	1164324.3	2066470.59	26.6	25.65	24.60
01.03.2020	1381351.351	4180823.529	26.11	25.45	26.41	1334729.7	2248235.29	26.39	25.06	26.41
01.04.2020	1122162.162	3171647.059	25.93	25.07	27.22	1253783.8	2054235.29	26.17	24.83	27.22
01.05.2020	1365000	3385058.824	23.78	23.03	26.81	1550405.4	2229882.35	23.98	23.06	26.81
01.06.2020	1605135.135	3766588.235	21.57	20.68	26.71	1787567.6	2600352.94	21.8	20.74	26.71
01.07.2020	1767972.973	4340000	21.54	20.68	27.31	1930540.5	3028117.65	21.89	20.64	27.31
01.08.2020	1964729.73	4526705.882	22.29	21.56	27.52	1964864.9	2951647.06	22.43	21.45	27.52
01.09.2020	1618918.919	4411647.059	22.73	22.06	27.98	1893783.8	3364705.88	23.08	22.48	27.98
01.10.2020	1537027.027	4300470.588	22.83	22.17	28.32	1892027	4221294.12	23.2	22.31	28.32
01.11.2020	1296486.486	3809411.765	23.06	22.14	28.31	1717432.4	3199058.82	23.16	22.15	28.31
01.12.2020	1327162.162	3724705.882	28.22	28.48	28.17	1857973	2926705.88	27.97	27.21	28.17
01.01.2021	1704054.054	3864941.176	24.99	24.13	28.22	1647432.4	2402235.29	24.67	23.75	28.22

	Варіант 19					Варіант 20				
	Тернопільська					Харківська				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	1286351.351	3788000	24.02	21.41	27.15	3305000	4000352.94	24.2	21.68	27.15
01.02.2017	1287063.763	3788614.801	24.68	22.26	27.03	3303876.7	4001632.73	24.69	22.27	27.03
01.03.2017	927972.973	2348941.176	24.68	22.26	27.00	4152973	5190117.65	24.69	22.27	27.00
01.04.2017	1217567.568	3076941.176	24.68	22.26	26.86	4601756.8	4588117.65	24.69	22.27	26.86
01.05.2017	1253513.514	2826705.882	24.68	22.26	26.42	4251216.2	5176235.29	24.69	22.27	26.42
01.06.2017	1263918.919	2684941.176	24.68	22.26	26.11	4123378.4	3646000	22.27	22.27	26.11
01.07.2017	1325810.811	2804352.941	24.68	22.26	25.97	4212567.6	4084235.29	24.69	22.27	25.97
01.08.2017	1529459.459	3207058.824	24.68	22.26	25.64	4538783.8	4434000	24.69	22.27	25.64
01.09.2017	1289459.459	2720470.588	24.68	22.26	26.11	4385675.7	4216588.24	24.69	22.27	26.11
01.10.2017	1156081.081	3081764.706	24.68	22.26	26.65	4288513.5	4835647.06	24.69	22.27	26.65
01.11.2017	1016216.216	2500000	24.68	22.26	26.71	4190540.5	4708000	24.69	22.27	26.71
01.12.2017	972432.4324	2143529.412	24.68	22.26	27.52	4005675.7	4211294.12	24.69	22.27	27.52
01.01.2018	972297.2973	2115058.824	29.23	26.55	28.43	3422297.3	3610941.18	29.22	27.51	28.43
01.02.2018	721216.2162	1715529.412	29.98	27.77	27.17	2998378.4	3104352.94	30.17	28.2	27.17
01.03.2018	766081.0811	1753411.765	29.98	27.77	26.34	3092567.6	3333058.82	30.17	28.2	26.34
01.04.2018	1121891.892	2526588.235	29.98	27.77	26.15	4751216.2	4388352.94	30.17	28.2	26.15
01.05.2018	966216.2162	2275882.353	29.98	27.77	26.18	4141891.9	4439529.41	30.17	28.2	26.18
01.06.2018	960000	2290000	29.98	27.77	26.20	4367837.8	4448117.65	30.17	28.2	26.20
01.07.2018	1028783.784	2426000	29.98	27.77	26.40	4523918.9	4720000	30.17	28.2	26.40
01.08.2018	1151891.892	2763411.765	29.98	27.77	27.48	4736351.4	4784823.53	30.17	28.2	27.48
01.09.2018	1002297.297	2457529.412	29.98	27.77	28.19	3822837.8	4054941.18	30.17	28.2	28.19
01.10.2018	907432.4324	2476117.647	29.98	27.77	28.13	4233783.8	5237411.76	30.17	28.2	28.13
01.11.2018	881621.6216	2286000	29.98	27.77	27.93	4281081.1	4666470.59	30.17	28.2	27.93
01.12.2018	789054.0541	1992823.529	29.98	27.77	27.79	3823378.4	3914705.88	30.17	28.2	27.79
01.01.2019	853648.6486	1909529.412	28.65	28.53	27.88	3437432.4	3306352.94	28.48	27.98	27.88

01.02.2019	702162.1622	1644352.941	28.39	28.09	27.16	3364054.1	3164588.24	28.23	27.61	27.16
01.03.2019	770405.4054	1763764.706	28.39	28.09	26.86	3630405.4	3290705.88	28.23	27.61	26.86
01.04.2019	938513.5135	2016117.647	28.39	28.09	26.81	4015810.8	3496705.88	28.23	27.61	26.81
01.05.2019	977702.7027	2059294.118	28.39	28.09	26.38	3964054.1	3397647.06	28.23	27.61	26.38
01.06.2019	1070000	2283411.765	28.39	28.09	26.50	4382973	3861882.35	28.23	27.61	26.50
01.07.2019	1143918.919	2633882.353	28.39	28.09	25.75	4490675.7	4236352.94	28.23	27.61	25.75
01.08.2019	1282567.568	2751647.059	28.39	28.09	25.25	4743648.6	4337411.76	28.23	27.61	25.25
01.09.2019	1077702.703	2388941.176	28.39	28.09	24.77	4554459.5	4290470.59	28.23	27.61	24.77
01.10.2019	1049189.189	2547176.471	28.39	28.09	24.81	4733378.4	4567058.82	28.23	27.61	24.81
01.11.2019	986081.0811	2328588.235	28.39	28.09	24.37	4735135.1	4277764.71	28.23	27.61	24.37
01.12.2019	951756.7568	2181058.824	28.39	28.09	23.61	4498513.5	4133294.12	28.23	27.61	23.61
01.01.2020	959864.8649	2167764.706	28.49	27.92	24.12	3908513.5	3463411.76	27.82	27.02	24.12
01.02.2020	856891.8919	2030117.647	27.09	26.07	24.60	4452432.4	3972705.88	26.69	26.02	24.60
01.03.2020	900540.5405	1956235.294	26.73	25.42	26.41	4892567.6	4012117.65	26.41	25.66	26.41
01.04.2020	722972.973	1562941.176	26.56	25.3	27.22	4261216.2	3115176.47	25.63	24.93	27.22
01.05.2020	964189.1892	1717411.765	24.64	23.4	26.81	5313648.6	3809058.82	23.19	22.44	26.81
01.06.2020	1140135.135	1797647.059	21.61	20.76	26.71	6336891.9	4620235.29	21.33	20.38	26.71
01.07.2020	1294594.595	2062470.588	21.85	20.89	27.31	6564324.3	5312823.53	21.47	20.51	27.31
01.08.2020	1397432.432	2487764.706	22.06	21.33	27.52	6715810.8	5338470.59	22.24	21.01	27.52
01.09.2020	1300270.27	2263176.471	22.53	21.96	27.98	6775000	5811176.47	23.11	21.82	27.98
01.10.2020	1301756.757	2566588.235	22.9	22.36	28.32	7017162.2	5891647.06	23.19	21.91	28.32
01.11.2020	1017027.027	2126588.235	23.09	22.55	28.31	6471486.5	5320941.18	23.42	21.88	28.31
01.12.2020	1065540.541	2285647.059	28.39	28.09	28.17	5631756.8	5031176.47	28.23	27.61	28.17
01.01.2021	1292432.432	2356000	25.56	25.3	28.22	5568783.8	4307529.41	25.71	24.36	28.22

