

Scientific journal
PHYSICAL AND MATHEMATICAL EDUCATION
Has been issued since 2013.

ISSN 2413-158X (online)
ISSN 2413-1571 (print)

Науковий журнал
ФІЗИКО-МАТЕМАТИЧНА ОСВІТА
Видається з 2013.

<http://fmo-journal.fizmatsspu.sumy.ua/>



Павленко Л.В., Павленко М.П., Хоменко В.Г., Хоменко С.В., Скурська М.М. Інноваційні підходи до вивчення статистики майбутніми ІТ-фахівцями на основі використання мови програмування R. Фізико-математична освіта. 2020. Випуск 1(23). С. 97-105.

Pavlenko L., Pavlenko M., Khomenko V., Khomenko S., Skurska M. Innovative approaches to the study of statistics by future IT-specialists based on the use of the programming language R. Physical and Mathematical Education. 2020. Issue 1(23). P. 97-105.

DOI 10.31110/2413-1571-2020-023-1-016
УДК 378.016:311.004.43

Л.В. Павленко¹, М.П. Павленко², В.Г. Хоменко³, С.В. Хоменко⁴, М.М. Скурська⁵
Бердянський державний педагогічний університет, Україна

¹*liliya.pavlenko@meta.ua*

ORCID: 0000-0001-7823-7399

²*pavlenko.2277@gmail.com*

ORCID: 0000-0003-0091-696X

³*v_g_homenko@ukr.net*

ORCID: 0000-0002-7361-2169

⁴*s_v_homenko@ukr.net*

ORCID: 0000-0002-9958-7272

⁵*mariyka09@ukr.net*

ORCID: 0000-0003-4572-6019

ІННОВАЦІЙНІ ПІДХОДИ ДО ВИВЧЕННЯ СТАТИСТИКИ МАЙБУТНІМИ ІТ-ФАХІВЦЯМИ НА ОСНОВІ ВИКОРИСТАННЯ МОВИ ПРОГРАМУВАННЯ R

АНОТАЦІЯ

Формулювання проблеми. У час «інформаційного вибуху» є проблема в статистичній освіті суспільства. Вивчення статистики є важливим компонентом освітніх програм підготовки фахівців у галузі ІТ. Кожного дня в світі генеруються великі обсяги різноманітних даних, що постійно збільшуються. Тому попит на ринку праці на аналітиків даних, дослідників даних постійно зростає. Удосконалення навчання студентів статистики потребує переходу від теоретичних методів навчання до практичного розв'язання завдань прикладного характеру та переміщення акценту з процесу статистичних розрахунків на аналіз та інтерпретацію одержаних результатів. Метою статті є теоретично обґрунтувати впровадження інноваційних підходів до вивчення статистики студентами на основі використання мови програмування R.

Матеріали і методи. Контент аналіз наукової й методичної літератури, узагальнення й систематизація. Анкетування студентів, первинна статистична обробка й узагальнення отриманих даних.

Результати. В роботі проаналізовані програмні засоби проведення статистичного аналізу даних, визначені особливості їх використання у навчальному процесі, запропоновано використовувати спеціалізовану мову програмування R у якості головного засобу навчання та програмні пакети MS Excel та Statistica у якості допоміжних засобів. Удосконалення курсу статистики для фахівців в галузі ІТ полягає, по-перше, в тому, що володіння математичною мовою і математичним моделюванням дозволить студенту краще орієнтуватися в прогнозуванні економічних, соціальних, технічних та інших процесів; по-друге, в тому, що статистика за своєю внутрішньою природою має багаті можливості для формування алгоритмічного мислення студентів.

Висновки. Проведене дослідження дозволило встановити, що навчання статистики має ґрунтуватися на реальних даних, які одержані в результаті статистичних досліджень. Розробка практичних та лабораторних робіт для майбутніх ІТ-фахівців має включати завдання, які будуть містити реальні дані для аналізу. У ході дослідження теоретично обґрунтовано впровадження інноваційних підходів до вивчення статистики. Визначено, що у якості головного методу навчання статистики виступає метод практичного навчання на основі програмування. Запропоновано використовувати мову та середовище програмування R, у якості головного засобу навчання.

КЛЮЧОВІ СЛОВА: мова програмування, підготовка ІТ-фахівців, аналіз даних, інноваційні підходи до навчання.

ВСТУП

Постановка проблеми. У час «інформаційного вибуху» є проблема в статистичній освіті суспільства. Вивчення статистики є важливим компонентом освітніх програм підготовки фахівців у галузі ІТ. Під час викладання курсів зі

статистики викладачі стикаються зі значною кількістю проблем, а саме: різний рівень знань, низький рівень вмотивованості, відсутність розуміння студентами необхідності вивчення статистики для використання в майбутній професії (Zieffler, A., Garfield J., Alt S., Dupuis D., Holleque K., and Chang B, 2008).

Протягом останнього часу виникає серйозне занепокоєння щодо майбутнього статистики, як навчальної дисципліни. Cox (1997), Moore (1997), Smith, Staetsky (2007) піднімають багато питань, щодо необхідності удосконалення цілей, змісту, методів та форм навчання курсу статистики.

Аналіз актуальних досліджень. Багатьма вченими проводилися дослідження, у яких вивчалися проблеми навчання статистики. Рекомендації щодо викладання статистики в навчальних закладах різного типу наведені в працях Shaughnessy (1992); Shaughnessy, Garfield, and Greer (1996); Ben-Zvi and Garfield (2004); Shaughnessy (2007); Garfield & Ben Zvi (2008); Langrall, Makar, Nilsson, & Shaughnessy (2017); Biehler, Frischmeier, Reading & Shaughnessy (2018). У дослідженнях Garfield & Ben-Zvi (2008) та Watson, Fitzallen & Carter (2013) було запропоновано використовувати перехід від теоретичного навчання до застосування статистичних методів у практичній діяльності під час навчання.

Des Nicholls (2001) зазначає, що за останні 50 років теоретичний зміст навчання статистики зазнав значного розвитку, але цей процес відбувався не керовано, шляхом введення в зміст нових понять. Як наслідок зміст навчання статистики, як дисципліни забезпечив викладачів великою кількістю теоретичних понять, спрямованих на вдосконалення теоретичної підготовки студентів. Усе це, не сприяє розвитку мотивації та зацікавленості студентів до вивчення статистики. Rumsey D. (2002), Gal I. та Garfield J. (1997) звертають увагу на проблеми навчання статистики, та пропонують змінити парадигму навчання, і спрямувати його орієнтацію на практичну галузь, ще під час навчання у ЗВО.

