

О.З. Гарпуль

кандидат фізико-математичних наук, доцент
Прикарпатський національний університет ім. В. Стефаника**ЗАСТОСУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ В ПРОЦЕСІ НАВЧАННЯ
ЕЛЕМЕНТІВ ТЕОРІЇ ЙМОВІРНОСТЕЙ ТА МАТЕМАТИЧНОЇ СТАТИСТИКИ СТУДЕНТІВ
СПЕЦІАЛЬНОСТЕЙ «МАТЕМАТИКА» ТА «ІНФОРМАТИКА»**

Анотація. У статті показано необхідність доповнення класичного курсу теорії ймовірностей та математичної статистики для спеціальностей «Математика» та «Інформатика» комп'ютерним практикумом, який спрямований на оволодіння студентами інструментарієм опрацювання статистичних даних з використанням засобів комп'ютерних технологій, що значно полегшує й прискорює обчислення статистичних показників, складання статистичних таблиць і побудову графіків, а також розширює можливості аналізу та наочного подання статистичних даних. На прикладі однієї лабораторної роботи в статті описані можливі шляхи набуття практичних навичок, умінь та досвіду роботи студентів в даному напрямі.

Ключові слова: теорія ймовірностей та математична статистика, комп'ютерний лабораторний практикум, функції електронних таблиць, мова програмування.

Сьогодні в нашій країні відбувається становлення нової системи освіти, орієнтованої на інтеграцію в світовий інформаційно-освітній простір. Цей процес супроводжується помітними змінами в організації процесу навчання, який повинен відповідати сучасним технічним можливостям. Сучасне суспільство ставить перед системою освіти нові завдання, пов'язані з розробленням педагогічної стратегії за умов комп'ютеризації та інформатизації всіх сфер життя суспільства.

Істотні зміни в інформаційному середовищі людини призвели до зниження ефективності використання традиційних підходів до навчання. Ці зміни пов'язані з впровадженням комп'ютерної техніки в різні сфери діяльності людини, що спричинює структурні зміни цієї діяльності. Можливості застосування комп'ютера в навчальному процесі перекривають традиційну сферу переважно алгоритмічної діяльності студентів, яка дотепер була базою формування математичної культури покоління, що підростає.

Нині важливого значення набувають проблеми інтенсифікації й оптимізації навчально-виховного процесу, активізації пізнавальної діяльності, розвитку творчого мислення. Використання сучасних інформаційних технологій навчання значною мірою сприяє розв'язуванню цих та інших завдань, які постають перед системою освіти. Науковцями розглядаються питання, пов'язані з використанням інформаційних технологій у навчанні математичних дисциплін у різних аспектах. Актуальними є питання розробки й впровадження комп'ютерного лабораторного практикуму «Розв'язування задач з теорії ймовірностей і математичної статистики з використанням програмних засобів обчислювальної техніки».

Теорія ймовірностей та математична статистика належить до математичних основ інформатики і є важливою складовою фундаментальної підготовки як фахівців з інформаційних технологій, так і педагогів. Крім того, ймовірнісно-статистичні методи використовуються в різних галузях науки, зокрема біології, медицині, моделюванні, розпізнаванні образів, прийнятті рішень, психології, педагогіці тощо. Для майбутніх учителів чи спеціалізованих фахівців (не тільки інформатики, фізики чи математики) знання методів теорії ймовірностей та математичної статистики є необхідними для здійснення аналізу даних у дослідженнях [1].

Метою розробки практикуму було сформулювати у студентів професійних знань, пов'язаних з використанням методів теорії ймовірності і математичної статистики в галузях прикладної математики та інформатики, отримати практичні навички застосування статистичних і емпіричних методів комп'ютеризації в галузі розробки експериментального дослідження програмних компонентів. Розвивати вміння створювати ефективне програмне забезпечення, на основі набутих теоретичних знань, формувати у студентів навички самостійної дослідницької роботи, що передбачає вивчення специфічних методів математичної статистики, інструментів і засобів, необхідних для виконання актуальних завдань оцінювання і прогнозування результатів різних видів діяльності. Потрібно показати також, що розвиток технічних програмних засобів обчислювальної техніки дає можливість говорити про нову концепцію в організації наукових досліджень – автоматизації експерименту.