	Варіант 21					Варіант 22				
	Херсонська					Хмельницька				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	1323378.378	1661647.059	24.18	21.42	27.15	1573378.4	3403176.47	23.82	21.59	27.15
01.02.2017	1323297.347	1661712.742	24.89	22.36	27.03	1572901.5	3404187	24.38	22.31	27.03
01.03.2017	1427837.838	1881058.824	24.89	22.36	27.00	1426621.6	3118235.29	24.38	22.31	27.00
01.04.2017	1912702.703	2263411.765	24.89	22.36	26.86	1947837.8	4379294.12	24.38	22.31	26.86
01.05.2017	2000945.946	2535529.412	24.89	22.36	26.42	1881351.4	4176000	24.38	22.31	26.42
01.06.2017	2202162.162	2760000	24.89	22.36	26.11	1941081.1	3957176.47	24.38	22.31	26.11
01.07.2017	2711216.216	3655647.059	24.89	22.36	25.97	2020270.3	4192470.59	24.38	22.31	25.97
01.08.2017	2970540.541	3692705.882	24.89	22.36	25.64	2276891.9	4843294.12	24.38	22.31	25.64
01.09.2017	2288918.919	2829294.118	24.89	22.36	26.11	1905135.1	3764588.24	24.38	22.31	26.11
01.10.2017	1961621.622	2805529.412	24.89	22.36	26.65	1791486.5	4647647.06	24.38	22.31	26.65
01.11.2017	1792162.162	2415176.471	24.89	22.36	26.71	1595135.1	3909411.76	24.38	22.31	26.71
01.12.2017	1655540.541	2121647.059	24.89	22.36	27.52	1466081.1	3084000	24.38	22.31	27.52
01.01.2018	1313918.919	1684941.176	30.15	27.08	28.43	1393648.6	2865411.76	29.01	26.93	28.43
01.02.2018	1172432.432	1537411.765	30.51	28.1	27.17	1113243.2	2471294.12	29.78	27.75	27.17
01.03.2018	1286081.081	1772470.588	30.51	28.1	26.34	1269189.2	2802470.59	29.78	27.75	26.34
01.04.2018	2033378.378	2492235.294	30.51	28.1	26.15	1796486.5	3853411.76	29.78	27.75	26.15
01.05.2018	1848513.514	2543058.824	30.51	28.1	26.18	1682973	4166000	29.78	27.75	26.18
01.06.2018	2116621.622	3067647.059	30.51	28.1	26.20	1683918.9	4130117.65	29.78	27.75	26.20
01.07.2018	2508108.108	3749764.706	30.51	28.1	26.40	1676486.5	4384352.94	29.78	27.75	26.40
01.08.2018	2641621.622	3729529.412	30.51	28.1	27.48	1810945.9	4568941.18	29.78	27.75	27.48
01.09.2018	1974324.324	2790000	30.51	28.1	28.19	1561621.6	4032470.59	29.78	27.75	28.19
01.10.2018	1748378.378	2527176.471	30.51	28.1	28.13	1485810.8	4699176.47	29.78	27.75	28.13
01.11.2018	1343513.514	1894117.647	30.51	28.1	27.93	1400270.3	3958823.53	29.78	27.75	27.93
01.12.2018	1552702.703	2070588.235	30.51	28.1	27.79	1280000	3270941.18	29.78	27.75	27.79
01.01.2019	1342432.432	1673411.765	28.98	28.65	27.88	1187973	2783764.71	28.51	28.74	27.88

01.02.2019	1330540.541	1722352.941	28.63	28.31	27.16	1075135.1	2611764.71	28.35	28.51	27.16
01.03.2019	1544054.054	1998705.882	28.63	28.31	26.86	1254189.2	3035882.35	28.35	28.51	26.86
01.04.2019	1661081.081	2093764.706	28.63	28.31	26.81	1390945.9	3111176.47	28.35	28.51	26.81
01.05.2019	1807432.432	2355294.118	28.63	28.31	26.38	1446891.9	2991058.82	28.35	28.51	26.38
01.06.2019	2126351.351	2883411.765	28.63	28.31	26.50	1561351.4	3490000	28.35	28.51	26.50
01.07.2019	2445270.27	3373647.059	28.63	28.31	25.75	1573513.5	3527882.35	28.35	28.51	25.75
01.08.2019	2440270.27	3329529.412	28.63	28.31	25.25	1750810.8	4007176.47	28.35	28.51	25.25
01.09.2019	1936891.892	2964823.529	28.63	28.31	24.77	1565675.7	3840000	28.35	28.51	24.77
01.10.2019	1860945.946	2751882.353	28.63	28.31	24.81	1555000	4146823.53	28.35	28.51	24.81
01.11.2019	1730675.676	2488823.529	28.63	28.31	24.37	1380810.8	3076235.29	28.35	28.51	24.37
01.12.2019	1677567.568	2309411.765	28.63	28.31	23.61	1346621.6	2812235.29	28.35	28.51	23.61
01.01.2020	1460405.405	2239882.353	28.71	27.82	24.12	1286216.2	3076470.59	28.12	27.8	24.12
01.02.2020	1482702.703	2249294.118	27.08	26.11	24.60	1265405.4	2999176.47	26.63	26.35	24.60
01.03.2020	1715405.405	2651764.706	27.08	25.78	26.41	1289189.2	3229529.41	26.27	25.77	26.41
01.04.2020	1439324.324	2116352.941	26.66	25.38	27.22	1136081.1	2786705.88	26.05	25.45	27.22
01.05.2020	1823378.378	2304941.176	24.06	22.97	26.81	1413243.2	3148470.59	23.93	23.59	26.81
01.06.2020	2325135.135	2883294.118	22.36	20.92	26.71	1658918.9	3284117.65	21.57	21.53	26.71
01.07.2020	2885000	3838470.588	22.49	21.04	27.31	1842837.8	3578823.53	21.59	21.47	27.31
01.08.2020	3055945.946	4118588.235	23.4	22.12	27.52	2017297.3	3984941.18	22.37	22.27	27.52
01.09.2020	2300675.676	3389764.706	24.12	22.79	27.98	1904864.9	3866470.59	22.99	22.76	27.98
01.10.2020	2164324.324	3031294.118	24.3	23.14	28.32	1931891.9	4009411.76	23.07	23.06	28.32
01.11.2020	1862297.297	2672000	24.53	23.2	28.31	1643378.4	3791058.82	23.15	22.86	28.31
01.12.2020	1825945.946	2490000	28.63	28.31	28.17	1723243.2	3418000	28.35	28.51	28.17
01.01.2021	1681216.216	2093294.118	26.82	25.17	28.22	1721351.4	2876470.59	25.17	25.17	28.22