Інвестиції в освіту розглядаються, як інвестиції у виробництво, де створюється людський капітал. За прогнозами Всесвітнього Економічного Форуму протягом наступних п'яти років у світі буде спостерігатися збільшення попиту майже в шість разів на фахівців у галузі статистичного аналізу даних.

Відповідно до опитування Modis (New STEM IQ survey: Americans view STEM as top-paying field, yet only 35% pursuing STEM jobs, 2018) 97,44 % респондентів (представники банків та промисловості) розглядають аналіз даних, як перспективу успішного розвитку в галузі продажів та маркетингу. Проте, інтерес до інтерпретації даних вище, ніж до проведення розрахунків. 42 % опитаних скаржаться на відсутність кваліфікованих фахівців, які володіють методами статистичного аналізу даних на ринку праці. 55 % респондентів зазначають, що складно знайти фахівців, які вміють проводити розрахунки та інтерпретацію отриманих результатів.

Кожного дня у світі генеруються великі обсяги різноманітних даних, які постійно збільшуються. Тому попит на ринку праці на аналітиків даних, дослідників даних постійно зростає. Економіст Hal Varian R. (2017) зазначає, що у 2020 році аналітик даних стане найбільш популярною професією

Отже, удосконалення навчання студентів статистики потребує переходу від теоретичних методів навчання до практичного розв'язування прикладних завдань та переміщення акценту з процесу статистичних розрахунків на аналіз та інтерпретацію одержаних результатів.

Щоб почати готувати інтелектуально-активного, збагаченого знаннями і вміннями фахівця, освіта має перейти від репродуктивного до інноваційного навчання. Інноваційне навчання – це творче поєднання традиційних і нових методів навчання, їхній вибір щодо кожної навчальної дисципліни, виходячи з її теоретичного змісту і практичної спрямованості (Камінська, 2014). Водночас, варто враховувати, що під час викладання навчального матеріалу в студентів важливо формувати не тільки певні професійні компетентності, а і скорегувувати їх на сучасні вимоги сьогодення. Мається на увазі, що майбутній фахівець повинен вміти висловлювати свої думки й концепції словами, розуміти мову символів, знаків, схем. Це вже не просто здатність до творчого мислення, а і здатність до прийняття неординарних рішень і дій.

Для організації інноваційного навчання статистики відповідно до сучасних вимог сьогодення доцільно застосовувати спеціальні програмні засоби для проведення статистичного аналізу даних. Однак, сьогодні є спеціалізовані мови програмування та середовища, за допомогою яких, можливо, швидко та якісно провести аналіз даних, виконати інтерпретацію результатів та підготувати висновки й звіти у різних форматах.

Отже, існує суперечність між традиційними підходами до навчання статистики та вимогами суспільства до рівня підготовки сучасного ІТ фахівця в галузі статистичного аналізу даних, а також між теоретичною спрямованістю змісту навчання статистики та необхідністю підготовки фахівця, який володіє прикладними засобами та методами статистичного аналізу даних.

Метою статті є теоретично обґрунтувати впровадження інноваційних підходів до вивчення статистики студентами на основі використання мови програмування R.

МЕТОДИ ДОСЛІДЖЕННЯ

У процесі дослідження були використані такі основні методи: контент аналіз наукової й методичної літератури, узагальнення й систематизація, для з'ясування стану розробленості проблеми; анкетування здобувачів вищої освіти й первинна статистична обробка отриманих результатів для уточнення поточного стану досліджуваної проблеми; узагальнення теоретичних і практичних даних для обґрунтування впровадження інноваційних підходів до вивчення статистики студентами на основі використання мови програмування R.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

Процес навчання студентів статистики пов'язаний із певними труднощами: навчальний матеріал з цього курсу містить велику кількість означень та формул. Водночас студентам необхідно не просто відтворювати їх, а й розуміти сенс і вміти застосовувати на практиці. Однак, за традиційної організації навчального процесу практичні завдання відірвані від реальних економічних, соціальних та інших процесів, які відбуваються в реальному житті. Дані, які аналізуються, є узагальненими та не дозволяють повною мірою сформулювати розуміння студентами необхідності та доцільності вивчення даної дисципліни та можливості реалізувати набуті компетентності у своїй подальшій професійній діяльності.

Тому більшість студентів засвоюють статистику фрагментарно, а отже, не формуються системні знання. Крім того, переважно вербальна подача інформації підвищує стомлюваність, у результаті знижується продуктивність навчального процесу (Фіцула, 2000).

Зменшується кількість статистично освічених людей, потенційним роботодавцям важко знайти фахівця, який зможе виконувати статистичні обчислення без попереднього навчання та пояснення. Тому постає необхідність у удосконаленні змісту навчання цієї дисципліни завдяки впровадження завдань практичного характеру.

Удосконалення змісту навчання курсу статистики потребує впровадження змін у методи та засоби її навчання використовуючи інноваційні технології.

Наукові інновації, які просувають уперед науковий прогрес, охоплюють усі галузі знань. Розрізняють соціально-економічні, організаційно-управлінські, техніко-технологічні інновації. Однією з різновидів соціальних інновацій є педагогічні інновації.

Педагогічна інновація – це нововведення в галузі педагогіки, цілеспрямовані прогресивні зміни, що вносять в освітнє середовище стабільні елементи (нововведення), що поліпшують характеристики, як окремих її компонентів, так і самої освітньої системи в цілому (Рапацевич, Є. С., 2006).

Педагогічні інновації можуть здійснюватися, як із використанням власних ресурсів освітньої системи (інтенсивний шлях розвитку), так і з залученням додаткових потужностей (інвестицій) – нових засобів, обладнання, технологій, капітальних вкладень і т. ін. (екстенсивний шлях розвитку).

Казакова В. (2006) зазначає, що поєднання інтенсивного й екстенсивного шляхів розвитку педагогічних систем дозволяє здійснювати так звані "інтегровані інновації", які будуються на стику різнопланових, різнорівневих педагогічних підсистем та їх компонентів.