Методи статистичного опрацювання отриманих результатів настільки різноманітні, що найчастіше досліднику дуже складно розібратися в них і вибрати адекватні поставленим завданням статистичні критерії. Цікавими є дослідження автора О.В. Гойко щодо аналізу основних програмних пакетів, придатних для статистичного опрацювання даних і можливість впровадження їх в навчальний процес [6]. У статті Т.І. Мамчич [7] наведено загальну характеристику статистичного опрацювання великих масивів даних за допомогою відповідних програмних продуктів. У роботі Д.А. Покришня та Є.Ю. Носенко [8] розглянуто з коротким описом кілька програмних продуктів різних виробників з різною методичною направленістю та наведено приклади використання математичних задач у навчанні інформатики за допомогою програмних продуктів GRAN1, Wolfram|Alpha, Microsoft Mathematics 4.0.

Шібзухов З.М. вважає, що Python, через зручний для читання синтаксис, доцільно використовувати для розв'язування математичних задач, тому що в ньому передбачено інструментарій для наукових обчислень й інженерних розрахунків, візуалізації моделей і даних, доступний на будь-якій платформі (Windows, Linux, OS X і т.д.) [10].

Чисельні розрахунки потребують багато часу, тому на сучасному етапі розвитку обчислювальної техніки актуальним постає питання розробки ефективних програм розв'язування математичних задач, за допомогою яких швидко здійснюються обчислення.

Оперативне, якісне і точне опрацювання великих масивів статистичних даних може бути виконане лише з використанням сучасних засобів обчислювальної техніки. За наявності потужних, надійних і разом з тим простих в експлуатації програмних продуктів статистичного аналізу дослідника звільняється від рутинних операцій, розширюється сфера застосування статистичних методів в різних галузях людської діяльності, що сприяє появі якісно нових можливостей статистичного аналізу і моделювання даних. Використання пакетів прикладних програм – це єдиний реальний практичний інструмент розв'язування задач багатofакторного кореляційно-регресійного аналізу в багатовимірному просторі.

Програмне забезпечення статистичних досліджень досить розвинуте. На сучасному ринку програмних продуктів пропонуються різноманітні пакети програм для статистичного опрацювання даних [2, 3]. Всесвітньо відомі статистичні пакети для комплексного опрацювання даних: BMDP, SPSS, SAS, Systat, Minitab, S-Plus, Statgraphics Statistica та інші [4, 5].

Використання згаданих пакетів програм дає змогу автоматизувати процес статистичного дослідження в таких напрямках: створення файлів даних і таблиць; групування даних; графічний аналіз даних; розрахунок варіаційних характеристик вибірових сукупностей; побудова рядів розподілу; аналіз рядів динаміки і прогнозування їх майбутніх рівнів; кореляційно-регресійний аналіз; багатомірний аналіз.

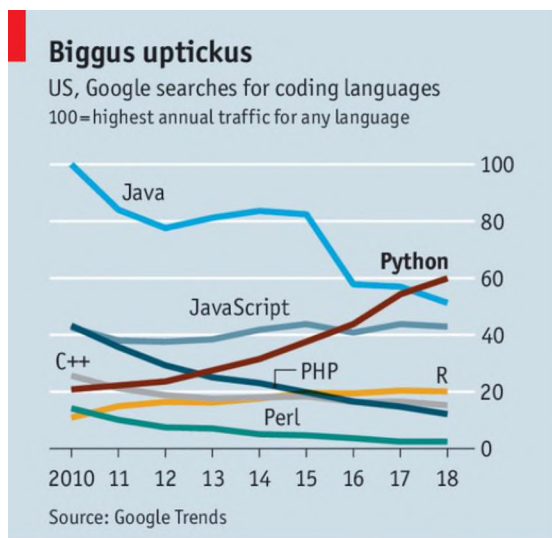
Тому вивчати можливості і володіти навичками аналізу та опрацювання статистичних даних з використанням сучасного програмного забезпечення потрібно у ВНЗ під час професійної підготовки майбутніх фахівців.

