

До питання навчання комп'ютерного опрацювання даних студентів гуманітарних спеціальностей

Життя і успішна практична діяльність сучасної людини значною мірою залежать від ефективної організації опрацювання даних. Процеси опрацювання даних реалізуються в усіх сферах людської діяльності: науці, техніці, економіці, технології, медицині та ін. Дані дедалі частіше розглядають як життєво важливі ресурси, які необхідно організувати так, щоб ними можна було легко користуватися.

Великі обсяги даних практично неможливо опрацьовувати без спеціальних комп'ютеризованих засобів. Якісне та швидке опрацювання даних є актуальною проблемою досліджень сучасних науковців в різних галузях науки. Особливо значимою ця проблема є при підготовці студентів гуманітарних спеціальностей, до яких слід віднести спеціальності: 6.010100 Дошкільне виховання, 6.010100 Початкове навчання, 6.010100 Педагогіка і методика середньої освіти. Українська мова і література. Студенти цих спеціальностей, зазвичай, у процесі навчальної або науково-дослідницької діяльності майже не звертаються до програмних засобів статистичного опрацювання даних, оскільки не мають навичок роботи з ними. Разом з тим є необхідність показати їм переваги роботи з такими засобами, що можуть стати у нагоді студентам гуманітарних спеціальностей при навчанні відповідних дисциплін і проведенні науково-дослідних робіт.

Наукове дослідження – це процес вивчення певного об'єкта (предмета або явища) з метою встановлення закономірностей його виникнення, розвитку і перетворення в інтересах раціонального використання у практичній діяльності людей [7]. Наукові дослідження проводяться у різних сферах життєдіяльності людей і використовуються педагогами, психологами, соціологами, науковцями різних галузей, що й визначає специфіку експериментально-дослідницької роботи.

Методику проведення наукових досліджень у педагогічній сфері вивчають у курсі «Основи науково-педагогічних досліджень» студенти наведених вище гуманітарних спеціальностей. Метою навчання дисципліни є розширення, поглиблення та систематизація підготовки фахівця до організації наукових досліджень, впровадження їх результатів у практику педагогічної роботи [8]. В одному із змістових модулів навчальної програми даної дисципліни «Технологія опрацювання та оформлення результатів наукового дослідження» передбачається вивчення таких тем:

- застосування методів математичної статистики в дослідженні проблем педагогіки і психології;
- міри зв'язку та їх застосування;
- використання ЕОМ у психолого-педагогічних дослідженнях;
- інтерпретація наукових даних та оформлення результатів досліджень.

Беззаперечним є той факт, що вивчення цього модуля потребує застосування комп'ютерних засобів навчання, спеціального програмного забезпечення, зокрема статистичних пакетів опрацювання даних.

Статистичні пакети опрацювання даних – це спеціальні комп'ютерні засоби, за допомогою яких опрацьовують дані та обчислюють результати експериментальних досліджень. Зазвичай їх використовують соціологи, політологи, педагоги та інші науковці при проведенні масштабних досліджень. В такі пакети вбудовано адаптований інтерфейс та довідкову систему, що дає змогу опанувати основи роботи з ними на достатньому рівні.

Опрацюванню даних, важливій стадії проведення наукового дослідження, присвячено значний обсяг спеціальної літератури. Зокрема, проблемам проведення науково-дослідницької діяльності присвячені роботи Бусигіна А.Л., Відерникова Л.В., Горностаєв П.В., Дайв Р., Кнепер К., Левитська Е.Ю., Легран П., Матушанський Г.У., Менг Т.В., Соломко Л.І. та ін. Питання системного підходу у дослідницькій роботі, де передбачається опрацювання результатів педагогічних досліджень, вперше поставили Архангельський С.І., Данилов М.А., Корольов Ф.Ф., Юдан Е.Г. У подальших дослідженнях системний підхід до аналізу дослідницької діяльності застосовували Андрєєв В.І., Вяткін Л.Г., Ільїн В.С., Леднев В.С. Махмутов М.І., Поляков В.А., Сластьонін В.А., Тализіна Н.Ф. ін. Питанням застосування методів математичної статистики у педагогічних дослідженнях приділяють увагу Греков А.А., Загвязинський В.І., Кузьміна Н.В., Мамчур Е.А., Стьопін В.С., Швирєв В.С. та ін.

У більшості робіт цих дослідників передбачене опрацювання даних за класичними методами, без використання комп'ютерів. Разом з тим використання сучасних інформаційних технологій може суттєво полегшити цей процес, уніфікувати його та прискорити.

