

Іщенко О.С., PhD студентка, ORCID 0000-0003-3315-4842
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

ЗАСТОСУВАННЯ ЗАСОБІВ R STUDIO ДЛЯ АНАЛІЗУ ФАКТОРІВ ЯКІ ВПЛИВАЮТЬ НА ЕНЕРГОСПОЖИВАННЯ

Зменшення рівня енергоспоживання є одним із пріоритетних напрямів розвитку для всіх енергоспоживачів. Вибір ефективної та надійної мови програмування дозволить забезпечити якісний аналіз всіх можливих впливів, а відповідно і більш раціональне енергоспоживання споживачами та підвищити якість управління енергоспоживанням. Основною метою даної роботи є визначення засобів R Studio саме для виявлення найбільш впливових на енергоспоживання факторів. Інформаційні технології дозволяють обробляти великі бази даних, а також застосовувати математичні засоби для виявлення факторів, які впливають на рівень енергоспоживання. R Studio є відкритим джерелом інтегрованої оболонки і має зручний інтерфейс, що спрощує роботу з R. Ряд особливостей, таких як підсвічування і автоматичне завершення коду, зручна навігація по скрипту і інші, роблять R Studio привабливою для аналізу статистичних баз даних з великою кількістю змінних. У даній статті докладно описаний вміст головного вікна R Studio та його основний функціонал. На прикладі бази даних підприємства було розглянуто математичні інструменти R Studio: побудовано матрицю парних коефіцієнтів кореляції для факторних ознак, проведено кореляційний аналіз впливу факторів на енергоспоживання.

Ключові слова: R Studio, кореляційний аналіз, програмне забезпечення, енергоспоживання, бази даних, інформаційні технології.

Вступ. Для прийняття оптимальних рішень при визначенні допустимого зменшення енергоспоживання підприємства або домогосподарства широкого розповсюдження зазнав метод факторного аналізу. Знаходження залежностей між енергоспоживанням та впливаючими факторами, виявлення найбільш значущого фактору та відсіювання менш значущих для побудови регресійної моделі можна назвати першим кроком при оцінці та побудові прогнозу [1-5].

Застосування факторного аналізу в інформаційних технологіях (ІТ) дозволить впливати на прискорений процес аналізу, обробки, поширення і використання великої бази інформації, а також своєчасно приймати необхідні рішення і планувати діяльність відповідно до отриманих результатів. На сьогоднішній день інформаційні технології виступають каталізатором еволюції в світі. Тому застосування ІТ для аналізу факторів призводить до розвитку науково-технічного прогресу і до широкого використання результатів в сфері енергетики [6].

У зв'язку з цим застосування сучасних комп'ютерних технологій і розробка нових методів з їх використанням обумовлена низкою причин, серед яких:

- зростання обсягів інформації;
- складність структури даних;
- поява сучасних методів прогнозування, їх розвиток і впровадження;
- складність алгоритмів розрахунку та інтерпретації результатів;
- високі вимоги до якості прогнозів;
- необхідність використання результатів прогнозування для вирішення задач планування і управління.

Безліч рядів техніко-економічних показників, що відображають зростання підприємств і компаній, мають нелінійний характер, що ускладнює їх аналіз та прийняття управлінських рішень на підставі статистичних даних.

R - мова, що дозволяє користувачеві використовувати оператори циклів, щоб послідовно аналізувати кілька наборів даних. Також мова R надає можливість об'єднати в окрему програму різні статистичні функції для проведення більш складного аналізу. R - мова з багатьма функціями для виконання статистичного аналізу і графічного відображення результатів, які візуалізуються у власному вікні і можуть бути збережені в різних форматах. Результати статистичного аналізу можуть бути відображені на екрані, а проміжні результати (P-values, коефіцієнт регресії і т. п.) можуть бути збережені в файлі і використовуватися для подальшого аналізу.

R Studio - це вільне та відкрите джерело інтегрованого середовища розробки для R, статистичних обчислень мова для аналізу даних. Є два продукта R Studio. Один - це R Studio Server, який може бути доступним через веб-браузер з віддаленого сервера Linux. Інший продукт, який працює локально - R Studio Desktop, який доступний для Microsoft Windows, Mac OS X та Linux [7-11].