На основі аналізу навчальних планів та програм для студентів напрямів підготовки «Математика» та «Інформатика» можна зробити висновок про необхідність впровадити у навчальний предмет ВНЗ комп'ютерний практикум «Розв'язування задач з теорії ймовірностей і математичної статистики з використанням програмних засобів обчислювальної техніки», який спрямований на оволодіння студентами інструментарієм опрацювання статистичних даних на основі засобів комп'ютерних технологій, використання яких значно полегшує й прискорює обчислення статистичних показників, складання статистичних таблиць і побудову графіків, а також розширює можливості для аналізу та наочного подання статистичних даних. В якості програмного забезпечення курсу вибрана програма Microsoft Excel, яка є доволі розповсюдженою та однією з найпопулярніших сьогодні програм роботи зі статистичними даними, а також пакет статистичного аналізу STATISTICA. MS Excel – це електронна таблиця з досить потужним математичним інструментарієм, де деякі статистичні функції є просто додатково вбудованими формулами. Але з використанням MS Excel неможливо побудувати якісні наукові графіки. MS Excel зручно використовувати для накопичення даних, проміжних перетворень, попередніх статистичних обчислень, для побудови деяких видів діаграм. Проте остаточний статистичний аналіз необхідно робити з використанням програм, які спеціально створені для цих цілей. Існують макроси-доповнення для MS Excel, до яких включаються додаткові статистичні функції, які в основних випадках є достатніми для звичайного застосування.

Для досягнення результативності стосовно статистичного пакету повинні задовільнятися певні вимоги: модульність; можливість асистування при доборі способу опрацювання даних; використання простої проблемно-орієнтованої мови для формулювання завдання користувача; автоматична організація процесу опрацювання даних та зв'язків з модулями пакета; ведення банку даних

користувача і складання звіту про результати зробленого аналізу; контрольований режим роботи користувача з пакетом; сумісність з іншими програмами. Тому, для реалізації цих вимог було запропоновано студентам використати мову програмування Python.

Звичайно, обмеження щодо вибору мови програмування не ставилось, але, як показують результати опитувань серед студентів факультету, саме цю мову програмування обирає більшість. Крім того, на основі даних google trends можна зробити висновок, яка мова останнім часом випереджає у пошуку всі інші:



Мовою Python користуються 40% професійних розробників. Ще 25% хочуть користуватися нею, але поки не можуть з організаційних причин.

Але що найцікавіше – 45 мільйонів аматорів освоїли чи освоюють її тільки через сайт Code Academy, де ведеться подібна статистика.

(Python) – це потужна мова програмування, якою легко оволодіти. В ній передбачено ефективні структури даних високого рівня та простий, але ефективний підхід до об'єктно-орієнтованого програмування. Елегантний синтаксис Python, динамічне опрацювання типів, а також те, що це інтерпретована мова, робить її ідеальною для написання скриптів та швидкої розробки прикладних програм у багатьох галузях на більшості платформ [10, 11].

Python є розповсюдженою мовою програмування з великою кількістю вільно поширюваних модулів і бібліотек для створення програм майже будь-якого типу (вебдодатки, робота з базами даних, наукові обчислення, інженерні розрахунки, візуалізація даних, розробка користувацького інтерфейсу і т. д.). Статистичні задачі доцільно розв'язувати за допомоги вбудованої відкритої, вільнопоширюваної бібліотеки, описаної мовою програмування Python – SciPy.

Можна вказати такі математичні можливості використання мови програмування Python: 1) існує чотири вбудованих типи числових символів: булеві значення, цілі числа, числа з плаваючою точкою, комплексні числа; 2) підтримуються всі основні математичні операції і дії; 3) наявні математичні модулі: decimal, fractions, math, numbers, random та ін.; 4) наявні бібліотеки для складних обчислень та візуалізації даних: SymPy, Numeric, NumPy, Pandas, Matplotlib, Dislin та ін. [11].

Серед величезної кількості завдань, що зустрічаються в статистиці, доцільно у лабораторному практикумі виокремлених кілька типів, зацікавлення до яких викликане або практичним значенням задачі, або якимись цінними її властивостями, що становить особливий інтерес для дослідження (табл. 1).

Таблиця 1.