Основними шляхами й об'єктами інноваційних перетворень у навчанні статистики є:

- розробка концепцій і стратегій розвитку статистичної освіти (Tishkovskaya, S., & Lancaster, G. A., 2012);
- оновлення змісту навчання статистики; зміна й розробка нових технологій навчання;
- поліпшення підготовки ІТ фахівців в галузі проведення статистичного аналізу даних;
- проектування нових моделей освітнього процесу для навчання статистики;
- удосконалення моніторингу освітнього процесу та навчання студентів;
- розробка електронних засобів навчання нового покоління.

Інновації можуть здійснюватися на різних рівнях. До вищого рівня належать інновації, які впливають на всю педагогічну систему.

Куліненко Л. (2013) зазначає, що при організації інноваційної діяльності слід враховувати, що:

- інноваційні ідеї повинні бути чіткими, переконливими й адекватними реальним освітнім потребам людини й суспільства, вони повинні бути трансформовані в конкретні цілі, завдання та технології;
- інноваційна діяльність повинна морально й матеріально стимулюватися, необхідно правове забезпечення інноваційної діяльності;
- у педагогічній діяльності важливі не тільки результати, а і способи, засоби, методи їх досягнення.

Актуальні проблеми викладання статистики в сучасному ЗВО полягають у перегляді досвіду, пов'язаного з активізацією навчання. Одне з головних завдання викладача – навчити студентів здобувати потрібну інформацію самостійно, навчити їх усвідомлено опрацьовувати отриману інформацію. Для того, щоби вони могли самостійно вивчати навчальні матеріали, потрібно, щоби вони були розроблені насамперед для студентів, а не для викладачів.

Можливості дисципліни «Статистика» для фахівців у галузі ІТ, полягає, по-перше, у тому, що володіння математичною мовою й математичним моделюванням дозволить студенту краще орієнтуватися в прогнозуванні економічних, соціальних, технічних та інших процесів; по-друге, у тому, що статистика за своєю внутрішньою природою має багаті можливості для формування алгоритмічного мислення студентів.

Майбутні ІТ фахівці повинні не лише знати теоретичні основи, але і вміти застосовувати засоби автоматизації проведення статистичного аналізу. До таких засобів належать спеціалізовані статистичні пакети програм та мови програмування.

Статистичні пакети за ознаками функціональності можуть бути розділені на 3 основні групи.

1. Універсальні статистичні пакети Statistica, SPSS, Statgraphics, STATA, Stadia, SYSTAT, S-PLUS, MS Excel. Дані пакети не орієнтовані на специфічну предметну галузь і можуть застосовуватися для аналізу даних із різних галузей діяльності. Як правило, вони пропонують широкий діапазон статистичних методів і мають порівняно простий інтерфейс. З такими пакетами рекомендується працювати користувачам-початківцям, які володіють лише базовими знаннями в галузі статистики, а також досвідченим користувачам на початкових етапах роботи з даними, коли ще чітко не визначені статистичні методи, які будуть застосовуватися для вирішення того чи іншого питання. Багатофункціональність універсального пакета дозволяє провести пробний аналіз різних типів даних із використанням широкого діапазону статистичних методів. Переважна більшість існуючих універсальних пакетів мають багато спільних функціональних можливостей та схожі за складом вбудованих статистичних процедур.

2. Професійні статистичні пакети, такі, як SAS або BMDP. Професійні пакети на відміну від універсальних дозволяють працювати з надвеликими обсягами даних, застосовувати вузькоспеціалізовані методи аналізу, створювати власну систему обробки даних. Як правило, подібні пакети складні і їх не варто використовувати в навчальному процесі.

3. Спеціалізовані статистичні пакети BioStat, Datastream, Datascope та ін. призначені для статистичного аналізу в специфічних галузях діяльності, у яких застосовуються особливі методи статистичного аналізу, як правило, не представлені в універсальних пакетах.

Спеціалізовані пакети дозволяють проводити аналіз із використанням обмеженої кількості спеціалізованих статистичних методів або застосовуються у спеціалізованій предметній галузі. Як правило, з подібними статистичними пакетами працюють фахівці, добре знайомі з методами аналізу даних в тій галузі, на яку орієнтований пакет. Наприклад, статистичний пакет BioStat створений для аналізу даних в галузі біології і медицини.

Більшість існуючих статистичних пакетів мають гнучку модульну структуру, яка може поповнюватися і розширюватися завдяки призначенням для користувача модулів, які додатково закуповуються або перебувають у вільному доступі в мережі Інтернет. Подібна гнучкість дозволяє адаптувати пакети до потреб конкретного користувача.

Статистичні пакети – це всього лише інструменти для досвідченого фахівця. Якщо фахівець не володіє достатніми знаннями й компетенціями, то, навіть найдосконаліший, програмний продукт не дозволить провести якісний аналіз даних. Проте неправильно підібраний програмний засіб, який не містить потрібний набір статистичних процедур, здатний ускладнити роботу, навіть досвідченого, фахівця.

Тому, під час підготовки ІТ фахівців необхідно знайомити здобувачів вищої освіти з наявними статистичними пакетами та їхніми характеристиками, але використання спеціалізованих мов програмування, є більш близьким та зрозумілим для студентів під час проведення статистичного аналізу даних.

Для проведення статистичного аналізу даних, можливо та доцільно використовувати мови програмування R та Python.

Розглянемо особливості мови програмування R. Мова R – це потужна високорівнева об'єктно-орієнтована мова програмування й середовище для статистичних обчислень і візуалізації вихідних і розрахункових даних, яка дозволяє вирішувати багато завдань у галузі обробки даних; це безкоштовна програма з відкритим кодом (GNU GPL), призначена для роботи під управлінням поширених операційних систем (Microsoft Windows, Mac OS, Linux і Unix). Для цієї мови розроблені декілька десятків тисяч спеціалізованих модулів та утиліт. Однією з найважливіших особливостей мови програмування R є ефективна реалізація векторних операцій, що дозволяє використовувати компактний запис під час обробки великого обсягу даних. Усе це робить R ефективним засобом для отримання корисної інформації з великих обсягів різноманітних статистичних даних, у тому числі і з Big Data. Мова R є зручним і ефективним інструментом для навчання статистичного аналізу, обробки і візуалізації даних.

В галузі аналізу даних та інтерактивних науково-дослідних розрахунків із візуалізацією результатів також, можливо, використовувати мову програмування Python. Python – це об'єктно-орієнтована мова програмування з відкритим вихідним кодом. Порівняно недавня поява поліпшених бібліотек для Python (насамперед, pandas) зробило її серйозним конкурентом мови R для проведення статистичного аналізу даних. У поєднанні з перевагами Python, як універсальної мови програмування це робить її відмінним вибором для створення додатків обробки даних.