	Варіант 23					Варіант 24				
	Черкаська					Чернівецька				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	2177567.568	2854235.294	23.93	21.49	27.15	1503243.2	4134823.53	23.6	21.21	27.15
01.02.2017	2176818.901	2854857.524	24.5	22.2	27.03	1503653.5	4134973.04	24.01	21.76	27.03
01.03.2017	2007027.027	3023294.118	24.5	22.2	27.00	1384864.9	4015058.82	24	21.75	27.00
01.04.2017	2902702.703	3898000	24.5	22.2	26.86	1662702.7	4668588.24	24.48	22.02	26.86
01.05.2017	2830270.27	3607294.118	24.5	22.2	26.42	1675675.7	4337882.35	24.85	22.59	26.42
01.06.2017	3145000	3641294.118	24.5	22.2	26.11	1704729.7	4078117.65	24.73	22.28	26.11
01.07.2017	3462027.027	4015647.059	24.5	22.2	25.97	1745675.7	4218000	24.42	21.76	25.97
01.08.2017	3811756.757	4311294.118	24.5	22.2	25.64	1933783.8	4842117.65	24.42	21.56	25.64
01.09.2017	2992027.027	3840470.588	24.5	22.2	26.11	1704324.3	4387411.76	24.79	21.85	26.11
01.10.2017	2701621.622	3966705.882	24.5	22.2	26.65	1642973	5283647.06	26.33	23.75	26.65
01.11.2017	2455405.405	3670117.647	24.5	22.2	26.71	1480675.7	4518705.88	27.29	24.92	26.71
01.12.2017	2301351.351	3187176.471	24.5	22.2	27.52	1437297.3	3830941.18	27.83	25.46	27.52
01.01.2018	1991756.757	2633411.765	29.16	26.53	28.43	1273783.8	3390235.29	29.13	24.98	28.43
01.02.2018	1715675.676	2480352.941	29.69	27.13	27.17	1071216.2	2914000	29.4	26.49	27.17
01.03.2018	1789459.459	2681411.765	29.69	26.23	26.34	1231621.6	3346705.88	28.57	26.49	26.34
01.04.2018	2615000	3569058.824	29.69	26.27	26.15	1593783.8	4323882.35	28.38	26.49	26.15
01.05.2018	2619594.595	3701764.706	29.69	26.36	26.18	1506891.9	4147411.76	28.46	26.49	26.18
01.06.2018	2834054.054	3855529.412	29.69	26.46	26.20	1446486.5	3855294.12	28.54	26.49	26.20
01.07.2018	3183243.243	4336705.882	29.69	26.41	26.40	1572973	4133529.41	28.47	26.49	26.40
01.08.2018	3412972.973	4391764.706	29.69	26.73	27.48	1682432.4	4279176.47	28.74	26.49	27.48
01.09.2018	2630540.541	4106705.882	29.69	28.43	28.19	1397027	3827294.12	30.36	26.49	28.19
01.10.2018	2472837.838	4789647.059	29.69	31.09	28.13	1328513.5	4205882.35	32.23	26.49	28.13
01.11.2018	2199189.189	4082705.882	29.69	31.34	27.93	1254594.6	3648941.18	31.64	26.49	27.93
01.12.2018	1750270.27	3264235.294	29.69	29.98	27.79	1186621.6	3104705.88	29.97	26.49	27.79
01.01.2019	1829189.189	2678352.941	28.27	28.59	27.88	1169594.6	3053529.41	28.54	28.56	27.88

01.02.2019	1683648.649	2614352.941	27.87	27.97	27.16	1112297.3	2874117.65	28.17	27.79	27.16
01.03.2019	2074459.459	3075764.706	27.87	27.97	26.86	1253378.4	3128235.29	28.14	27.79	26.86
01.04.2019	2224189.189	3232000	27.87	27.97	26.81	1310540.5	3266235.29	28.37	27.92	26.81
01.05.2019	2551486.486	3221176.471	27.87	27.97	26.38	1394459.5	3330235.29	29.86	29.46	26.38
01.06.2019	2889189.189	3572823.529	27.87	27.97	26.50	1401756.8	3428470.59	29.37	28.58	26.50
01.07.2019	3063243.243	4461764.706	27.87	27.97	25.75	1248648.6	3351529.41	28.49	27.61	25.75
01.08.2019	3283918.919	4080000	27.87	27.97	25.25	1378108.1	3672000	28.11	27.35	25.25
01.09.2019	2656756.757	3953294.118	27.87	27.97	24.77	1358378.4	3484941.18	28.05	27.39	24.77
01.10.2019	2665135.135	4194470.588	27.87	27.97	24.81	1326891.9	3765529.41	28.12	27.39	24.81
01.11.2019	2486621.622	3964941.176	27.87	27.97	24.37	1275135.1	3505411.76	27.97	27.25	24.37
01.12.2019	2355810.811	3626823.529	27.87	27.97	23.61	1265270.3	3342000	27.68	26.99	23.61
01.01.2020	2287027.027	3826000	27.51	27.34	24.12	1205540.5	3292941.18	27.68	26.99	24.12
01.02.2020	2196756.757	4033294.118	26.56	26.08	24.60	1274189.2	3465294.12	26.68	25.66	24.60
01.03.2020	2381486.486	4304352.941	26.22	25.78	26.41	1224594.6	3111294.12	26.23	25.05	26.41
01.04.2020	2101891.892	3352117.647	25.91	25.39	27.22	928243.24	2210823.53	26.12	25.05	27.22
01.05.2020	2751081.081	3592705.882	23.36	23.04	26.81	1162973	2438000	23.87	22.97	26.81
01.06.2020	3237972.973	4320705.882	21.1	20.87	26.71	1300945.9	2876117.65	21.67	20.69	26.71
01.07.2020	3663378.378	5065764.706	21.05	20.63	27.31	1733783.8	3741764.71	21.64	20.65	27.31
01.08.2020	4014864.865	5264000	22.08	21.61	27.52	1774864.9	3902823.53	22.2	21.25	27.52
01.09.2020	3244189.189	4962588.235	22.98	22.4	27.98	1564729.7	3697294.12	22.47	21.42	27.98
01.10.2020	3334459.459	5549176.471	22.86	22.37	28.32	1546891.9	3951176.47	22.87	21.44	28.32
01.11.2020	2941891.892	4713411.765	22.99	22.37	28.31	1409729.7	3692352.94	22.97	21.57	28.31
01.12.2020	2755135.135	4809647.059	27.87	27.97	28.17	1442702.7	3895294.12	27.68	26.99	28.17
01.01.2021	2759189.189	4080352.941	25.02	24.72	28.22	1472027	3548588.24	24.64	23.3	28.22

	Варіант 25					Варіант 26				
	Чернігівська					м. Київ				
	sales gasoline	sales diesel	price gasoline	prise diesel	exchange	sales gasoline	sales diesel	price gasoline	prise diesel	exchange
01.01.2017	801081.0811	1183882.353	22.22	20.09	27.15	11711081	6097647.06	24.36	22.04	27.15
01.02.2017	801325.0832	1183313.325	22.68	20.98	27.03	11708234	6096393.3	24.86	22.96	27.03
01.03.2017	742027.027	1159647.059	22.68	20.98	27.00	12586351	6681647.06	24.68	22.7	27.00
01.04.2017	1218243.243	1656000	22.68	20.98	26.86	13356892	6936352.94	24.85	22.79	26.86
01.05.2017	1202297.297	1466000	22.68	20.98	26.42	13739730	7229764.71	25.15	23.26	26.42
01.06.2017	1237702.703	1330117.647	22.68	20.98	26.11	14038243	7254117.65	24.84	22.62	26.11
01.07.2017	1280945.946	1392235.294	22.68	20.98	25.97	13434595	7034941.18	24.68	22.18	25.97
01.08.2017	1336216.216	1509529.412	22.68	20.98	25.64	15386486	8014588.24	24.88	22.23	25.64
01.09.2017	1442702.703	1760000	22.68	20.98	26.11	15339324	8284941.18	25.35	22.59	26.11
01.10.2017	1323918.919	1790117.647	22.68	20.98	26.65	15219865	8526117.65	27.16	24.29	26.65
01.11.2017	1045000	1578941.176	22.68	20.98	26.71	14267703	8188470.59	28.14	25.7	26.71
01.12.2017	965540.5405	1219294.118	22.68	20.98	27.52	13234730	7939764.71	28.93	26.33	27.52
01.01.2018	915135.1351	1186235.294	27.09	27.69	28.43	10843378	6268470.59	30.43	27.69	28.43
01.02.2018	769459.4595	1007411.765	27.98	28.66	27.17	10214189	5992705.88	31.1	28.66	27.17
01.03.2018	819594.5946	1108941.176	27.98	27.7	26.34	10192162	5979647.06	30.12	27.7	26.34
01.04.2018	1134459.459	1452941.176	27.98	27.49	26.15	14159324	8176235.29	29.98	27.49	26.15
01.05.2018	1121486.486	1585294.118	27.98	27.7	26.18	11073784	6869176.47	30.32	27.7	26.18
01.06.2018	1090675.676	1431294.118	27.98	28.08	26.20	11015135	6893764.71	30.57	28.08	26.20
01.07.2018	975540.5405	1346470.588	27.98	27.71	26.40	10753378	6778235.29	30.2	27.71	26.40
01.08.2018	1109594.595	1517647.059	27.98	28.18	27.48	11210946	7024941.18	30.65	28.18	27.48
01.09.2018	588783.7838	883176.4706	27.98	29.75	28.19	10747838	7172000	32.44	29.75	28.19
01.10.2018	864729.7297	1481411.765	27.98	32.06	28.13	13628514	8696588.24	33.89	32.06	28.13
01.11.2018	790270.2703	1377764.706	27.98	32.3	27.93	14130946	9271294.12	33.01	32.3	27.93
01.12.2018	709729.7297	1041647.059	27.98	30.47	27.79	13148108	8617529.41	30.88	30.47	27.79
01.01.2019	648378.3784	955058.8235	26.94	27.21	27.88	11490000	7900823.53	29.09	29.06	27.88

