

К.С. Жижин

Проблемы IT-подготовки кадров непрофильных специальностей

В настоящее время информационные технологии применяются во всех областях деятельности. Однако существуют области (например, управление, финансы, образование), которые в связи с этими технологиями затрагиваются в периодической и другой научной и публицистической литературе намного чаще, чем остальные. Но применение компьютерной техники и современных технологий в транспорте, сельском хозяйстве, медицине (именно этому посвящена предлагаемая статья) — все это не менее, а, может быть, и более важные темы для дискуссий и научных работ.

Сегодня Россия вступила в полосу самых активных преобразований в экономико-промышленном переустройстве общества. Однако взяв за основу принципы капиталистического переустройства промышленности, реформаторы, видимо, не смогли учесть особенности тех товарно-денежных отношений, которые капиталистический мир шлифовал более 200 лет. И как итог этого — главный камень преткновения для сегодняшних промышленных предприятий это нехватка квалифицированных рабочих рук.

Социалистический принцип хозяйствования не предъявлял жестких требований к качественным характеристикам специалиста. Большей частью в массовом производстве дефицит квалифицированных кадров перекрывал вал рабочей силы: было из кого выбирать.

В сегодняшних условиях к товарно-денежным отношениям совсем другой подход, а, следовательно, изменился и подход к производителю благ: требуется чуть ли не «поштучное» производство специалистов. К тому же ситуацию крайне усложняет чрезвычайно напряженная демографическая ситуация в стране.

В таких условиях следует задуматься о том, как ускорить и повысить качество обучения специалистов, и о том, чему их учить, чтобы они были востребованы в нашей стране.

В государственных образовательных учреждениях среднего профессионального образования все активнее внедряются информационные технологии (IT) обучения для медицинских специальностей, реализуемые с помощью персональных ЭВМ.

Однако на практике большей частью в лечебно-профилактических учреждениях ЭВМ, как основной элемент IT, используется не более чем усовершенствованная пишущая машинка. И это несмотря на то что с помощью IT можно реализовать широчайшие возможности и для совершенствования процессов диагностики и лечения, и для обработки результатов наблюдений. В 60–70-е годы XX столетия использование ЭВМ даже для анализа итогов работы, не говоря уже о разработке прогнозов и планирования, было сопряжено с рядом методических и методологических трудностей для рядового пользователя.

В настоящее же время проблема эта снята: для людей далеких от тонкостей технологий программирования создано много готовых для использования программ. В частности наиболее популярными статистическими софтами на рынке IT сегодня являются американские программы SPSS (Statistical Package for Social Science — Статистический пакет для социальных наук) и STATISTICA. При всех технологических отличиях в обработке эмпирического мате-

риала их роднит широкий спектр, используемых модулей, достаточная простота даже для слабо подготовленных в статистике медицинских работников. И тем не менее, простота эта кажущаяся. Поэтому считаем необходимым сделать несколько предварительных замечаний для пользователей упомянутых программных статистических продуктов.

Первый шаг к ИТ и статистическому анализу — это ввод данных в программу. Для управления ими используется редактор данных. Окно редактора представляет собой электронную таблицу, в которой столбцы отражают переменные, а строки — объекты (случаи). На пересечении строки и столбца находится ячейка, в которой сохраняется значение переменной (столбца) для данного объекта (строки). Для работы с таблицами, не умевающими в пределах окна, предусмотрены вертикальные и горизонтальные полосы прокрутки, а также автоматические функции перехода к заданной переменной или объекту.

Перед тем как начать ввод данных, необходимо определить структуру будущего файла. Для этого надо знать как будут использоваться в анализе те или иные переменные. К сожалению, многие пользователи, слабо владеющие ИТ, начинают задумываться об этом гораздо позже, чем следует, и в этом кроется причина большинства неудач при подведении итогов работы. Если организовывать эксперимент по принципу «соберем как можно больше фактов, а потом, как-нибудь их обработаем», то при таком подходе фактический материал порой оказывается совершенно непригодным для анализа, и даже опытный аналитик и статистик тут ничем не поможет.

Следует отметить, что структура файла данных должна строго соответствовать плану исследования. Вместе с тем, план исследования должен быть составлен так, чтобы его исходные данные можно было обработать в соответствии с задачами и гипотезами исследования. Самый оптимальный и простой путь обеспечения этих соответст-

вий — определение структуры данных на этапе планирования исследования, еще до сбора данных. Это позволит избежать большинства типичных ошибок, относящихся как к планированию исследования, так и к организации данных и приемам их обработки.

Поскольку средний медицинский работник при сборе и первичной обработке статистического материала на своем рабочем месте обычно имеет дело с *выборкой* — частью, долей значительно большей по численности группы, называемой *генеральной совокупностью*, то ему надлежит вникнуть в некоторые базовые (обязательные) положения статистического анализа, которыми пренебрегать нельзя. Конечной целью статистической обработки опытных, отчетных данных является переложение выводов, полученных в выборке, на всю изучаемую генеральную совокупность, и некорректное использование статистических приемов может не только не помочь, а даже навредить.