Тематика лабораторних робіт

Лабораторна робота №1	Варіаційні ряди та статистичні розподіли. Обчислення емпіричних з використанням даних статистичних засобів первинного опрацювання
Лабораторна робота №2	Міри центральної тенденції (МЦТ). Міри мінливості (ММ). Знаходження основних показників вибірки
Лабораторна робота №3	Основи статистичного опрацювання неперервних даних. Побудова гістограм для нормального та показникового розподілів вибірки
Лабораторна робота №4	Обчислення дисперсії, середнього квадратичного відхилення, коефіцієнтів асиметрії та ексцесу
Лабораторна робота №5	Основи програмування перевірки статистичних гіпотез
Лабораторна робота №6	Програмна реалізація критерію узгодженості Пірсона
Лабораторна робота №7	Проста лінійна регресія
Лабораторна робота №8	Елементи кореляційного аналізу та їх програмна реалізація

Виконання лабораторних робіт даного практикуму спрямовано на поглиблення засвоєння лекційного і опрацьованого самостійно матеріалу та набуття студентами практичних навичок опрацювання емпіричних даних. Розглядаються основи описової статистики, дискретні і неперервні розподіли ймовірностей, методи оцінювання параметрів регресійних залежностей, кореляції,

статистичні тести, найчастіше використовувані в цих галузях, а також методи планування експерименту і перевірка гіпотез, застосування емпіричних методів для аналізу продуктивності і надійності програмних систем тощо.

На лабораторних заняттях студенти розв'язують задачі щодо опрацювання даних з використанням базових функцій електронних таблиць Excel і вбудованого в Excel статистичного пакету "Аналіз даних", а також розробляють алгоритми і програми опрацювання даних за статистичними методами, реалізуючи їх у програмних середовищах [9].

Основними завданнями виконання циклу лабораторних робіт є набуття студентами знань та умінь:

- здійснювати опрацювання емпіричних даних за допомогою базових функцій електронних таблиць Excel і статистичного пакету "Аналіз даних";
- розробляти алгоритми і створювати програмні засоби для опрацювання емпіричних даних.

Виконання лабораторних робіт сприяє формуванню самостійності у аналізі проведених обчислень, дослідженні практичних задач, які є необхідною складовою підвищення технічного рівня підготовки студента.

До лабораторних робіт (комп'ютерного практикуму) дисципліни включаючи короткі теоретичні відомості, приклади розв'язування задач, індивідуальні та контрольні завдання, перелік основних теоретичних питань стосовно вказаної теми, тестування.

Під час виконання лабораторної роботи студент повинен:

- Ознайомитися з теоретичним матеріалом до теми та прикладами розв'язування задач.
- Розробити алгоритм розв'язування обраної задачі (у вигляді математичної моделі).
- Подати алгоритм через опис програми будь-якою мовою програмування (перевага надається запропонованій мові програмування).

Наведемо приклад виконання одного із завдань практикуму.

Задача. Написати програму, для генерування випадкової послідовності даних з $(N + 10)$ елементів, які набувають значень із набору $(1, 2, 3, \dots, (N+1))$. Тут N – номер студента в журналі старости. Розподіл ймовірностей рівномірний. Вивести початкову та впорядковану послідовність. Обчислити моду, медіану, середнє арифметичне сукупності, дисперсію та стандартне відхилення вибірки. Процедури обчислення кожної з цих величин мають бути реалізовані у вигляді окремого блоку програми (функції/процедури/класу/методу класу залежно від використовуваного середовища для виконання робіт).

Математична модель задачі реалізується через визначення формули для обчислення статистичних величин, таких як:

Мода M_o – це значення, яке найчастіше трапляється серед емпіричних даних.

Медіана M_d – це значення, яке знаходиться в середині упорядкованої послідовності емпіричних даних.

Середнє арифметичне сукупності n значень дорівнює $\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n}$. (1)

Дисперсія вибірки обсягом n визначається як: $s_x^2 = \frac{\sum(x_i - \bar{x})^2}{n-1}$, (2), де \bar{x} – середнє арифметичне вибірки.