Дані – це характеристики певних об'єктів чи явищ, зафіксовані (закодовані) у певній формі, придатній для подальшого опрацювання, зберігання і використання [1].

У вузькому розумінні термін «дані» відноситься тільки до даних з реєстраційних документів (анкет, бланків інтерв'ю, протоколів спостереження тощо). В якості даних виступають як опрацьовані, так і не опрацьовані на комп'ютері результати досліджень [2].

У широкому розумінні дані – це факти, зібрані у ході емпіричного дослідження.

Дані є основним об'єктом наукової діяльності людини. Саме з ними працює дослідник під час дослідження: визначає методи отримання даних, їх накопичення, опрацювання, передавання, формулює висновки на їх основі. Яскравим прикладом здійснення цих операцій є проведення соціологічних досліджень та опрацювання даних за допомогою методів математичної статистики.

У контексті нашого дослідження доцільно використовувати таке визначення.

Дані – це первинні відомості, що отримані в результаті соціологічного дослідження, відповіді респондентів, оцінки експертів, результати спостережень тощо [2].

Поняття «соціологічні дані» і «емпіричні дані» дані у підручниках, словниках, як правило, спеціально не визначаються і зазвичай вважаються синонімами [2].

Опрацюванням даних називають математично-статистичний аналіз даних, їх компактне, придатне для аналізу та інтерпретації подання. Над даними можна виконувати такі операції:

- 1) готувати їх до опрацювання; шифрувати, кодувати тощо;
- 2) опрацьовувати (вручну або за допомогою комп'ютера); табулювати, розраховувати багатовимірні розподіли ознак, класифікувати тощо;
- 3) аналізувати;
- 4) інтерпретувати.

Стандартні статистичні методи опрацювання даних включені до складу електронних таблиць, таких, як Excel, Lotus 1-2-3, QuattroPro, та в математичні пакети загального призначення – Mathcad, Matlab, Maple та ін. Потужніші можливості статистичного опрацювання даних можна отримати за допомогою спеціалізованих пакетів прикладних програм, як вітчизняних – STADIA, МЕЗОЗАБР, СИГАМД, СТОД, САНИ, ОЛІМП: СтатЕксперт та ін., так і зарубіжних – StatPlus, STATCRAPHS, SPSS, SAS, BMDP, STATISTICA та ін.

Протягом останніх років значної популярності у проведенні статистичного опрацювання даних набув табличний процесор Microsoft Excel, що пояснює його інтеграцію до пакету Microsoft Office. Excel включає в себе програмну надбудову «Пакет аналізу» і бібліотеку з 78 статистичних функцій. У повсякденній діяльності такого набору інструментів буває, як правило, цілком достатньо для проведення досить повного і якісного статистичного аналізу даних. Якщо ж користувача не задовольняють подібні можливості використання Excel, тоді можна звернутися до потужних спеціалізованих пакетів статистичного аналізу даних, зокрема до пакету StatPlus фірми AnalystSoft Inc. [4], SPSS for Windows, Statistica Plus Multilanguage, STATISTICA, MATHLAB, SYSTAT.

Наразі існує велика кількість літературних джерел стосовно роботи з Excel різних версій, що переважно адресовані користувачам, які вже мають навички роботи з подібними програмними засобами. Тоді як новачкам, особливо гуманітарних спеціальностей, досить важко самостійно опанувати з систему опрацювання даних в Excel. Можливості використання Excel досить об'ємні, це універсальний табличний процесор, що може стати у нагоді при проведенні експериментальних досліджень студентами гуманітарних спеціальностей, як у процесі науково-дослідної, так і навчальної діяльності.

Розглянемо особливості роботи з програмним засобом статистичного опрацювання даних StatPlus, в який також включає потужний набір статистичних інструментів і графічних методів аналізу, простий та зручний інтерфейс. Область застосування засобу StatPlus практично необмежена – соціологія, фінансовий аналіз, економіка, страхування, охорона здоров'я та клінічні дослідження тощо. Це лише кілька галузей, де він вже використовується [4].

Проілюструємо можливості та особливості використання зазначених програмних засобів Excel і StatPlus для опрацювання даних на конкретному прикладі при вивченні теми «Застосування методів математичної статистики в дослідженні проблем педагогіки і психології» у курсі «Основи науково-педагогічних досліджень».