Выборочные совокупности (выборки) бывают *независимыми* (несвязанными), если процедура оценки результатов измерения в них не оказывается взаимопроникающей, не влияющей на результаты другой выборки. В том случае, если такое влияние имеет место, выборки называются *зависимыми* (связанными).

Выборка обязательно должна быть *репрезентативной*, т.е. обладать способностью адекватно представлять генеральную совокупность по всем статистическим характеристикам: среднее значение признака, стандартное среднеквадратическое отклонение, стандартная ошибка среднего значения признака и др.

Добиться репрезентативности в силу разных причин очень трудно. Для облегчения ситуации разработан ряд приемов:

- надо заранее заложить одинаковые пропорции основных признаков в выборке и в генеральной совокупности. Главная трудность — практически всегда отсутствует ин-

формация, какие признаки являются важными для изучаемого явления, а какие нет, и самое главное о том, каковы закономерности распределения этих признаков в генеральной совокупности;

- рандомизация (перемешивание), т.е. случайный отбор наблюдений, фактов из генеральной совокупности. Надо всегда помнить, что при случайном отборе в выборку *с равной вероятностью в обязательном порядке* попадают буквально все компоненты совокупности (пациенты, лабораторные животные, препараты и т.д.), имеющие и не имеющие существенные для изучаемого явления признаки.

Безусловно, при выборочном отборе признаков не стоит доходить до абсурда, когда объем выборки вдруг окажется равным генеральной совокупности! Спасать положение может априорное знание закона распределения основных (изучаемых) признаков в генеральной совокупности, но и тут исследователя подстерегает трудность (особенно в медицине), так как многое в медико-биологических процессах пока еще не может быть формализовано и переведено на лаконичный математический язык.

На этапе планирования исследования структура данных может быть задана в виде предварительного списка переменных с указанием их типов и диапазонов возможных значений. Каждая переменная — это имеющее значение для исследователя основание, позволяющее отличать объекты друг от друга. На предварительном этапе следует выделять два типа переменных: количественные и категориальные (номинативные).

Количественная переменная предпочтительнее, она позволяет четко различать объекты по уровню выраженности некоторого свойства, например средний балл отметки, тестовый показатель и пр. И идентификация количественных переменных на предварительном этапе не составляет труда: обычно они соответствуют тому, что исследователь намеревается измерить.

Категориальные (номинативные) переменные обычно используются для деления объектов (испытуемых) на группы или категории: пол, класс, возрастная категория, уровень дохода и пр. Надо стараться избегать типичной ошибки начинающих исследователей — игнорирование оснований для деления объектов на группы в качестве самостоятельных номинативных переменных в структуре данных.

Важным свойством номинативных переменных является возможность их представления в виде набора целых чисел. Числовое представление данных в компьютерных программах всегда предпочтительнее символьного, поскольку обработка чисел происходит быстрее, проще и меньше вероятности ошибок.

Кроме того, числовое представление легко модифицировать: вы можете переназначить числа, соответствующие созданным элементам, а также (что часто требуется на практике) без проблем включить в анализ новые элементы.

Порядок создания переменных также важен при вводе данных. Здесь следует придерживаться простого правила: наиболее важные и часто используемые переменные должны помещаться в начало файла, для остальных данных вопрос порядка следования не столь важен, однако рекомендуется объединять их в группы по физическим признакам. Чаще всего в начало файла следует поместить категориальные переменные, которые далее предполагается использовать для деления объектов (испытуемых) на группы. Далее можно перечислять остальные сведения, а логическое объединение переменных производить в зависимости от того, какие аспекты они отражают.

Очень часто серьезной проблемой медико-биологических (диагностических, гигиенических) исследований является *фактическое отсутствие естественных интервальных шкал*, и тем более шкал отношений для оцениваемых переменных. Специалистами в вопросах математической статисти-

ки для этих целей разработаны специальные процедуры, позволяющие построить интервальные шкалы. И, несмотря на то что такие приемы имеют искусственное происхождение, эти издержки перекрываются обилием математических методов, имеющих глубокое теоретическое обоснование. В современной статистике наиболее используемыми видами таких шкал являются:

- шкала стэнов (англ. *sten* — десятка) — десятибалльная шкала, со средним $\bar{x} = 5,5$ и стандартным отклонением $\sigma = 2$;
- шкала стэнайнгов (англ. *stanine* — девятка) — девятибалльная шкала, имеющая среднее значение $\bar{x} = 5$ и $\sigma \approx 2$;
- процентильные шкалы.

Итак, с позиций классической математической статистики измеряемые признаки обязательно должны быть случайными величинами, только тогда для изучения закономерных изменений таких величин могут использоваться хорошо себя зарекомендовавшие законы распределения вероятностей.

Основным законом распределения является нормальный закон (закон Гаусса), в котором плотность распределения вероятностей графически выражается колоколообразной или «нормальной» кривой и задается следующей формулой:

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{[x-\bar{x}]^2}{2\sigma^2}},$$

где $f(x)$ — функция плотности распределения вероятностей;

σ — стандартное отклонение;

\bar{x} — среднее значение признака x .

