

УДК 61:681.03:57.089:311.21

АНАЛІЗ СУЧАСНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ СТАТИСТИЧНОГО ОБРОБЛЕННЯ Й АНАЛІЗУ БІОМЕДИЧНИХ ДОСЛІДЖЕНЬ

О. В. Гойко, С. І. Мохначов

Національна медична академія післядипломної освіти імені П. Л. Шупика

У статті наведена базова інформація про основні програмні пакети, придатні для статистичного оброблення біомедичних даних, подаються їх можливості та характеристики, а також деякі рекомендації щодо їх вибору для аналізу інформації.

Ключові слова: статистичне оброблення, аналіз біомедичних даних, програмне забезпечення.

АНАЛИЗ СОВРЕМЕННОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ И АНАЛИЗА БИОМЕДИЦИНСКИХ ИССЛЕДОВАНИЙ

О. В. Гойко, С. И. Мохначов

Национальная медицинская академия последипломного образования имени П. Л. Шупика

В статье представлена базовая информация об основных программных пакетах, предназначенных для статистической обработки биомедицинских данных, подаются их возможности и характеристики, а также некоторые рекомендации по выбору для анализа информации.

Ключевые слова: статистическая обработка, анализ биомедицинских данных, программное обеспечение.

ANALYSIS OF ADVANCED SOFTWARE FOR STATISTICAL PROCESSING AND ANALYSIS OF BIOMEDICAL RESEARCH

O. V. Goyko, S. I. Mokhnachov

National Medical Academy of Post-Graduate Education by P. L. Shupyk

This article provides the basic information about the basic software package intended for statistical processing of biomedical data, submitted by their features and capabilities, as well as some recommendations on their choice for the analysis of information.

Key words: statistical treatment, analysis of biomedical data, software.

Вступ. В останні 20-30 років медицина і біологія вступили у нову фазу свого розвитку. Нагромадження величезних масивів кількісних даних і доступність обчислювальної техніки посилили математизацію біології та медицини. Сьогодні вже неможливо собі уявити наукові дослідження в галузі біології та медицини, які виконані без статистичного аналізу отриманих фактичних даних. Біостатистика стала суттєвим кроком у русі медичного дослідження від описів окремих випадків до експериментів з групами контролю, і, нарешті, до великомасштабних рандомізованих контрольованих досліджень, яким тепер надається перевага як стандартам наукового доказу.

Разом з тим, методи статистичної обробки отриманих результатів настільки різноманітні, що найчастіше досліднику дуже складно розібратися у них і вибрати адекватні поставленим завданням статистичні критерії. Тривалий час аналіз медичних даних могли здійснювати лише високопрофесійні фахівці, оскільки це вимагало серйозної попередньої підготовки. З появою обчислювальної техніки і вдосконаленням сучасних програм оброблення й аналізу даних статистична обробка піднялася на новий рівень. Тепер дослідник-медик може й не мати математичної підготовки. Досить оперувати статистичними поняттями і, найголовніше, правильно вибрати метод ана-

лізу [5], володіти навичками роботи на персональному комп'ютері та освоїти новітні пакети програмного забезпечення для статистичного оброблення й аналізу біомедичних досліджень.

Метою роботи є аналіз сучасного програмного забезпечення для статистичного оброблення статистичних даних біомедичних досліджень, а також деякі рекомендації щодо їх вибору.

Результати та їх обговорення. В основі оброблення й аналізу даних лежать математичні методи, які за останні півстоліття істотно не змінилися, проте відповідне програмне забезпечення за цей час зазнало істотних змін. Зі зміною поколінь ЕОМ мінялися й програмні засоби обробки та аналізу даних. І якщо можливості перших ЕОМ з аналізу даних не перевершували можливості сучасних середніх калькуляторів, то в 70-і роки вже з'явилися пакети прикладних програм, які містили практично всі ті математичні методи оброблення, які входять до складу й сучасних пакетів. Розвиток пакетів здійснювався шляхом вдосконалювання технології й аналізу (табл. 1).

На сьогоднішній день число пакетів для оброблення інформації досягає кількох десятків, серед яких зарубіжні пакети, такі як: SYSTAT, STATGRAPHICS, BMDP, SPSS, SAS, CSS, Statistica, а також вітчизняні: STADIA, ЕВРІСТА, МЕЗОЗАВР, САНІ, КЛАСС-МАСТЕР, СІГАМД тощо.

Більшість комп'ютерних статистичних програм не є чисто медичними прикладними програмами, оскільки більшість методів статистичного аналізу є універсальними й можуть застосовуватися не лише в різних галузях медичної статистики, але й у найрізноманітніших галузях людської діяльності. Наприклад, з погляду формальної логіки статистичний прогноз інфекційної захворюваності й прогноз курсу долара — це та ж сама задача, а тому вона може вирішуватися за допомогою одних і тих же пакетів прикладних програм.

Основну частину наявних пакетів для оброблення даних можна віднести до трьох категорій: спеціалізовані пакети, пакети загального призначення і професійні пакети.