Приклад. У двох групах студентів (Ігр. і IIгр.) проведено тестування з інформатики (рис. 1). Визначити за його результатами (кількістю правильних відповідей), чи суттєво різняться успішності у даних групах.

| № пит. | 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 |
|---------------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| І гр. | 20 | 13 | 18 | 20 | 15 | 11 | 19 | 20 | 15 | 12 | 19 | 19 | 13 | 18 | 20 | 11 | 17 | 10 | 16 | 15 | 20 | 11 | 18 | 10 | 15 |
| II гр. | 14 | 16 | 14 | 18 | 18 | 15 | 13 | 15 | 14 | 13 | 18 | 18 | 17 | 14 | 15 | 15 | 14 | 15 | 14 | 19 | 17 | 14 | 14 | 15 | 14 |

Рис 1. Результати тестування з інформатики

Розв'язування. Для порівняння за певними ознаками досліджуваних сукупностей використовують спеціальні статистичні критерії: F-критерій (або критерій Фішера), t-критерій (критерій Стьюдента) та інші, що ґрунтуються на порівнянні різних параметрів досліджуваних вибірок (середніх значень, дисперсій та ін.).

Дисперсія – це показник, що характеризує розсіювання значень елементів сукупності (вибірки) навколо її середнього значення.

Для визначення показника розсіювання успішності груп студентів скористаємось формулою стандартного середнього квадратичного відхилення:

$$\sigma = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}},$$

де σ – стандартне середнє квадратичне відхилення (чим більше значення σ , тим більше різномірно розсіяна сукупність);

x_i – значення окремих елементів вибірки;

n – кількість досліджуваних об'єктів.

Для порівняння двох вибірок обчислимо F_{emp} за формулою:

$$F_{emp} = \frac{\sigma_1^2}{\sigma_2^2},$$

де σ_1 – середнє квадратичне відхилення першої сукупності,

σ_2 – середнє квадратичне відхилення другої сукупності.

Для розв'язування даного прикладу скористаємось програмним засобом Excel (рис. 2).

| | A | B | C | D | E | F | G |
|---|--|---|---------------------|---|----------|---|---------------------|
| 1 | I група | К-ть правильно виконаних завдань, X_i | x_i^2 | | II група | К-ть правильно виконаних завдань, X_i | x_i^2 |
| 2 | 1 | 20 | 400 | | 1 | 14 | 196 |
| 3 | 2 | 13 | 169 | | 2 | 16 | 256 |
| 4 | 23 | 18 | 324 | | 23 | 14 | 196 |
| 5 | 24 | 10 | 100 | | 24 | 15 | 225 |
| 6 | 25 | 15 | 225 | | 25 | 14 | 196 |
| 7 | n = 25 | $\sum x_i = 395$ | $\sum x_i^2 = 6545$ | | n = 25 | $\sum x_i = 383$ | $\sum x_i^2 = 5943$ |
| 8 | $\sigma = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{6545 - \frac{395^2}{25}}{24}} = \sqrt{12667} = 3,559$ | | | $\sigma = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}} = \sqrt{\frac{5943 - \frac{383^2}{25}}{24}} = \sqrt{3143} = 1,773$ | | | |
| 9 | $F_{emp} = \frac{\sigma_1^2}{\sigma_2^2} = 4,03$ | | | | | | |

Рис. 2. Приклад статистичного опрацювання даних в Excel

Для цього для знаходження показника розсіювання даних для студентів I-ї групи введемо у клітинку C33 формулу $=КОРЕНЬ(((C28-(B28*B28/25))/24))$, а для знаходження показника розсіювання даних для студентів II-ї групи у клітинку F33 $=КОРЕНЬ(((G28-(F28*F28/25))/24))$. Далі для знаходження F_{emp} у клітинку D36 введемо формулу $=(C33*C33)/(F33*F33)$. Результати обчислень подані на рис. 2. Відмітимо, що для знаходження цих параметрів в Excel можна скористатись також вбудованими статистичними функціями або засобом Аналізу даних.

Далі продемонструємо розв'язування даного прикладу за допомогою спеціалізованого програмного засобу StatPlus, в якому, на відміну від попереднього методу розв'язування, для отримання різних статистичних показників користувачеві достатньо ввести тільки дані відповідних статистичних вибірок.

Для обчислення F-критерію у середовищі StatPlus скористаємось вказівкою *Статистика* → *Основная статистика/Таблицы* → *F-тест для дисперсии* і введемо дані вибірок тестування студентів, як показано на рис. 3.