Отже, використання спеціалізованої мови програмування, як засобу навчання сприяє розвитку вмінь проведення статистичного аналізу даних та розвитку алгоритмічного мислення майбутніх ІТ фахівців.

З метою вивчення актуальності проблеми наукового дослідження було проведено анкетування серед студентів ІТ спеціальностей. Були вивчені питання, які дозволяють з'ясувати думку здобувачів вищої освіти, щодо проблеми удосконалення методики навчання статистики майбутніх ІТ фахівців.

Результати дослідження представлені у відсотках та вказують кількість позитивних відповідей на запитання. Опитування було організовано з використанням Google Forms. В дослідженні приймали участь 83 студенти, майбутні ІТ фахівці, що вивчають статистику.

Заявлений інтерес студентів до вивчення курсу статистики

У цьому блоці було задано студентам два запитання. Результати відповідей на перше запитання опитування на рис. 1. Аналіз відповідей дозволяє встановити рівень обізнаності студентів у затребуваності спеціалістів на ринку праці, які вміють проводити аналіз даних.

Аналіз відповідей студентів дозволяє зробити висновки, що більшість з опитаних респондентів, 42,17 % вважають, що фахівець з аналізу даних затребуваний на ринку праці. Це підтверджує актуальність та необхідність вивчення курсу статистики для фахівців у сфері ІТ.

У другому запитанні було уточнено, які саме спеціальності з аналізу даних студенти вважають найбільш актуальними сьогодні. Результати опитування студентів наведені на рис. 2.

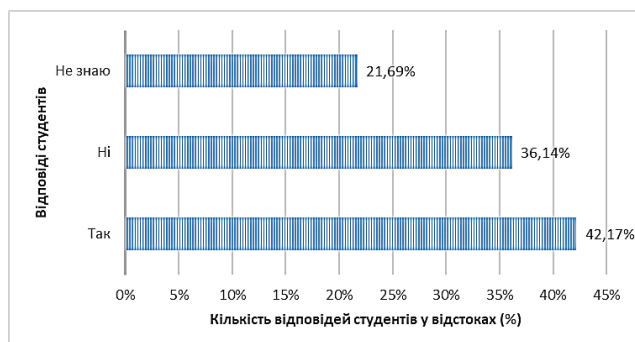


Рис. 1. Результати відповідей на запитання щодо обізнаності студентів про попит на фахівців з аналізу даних на ринку праці

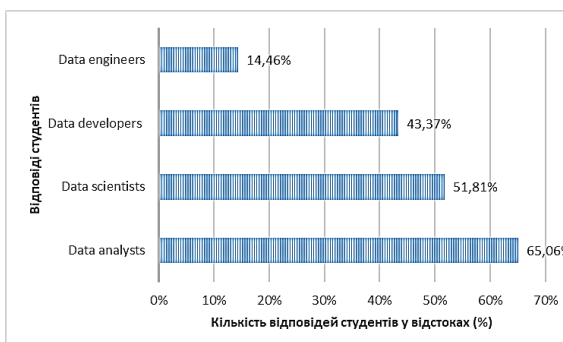


Рис. 2. Результати анкетування щодо обізнаності студентів у сучасних професіях з аналізу даних на ринку праці

Серед студентів, майбутніх програмістів, найбільш відомою є професія data analysts (65,06 %), на другому місці перебуває професія data scientists (51,81 %). Дані професії знають понад 50 % студентів, що говорить про їхню обізнаність та зацікавленість у даній галузі.

Отже, за результатами вивчення відповідей на запитання цього блоку, можемо зробити наступний висновок. Навчання статистики майбутніх ІТ фахівців є актуальним, адже студенти обізнані про існування професій у галузі аналізу даних та вважають, що статистика знадобиться їм у майбутній професійній діяльності.

Думка студентів про необхідність наповнення змісту завданнями прикладного характеру.

Студентам було запропоновано відповісти на запитання відкритого типу: «Аналіз даних, з якої предметної галузі Вам цікаво проводити?». Відповіді студентів показали, що найбільш популярними даними для опрацювання, є дані із соціології, медицини, інженерії, економіки, біології.

Також, було вивчено думку про те, з якими даними студентам цікаво працювати на практичних заняттях. Результати відповідей на запитання наведені на рис. 3.

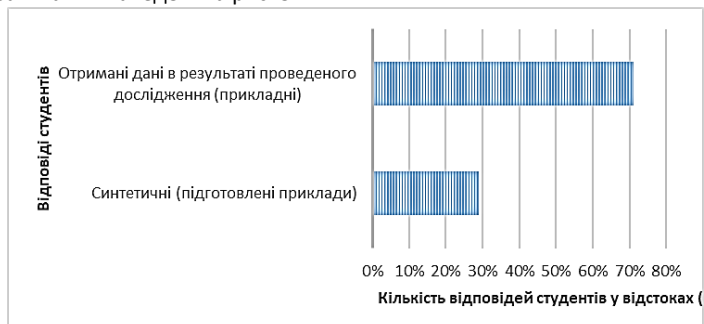


Рис. 3. Думка студентів щодо походження даних для практичних завдань

Серед опитаних респондентів 71,08 % вважають, що найбільш привабливими для них є дані, які отримані в результаті практичного дослідження та мають прикладний характер. Це говорить про необхідність розробки практичних та лабораторних робіт на основі реальних даних одержаних у результаті статистичних досліджень.

Інтерес студентів до використання мов програмування та програмних засобів для статистичного аналізу даних.

Мета третього блоку запитань полягала у вивченні думки респондентів про необхідність та доцільність застосування програмних засобів та мов програмування щодо проведенні статистичного аналізу даних.

Студентам були поставлені такі запитання: «Чи знаєте Ви мови програмування, з допомогою яких, можливо проводити статистичний аналіз даних (впишіть)?», «Інтерфейс якого програмного продукту для Вас більш зручний у використанні?», «Як Вам цікавіше проводити аналіз даних використовуючи спеціальні програмні засоби чи за допомогою мови програмування?»

Відповідно до першого запитання думки респондентів розділилися таким чином: мову програмування R вказали 55,42 %, мову програмування Python – 28,92 %. Також було вказано такі мови програмування, як C++ (9,64 %) та Java (6,02 %) (рис. 4).