© Іщенко О.С., 2019

Мета та завдання дослідження

Метою роботи є визначення засобів R Studio для оцінки найбільш суттєвих факторів, що впливають на обсяги енергоспоживання.

Для досягнення поставленої мети були визначені наступні завдання:

1. Розгляд вмісту головного вікна та існуючих функцій в інтегрованому середовищі R Studio.
2. Оцінка колінеарності факторів за допомогою побудови матриці парних коефіцієнтів кореляції засобами R Studio.
3. Аналіз впливу факторних ознак на результативну ознаку.

Матеріали та результати досліджень

1. Розглянемо вміст головного вікна програми R Studio. У верхньому рядку знаходяться основні меню: File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help. У другому рядку знаходиться панель швидкого доступу. Перша іконка дозволяє створити файли різних типів, друга іконка також дозволяє створити проєкт в R, третя іконка, Open an existing file, дозволяє відкрити вже існуючий файл (у першій та другій іконки з правої сторони знаходиться стрілка, при натисканні на яку з'являється випадаючий список). Наступні дві іконки дозволяють зберегти поточний файл або всі відкриті файли. За допомогою іконки Print the current file можна роздрукувати поточний файл (поточний файл відображається у верхньому лівому вікні). Go to file/function дозволяє швидко здійснити перехід до необхідного файлу чи функції.

У верхній лівій чверті знаходиться вікно редактора поточних файлів. З допомогою вкладок зверху можна здійснити перехід між всіма відкритими файлами. Іконки у правому верхньому куті дозволяють згорнути або розгорнути дане вікно. Ще нижче знаходить панель для швидкого доступу у даному вікні (див. Рис. 4). Дві стрілки на початку панелі дозволяють перемикатися між різними відкритими файлами. За допомогою наступної іконки можна зберегти поточний відкритий файл. Source on Save дозволяє налаштувати синхронізацію файла зі своїм джерелом. Іконка File/Replace дозволяє знайти або замінити необхідну інформацію у файлі. Code Tools містить у собі функції для роботи з кодом. Compile Report дозволяє скласти звіт, у якому буде надано сам код, а також вихідний результат у бажаному форматі (MS Word і т.ін.). Перша іконка у лівому куті Run the current line or section запускає обраний рядок або фрагмент. Друга іконка Re-Run previous code region, запускає попередній фрагмент коду. Іконка Source the contents of the active document показує весь вміст активного документу (з правої сторони знаходиться стрілка, при натисканні на яку з'являється випадаючий список). Знизу даного вікна у правому куті відображається місцезнаходження текстового курсора, наприклад 52:1 означає, що курсор вводу стоїть у 51 рядку перед 1-м знаком. У нижньому правому куті знаходить кнопка, для вибору мови на яку можна конвертувати код в R [9, 11].

Вкладка Console знаходиться у нижній лівій чверті головного вікна. Його можна назвати найважливішим для роботи в R. У Console можна прописувати свої команди, а також продивлятися результати роботи програми. Вкладка Terminal забезпечує доступ до системної оболонки зсередини RStudio IDE (Integrated Development Environment).

У верхній правій чверті головного вікна знаходиться вкладка Environment, що містить інформацію про змінні, дані, значення, функції і т.ін. (див. Рис. 1). Дана вкладка дуже корисна коли є велика кількість змінних у пам'яті комп'ютера. На панелі швидкого доступу вкладки Environment представлений найбільш важливий функціонал. Перша іконка забезпечує завантаження робочого простору R (при зберіганні робочого простору зберігаються напрацювання поточної сесії, назви змінних, всі об'єкти, моделі і т.ін.). Друга іконка дозволяє зберегти робочий простір. За допомогою третьої іконки можна імпортувати бази даних в R. Наступна іконка дозволяє очистити робочий простір. За допомогою першої іконки у правому верхньому куті можна налаштувати як буде відображатися дане вікно: список, матриця. Наступна іконка дозволяє оновити інформацію. Наступна вкладка History відповідає вивід інформації про історію команд (див. Рис. 4). Дві перші іконки відповідають за завантаження та збереження історії відповідно. За допомогою іконки To Console можна перемістити обрану команду у вікно Console. Команда To Source копіює обрану команду до поточного файлу скрипта. За допомогою останніх двох іконок можна видалити виділені рядки або ж взагалі всі команди. Через поле вводу вікна у верхньому куті можна здійснити пошук по історії команд. Вкладка Connections дозволяє легко підключатися до різних джерел даних та досліджувати об'єкти та дані всередині з'єднання. Він призначений для роботи з різними інструментами для роботи з базами даних в R.