01.02.2019	581621.6216	825411.7647	26.45	26.78	27.16	11741486	7283294.12	28.93	28.58	27.16
01.03.2019	758243.2432	988823.5294	26.45	26.78	26.86	14261351	8513882.35	28.91	28.61	26.86
01.04.2019	860000	1047294.118	26.45	26.78	26.81	14546892	8701294.12	29.64	28.84	26.81
01.05.2019	922027.027	979294.1176	26.45	26.78	26.38	14510405	8723411.76	30.45	29.52	26.38
01.06.2019	1004459.459	1064117.647	26.45	26.78	26.50	15441622	9085882.35	30.51	29.34	26.50
01.07.2019	1068918.919	1203882.353	26.45	26.78	25.75	14450270	8708117.65	29.88	28.63	25.75
01.08.2019	1180405.405	1261882.353	26.45	26.78	25.25	15012703	8809764.71	29.69	28.32	25.25
01.09.2019	978243.2432	1194941.176	26.45	26.78	24.77	15416757	9136941.18	29.61	28.3	24.77
01.10.2019	972972.973	1309764.706	26.45	26.78	24.81	15789324	9541058.82	29.61	28.27	24.81
01.11.2019	915675.6757	1053294.118	26.45	26.78	24.37	16123784	9816117.65	29.59	28.14	24.37
01.12.2019	873513.5135	943176.4706	26.45	26.78	23.61	15740811	9790000	28.81	27.76	23.61
01.01.2020	853783.7838	1123058.824	25.89	25.74	24.12	14627703	9222705.88	28.81	27.76	24.12
01.02.2020	815675.6757	1043176.471	25.03	24.76	24.60	16645541	10410705.9	27.38	26.3	24.60
01.03.2020	896486.4865	1212470.588	24.29	23.69	26.41	15694595	9947294.12	27.16	26.01	26.41
01.04.2020	874054.0541	1190117.647	23.84	23.15	27.22	11871081	7472000	26.77	25.39	27.22
01.05.2020	1098783.784	1371764.706	20.99	20.73	26.81	15317568	8953764.71	24.43	23.23	26.81
01.06.2020	1262027.027	1702705.882	18.78	18.33	26.71	17994730	10142705.9	22.63	21.47	26.71
01.07.2020	1324864.865	1767529.412	18.71	18.2	27.31	17757838	10650941.2	22.63	21.59	27.31
01.08.2020	1370270.27	1878352.941	18.92	18.4	27.52	18330135	10627294.1	23.73	22.48	27.52
01.09.2020	1279594.595	1977647.059	19.37	18.87	27.98	19139324	11242588.2	24.65	23.29	27.98
01.10.2020	1184459.459	2076941.176	19.43	18.87	28.32	18990946	11341058.8	24.7	23.5	28.32
01.11.2020	1027567.568	1965294.118	19.64	18.87	28.31	16963649	10417058.8	24.83	23.59	28.31
01.12.2020	993783.7838	1696823.529	26.45	26.78	28.17	16706081	10832117.6	28.81	27.76	28.17
01.01.2021	1116351.351	1368235.294	21.05	21.18	28.22	15859189	9828588.24	27.23	26.05	28.22

*У разі якщо кількість осіб у групі більша за 26, 27-ий студент за списком обирає 1 варіант.

ЧАСТИНА 3

Тестовий практикум

1. Які із цих завдань можна вирішити за допомогою лінійної регресії:

а) оцінка ймовірності падіння великого метеорита на Землю;

б) розпізнавання осіб на фотографії;

в) визначення факторів, що впливають на ціну квартири;

г) встановлення ринкової ціни старої машини.

2. У більшості випадків залежності між економічними змінними є:

а) стохастичними;

б) функціональними;

в) гетероскедастичними;

г) мультиколінеарними.

3. Просторовими даними є:

а) дані за декількома об'єктами за один і той саме період або момент часу;

б) дані, що характеризують один і той саме об'єкт, але за різні періоди або моменти часу;

в) дані за декількома змінними за різні періоди або моменти часу;

г) правильна відповідь відсутня.

4. Часовими рядами є:

а) дані за декількома об'єктами за один і той саме період або момент часу;

б) дані, що характеризують один і той саме об'єкт, але за різні періоди або моменти часу;

в) дані за декількома змінними за різні періоди або моменти часу;

г) правильна відповідь відсутня.

5. Вибір структури моделі на основі апріорної інформації називається:

а) ідентифікацією;

б) верифікацією;

в) специфікацією;

г) валідацією.

6. Аналітичний метод підбору виду рівняння регресії заснований на:

а) вивченні природи зв'язку ознак;

б) вивченні поля кореляції;

в) порівнянні величини залишкової дисперсії за різними моделями;

г) правильна відповідь відсутня.

7. Графічний метод підбору виду рівняння регресії заснований на:

а) вивченні поля кореляції;

б) вивченні природи зв'язку ознак;

в) порівнянні величини залишкової дисперсії за різними моделями;

г) правильна відповідь відсутня.

8. Експериментальний метод підбору виду рівняння регресії заснований на:

а) порівнянні величини залишкової дисперсії за різними моделями;

б) вивченні поля кореляції;

в) вивченні природи зв'язку ознак;

г) правильна відповідь відсутня.

9. До якого типу моделей, що класифікуються за критерієм пояснюваної здібності, належить економетрична модель роздрібного товарообігу, що характеризує залежність між роздрібним товарообігом та доходами населення:

а) казуальна;

б) нсказуальна;

в) часова;

г) дисперсійна.

10. До якого типу моделей, що класифікуються за часовим критерієм, належить економетрична модель рентабельності виробництва восьми підприємств 2019 р., яка характеризує залежність між рентабельністю виробництва і продуктивністю праці:

а) статична;

б) динамічна;

в) просторова;

г) візуальна.

11. До вимог, що накладаються на дані для кореляційно-регресійного аналізу, відносять:

а) кількість спостережень має бути в 7–8 разів більше ніж кількості факторів;

б) кількість спостережень має бути в 5–6 разів більше ніж число факторів;

в) кількість спостережень має бути в 2–3 рази більше ніж кількість факторів;

г) кількість спостережень має бути в 3–4 рази більше ніж кількість факторів.

12. Регресійний аналіз – метод, який дозволяє:

а) встановити залежність між двома змінними у вигляді рівняння;

б) отримати кількісну оцінку тісноти зв'язку між змінними;

в) провести класифікацію вихідних даних;

г) правильна відповідь відсутня.

13. Кореляційний аналіз – метод, який дозволяє:

а) провести класифікацію вихідних даних;

б) встановити залежність між двома змінними у вигляді рівняння;

в) отримати кількісну оцінку тісноти зв'язку між змінними;

г) правильна відповідь відсутня.

14. Економетрична модель описує:

а) стохастичні зв'язки між змінними;

б) функціональні зв'язки між змінними;

в) набір цифрових даних;

г) склад змінних.

15. Коефіцієнт кореляції є величиною:

а) безрозмірною;

б) розмірною;

в) має ту саму одиницю виміру, що і випадкова величина;

г) правильна відповідь відсутня.

16. Рівняння парної регресії доповнюється коефіцієнтом парної кореляції тому, що:

а) необхідно знати тісноту зв'язку в лінійній формі;

б) це потрібно для отримання оцінок коефіцієнтів регресії;

в) це необхідно для розрахунку величини залишкової дисперсії;

г) правильна відповідь відсутня.

17. Які з перерахованих числових характеристик використовуються для аналізу ступеня взаємозв'язку випадкових величин:

а) коваріація і коефіцієнт кореляції;

б) ймовірність;

в) коефіцієнт кореляції;

г) ймовірність і коефіцієнт кореляції.

18. Вибірковою коваріацією двох змінних називається:

а) середня величина добутку відхилень цих змінних від своїх середніх;

б) сума добутків відхилень цих змінних від своїх середніх;

в) математичне сподівання добутку відхилень цих величин від своїх середніх значень;

г) правильна відповідь відсутня.