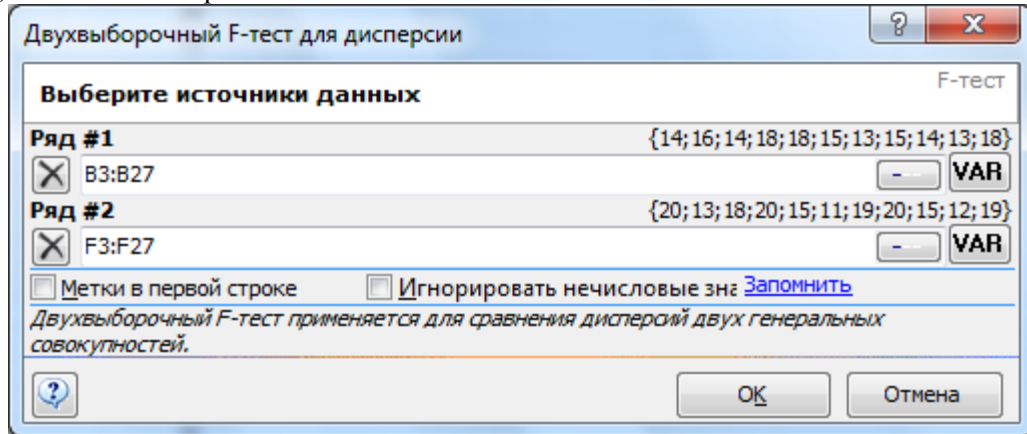


Рис. 3. F-тест для дисперсії

В результаті отримаємо розв'язок (рис.4), що збігається з попередніми результатами.

| | A | B | C | D | E |
|----|--|----------------|---|-------------------------------|-----------|
| 1 | Двухвыборочный F-тест для дисперсии | | | | |
| 2 | Описательная статистика | | | | |
| 3 | VAR | Размер выборки | Среднее | Стандартная ошибка (среднего) | Дисперсия |
| 4 | I группа | 25 | 15,8 | 0,7118 | 12,6667 |
| 5 | II группа | 25 | 15,32 | 0,3546 | 3,1433 |
| 6 | | | | | |
| 7 | Итоги | | | | |
| 8 | Значение F-критерия | 4,03 | P(F<=f)(Вероятность, соответствующая критерию Фишера) | | 0,0006 |
| 9 | | | | | |
| 10 | | | | | |
| 11 | | | | | |
| 12 | | | | | |
| 13 | | | | | |
| 14 | | | | | |
| 15 | | | | | |
| 16 | | | | | |

Рис. 4. Приклад статистичного опрацювання даних у StatPlus.

Аналізуючи отримані результати, можна стверджувати, що існують відмінності успішності між вибірками з двох цих груп студентів. За результатами тестування I-ша група має досить високий показник розсіювання, тобто є різномірно розсіяною сукупністю, а II-га група має менший показник розсіювання, тобто є одноріднішою сукупністю.

Виконання розрахунків з використанням програмних засобів статистичного опрацювання даних не потребує значних часових витрат, сприяє уникненню математичних та обчислювальних помилок. Оволодіння сучасними програмними засобами дозволяє якісно та швидко здійснювати статистичне опрацювання даних. Опанування методикою роботи з такими засобами студентами-гуманітаріями при проведенні експериментальних досліджень сприяє пошуку раціональних шляхів вирішення поставлених завдань.

Література

1. Рамський Ю.С., Цибко Г. Ю. Проектування й опрацювання баз даних: Посібник для вчителів. – Тернопіль: Навчальна книга – Богдан, 2005. – 116 с.
2. Добренское В.И., Кравченко А.И. Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2004. – 768 с.
3. Ядов В.А. Стратегия социологического исследования: Описание, объяснение, понимание

социальной реальности. М.: Добросвет, 1998.

4. AnalystsoftSoft Inc. – Analysis Made Easy... – [Електронний ресурс] – Режим доступу: <http://www.analystsoft.com/StatPlus>. User-friendly statistical software.

5. Макарова Н.В., Трофимец В.Я. Статистика в Excel: Учеб. пособие. – М.: Финансы и статистика, 2002. – 368 с.

6. Нові технології навчання: Наук.-метод. зб. /Кол. авт. – К.: Наук.-метод. центр вищої освіти, 2004. – Вип. 39. – 203 с.

7. Білуха М. Т. Методологія наукових досліджень: Підручник. – К.: АБУ, 2002. – 480 с.

8. Зінченко В. П. Програма навчального курсу «Основи науково-педагогічних досліджень» (за вимогами кредитно-модульної системи). Для студентів спеціальності 6.010101 Дошкільна освіта. – Глухів 2010 р. – 33 с.