Спеціалізовані пакети, як правило, містять методи з одного-двох розділів статистики або методи,

що використовуються в конкретній предметній галузі (наприклад, *Мезозавр* — програма аналізу часових рядів). Спеціалізовані пакети застосовуються для вирішення вузького кола завдань з використанням спеціальних методів статистичного аналізу. Експлуатація цих програм вимагає високого рівня підготовки користувача в галузях певних розділів статистики.

Професійні пакети призначені для користувачів, які мають справу із надзвичайно великими обсягами даних або вузькоспеціалізованими методами аналізу.

Пакети загального призначення або універсальні (Statgraphics, SPSS, Statistica, Excel, STADIA тощо) є найбільш зручними для користувача-початківця завдяки відсутності орієнтації на специфічну предметну галузь, широкому діапазону статистичних методів і дружньому інтерфейсу користувача. Вони більш доступні для практики й можуть використовуватися широким колом фахівців різного профілю. Практично всі задачі, які стосуються оброблення й аналізу медико-біологічних досліджень, можуть бути вирішені за допомогою універсальних пакетів.

Пакет Statgraphics розроблявся ще для роботи в середовищі DOS, а потім був адаптований до операційної системи Windows і отримав нову назву Statgraphics Plus. Сучасний пакет STATGRAPHICS PLUS - це досить потужна статистична програма, яка містить більше 250 статистичних функцій. За своїми характеристиками пакет займає проміжне місце між SPSS і Statistica.

Пакет SPSS (Statistical Package For Social Science) - це один із найбільш часто використовуваних пакетів статистичної обробки медико-біологічних даних (<http://www.spss.com/>). Цей пакет створювався ще для «великих» електронно-обчислювальних машин і послідовно переводився для роботи в середовищі DOS, а потім Windows.

Пакет SPSS досить потужний і добре відпрацьований, наближається за своїми можливостями до професійних пакетів, і реалізація статистичних процедур добре пристосована до практичної роботи.

Поряд з пакетом SPSS, великою популярністю в освітніх та наукових закладах США користується

Таблиця 1. Хронологія розвитку пакетів прикладних програм з обробки й аналізу даних

Роки	Основні пакети аналізу даних	Операційні системи
1970-1985	SSP, BMDP, SAS	
1985-1995	Statgraphics, STATA, SAS, Systat, STADIA, МЕЗОЗАВР, САНІ, Евріста, Клас-майстер тощо	DOS
1995-2012	Statgraphics Plus, SAS, SPSS, Statistica, Excel тощо	Windows

пакет STATA. Це професійний статистичний програмний пакет з data-management system, який досить часто застосовують для біомедичних цілей. Програма добре документована, видається спеціальний журнал для користувачів системи. Офіційний сайт <http://www.stata.com/>.

Не менш популярним інструментом розробки користувачьких додатків і не лише в медицині, а й в бізнесі, економіці, фінансах, промисловості тощо є інтегрована система аналізу та управління даними STATISTICA [1].

Пакет Statistica спеціально створювався для роботи в середовищі Windows і відповідає всім стандартам Windows, що дозволяє зробити аналіз високоінтерактивним. STATISTICA складається з набору модулів, в кожному з яких зібрані тематично пов'язані групи процедур, що мають високу швидкість і точність обчислень. Система STATISTICA містить повний набір класичних методів аналізу даних [2]: від основних методів статистики до просунутих методів, що дозволяє гнучко організувати аналіз. Дані системи STATISTICA легко конвертувати в різні бази даних і електронні таблиці. Ця система відрізняється найбільш розвиненим інтерфейсом із користувачем і багатими графічними можливостями, підтримує високоякісну графіку, що дозволяє ефектно візуалізувати дані і проводити графічний аналіз. Гнучка і потужна технологія доступу до даних дозволяє ефективно працювати як з таблицями даних на локальному диску, так і з віддаленими сховищами даних.

Система STATISTICA є відкритою системою: містить мови програмування, які дозволяють розширювати систему, запускати її з інших Windows-додатків, наприклад, з Excel.

Електронна таблиця Excel найпоширеніша, оскільки повністю русифікована і найбільш доступна, вона встановлюється автоматично при інсталяції пакета MS Office. Електронна таблиця Excel тісно інтегрується з іншими програмами пакета MS Office, наприклад, MS Word і PowerPoint, а тому найчастіше використовується при оформленні результатів роботи. MS Excel - це електронна таблиця з достатньо потужними математичними можливостями, проте деякі статистичні функції є просто додатковими, а тому розрахунки, зроблені за допомогою вбудованих окремих формул, не завжди визнаються авторитетними біомедичними журналами [4]. MS Excel, як правило, використовується при найпростішому статистичному аналізі даних. Окрім того, в MS Excel неможливо побудувати якісні наукові графіки. MS Excel добре підходить для накопичення даних, попе-

редніх статистичних прикидок, для побудови деяких видів діаграм, проте остаточний статистичний аналіз бажано робити в програмах, які спеціально створені для цих цілей. Існує макрос-додаток XLSTAT-Pro (<http://www.xlstat.com>) для MS Excel, який включає в себе більше 50 статистичних функцій, включаючи аналіз виживаності.