19. Розрахований за вибіркою коефіцієнт кореляції виявився рівним –1. Це означає, що:

а) між досліджуваними змінними є функціональний лінійний обернений зв'язок;

б) між досліджуваними змінними є слабкий обернений лінійний зв'язок;

в) між досліджуваними змінними відсутній зв'язок;

г) між досліджуваними змінними є зв'язок, але він не є лінійним.

20. Якщо коефіцієнт кореляції становить 0,31:

а) рівень тісноти зв'язку між змінними є низьким;

б) рівень тісноти зв'язку між змінними є середнім;

в) рівень тісноти зв'язку між змінними є високим;

г) правильна відповідь відсутня.

21. Рівнем значущості називається:

а) ймовірність помилки;

б) сукупність значень критерію перевірки, при яких нульову гіпотезу відхиляють;

в) сукупність значень критерію перевірки, при яких нульову гіпотезу не відхиляють;

г) правильна відповідь відсутня.

22. Значення t-статистики коефіцієнта кореляції, що не перевищує критичне значення, свідчить про:

а) неправильно обчислений коефіцієнт;

б) незначущість коефіцієнта в моделі;

в) гетероскедастичність у моделі;

г) гомоскедастичність у моделі.

23. Табличне значення критерію Стьюдента залежить:

а) від рівня значимості, від кількості спостережень;

б) тільки від кількості спостережень;

в) тільки від рівня значимості;

г) правильна відповідь відсутня.

24. Рівняння регресії – це:

а) формула статистичного зв'язку між змінними;

б) рівняння тісноти зв'язку між змінними;

в) рівняння математичного сподівання випадкової величини;

г) різновид прогресії.

25. Відповідно до змісту регресії фактична величина залежної змінної складається з:

а) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишків;

б) теоретичного значення залежної змінної, знайденого за рівнянням регресії, скоригованого на величину стандартної помилки;

в) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишкової дисперсії;

г) правильна відповідь відсутня.

26. Чи відповідають теоретичним уявленням знаки і значення оцінок параметрів економетричної моделі залежності споживання (у) від доходу (х): $y = 145,65 + 0,825x$:

а) так;

б) ні;

в) частково відповідають;

г) правильна відповідь відсутня.

27. Як називається різниця між фактичним і розрахованим значеннями залежної змінної?

а) різниця;

б) білий шум;

в) залишки;

г) гетероскедастичність.

28. Класичний підхід до оцінювання коефіцієнтів регресії заснований на:

а) методі найменших квадратів;

б) графічній оцінці;

в) методі максимальної правдоподібності;

г) кореляційному аналізі.

29. Сутність методу найменших квадратів полягає у:

а) мінімізації суми квадратів відхилень фактичних значень результативної ознаки від теоретичних;

б) мінімізації суми квадратів коефіцієнтів регресії;

в) мінімізації суми квадратів значень залежної змінної;

г) правильна відповідь відсутня.

30. Оцінка параметра β показує:

а) на скільки відсотків зміниться результат у разі зміни фактора на 1 %;

б) на скільки одиниць зміниться фактор у разі зміни результату на одну одиницю;

в) на скільки одиниць зміниться результат у разі зміни фактора на одну одиницю;

г) у скільки разів зміниться результат у разі зміні фактора на одну одиницю.

31.Якщо за однією і тією самою вибіркою розраховані регресії *y* на *x* і x на *y*, то чи збігаються в цьому випадку лінії регресії:

а) так, оскільки оцінки параметрів регресії однакові;

б) так, оскільки оцінки параметрів регресії різні;

в) ні, оскільки оцінки параметрів регресії різні;

г) правильна відповідь відсутня.

32. Оцінка параметра *β* у лінійній регресії сукупного попиту на мобільні телефони (у тис. грн) за ціною (у грн) виявилася рівно –1. Це означає, що:

а) збільшення ціни на 1 грн знижує попит на мобільні телефони на одну тисячу грн;

б) збільшення ціни на 1 % знижує попит на мобільні телефони на 1 %;

в) збільшення ціни на 1 грн знижує попит на мобільні телефони на 1 %;

г) збільшення ціни на 1 % знижує попит на мобільні телефони на одну тисячу грн.

33. Для визначення значущості оцінок параметрів регресії розраховують:

а) t-статистику Стьюдента;

б) F-критерій Фішера;

в) коефіцієнт детермінації;

г) коефіцієнт кореляції.

34. Рівняння регресії є якісним, якщо:

а) t-статистика, F-статистика більші за критичні значення;

б) t-статистика, F-статистика, DW-статистика високі;

в) коефіцієнт детермінації більше 0,8;

г) правильна відповідь відсутня.

35. Як називається інтервал, в якому з наперед заданою ймовірністю міститься істинне значення параметра:

а) довірчий;

б) предикативний;

в) дескриптивний;

г) правильна відповідь відсутня.

36. Коефіцієнт еластичності показує:

а) на скільки відсотків зміниться результат у разі зміни фактора на 1 %;

б) на скільки одиниць зміниться фактор у разі зміни результату на одну одиницю;

в) на скільки одиниць зміниться результат у разі зміни фактора на 1 одиницю;

г) у скільки разів зміниться результат у разі зміни фактора на одну одиницю.

37.Пояснена сума квадратів відхилень відображає вплив на відхилення y:

а) досліджуваного фактора *x*;

б) неврахованих у моделі факторів;

в) досліджуваного фактора *x* та інших факторів;

г) правильна відповідь відсутня.

38.Залишкова сума квадратів відхилень відображає вплив на відхилення y:

а) досліджуваного фактора *x*;

б) неврахованих у моделі факторів;

в) досліджуваного фактора *x* та інших факторів;

г) правильна відповідь відсутня.

39.Залишкова сума квадратів дорівнює нулю в тому разі, коли:

а) *y* пов'язаний з x функціонально;

б) значення *y*, розраховані за рівнянням регресії, дорівнюють середньому значенню *y*;

в) уся загальна дисперсія обумовлена впливом інших факторів;

г) правильна відповідь відсутня.

40.Загальна сума квадратів відхилень збігається із залишковою, коли:

а) фактор *x* не чинить впливу на результат;

б) інші чинники не впливають на результат;

в) фактор *x* та інші фактори однак впливають на результат;

г) правильна відповідь відсутня.

41. Коефіцієнт детермінації характеризує:

а) частку «поясненої» регресією дисперсії в загальній сумі дисперсії результативної ознаки;

б) співвідношення факторної і залишкової дисперсій;

в) частку залишкової дисперсії в загальній дисперсії результативної ознаки;

г) правильна відповідь відсутня.

42. Коефіцієнт детермінації існує на інтервалі:

а) [0; 1];

б) [0; 1);

в) (0; 1];

г) (0; 1).

43. Оцінка адекватності рівняння регресії в цілому проводиться за:

а) F-критерієм Фішера;

б) коефіцієнтом детермінації;

в) стандартною помилкою регресії;

г) правильна відповідь відсутня.

44. F-критерій характеризує:

а) співвідношення факторної і залишкової дисперсій;

б) частку факторної дисперсії в загальній дисперсії результативної ознаки;

в) частку залишкової дисперсії в загальній дисперсії результативної ознаки;

г) правильна відповідь відсутня.

45. Якщо коефіцієнт детермінації дорівнює нулю, то критерій Фішера дорівнює:

а) нулю;

б) одиниці;

в) більше або дорівнює одиниці;

г) правильна відповідь відсутня.

46. За результатами бюджетного обстеження п'яти сімей, побудовано економетричну модель залежності заощаджень від доходу: y = 3 + 0,5x. Спрогнозуйте накопичення родини, що має дохід 40 тис. грн:

а) 20 003;

б) 20 000;

в) 10 000;

г) 11 000.

47. На якому етапі економетричного моделювання будують економетричну модель:

а) перевірка моделі на адекватність;

б) оцінка параметрів моделі;

в) вибір загального вигляду моделі (специфікація);

г) прогноз залежної змінної.

48. Коефіцієнт кореляції є величиною:

а) має ту саму одиницю виміру, що і випадкова величина;

б) безрозмірною;

в) розмірною;

г) правильна відповідь відсутня.

49. Відповідно до змісту регресії, фактична величина залежної змінною складається з:

а) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишкової дисперсії;

б) теоретичного значення залежної змінної, знайденого за рівнянням регресії і залишків;

в) теоретичного значення залежної змінної, знайденого за рівнянням регресії, скоригованого на величину стандартної помилки;

г) теоретичного значення незалежної змінної, знайденого за рівнянням регресії і залишків.