Одержані результати дозволяють стверджувати, що мова R є найбільш відомою, як засіб проведення статистичного аналізу даних. Отже, для розв'язування завдань прикладного характеру будемо використовувати саме цю мову програмування.

У виборі зручності інтерфейсу програмного пакету респонденти віддали перевагу MS Excel (56,63 %), на другому місці програмний пакет Statistica (28,92 %), наступний – SPSS (14,46 %) (рис. 5).

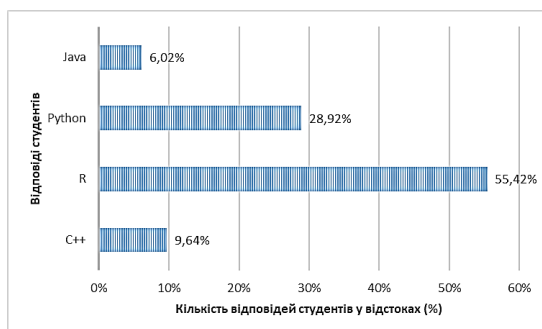


Рис. 4. Відповіді респондентів на запитання про зручність інтерфейсу програмних пакетів

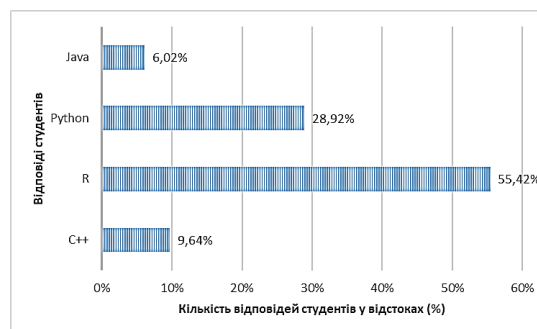


Рис. 5. Вибір програмних пакетів для статистичного аналізу даних

Отже, для проведення практичних розрахунків студентам буде запропоновано використовувати MS Excel та Statistica.

За результатами відповідей студентів на третє запитання цього блоку для організації навчання статистичного аналізу даних у якості головного засобу студентами обрано мову програмування (57,83 %) (рис. 6).

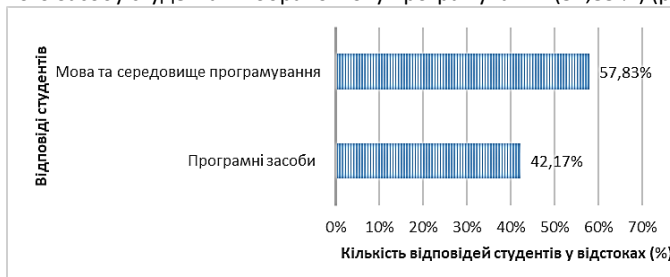


Рис. 6. Відповіді респондентів про вибір засобу для розв'язування завдань зі статистичного аналізу даних

Отже, для проведення практичних розрахунків студентам на заняттях буде запропоновано використовувати мову програмування R, як головний засіб. MS Excel та Statistica будуть використовуватися, як допоміжні засоби в проведенні статистичного аналізу.

ОБГОВОРЕННЯ

Враховавши та узагальнивши результати проведеного дослідження, на нашу думку, доцільно будувати зміст та структуру курсу з урахування побажань студентів. На практичних заняттях мають бути розглянуті завдання, які несуть реальний прикладний характер та ґрунтуються на реальних статистичних даних. Одним із головних методів навчання має виступати практичний метод навчання на основі програмування. Засобами статистичного аналізу даних на практичних заняттях можуть виступати, як програмні засоби аналізу даних (MS Excel та Statistica), так і мова та середовище програмування R.

Розглянемо приклад для навчання проведення статистичного аналізу в середовищі R. Для проведення аналізу візьмемо дані із сайту <https://abit-poisk.org.ua>, а саме дані щодо абітурієнтів за 2017 рік. На даному сайті розташовані великі обсяги даних, для нашого прикладу візьмемо лише абітурієнтів, які вступали на факультет фізико-математичної комп'ютерної та технологічної освіти Бердянського державного педагогічного університету на спеціальності «Професійна освіта (Комп'ютерні технології)» та «Професійна освіта (Енергетика)», рівень «бакалавр». Усього на ці спеціальності було подано 31 заяву. Далі проведемо аналіз цих даних, застосовуючи описові статистики в R та представлення отриманих результатів з допомогою найбільш поширених графіків в R при аналізі цих даних.

Крок 1. Задаємо прізвище, спеціальність, id, загальний бал ЗНО, статус (бюджет/контракт), після чого дані заносимо до таблиці. Задавати значення будемо у вигляді векторів командою <- c(«значення_вектору1, значення_вектору2, ...»). Таблицю з отриманих векторів будемо з допомогою команди > studentdata

```
> last_name <- c("Shvachko", "Dybiaga", "Kartashov", "Sytoenko",
"Filipenko", "Klimenko", "Vertelnik", "Diakov", "Salionov", "Bag
nuk", "Kombarov", "Baranovsky", "Kiselov", "Sakun", "Bova", "Pota
pova", "Kobzar", "Semestsov", "Cybulka", "Feylov", "Kutyshtsin", "K
artashov", "Gavrylenko", "Trotsenko", "Banchukov", "Kryshynskiy", "S
agirov", "Korobov", "Shatalina", "Trchovod", "Popov")
> specialist
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
27 28 29 30 31
> id <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,
21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31)
> zag_bal <- c(186, 184, 180, 179, 173, 173, 170, 168, 167, 166, 163, 16
2, 160, 156, 148, 145, 145, 142, 142, 140, 140, 139, 135, 131, 129, 123, 1
4, 146, 140, 136, 128)
> status <- c(1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
0, 0, 1, 1, 1, 0, 0)
> studentdata <- data.frame(id, last_name, zag_bal, status)
```

Рис. 7. Команди для створення таблиці з інформацією про абітурієнтів

Крок 2. Розрахуємо основні статистичні величини: середнє значення, медіана, стандартне квадратичне відхилення, мінімальне та максимальне значення

```
> y <- mean(zag_ba1)
> y
[1] 153
> sd <-sd(zag_ba1)
> sd
[1] 18.03145
> var <-var(zag_ba1)
> var
[1] 325.1333
> mad <-mad(x)
Error in mad(x) : object 'x' not found
> mad <-mad(zag_ba1)
> mad
[1] 22.239
> mad <-mad(zag_ba1)
> mad
[1] 22.239
> min <-min(zag_ba1)
> min
[1] 123
> max <-max(zag_ba1)
> max
[1] 186
```

Рис. 8. Результати розрахунку основних статистичних величин

За результатами проведених розрахунків отримали такі дані: середній бал абітурієнтів з ЗНО становить – 153, середня різниця між балами різних абітурієнтів складає 22 бали, найбільш низький результат (значення min) – 123 бали, найкращий результат (значення max) – 186 балів.