У нижній правій чверті знаходиться вікно з вкладками: Files (містить список файлів у обраній теці), Plots (вивід графіків), Packages (містить інформацію про всі пакети, також можна видалити або завантажити потрібний пакет), Help (відображення сторінки довідки про потрібну функцію), Viewer (використовується для перегляду локального веб-вмісту).

2. Для оцінки факторів впливу на енергоспоживання підприємства шляхом побудови регресійної моделі засобами R Studio обрано базу даних показану на Рис. 1. Статистичні дані загального споживання

електроенергії (tec) зібрані за дванадцять місяців, також обрані наступні фактори впливу: «Градусо-доба» (dd), «Кількість робочих годин» (wh), «Тариф» (et).

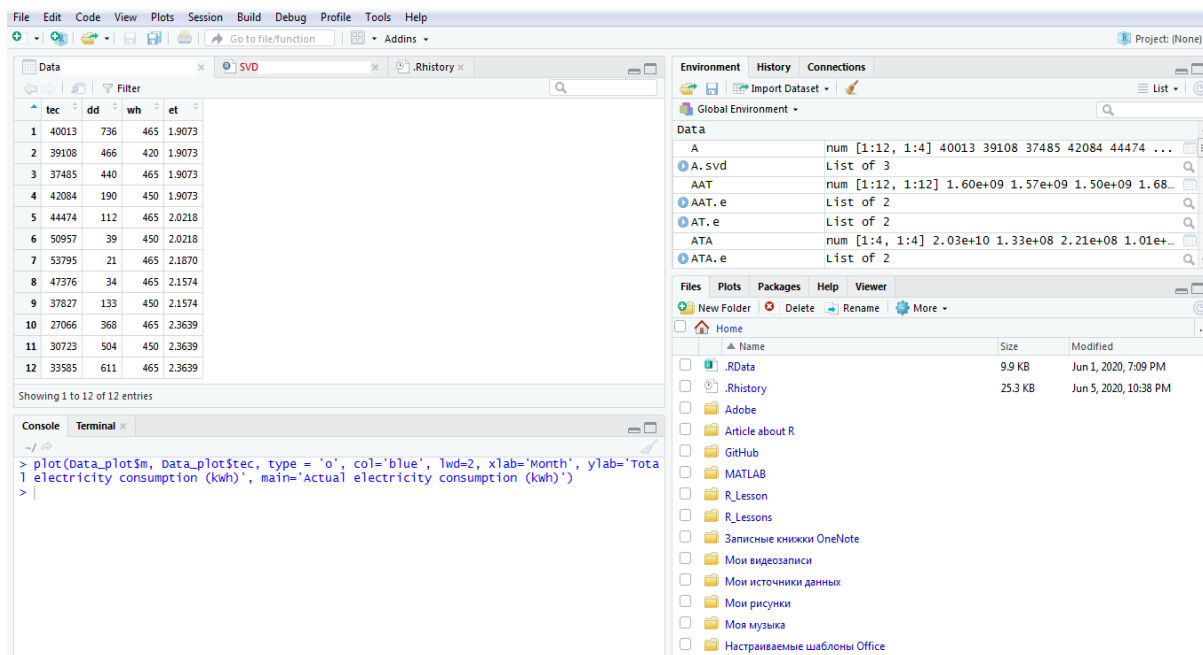


Рисунок 1 – Представлення бази даних у головному вікні R Studio

Як було сказано вище, R Studio має у своєму функціоналі можливість побудови графіків. На Рис. 2 у нижній правій чверті представлено графік загального енергоспоживання за досліджуваний період.