50. Оцінка адекватності рівняння регресії в цілому проводиться за:

а) стандартною помилкою регресії;

б) F-критерієм Фішера;

в) d-статистикою;

г) BG-статистикою.

51. Чим скоригований коефіцієнт детермінації відрізняється від звичайного:

а) скоригований коефіцієнт детермінації завжди менший, ніж звичайний;

б) скоригований коефіцієнт детермінації містить поправку на кількість ступенів свободи для отримання незміщеної оцінки дисперсії;

в) скоригований коефіцієнт детермінації більший, ніж звичайний;

г) скоригований коефіцієнт детермінації обчислюється набагато простіше, ніж звичайний.

52. Вкажіть помилкове твердження:

а) при наявності автокореляції значення коефіцієнта детермінації завжди буде істотно нижче одиниці;

б) статистика d лежить у межах від 0 до 4;

в) статистика d не використовується в авторегресійних моделях;

г) правильна відповідь відсутня.

53. Критерій Дарбіна-Уотсона застосовується для:

а) визначення статистичної значущості моделі в цілому;

б) перевірки моделі на автокореляцію залишків;

в) визначення економічної значущості моделі в цілому;

г) порівняння двох альтернативних варіантів моделі.

54. Ознакою мультиколінеарності є:

а) усі оцінки параметрів моделі статистично незначимі, а коефіцієнт детермінації – високий;

б) високе значення d;

в) високе значення F-статистики;

г) низьке значення статистики BG.

55. Тест Гольдфелда-Кванта заснований на використанні:

а) t-статистики розподілу Стьюдента;

б) F-статистики розподілу Фішера;

в) статистики Дарбіна-Уотсона;

г) коефіцієнта рангової кореляції Спірмена.

56. Економетрична модель описує:

а) набір цифрових даних;

б) стохастичні зв'язки між змінними;

в) функціональні зв'язки між змінними;

г) склад змінних.

57. Значення t-статистики коефіцієнта кореляції, що не перевищує критичне значення, свідчить про:

а) неправильно обчислений коефіцієнт;

б) незначущість коефіцієнта в моделі;

в) гетероскедастичність у моделі;

г) гомоскедастичність у моделі.

58. Для оцінки значущості оцінок параметрів регресії розраховують:

а) F-критерій Фішера;

б) коефіцієнт детермінації;

в) t-статистику Стьюдента;

г) χ-квадрат-критерій Пірсона.

59. Коефіцієнт детермінації характеризує:

а) частку «поясненої» регресією дисперсії в загальній сумі дисперсії результативної ознаки;

б) співвідношення факторної і залишкової дисперсій;

в) частку залишкової дисперсії в загальній дисперсії результативної ознаки;

г) взаємозв'язок між показниками.

60. F-критерій характеризує:

а) співвідношення факторної і залишкової дисперсій;

б) частку факторної дисперсії в загальній дисперсії результативної ознаки;

в) частку залишкової дисперсії в загальній дисперсії результативної ознаки;

г) частку залишкової дисперсії в загальній дисперсії факторної ознаки.

61. Оцінки параметрів у разі наявності автокореляції залишків:

а) незміщені й ефективні;

б) зміщені й ефективні;

в) незміщені й неефективні;

г) зміщені й неефективні.

62. Автокореляцією залишків називають:

а) наявність взаємозв'язку між ознаками-факторами;

б) значення аналітичної функції, що характеризує залежність рівнів ряду від часу;

в) значення переходу;

г) наявність взаємозв'язку між послідовними елементами часового чи просторового ряду даних.

63. Яке з тверджень правильне (стосовно гетероскедастичності):

а) гетероскедастичність проявляється через низьке значення статистики d;

б) оцінки внаслідок гетероскедастичності перестають бути спроможними;

в) оцінки і дисперсії оцінок залишаються незміщеними;

г) висновки за статистиками є ненадійними (стосовно гетероскедастичності).

64. Як десятковий розділювач у R використовується:

а) кома;

б) крапка;

в) двокрапка;

г) крапка з комою.

65. Послідовністю команд мови R, яка виконує певні дії, називається:

а) глобал енвайронмент;

б) консоль;

в) вокспейс;

г) скрипт.

66. Команда, яка використовується в R для побудови графіків:

а) lm;

б) summary;

в) qplot;

г) qt.

67. Команда, яка використовується в R для знаходження середнього:

а) qf;

б) mean;

в) predict;

г) gqtest.

68. Команда, яка використовується в R для побудови економетричної моделі:

а) lm;

б) summary;

в) qplot;

г) confint.

69. Команда, яка використовується в R для знаходження довірчих інтервалів:

а) qf;

б) predict;

в) gqtest;

г) confint.

70. Команда, яка використовується в R для знаходження точкового прогнозу:

71. а) mean;

72. б) predict;

73. в) gqtest;

74. г) confint.

75.Команда, яка використовується в R для діагностики моделі на наявність гетероскедастичності на основі тесту Гольдфельда-Квандта:

а) qf;

б) mean;

в) predict;

г) gqtest.

76.Команда, яка використовується в R для знаходження критичного значення статистики Фішера:

а) qf;

б) mean;

в) predict;

г) confint.

77.Команда, яка використовується в R для знаходження критичного значення статистики Стьюдента:

а) confint;

б) mean;

в) qt;

г) gqtest.

78.Команда, яка використовується в R для діагностики моделі на наявність автокореляції залишків за критерієм Дарбіна-Уотсона:

а) confint;

б) predict;

в) dwt;

г) gqtest.

79.Команда, яка використовується в R для діагностики моделі на наявність автокореляції залишків за критерієм Бреуша-Годфрі:

а) bgtest;

б) mean;

в) predict;

г) gqtest.

80. Команда, яка використовується в R для знаходження критичного значення χ-квадрат критерію:

а) confint;

б) predict;

в) qchisq;

г) gqtest.

81. Команда, яка використовується в R для діагностики моделі на наявність мультиколінеарності:

а) confint;

б) vif;

в) qchisq;

г) gqtest.

82. Значення експоненти оцінки у виробничій функції можна знайти в R за допомогою команди:

а) exp(coef(model)[1]);

б) exp(summary(model)[2]);

в) log(coef(model)[1]);

г) mean(coef(model)[1]).

83. Команда, яка використовується в R для знаходження кореляційної матриці ознак-факторів:

а) confint;

б) mean;

в) cor;

г) gqtest.

84. Команда, яка використовується в R для критерію крос-валідації GCV:

а) confint;

б) mean;

в) cor(X0);

г) cv.glmnet.

85. Команда, яка використовується в R для побудови рідж-регресії:

а) confint;

б) cor;

в) cv.glmnet(X0,y,alpha1);

г) glmnet.

86. Команда, яка використовується в R для знаходження коефіцієнтів моделі з мультиколінеарністю, які відповідають мінімальній λ:

а) exp(summary(model)[2]);

б) log(coef(model)[1]);

в) mean(coef(model)[1]);

г) coef(m_rr,s"cv$lambda.min").

87. Що робити, якщо при копіюванні команд та вставці у робочу зону R текст українською мовою відображається символами, які неможливо прочитати:

а) у вкладці File натиснути кнопку Reopen with encoding та обрати кодування UTF-8;

б) у вкладці File натиснути кнопку Reopen with encoding та обрати кодування BIG5;

в) у вкладці File натиснути кнопку Reopen with encoding та обрати кодування GB2312;

г) у вкладці File натиснути кнопку Reopen with encoding та обрати кодування SHIFT-JIS.

88. Ознакою помилки в R є повідомлення:

а) mistake;

б) error;

в) fault;

г) stumble.

89.Яка команда в R дозволяє викликати протокол економетричної моделі:

а) resume;

б) brief;

в) summary;

г) abstract.

ТЕРМІНОЛОГІЧНИЙ СЛОВНИК

Автокореляція – кореляція між післяпослідовними значеннями часового ряду.

Авторегресія інтегрованого змінного середнього – метод побудови моделі часового ряду для складання прогнозу.

Аналіз головних компонент – метод багатовимірного аналізу «проєктує» багатовимірну хмару на площину компонент.