Стандартне відхилення вибірки визначається як $s_x = \sqrt{s_x^2}$. (3)

Розрахунки показників МЦТ та ММ можна здійснити з використанням MS Excel за трьома способами:

- виконання математичних операцій за відповідними формулами МЦТ і ММ;
- використання вбудованих статистичних функцій:

Обсяг вибірки	=СЧЁТ()	Дисперсія	=ДИСП()
Середнє	=СРЗНАЧ()	Ст. відхилення	=СТАНДОТКЛОН()
Мода	=МОДА()	Асиметрія	=СКОС()
Медіана	=МЕДИАНА()	Експес	=ЭКССЕСС()

- спеціального розділу «Описова статистика» пакету «Аналіз даних»: виконати команди головного меню Excel Аналіз даних із закладки Дані та викликати допоміжне вікно «Описова статистика». Встановити в описовому вікні «Описова статистика» вхідні дані та параметри виведення, отримати результат.

Для розв'язування задачі засобами мови Пайтон використовують такі функції:

mean(data)	для визначення середнього значення масиву даних	statistics.mean([11, 2, 13, 14, 44]) # returns 16.8
mode()	обчислення моди	statistics.mode(data_points)
median()	медіана (середнє значення) масиву даних	statistics.median(data_points)
stdev(data, xBar=None)	стандартне відхилення вибірки	statistics.stdev(data)
variance()	дисперсія даних	variance(data)
correlation(a, v)	обчислення коефіцієнта кореляції	
random	генерування широкого спектру випадкових чисел	random.seed([x])

Програмний код для визначення середнього значення і медіани, для прикладу, буде мати вигляд:

```

01 import statistics
02 from fractions import Fraction as F
03 from decimal import Decimal as D
04
05 statistics.mean([11, 2, 13, 14, 44])
06 # returns 16.8
07
08 statistics.mean([F(8, 10), F(11, 20), F(2, 5), F(28, 5)])
09 # returns Fraction(147, 80)
10
11 statistics.mean([D("1.5"), D("5.75"), D("10.625"), D("2.375")])
12 # returns Decimal('5.0625')
```

```

01 import random
02 import statistics
03
04 data_points = [ random.randint(1, 100) for x in range(1,50) ]
05 statistics.median(data_points)
06 # returns 53
07
08 data_points = [ random.randint(1, 100) for x in range(1,51) ]
09 statistics.median(data_points)
10 # returns 51.0
11
12 data_points = [ random.randint(1, 100) for x in range(1,51) ]
13 statistics.median(data_points)
14 # returns 49.0
15
16 data_points = [ random.randint(1, 100) for x in range(1,51) ]
17 statistics.median_low(data_points)
18 # returns 50
19
20 statistics.median_high(data_points)
21 # returns 52
22
23 statistics.median(data_points)
24 # returns 51.0
```

На основі проведеного аналізу щодо впровадження комп'ютерного практикуму можна стверджувати, що виконання лабораторних робіт сприяє набуттю студентами навичок володіння інструментарієм опрацювання статистичних даних за допомогою засобів комп'ютерних технологій, використання яких значно полегшує й прискорює обчислення, формує самостійність у аналізі проведених обчислень, дослідженні практичних задач, які є необхідною складовою підвищення технічного рівня підготовки студента.

У перспективах даного дослідження для ефективної роботи планується розробка сайту-навігатора для підтримки дистанційного навчання даного курсу та організація навчального процесу через віртуальні класи.

Список використаних джерел:

1. Кобильник Т.П. Використання web-сервісу wolfram|alpha для розв'язування задач з теорії ймовірностей. Information Technologies in Education. 2015. № 24.
2. Василенко Ж.В. Програмное обеспечение по статистическому анализу данных. Методология сравнительного анализа. URL: http://www.giac.unibel.by/sm_full.aspx?guid=8313 (дата обращения 01.07.2017).
3. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учеб.пособ. 2-е изд., испр. и доп. М.: Форум, 2008. 464 с.