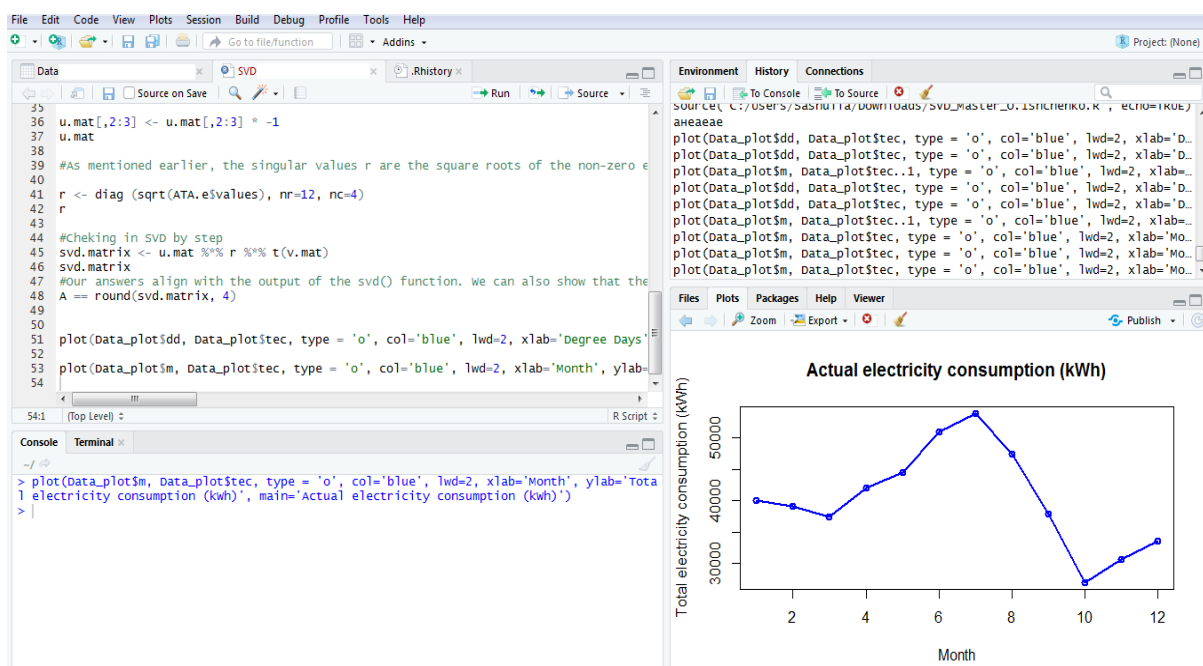


Рисунок 2 – Код та вивід даних для побудови графіка енергоспоживання підприємства за досліджуваний період

Наступним кроком після збереження бази даних у програмному забезпеченні R Studio була перевірка факторів на наявність колінеарності між ними, це забезпечить перевірку на незалежність факторів між собою, що в свою чергу впливає на побудову якісної регресійної моделі. Для оцінки наявності зв'язку між факторними ознаками було використано інструмент кореляції (cor) у R Studio [12-14].

На Рис. 3 виділено код для виводу досліджуваних даних, а потім вже для побудови матриці кореляційного зв'язку між факторами. Можна побачити, що при запиті на вивід досліджуваної бази даних виводиться лише частина – дані за 6 місяців, але це специфіка характерна для виводу досліджуваних даних при розрахунках, насправді досліджується повна база даних.