Багатовимірне шкалювання, аналіз головних координат – будує подобу географічної карти на основі матриці відстаней.

Біноміальний тест – дозволяє перевірити, чи розрізняються пропорції.

Графік «стебло-з-листям» – псевдографіка, що показує частоту поширення значень у вибірці.

Графік «ящик-з-вусами» – діаграма для відображення основних характеристик однієї або декількох вибірок.

Графік порівняння квантилів – графік, що показує співвідношення квантилів в двох вибірках.

Графік-гістограма – діаграма для відображення частоти появ значень у вибірці.

Графік-коррелограма, діаграма розсіювання – графік, який показує співвідношення двох ознак у вибірці.

Дисперсійний аналіз – сімейство параметричних тестів, які використовуються для порівняння декількох вибірок.

Дисперсія – усереднена різниця між середнім значенням і всіма іншими значеннями вибірки.

Довірчий інтервал – проміжок, в якому може бути значення будь-якого параметра.

Інформаційний критерій Акаіке – критерій, що оцінює оптимальність моделі; найбільш оптимальна модель зазвичай відповідає мінімальному AIC.

Квантиль – повертає значення квантилів (за замовчуванням 0, 25, 50, 75 і 100 %) для вибірки.

Кореляційна матриця – повертає коефіцієнти кореляції для всіх пар вибірок.

Кореляційний аналіз – група методів, що слугує для визначення сили відповідності між двома вибірками.

Лінійний дискримінантний аналіз – метод багатовимірного аналізу, що дозволяє створити класифікацію на основі тренувальної вибірки.

Лінійний регресійний аналіз – досліджує характер лінійного зв'язку між об'єктами.

Матриця відмінностей, матриця відстаней – обчислює відстань (відмінність) між об'єктами набору ознак за заданим алгоритмом.

Медіана – значення, що відтинає половину впорядкованої вибірки.

Міжквартильний розкид – відстань між другим і четвертим квартилями, стійкий (робастний) метод вимірювання розкиду даних.

Непараметричний – не пов'язаний припущеннями про тип розподілу, придатний для аналізу довільно розподілених даних.

Нормальний розподіл – головний розподіл статистики, виникає, якщо довго стріляти в мішень, а йотом підрахувати всі відстані до «десятки», основа параметричних методів.

Параметричний – відповідає нормальному розподілу, придатний для аналізу нормально розподілених даних.

Розподіл – «зовнішній вигляд» даних; теоретичний розподіл вказує, як повинні виглядати дані; розподіл вибірки – як дані виглядають насправді.

Середнє арифметичне – сума всіх значень вибірки, поділена на її обсяг.

Стандартна помилка – стандартне відхилення, нормалізоване за обсягом вибірки.

Стандартне відхилення, середнє квадратичне відхилення – корінь квадратний з дисперсії.

Стовпчасті діаграми – діаграма для відображення декількох числових значень.

Тест (критерій) Вілкоксона (Уілкоксона) – використовується для порівняння медіан однієї або двох вибірок, непараметричний аналог t-тесту.

Тест (критерій) Колмогорова-Смірнова – використовується для порівняння двох розподілів, порівняння розподілу вибірки з нормальним.

Тест (критерій) Стьюдента – сімейство параметричних тестів, які використовуються для порівняння середніх однієї або двох вибірок.

Тест (критерій) Фішера, F-тест – параметричний тест, який використовується для порівняння дисперсій у двох вибірках.

Тест (критерій) хі-квадрат, критерій згоди, тест Пірсона – перевіряє, чи є відповідність між рядками і стовпцями в таблиці спряженості.

Тест (критерій) Шапіро-Вілка – тест для перевірки гіпотези про нормальність розподілу вибірки.

Тест пропорцій – перевіряє, зокрема, чи однакові пропорції.

Типи даних – різні типи даних, які використовують у статистичному аналізі.

Формули моделей – спосіб короткого опису моделі для статистичного аналізу.

СПИСОК ВИКОРИСТАНОЇ ДЖЕРЕЛ

Основна

1. Data Mining. Теория и практика / под ред. И.Н. Брянцева. – М.: Издательская группа «БДЦ-пресс», 2006. – 208 с.

2. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А.А. и др. – 2-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2007. – 384 с.

3. Интеллектуальные средства диагностики и прогнозирования надежности авиадвигателей: монография / Дубровин В. И., Субботин С. А., Богуслаев А. В., Яценко В. К. – Запорожье: ОАО "Мотор-Сич", 2003. – 279 с.

4. Дюк В. Data Mining : учебный курс / В. Дюк, А. Самойленко. – СПб: Питер, 2001. – 368 с.

5. Базы данных. Интелектуальная обработка информации / Корнеев В. В., Гареев А. Ф., Васютин С. В., Райх В. В. – М.: Нолидж, 2000. – 352 с., ил.

6. Кричевский М. Л. Интеллектуальные методы в менеджменте. – СПб., 2005. – 304 с.

7. Майборода Р. Є. Регресія: Лінійні моделі: навчальний посібник / Р. Є. Майборода. – К.:ВПЦ «Київський університет», 2007. – 296 с.

8. Марченко О. О. Актуальні проблеми Data Mining : навчальний посібник для студентів факультету комп'ютерних наук та кібернетики / О. О. Марченко, Т. В. Россада. – Київ. – 2017. – 150 с.

9. Осипов О. Ю. Теорія ймовірностей та математична статистика : навчальний посібник для студентів фізичного факультету / О. Ю. Осипов. – Запоріжжя : ЗНУ, 2011. – 190 с.

10. Прокопенко І. Ф. Комп'ютеризація економічного аналізу (теорія, практика): навч. посіб. / І. Ф. Прокопенко, В. І. Ганін, В. В. Москаленко. – К.: ЦНЛ, 2005. – 340 с.

11. Кабаков Р. И. R в действии. Анализ и визуализация данных в программе R / пер. с англ. Полины А. Волковой. – М.: ДМК Пресс, 2014. – 588 с.

12.Ситник В. Ф. Інтелектуальний аналіз даних (дейтамайнінг) : навчальний посібник / В. Ф. Ситник, М.Т. Краснюк. – К. : КНЕУ, 2007. – 376 с.

13.Ситник В. Ф. Системи підтримки прийняття рішень: навч.посібник / В. Ф. Ситник. – К.:КНЕУ, 2003.

14.Ситник В. Ф., Краснюк М.Т. Інтелектуальний аналіз даних (дейтамайнінг): навч. посібник / В. Ф. Ситник. – К.: КНЕУ, 2007. – 376 с.

15.Чубукова И.А. Data Mining: учебное пособие / И. А. Чубукова. – М.: Интернет-университет информационных технологий: БИНОМ: Лаборатория знаний, 2006. – 382 с.

Допоміжна

16.Прикладной статистический анализ данных. Теория. Компьютерная обработка. Области применения / Алексахина С. В., Балдина А. В., Николаева А. Б., Строганова В. Ю. – М.: ПРИОР, 2002. – 688 с.

17.Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP / Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И.. – СПб.: БХВ Петербург, 2007. – 384 с.

18. Онтологічні схеми інтелектуалізації систем підтримки прийняття рішень (СППР) / В. І. В'юн, Г. Є. Кузьменко // Мат. машини і системи. – 2008. – № 3. – С. 115–120

19.Вітлінський В. В., Катуніна О. С. Методологічні аспекти моделювання розвитку та життєздатності систем і контрагентів цифрової економіки / Вітлінський В. В., Катуніна О. С. // "Проблеми економіки". – 2018. – № 1. – с. 333–341.

20.Гафіяк А. М. Автоматизована інформаційна система управління організаційно-технологічними процесами на підприємстві / А. М. Гафіяк, Р. М. Костирко // Тези 68-ї наукової конференції професорів, викладачів, наукових працівників, аспірантів та студентів університету. – Том 2. Полтава, 15 квітня – 15 травня 2016 р. – Полтава: ПолтНТУ, 2016. – С. 167–168.

21.Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – М.: Физматлит, 2006. – 816 с.

22. М'якшило О. М. Планування собівартості продукції харчового підприємства на основі аналітичних моделей OLAP-кубів [Текст] / О. М. М'якшило, О.В. Харкянен // Харчова промисловість. – 2011. – № 10–11. – С. 332–337.