Крок 3. Побудуємо гістограму частот для балів ЗНО з допомогою команди `> barplot`.

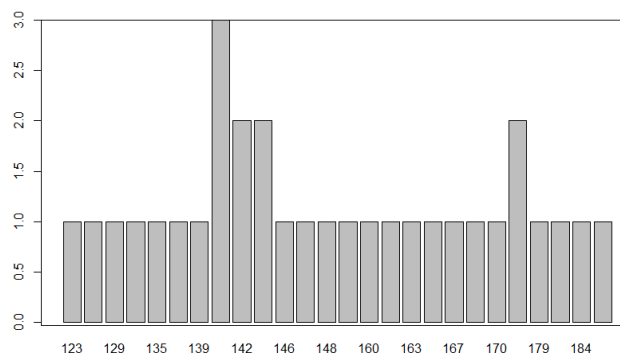
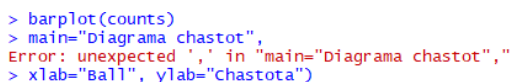


Рис. 9. Гістограма частот балів ЗНО

Гістограма частот показує, що найбільша кількість абітурієнтів має результат від 139 до 142 балів, а також те, що абсолютна більшість має унікальний бал з ЗНО, який більше не повторюється.

Крок 4. Побудуємо гістограми балів/частот із нормальною кривою розподілу. Для цього використаємо команду > box. Будувати будемо: за віссю x – параметр zag_bal, та за віссю y – частота зустрічі значення балу в таблиці.

```
> box()
> library(plotrix)
> x <- studentdata$zag_bal
> h <- hist(x, breaks = 12, col = "red",
+         xlab = "Ball ZNO",
+         main = "Histogram with normal curve and box")
> xfit <- seq(min(x), max(x), length = 40)
> yfit <- dnorm(xfit, mean = mean(x), sd = sd(x))
> yfit <- yfit * diff(h$mids[1:2]) * length(x)
> lines(xfit, yfit, col = "blue", lwd = 2)
> box()
```

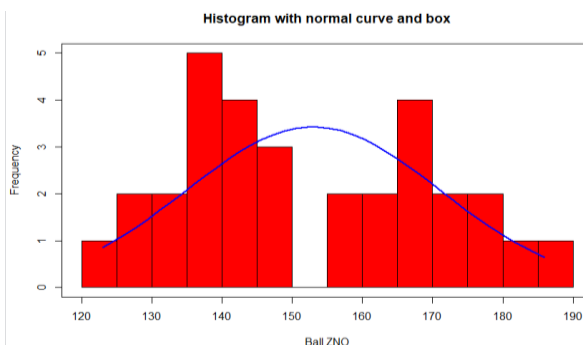


Рис. 10. Команди для побудови гістограм частот із кривою розподілу та результат виконання команд.

Гістограма розподілу показує, що дані за балами абітурієнтів не підпорядковуються нормальному закону розподілу. Маємо дуже багато «середніх» абітурієнтів, тобто тих, хто склав ЗНО від 135 до 145 балів, також вирізняються ті, хто склав на 165 балів, тобто абітурієнти з «достатнім» рівнем. Дуже мало тих, хто склав більше ніж на 180 балів.

Крок 5. Побудуємо діаграму ядерної оцінки щільності значень для балів ЗНО за допомогою команди > box.

```
> view(studentdata)
> view(studentdata)
> box()
> par(mfrow=c(2,1))
> d <- density(studentdata$zag_bal)
> plot(d)
```

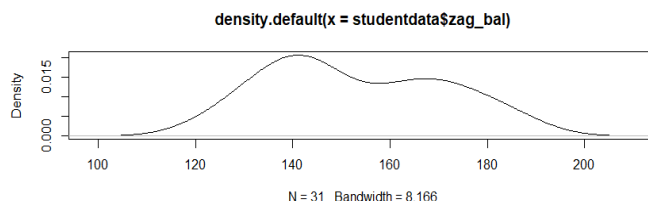


Рис. 11. Команди для побудови діаграми ядерної оцінки щільності та результат їх виконання

Діаграма ядерної оцінки щільності показує, що найбільша щільність спостерігається в проміжку від 130 до 155 балів. Тобто в цьому проміжку, виходячи з графіку, значення відрізняються на 25 балів, тоді, якщо брати повну таблицю, то вони відрізняються на 22 (див. середньоквадратичне відхилення).

У результаті розв'язування завдань прикладного характеру з використанням теоретичних знань із різних розділів статистики студенти оволодіють не лише навичками використання статистичних методів, а і сформулюють уміння проведення інтерпретації результатів та здійснювання прогнозування досліджуваних процесів. Необхідно наголосити, що використання програмування, як практичного методу навчання, дозволить студентам поліпшити свої знання та уміння в галузі програмування, використання алгоритмів та шаблонів проектування.

Використовуючи реальні дані для статистичного аналізу, студенти зможуть зрозуміти необхідність та доцільність статистичних досліджень у повсякденному житті.

Для інформаційно-методичного супроводу курсу статистики для студентів програмістів розроблено електронний засіб навчання. Розробка електронного засобу враховує вікові особливості та рівень підготовки студентів. Розроблений засіб навчання містить теоретичні матеріали, завдання до практичного виконання, наочні відеоматеріали з прикладами використання мови програмування R, довідник команд мови R та перелік рекомендованої літератури. Електронний навчальний засіб розміщений у мережі Інтернет за посиланням – <http://rstatistics.000webhostapp.com/>.

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШОГО ДОСЛІДЖЕННЯ

Проведене дослідження дозволило встановити, що навчання статистики має ґрунтуватися на реальних даних, які одержані в результаті статистичних досліджень, що є більш актуальним для студентів у порівнянні з синтетичними прикладами. Отже, розробка практичних та лабораторних робіт для майбутніх IT-фахівців має включати завдання, які будуть містити реальні дані з таких предметних галузей: соціологія, медицина, інженерія, економіка, біологія.