С помощью модуля «критерий Вилкоксона» статистического пакета *SPSS* можно доказать, что после вакцинации против гриппа показатели полярографической активности фильтрата сыворотки крови двух групп пациентов — контрольной и опытной (*A* и *B*) — отличаются друг от друга. Данный

модуль с непараметрическим тестом Вилкоксона был выбран не случайно, его использование наиболее целесообразно в случаях малых выборок.

Пакет *SPSS* предназначен для выполнения всех этапов статистического анализа данных. Он разработан в США в 60-е годы XX столетия в стенах Чикагского университета, и первоначально предназначался для реализации сугубо специфических проблем психологии, социологии, антропологии.

В настоящее время разработано 13 версий, и как показала практика — все они многоцелевые, взаимопроникающие и пригодны для анализа данных в областях науки и практики, достаточно далеких от психологии и социологии, на которые пакет *SPSS* был первоначально ориентирован.

В *SPSS* существует несколько десятков модулей, дополнительно подключаемых к пакету. Появляется возможность не только непосредственно обрабатывать данные, но и формировать отчеты о результатах обработки. Он имеет значительный арсенал графического представления результатов. К достоинствам пакета можно отнести русификацию версий, есть подробные руководства на web-странице Internet. Относительным недостатком пакета является узкий набор статистических методов, входящих в базовую версию и полное отсутствие ряда статистических алгоритмов, широко используемых в России. Графика (в сравнении с тем же пакетом *STATISTICA*) — слабое звено *SPSS*. И, тем не менее, он в числе лидеров на рынке статистических услуг как в России, так и во всем мире.

В табл. 1 представлены лабораторные данные колебаний показателей активности сыворотки крови при введении в организм человека противогриппозного биологического препарата, полученные в реальном клиническом опыте изучения эффективности вакцинации против гриппа.

Тем, у кого еще не установлен пакет *SPSS*, следует ознакомиться с руководствами по его использованию [1–3].

Таблица 1

Динамика показателей активности сыворотки крови при введении в организм человека противогриппозной вакцины (А — не привитые, В — привитые)

Номер пробы	А	В
1	18,5	25,5
2	16,0	21,0
3	20,5	22,5
4	18,0	22,0
5	15,5	19,5
6	16,0	19,5
7	17,5	26,0
8	15,5	24,0
9	16,6	22,2
10	18,0	21,1
11	19,9	19,9
12	17,7	22,2

Алгоритм решения задачи следующий:

1. Выдвигаем альтернативные гипотезы:

- нулевая — нет сдвига;
- ненулевая — есть сдвиг в показателях активности фильтрата сыворотки крови.

2. Запустив пакет *SPSS*, введем исходные данные в таблицу для анализа:

- выборка А — первый столбец;
- выборка В — второй.

3. Алгоритм решения задачи, начинаем с проверки на нормальность распределения по тесту Шапиро-Уилкса, вызвав и активизировав соответствующий модуль. Этот шаг обязателен, особенно для случаев, когда в разработку взяты малые выборки.

4. Далее для контроля правильности выбранного варианта обработки эмпирических данных, вызываем модуль теста Стьюдента (в некоторой степени он «конкурент» тесту Вилкоксона) и проверяем достоверность сдвига измеряемых параметров. Получаем

ответ: изменение активности фильтрата на грани «достоверность — недостоверность».

5. Модуль критерия Вилкоксона напротив показывает, что значение сдвига в оценке показателей достаточно высоко — 3,300 при двухстороннем уровне достоверности $p = 0,001$. Односторонний уровень достоверности еще выше: $p = 0,0005$, т. е. с высокой степенью достоверности можно утверждать, что нулевая гипотеза об отсутствии различий сравниваемых выборок должна быть отвергнута. Иными словами, вакцинация дает стойкий и однозначный эффект.

Из сказанного следует, что параллельное применение параметрического критерия Стьюдента в данном случае помощи исследователю, в силу малого объема выборки, оказать не смогло, а непараметрический критерий Вилкоксона, оказался мощнее. И позволил дать корректный ответ, что может служить залогом:

- успешности исследования на большем по объему экспериментальном материале;
- оправданность финансовых и труд затрат экспериментатора.

Резюмируя сказанное, отметим, что использование данного приема обработки эмпирического материала позволяет практически любому медицинскому работнику со среднемедицинским образованием получить результат анализа не только с высокой точностью, но и в предельно короткий срок — в течение 60 с, не считая времени ввода данных в память ЭВМ, тогда как ручная обработка занимает почти два часа и не исключает наличие ошибок.

Список литературы

1. Яноши Л. Теория и практика обработки результатов измерений. М.: Мир, 1968.
2. Bühl A., Zöfel P. SPSS Version 10. Einführung in die moderne Datenanalyse unter Windows / Бюль А., Цефель П. SPSS: искусство обработки информации / пер. с нем. М., СПб.: ООО DiaSoft, 2002.
3. SPSS Base 8.0. Руководство пользователя. СПСС Русь, 1998.