23. Мастицкий С. Э. Статистический анализ и визуализация данных с помощью R / С. Э. Мастицкий, В. К. Шитиков. – М.: ДМК Пресс., 2014. – 401 с.

24. Опанасенко В. М. Два підходи до формалізації процесу проєктування проблемно-орієнтованих пристроїв / В. М. Опанасенко, О. М. Лісовий // Радіоелектронні і комп'ютерні системи. – 2010. – № 5. – С. 251–257.

25. Томашевський О. В. Комп'ютерні технології статистичної обробки даних: навч. посібник для студ. вищ. навч. закл. / О. В. Томашевський, В. П. Рисіков.–Запоріжжя: ЗНТУ, 2006. – 174 с.

26. Тюрин Ю. Н. Анализ данных на компьютере / Ю. Н. Тюрин, А. А. Макаров; под ред. В. И. Фигурнова. – М. : ИНФА-М., 2003. – 544 с.

27. Наглядная статистика. Используем R! / А. Б. Шипунов, Е. М. Балдин, П. А. Волкова, А. И. Коробейников, С. А.Назарова, С. В. Петров, В. Г. Суфиянов. – М.: ДМК Пресс, 2012. – 298 с.

Інформаційні ресурси інтернет

28. R-Studio: офіційний сайт [Електронний ресурс]. – Режим доступу: http://www.rstudio.com/

29. Аналіз даних та статистичне виведення на мові R: інтернет-курс [Електронний ресурс]. – Режим доступу: https://courses.prometheus.org.ua/courses/IRF/Stat101/2016_T3/about

30. Аналіз даних за допомогою R: дистанційний курс [Електронний ресурс] // Система дистанційної освіти Moodle. – Режим доступу: http://moodle.nusta.edu.ua/course/view.php?id=1773

Міжнародні видання

31. Adelchi Azzalini. Data analysis and data mining / Adelchi Azzalini, Bruno Scarpa. – Oxford University Press, 2012. – 278 p.

32. Gisele L. Pappa. Automating the Design of Data Mining Algorithms / Gisele L. Pappa, Alex A. Freitas. – Springer, 2010. – 187 p.

33. John K. Kruschke. Doing Bayesian Data Analysis / John Kruschke. – Elsevier, 2015. – 759 p.

34. Matloff N. The Art of R Programming: A Tour of Statistical Software Design. - No Starch Press, 2011. – 154 p.

35. Siegmund Brandt. Data Analysis / Siegmund Brandt. – Springer, 2014. – 523 p.

36. Stephane Tuffery. Data mining and statistics for decision making / Stephane Tuffery. – Wiley, 2011. – 689 p.

37. Teetor P. R Cookbook. – O'Reilly Media, 2011. – 436 p.

38. Ugarte M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. – Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. – 700 p.

39. Venables W.N., Ripley B.D. Modern applied statistics with S. / Venables W.N., Ripley B.D. – Springer, 2002. – 495 p.

40. Yanchang Zhao. Data mining applications with R / Yanchang Zhao, Yonghua Cen. – Elsevier, 2014. – 471 p.

Навчальне видання

СЕРІЯ «НА ДОПОМОГУ СТУДЕНТУ УДФСУ»

АНАЛІЗ ДАНИХ ЗА ДОПОМОГОЮ R

Навчальний посібник

Відповідальний за випуск *Д.Ф. Салахова*

Редактор *Н.І. Грицюк*

Форматування та
комп'ютерна верстка *Н.М. Шамардак*

Здано до друку 25.05.2021. Формат 60×84/16
Папір офсетний № 1. Гарнітура "Times New Roman"
Друк. арк. 18.6
Наклад 300 прим. Замовлення № 960

Підготовлено до друку Видавничо-поліграфічним центром
Університету ДФС України
08205, вул. Університетська, 31, м. Ірпінь, Київська обл., Україна

Свідоцтво про внесення суб'єкта видавничої справи
до державного реєстру видавців, виготовлювачів і
розповсюджувачів видавничої продукції
Серія ДК № 5104 від 20.05.2016

СЕРІЯ «НА ДОПОМОГУ СТУДЕНТУ УДФСУ»

Серія «На допомогу студенту УДФСУ» заснована 2016 року.

Редакційна колегія:
Шевчук О. А., д.е.н.
Топчій В. В., д.ю.н.
Краєвський В. М., д.е.н.
Куженлєв М. О., д.е.н.
Шевчук С. В., д.е.н.
Горбовий А. Ю., д.т.н.
Суліма Є. М., д.філософ.н.,член-кореспондент НАПН України
Чмелюк В. В., к.ю.н.
Малиновський І. Й., к.н.фіз.вих.
Шевчук В. А., к.ю.н.

У СЕРІЇ «ПА ДОПОМОГУ СТУДЕПТУ УДФСУ» ВИЙШЛИ ДРУКОМ:

2016

«Методичні основи спеціальної фізичної та технічної підготовки студентів за розділом «Легка атлетика»
«Самостійпа робота студепта як одпа з форм впливу па фупкціопальпу, фізичну та психологічну підготовленість»
«Організація роботи командира механізованого взводу»
«5.45-мм автомати Калашникова (АК-74, АКС-74, АК-74П, АКС-74П) та 5.45-мм ручні кулемети Калашникова (РПК-74, РПКС-74, РПК-74Н, РПКС-74Н)»
«Гранатомет підствольний ГП-25»
«Ручні гранати»
«Кулемети Калашпикова – 7.62, ПК, УЖМ, ПКТ»
«Ручний протитанковий гранатомет РПГ-7»
«9-мм пістолет Макарова (ПМ)»

2017

«Вища та прикладпа математика»
«Цивільний захист»
«Програмування мовою JAVA : практикум»
«Інформаційні системи і технології в юридичній практиці»
«Дослідження операцій : практикум»
«Чисельні методи»
«English for Students of Finance»
«Основи військової розвідки»

2018

«CASE-технології. Міждисциплінарне інформаційне моделювання»
«Економічна інформатика: практикум»
«Економічна теорія (політекономія, мікроекономіка, макроекономіка). Політекономія»
«Економічна теорія (політекономія, мікроекономіка, макроекономіка). Мікроекономіка»
«Екопомічпа теорія (політекопомія, мікроекопоміка, макроекопоміка). Макроекопоміка»
«Охорона праці»
«Екопоміка і організація діяльпості об'єдпань підприємств»
«Основи християнської культури»
«Економіка підприємства»
«Фізика»
«Трудове право України»

2019

«Основи тактичної медицини»
«Аудит»
«Збірник задач. Вища та прикладна математика»
«Міжнародні розрахунки та валютні операції»
«Підготовка озброєння механізованого взводу до бойового застосування»
«Контролінг в управлінні підприємством»
«Актуальні питання судової експертизи (у питаннях і відповідях)»
«Інноваційний менеджмент»
«Організація навчальних занять з фізичного виховання за методом колового тренування»
«Теорія судових доказів (у таблицях і схемах)»
«Статистика»
«Фінансовий менеджмент проектів і програм»
«Ручний протитанковий гранатомет РПГ-7 (РПГ-7Д)»
«Організація роботи командира механізованого відділення»
«Психологія управління»
«Deutsch und Wirtschaft»
«Основи тактичної підготовки працівників правоохоронних органів»
«Основи кінології»
«Моделювання систем»
«UML. Уніфікована мова моделювання інформаційних систем»
«Історія держави і права України»
«Економічна теорія»

2020

«Менеджмент»
«Бізнес-аналітика та моделювання»
«Безпека життєдіяльності»
«Статистичний аналіз даних»
«Хімія»
«Мікробіологія»
«Медичне право»
«Бухгалтерський облік в управлінні підприємством»
«Загальна психологія»
«Хортинг у закладах вищої освіти»
«Організація бухгалтерського обліку»
«Організація охорони державної таємниці в Україні»
«Цивільне право України (особлива частина) »
«Цивільний процес України»
«Оздоровчий туризм»
«English Grammar for Philology Students»
«Правове регулювання зовнішньоекономічної діяльності»
«Конкурентна розвідка»
«Теорія та методика бойового хортингу»
«Фітнес-технології фізичного виховання у закладах вищої освіти»

2021

«Теоретичні основи товарознавства»
«Психологія особистості»