У ході дослідження теоретично обґрунтовано впровадження інноваційних підходів до вивчення статистики. Визначено, що у якості головного методу навчання статистики виступає метод практичного навчання на основі програмування. Отже, пропонуємо використовувати мову та середовище програмування R, у якості головного засобу навчання. Програмні пакети MS Excel та Statistica, варто використовувати у якості допоміжних засобів навчання.

У подальших дослідженнях планується розробити методику впровадження та використання мов програмування R та Python для проведення статистичного аналізу даних.

Список використаних джерел

1. Казаков В.Г. Новый час - новые технологии профессиональной подготовки. *Професійна освіта*. 2006. №1. С. 12.
2. Камінська А.В. Формування готовності майбутніх викладачів до інноваційної діяльності у вищому навчальному закладі. *Науковий вісник Донбасу*, 2011. №1. URL: <http://nvd.luguniv.edu.ua/archiv/NN13/11kavvz.pdf> (Дата звернення 15.02.2020).

3. Куліненко Л.Б. Технології інноваційного освітнього середовища. *Науковий часопис Національного педагогічного університету імені М.П.Драгоманова. Серія 7. Релігієзнавство. Культурологія. Філософія*. Київ, Видавництво НПУ імені М. П. Драгоманова, 2013. Вип. 29 (41). С. 189-196
4. *Психолого-педагогічний словарь* / сост. Е. С. Рапацевич. Минск: Современное слово, 2006. 928 с.
5. Фіцула М.М. *Педагогіка* : навч. посіб. для студентів пед. навч. закладів. Київ, 2000. 542 с.
6. Ben-Zvi D. Garfield, J. *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Springer, 2004. 423 p.
7. Biehler R., Frischmeier D., Reading C., Shaughnessy, J. M. Reasoning about data. *International Handbook of Research in Statistics Education*. Cham, Switzerland: Springer International. 2018. Pp. 139-192.
8. Cox D. R. The Current Position of Statistics: a Personal View. *International Statistical Review*. 1979. 65(3). Pp. 262-276.
9. Gal I., Garfield J. Curricular Goals and Assessment Challenges in Statistics Education. *In The Assessment Challenge in Statistics Education*. Amsterdam, The Netherlands, The International Statistical Institute, 1997. Pp. 1-13.
10. Garfield J., Ben-Zvi D. *Developing Students' Statistical Reasoning: Connecting Research and Teaching Practice*. Springer Science & Business Media, 2008. 572 p.
11. Hal Varian R. Nel 2020 il data analyst sarà la professione più ricercata. URL: <https://www.giornaledibrescia.it/rubriche/impresa-4-0/nel-2020-il-data-analyst-sar%C3%A0-la-professione-pi%C3%B9-ricerca-1.3182021>. (Дата звернення 15.02.2020).
12. Langrall C. W., Makar K., Nilsson P., Shaughnessy J. M. Teaching and learning probability and statistics: An integrated perspective. *Compendium for Research in Mathematics Education*. Reston, VA: NCTM, 2017. Pp. 490-525.
13. Moore, D. S. New Pedagogy and New Content: The Case of Statistics, *International Statistical Review*, 65(2), Pp. 123-165.
14. New STEM IQ survey: Americans view STEM as top-paying field, yet only 35% pursuing STEM jobs. URL: <https://www.modis.com/en-us/resources/employers/stem-iq-survey-2018/>. (Дата звернення 15.02.2020).
15. Nicholls D.F. Future Directions for the Teaching and Learning of Statistics at the Tertiary Level. *International Statistical Review*, 2001. 69(1). Pp. 11-15.
16. Rumsey D. Statistical Literacy as a Goal for Introductory Statistics Courses, *Journal of Statistics Education*, 2002. 10(3). URL: www.amstat.org/publications/jse/v10n3/rumsey2.html. (Дата звернення 15.02.2020).
17. Shaughnessy J. M. Research in probability and statistics: Reflections and directions. *Handbook of Research on Mathematics Teaching and Learning*, Macmillan Publishing Co, Inc, 1992. Pp. 465-494.
18. Shaughnessy J. M. Research on statistics learning and reasoning. *Second Handbook of Research on Mathematics Teaching and Learning*. Reston, VA: NCTM, 2007. Pp.957-1009.
19. Shaughnessy J.M., Garfield J., Greer B. Data Handling. *International Handbook of Mathematics Education*. Kluwer International Handbooks of Education, 1996. Vol 4. Springer, Dordrecht, Pp. 205-237.
20. Smith T. M. F., Staetsky L. The Teaching of Statistics in UK Universities. *Journal of the Royal Statistical Society*, 2007. Series A, 170, Pp.581-622. DOI: 10.1111/j.1467-985x.2007.00482.x.
21. Tishkovskaya S., Lancaster G. A. Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2). 56 p. DOI: 10.1080/10691898.2012.11889641.
22. Watson J. M., Fitzallen N., Carter P. *Top Drawer Teachers: Statistics*. Adelaide, Australia: Australian Association of Mathematics Teachers and Services Australia. 2013. URL: <http://topdrawer.aamt.edu.au/Statistics>. (Дата звернення 15.02.2020).
23. Zieffler A., Garfield J., Alt S., Dupuis D., Holleque K., Chang B. What Does Research Suggest about the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature, *Journal of Statistics Education*, 2008.16(2). <http://www.amstat.org/publications/jse/v16n2/zieffler.html>. (Дата звернення 15.02.2020). DOI: 10.1080/10691898.2008.11889566