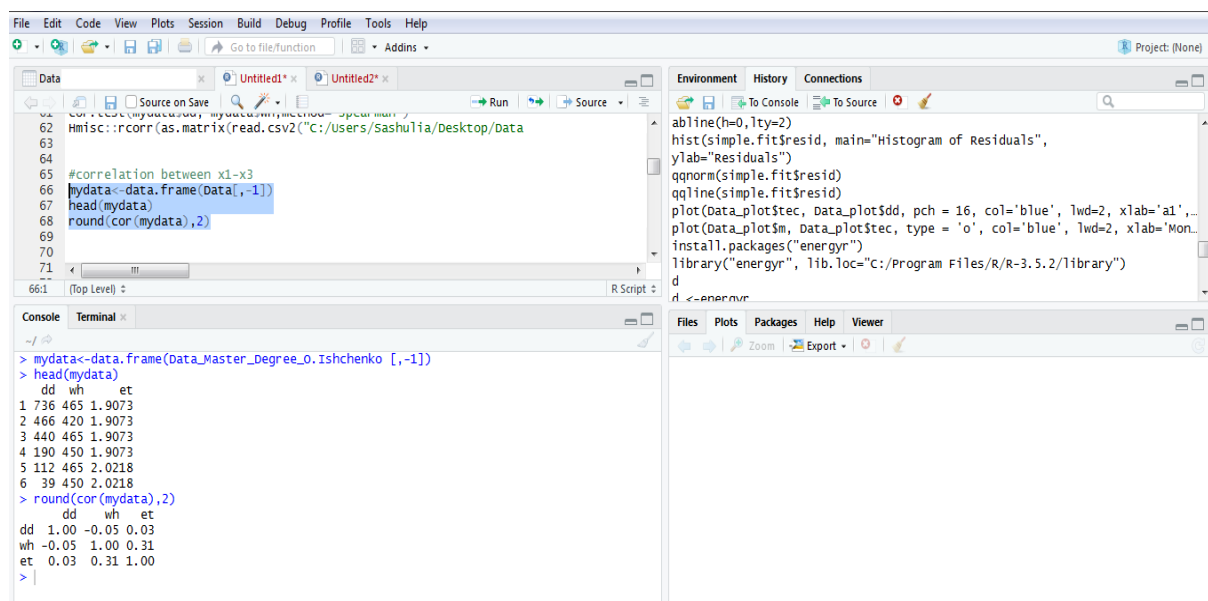


Рисунок 3 – Код та вивід даних при побудові матриці кореляційного зв'язку між факторами

Аналіз такої матриці дозволяє визначити сили зв'язку між факторними ознаками та наявність колінеарності. Для характеристики сили зв'язку прийнято використовувати шкалу, надану у Таблиці 1 [12-14].

Таблиця 1 - Характеристика сили зв'язку між ознаками

Значення коефіцієнту кореляції	0,1 – 0,3	0,3 – 0,5	0,5 – 0,7	0,7 – 0,9	0,9 – 0,99
Характеристика сили зв'язку	слабка	помірна	помітна	висока	дуже висока