Програма вітчизняної розробки STADIA включає в себе всі необхідні статистичні функції, призначені для статистичного аналізу даних. Проте ця програма фактично не змінюється з 1996 року, а тому графіки і діаграми, побудовані за допомогою STADIA, виглядають в сучасних презентаціях архаїчно. До позитивних якостей програми можна віднести російськомовний інтерфейс і наявність книг, що описують роботу [3]. Зі сторінки <http://www.protein.bio.msu.ru/~akula/index.htm> можна взяти демо-версію STADIA.

Слід зазначити, що всі ці пакети постійно оновлюються і з кожним роком з'являються їх нові версії.

При виборі пакета для аналізу даних можна виділити два аспекти: а) початковий вибір пакета аналізу; б) поточний вибір при переході на більш сучасний, більш потужний пакет. Підходи в обох випадках дещо відрізняються.

У першому випадку на вибір накладаються такі обмеження:

1. Можливості комп'ютера.
2. Можливості одержання установчої версії пакета.
3. Характеристики пакета.

Що стосується першого пункту, то варто вибирати найбільш сучасні версії пакетів із тих, що можуть бути встановлені на наявний комп'ютер. Другий пункт очевидний - вибирати можна з тих пакетів, що доступні. Що стосується характеристик пакета, то тут варто розглянути такі аспекти: а) обчислювальні можливості, б) зручність роботи, в) складність освоєння.

а) *Обчислювальні можливості.* У випадку, коли необхідно обробляти медичні дані помірних обсягів (до декількох тисяч спостережень) стандартними статистичними методами, то найкраще використовувати універсальні пакети. Якщо дивитися з позицій лікаря-дослідника, то всі сучасні універсальні статистичні пакети за своїми обчислювальними можливостями повністю відповідають можливим потребам (Statistica, SPSS, Statgraphics Plus та інші пакети, що працюють в операційній системі Windows). Проте завжди варто переконатися, що обраний пакет містить необхідні методи обробки.

б) *Зручність роботи.* Всі сучасні пакети досить зручні в роботі (коли вони вже освоєні).

в) *Складність освоєння.* За складністю освоєння пакети дещо розрізняються і тут варто віддати перевагу русифікованим пакетам або пакетам, з яких є доступна література або є ймовірність пройти курс навчання.

Варто зауважити, що без крайньої необхідності (неможливість забезпечити необхідну обробку даних) не бажано змінювати обраний і освоєний пакет аналізу, тому що це призведе до значного збільшення трудовитрат.

Що стосується заміни пакета на більш сучасну версію, то тут є дві крайності:

1. Прагнення до постійного відновлення, установки самих останніх версій пакетів, як правило, віднімає багато сил, не дозволяє виробитися корисним стереотипам дій, у той же час не приводить до суттєвого зростання можливостей.

2. З іншого боку, уподобання застарілих пакетів - найчастіше не дозволяє повною мірою використовувати можливості сучасної техніки і програмного забезпе-

чення. Існує деякий емпіричний оптимум, що може визначатися зразковим терміном експлуатації пакета в 2-3 роки, після закінчення котрого доцільно здійснювати перехід до більш сучасних пакетів. При цьому переважно вибирати чергову версію того ж пакета, що використовувався раніше. Запам'ятайте: спадкоємність значно полегшує процес освоєння пакета.

Висновки. Широке впровадження сучасних комп'ютерних технологій і застосування пакетів прикладних програм докорінно змінило процес оброблення й аналізу медичних даних. Завдяки використанню комп'ютерів і сучасного програмного забезпечення, оброблення й аналіз медико-біологічних даних став набагато легшим. При цьому для застосування основних статистичних методів оброблення медичних даних лікарю не потрібно заглиблюватися в складність математичних процедур, а варто зрозуміти, для чого і як ці методи використовуються, а також вміло використовувати обраний пакет прикладних програм.

Література

1. Боровиков В. Statistica. Искусство анализа данных на компьютере / В. Боровиков. - СПб.: Питер, 2001. - 656 с.
2. Гойко О. В. Практичне використання пакета STATISTICA для аналізу медико-біологічних даних : навчальний посібник для студентів вищих навчальних закладів / О. В. Гойко. - Київ, 2004. - 76 с.
3. Кулаічев А. П. Методи і засоби аналізу даних у середовищі Windows / А. П. Кулаічев. - М. : Інко, 2002. - 341 с.

4. Лапач С. Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С. Н. Лапач, А. В. Чубенко, П. Н. Бабич. - Издательство "Морион Лтд", 2000. - 320 с.
5. Мінцер О. П. Оброблення клінічних і експериментальних даних у медицині : навч. посібник / О. П. Мінцер, Ю. В. Вороненко, В. В. Власов. - К. : Вища школа, 2003. - 350 с.