4. Програми для математичної і статистичної обробки даних. URL: <http://chem-bio.com.ua/aspirant/grant/item/> (дата звернення 01.07.2017).
5. Функції Excel (за категоріями) URL: <https://support.office.com/uk-ua/article> (дата звернення 01.07.2017).
6. Гойко О. В. Практичне використання пакета STATISTICA для аналізу медико-біологічних даних: навчальний посібник для студентів вищих навчальних закладів. Київ, 2004. 76 с.
7. Мамчич Т.І., Оленко А.Я., Осипчук М.М., Шпортюк В.Г. Статистичний аналіз даних з пакетом STATISTICA. Дрогобич: Відродження, 2006. 208 с. ISBN 966-538-161-X.
8. Покришень Д.А., Носенко Є.Ю. ІКТ для розв'язування системи нерівностей. *Інформаційні технології і засоби навчання*. 2012. № 1. URL: <http://www.journal.iitta.gov.ua>.
9. Практикум з теорії ймовірностей та математичної статистики: Навч. посіб. для студ. вищ. навч. закл. / Р.К. Чорней та ін.; За ред. Р.К. Чорнея. К.: МАУП, 2003. 328 с.
10. Основы вычислительной аналитики на Python. Спецкурс профессора Шибзухова З.М. URL: <http://tidm.ru/osnovy-vychislitelnoy-analitiki-na-python>.
11. Лутц М. Изучаем Python, 4-е издание. Пер. с англ. СПб.: Символ-Плюс, 2011. 1280 с.

References:

1. Kobylnyk T.P. Vykorystannia web-servisu wolfram|alpha dlia rozviazuvannia zadach z teorii ymovirnostei. *Information Technologies in Education*. 2015. № 24.
2. Vasilenko Zh.V. Programmnoe obespechenie po statisticheskomu analizu danykh. Metodologija sravnitel'nogo analiza. URL: http://www.giac.unibel.by/sm_full.aspx?guid=8313 (data obrashhenija 01.07.2017).
3. Vukolov Je.A. Osnovy statisticheskogo analiza. Praktikum po statisticheskim metodam i issledovaniju operacij s ispol'zovaniem paketov STATISTICA i EXCEL: ucheb.posob. 2-e izd., ispr. i dop. M.: Forum, 2008. 464 s.
4. Prohramy dlia matematychnoi i statystychnoi obrobky danykh. URL: <http://chem-bio.com.ua/aspirant/grant/item/> (data zvernennia 01.07.2017).
5. Funktsii Excel (za katehoriiamy) URL: <https://support.office.com/uk-ua/article> (data zvernennia 01.07.2017).
6. Hoiko O. V. Praktychne vykorystannia paketa STATISTICA dlia analizu medyko-biolohichnykh danykh: navchalnyi posibnyk dlia studentiv vyshchych navchalnykh zakladiv. Kyiv, 2004. 76 s.
7. Mamchych T.I., Olenko A.Ia., Osypchuk M.M., Shportiuk V.H. Statystychnyi analiz danykh z paketom STATISTICA. Drohobych: Vidrodzhennia, 2006. 208 s. ISBN 966-538-161-X.
8. Pokryshen D.A., Nosenko Ye.Iu. IKT dlia rozviazuvannia systemy nerivnostei. Informatiini tekhnolohii i zasoby navchannia. 2012. № 1. URL: <http://www.journal.iitta.gov.ua>.
9. Praktykum z teorii ymovirnostei ta matematychnoi statystyky: Navch. posib. dlia stud. vyshch. navch. zakl. / R.K. Chornei ta in.; Za red. R.K. Chornea. K.: MAUP, 2003. 328 s.
10. Osnovy vychislitel'noj analitiki na Python. Speckurs professora Shibzuhova Z.M. URL: <http://tidm.ru/osnovy-vychislitelnoy-analitiki-na-python>.
11. Lutc M. Izuchaem Python, 4-e izdanie. Per. s angl. SPb.: Simvol-Pljus, 2011. 1280 s.

Application of software in the process of teaching components of the theory of probability and mathematical statistics of students of specialties "Mathematics" and "Informatics"

O.Z. Garpul

Abstract. The article shows the necessity add to classical course of probability theory and mathematical statistics for the specialties "Mathematics" and "Informatics" by a computer workshop aimed at mastering by students the tools of statistical information processing by means of computer technologies, the use of which greatly facilitates and accelerates the calculation of statistical indicators , compilation of statistical tables and plotting, as well as expanding the possibilities for analysis and visual representation of statistical data. The example of one laboratory work in the article describes the possible ways of acquiring practical skills, skills and experience of students in this direction.

Keywords: probability theory and mathematical statistics, computer laboratory practice, functions of spreadsheets, programming language.