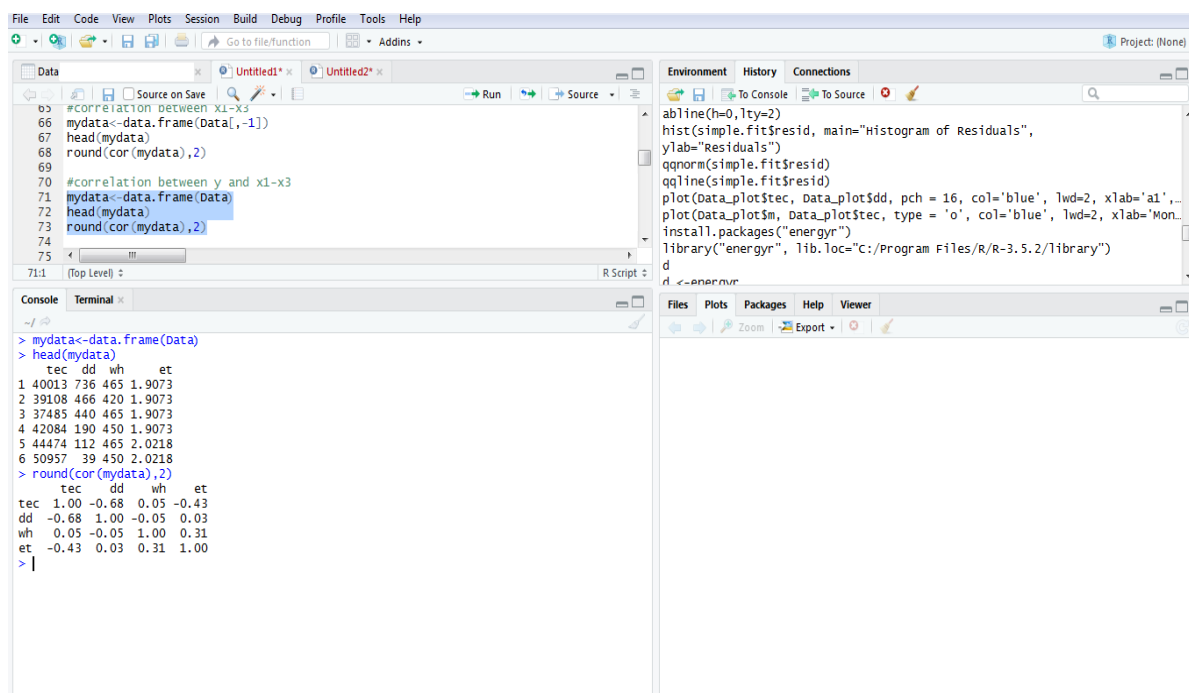


Рисунок 4 – Код та вивід даних при побудові матриці кореляційного зв'язку між результативною та факторними ознаками

Використовуючи Таблицю 1 для характеристики сили зв'язку можна визначити, що сила зв'язку між факторами «Градусо-доба» (dd) та «Кількість робочих годин» (wh) є слабкою ($r_{x_2x_1} = -0,05$), те що коефіцієнт кореляції має від'ємне значення означає те, що збільшення (зменшення) значення одного фактора призводить до закономірного зменшення (збільшення) значення іншого фактора; сила зв'язку між факторами «Градусо-доба» (dd) та «Тариф» (et) - слабка ($r_{x_3x_2} = 0,03$); сила зв'язку між факторами «Кількість робочих годин» (wh) та «Тариф» (et) - помірна ($r_{x_3x_1} = 0,31$). Отже, за результатом проведеного розрахунку для досліджуваної бази даних можна сказати, що три обрані факторні ознаки є незалежними та прийнятними для побудови енергетичної моделі підприємства.

3. Для оцінки впливу факторних ознак на результативну ознаку також було використано інструмент кореляції (cor) у R Studio (Рис. 4). Але слід зазначити, що при цьому аналізі слід звернути увагу на ті факторні ознаки, які мають найбільший вплив на результативну ознаку. В даному аналізі зв'язок повинен бути тісним, а отже значення коефіцієнта кореляції високе. З кореляційної матриці (перший стовбчик) представленої на Рис. 4 можна зробити висновок, що найбільший вплив на загальне енергоспоживання підприємства має фактор «Градусо-доба» (dd) ($r_{yx_1} = -0,68$), на другому місці – «Тариф» (et) ($r_{yx_3} = -0,43$), та мінімальний вплив на показник енергоспоживання досліджуваного підприємства має фактор «Кількість робочих годин» (wh) ($r_{yx_2} = 0,05$).

Висновки

Для ознайомлення з основними функціями і можливостями інтегрованого середовища розробки R Studio були розглянуті складові частини головного вікна та існуючий функціонал усіх вбудованих опцій.

Проведено оцінку колінераності між факторами, які впливають на показники енергоспоживання підприємства. Слабке та помірне значення коефіцієнтів кореляції доводить незалежність кожного з факторів. Матриця парних коефіцієнтів кореляції була побудована за допомогою функції «cor» у R Studio.

Також за допомогою функції «cor» у інтегрованому середовищі розробки R Studio було проаналізовано вплив трьох факторних ознак на результативну, але у цьому випадку тіснота зв'язку повинна бути максимальною. Найбільший показник коефіцієнта кореляції визначив найвпливовіший на обсяги енергоспоживання підприємства за досліджуваний період фактор – «Градусо-добу» (dd).