References

1. Kazakov, V.H. (2006). Novyi chas - novi tekhnologii profesiinoi pidhotovky [New time - new technologies of vocational training]. *Profesiina osvita – Professional education*, №1, 12 [in Ukrainian].
2. Kaminska, A. V. (2011). Formuvannia hotovnosti maibutnikh vykladachiv do innovatsiinoi diialnosti u vyshchomu navchalnomu zakladi [Formation of future teachers' readiness for innovation in higher education]. *Naukovyi visnyk Donbasu - Scientific Bulletin of Donbass*, (1). Retrieved from: <http://nvd.luguniv.edu.ua/archiv/NN13/11kavvzn.pdf/> [in Ukrainian].
3. Kulinenko, L.B. (2013). Tekhnologii innovatsiinoho osvitnoho seredovyscha [Technologies of innovative educational space]. *Naukovyi chasopys Natsionalnoho pedahohichnoho universytetu imeni M.P.Drahomanova - Scientific journal of M.P.Dragomanov National Pedagogical University*. Vol 29 (41). – 189-196 [in Ukrainian].
4. Rapatsevych, E. S. (Ed.) (2006). *Psykholoho-pedahohicheskyi slovar [Psychological-pedagogical dictionary]*. Minsk: Sovremennoe slovo [in Russian].
5. Fitsula, M.M. (2000). *Pedahohika navchalnyi posibnyk dlia studentiv pedahohichnykh navchalnykh zakladiv [Pedagogy educational guide for students of pedagogical institutions]*. Kyiv [in Ukrainian].
6. Ben-Zvi, D., & Garfield, J. B. (Eds.). (2004). *The challenge of developing statistical literacy, reasoning and thinking*. Dordrecht, The Netherlands: Kluwer academic publishers [in English].
7. Biehler, R., Frischmeier, D., Reading, C., & Shaughnessy, J. M. (2018). Reasoning about data. *In International handbook of research in statistics education*. Springer, Cham [in English].
8. Cox, D. R. (1997). The current position of statistics: a personal view. *International statistical review*, 65(3) [in English].
9. Gal, I. and Garfield, J. (1997). Curricular Goals and Assessment Challenges in Statistics Education. *In The Assessment Challenge in Statistics Education*. Amsterdam, The Netherlands: The International Statistical Institute [in English].
10. Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. Springer Science & Business Media [in English].

11. Hal Varian, R (2017). *Nel 2020 il data analyst sarà la professione più ricercata*. Retrieved from: <https://www.giornaledibrescia.it/rubriche/impresa-4-0/nel-2020-il-data-analyst-sar%C3%A0-la-professione-pi%C3%B9-ricercata-1.3182021> [in Italian].
12. Langrall, C. W., Makar, K., Nilsson, P., & Shaughnessy, J. M. (2017). Teaching and learning probability and statistics: An integrated perspective. In J. Cai (Ed.), *Compendium for Research in Mathematics Education*. Reston, VA: NCTM [in English].
13. Moore, D. S. (1997). New Pedagogy and New Content: The Case of Statistics, *International Statistical Review*, 65(2) [in English].
14. New STEM IQ survey: Americans view STEM as top-paying field, yet only 35% pursuing STEM jobs. Retrieved from: <https://www.modis.com/en-us/resources/employers/stem-iq-survey-2018/> [in English].
15. Nicholl, D. F. (2001). Future directions for the teaching and learning of statistics at the tertiary level. *International Statistical Review*, 69(1), 11-15 [in English].
16. Rumsey, D. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses, *Journal of Statistics Education*, 10(3). Retrieved from: www.amstat.org/publications/jse/v10n3/rumsey2.html [in English].
17. Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. In D. Grouws (Ed.). *Handbook of Research on Mathematics Teaching and Learning* [in English].
18. Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. Lester (Ed.). *Second Handbook of Research on Mathematics Teaching and Learning*, 957-1009. Reston, VA: NCTM [in English].
19. Shaughnessy J.M., Garfield J., Greer B. (1996) Data Handling. *International Handbook of Mathematics Education.*, vol 4. Springer, Dordrecht [in English].
20. Smith, T. M. F. and Staetsky, L. (2007). The Teaching of Statistics in UK Universities, *Journal of the Royal Statistical Society, Series A*, 170, DOI: 10.1111/j.1467-985x.2007.00482.x [in English].
21. Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations and strategies for reform. *Journal of Statistics Education*, 20(2). DOI: 10.1080/10691898.2012.11889641 [in English].
22. Watson, J. M., Fitzallen, N., & Carter, P. (2013). *Top Drawer Teachers: Statistics*. Adelaide, Australia: Australian Association of Mathematics Teachers and Services Australia. Retrieved from <http://topdrawer.aamt.edu.au/Statistics> [in English].
23. Zieffler, A., Garfield J., Alt S., Dupuis D., Holleque K., and Chang B. (2008). What Does Research Suggest about the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature, *Journal of Statistics Education*, 16(2). Retrieved from: <http://www.amstat.org/publications/jse/v16n2/zieffler.html>. DOI: 10.1080/10691898.2008.11889566 [in English].

INNOVATIVE APPROACHES TO THE STUDY OF STATISTICS BY FUTURE IT-SPECIALISTS BASED ON THE USE OF THE PROGRAMMING LANGUAGE R

Liliia Pavlenko, Maksym Pavlenko, Vitalii Khomenko, Svitlana Khomenko, Mariia Skurska
Berdiansk State Pedagogical University, Ukraine

Abstract.

Formulation of the problem. Today there is a problem in statistically educated citizens. The study of statistics is an important component of educational programs for training specialists in IT. Every day, large volumes of various data are generated in the world, they are constantly increasing. Therefore, the demand in the labor market for data analysts, data researchers is constantly growing. Improving the training of students in statistics requires a transition from theoretical teaching methods to the practical solution of applied problems. It is necessary to shift the focus from the process of statistical calculations to the analysis of the results and its interpretation. The purpose of the article is the theoretical and implementation of innovative approaches to the study of statistics by students based on the use of the programming language R.

Materials and methods. Content analysis of scientific and methodological literature, generalization and systematization. Student questionnaire, primary statistical processing and synthesis of the data obtained.

Results. The paper analyzes the software tools for statistical analysis of data, identifies the features of their use in the educational process. It is proposed to use the specialized programming language R as the main learning tool and the MS Excel and Statistica software packages as auxiliary tools.

Improving the statistics course for IT specialists consists, firstly, in that knowledge of the mathematical language and mathematical modeling will allow the student to better navigate the forecasting of economic, social, technical and other processes; secondly, the fact that statistics, by their internal nature, has rich possibilities for the formation of students' algorithmic thinking.

Conclusions. The conducted research has established that training of statistics should be based on real data obtained from statistical surveys. The development of hands-on and lab work for future IT professionals should include tasks that will contain real data. In the course of the research, the introduction of innovative approaches to the study of statistics is theoretically substantiated. It is determined that the main method of teaching statistics is the method of practical training based on programming. It is suggested to use R programming language as the main learning tool.

Keywords: programming language, IT-specialists, data analysis, innovative approaches to learning.