References

1. G.M. Huebner, I. Hamilton, Z. Chalabi, D. Shipworth, T. Oreszczyn Explaining domestic energy consumption – the comparative contribution of building factors, socio-demographics, behaviours and attitudes Appl Energy, 159 (2015), pp. 589-600
2. R.V. Jones, K.J. Lomas Determinants of high electrical energy demand in UK homes: socio-economic and dwelling characteristics Energy Build, 101 (2015), pp. 24-34
3. N. Fumo and M. A. R. Biswas, "Regression analysis for prediction of residential energy consumption", Renew. Sustain. Energy Rev., vol. 47, pp. 332-343, 2015.
4. M. G. Sobamowo, "On the Extension of Sarrus ***** Rule to Matrices: Development of New Method for the Computation of the Determinant of Matrix", Int. J. Eng. Math., vol. 2016, 2016.
5. F. Kaytez, M. C. Taplamacioglu, E. Cam and F. Hardalac, "Electrical Power and Energy Systems Forecasting electricity consumption: A comparison of regression analysis neural networks and least squares support vector machines", Int. J. Electr. Power Energy Syst., vol. 67, pp. 431-438, 2015.
6. Bublitz, Andreas, Dogan Keles & Wolf Fichtner (2016). "An analysis of the decline of electricity spot prices in Europe: Who is to blame?" Energy Policy (submitted).
7. R: data analysis and visualization. Available from: <http://r-analytics.blogspot.ru> (accessed: 10.02.2020).
8. R-FAQ. Available from: https://cran.r-project.org/doc/FAQ/R-FAQ.html#What-is-R_003f (accessed: 05.02.2020)
9. Buhovec A. G., Moskalev P. V., Bogatova V. P., Biryuchinskaya T. Y. Statisticheskii analiz dannyh v sisteme R: uchebnoe posobie [Statistical data analysis using R: the textbook]. Voronezh, 2010, 124 p.
10. Golemund G., Wickham H. R for data science. Available from: <http://r4ds.had.co.nz> (accessed: 02.02.2020).
11. Bunn A., Korpela M. R: A language and environment for statistical computing. 2013.
12. Wagner H. M. Linear programming techniques for regression analysis //Journal of the American Statistical Association. 1959. T. 54. №. 285. C. 206-212.
13. Browne, M. W., & Shapiro, A. (1986). The asymptotic covariance matrix of sample correlation coefficients under general conditions. Linear Algebra and Its Applications, 82, 169 –176. [http://dx.doi.org/10.1016/0024-3795\(86\)90150-3](http://dx.doi.org/10.1016/0024-3795(86)90150-3)
14. Cudeck, R., & O'Dell, L. L. (1994). Applications of standard error estimates in unrestricted factor analysis: Significance tests for factor loadings and correlations. Psychological Bulletin, 115, 475– 487. <http://dx.doi.org/10.1037//0033-2909.115.3.475>

O. Ishchenko, PhD student, ORCID 0000-0003-3315-4842
National Technical University of Ukraine
«Igor Sikorsky Kyiv Polytechnic Institute»

APPLICATION OF R STUDIO TOOLS FOR ANALYSIS OF FACTORS AFFECTING ENERGY CONSUMPTION

Reducing energy consumption is one of the priority areas of development for all energy consumers. The choice of an efficient and reliable programming language will provide a qualitative analysis of all possible impacts, and consequently a more rational energy consumption by consumers and improve the quality of energy management. The main purpose of this work is to determine the means of R Studio to identify the most influential factors in energy consumption. Information technology allows you to process large databases, as well as use mathematical tools to identify factors that affect the level of energy consumption. R Studio is an open source integrated shell and has a user-friendly interface that simplifies working with R. A number of features, such as backlighting and auto-completion, easy script navigation and others, make R Studio attractive for analyzing statistical databases with many variables. This article describes in detail the contents of the main R Studio window and its main functionality. On the example of the enterprise database the mathematical tools of R Studio were considered: the matrix of pair correlation coefficients for factor signs is constructed, the correlation analysis of influence of factors on energy consumption is carried out.

Key words: R Studio, correlation analysis, software, power consumption, databases, information technology.

Надійшла 17.10.2019
Received 17.10.2019