

Использование информационных технологий в преподавании статистики в гуманитарных вузах.

О.П. Страхова, А.А.Каблуков

Запорожский государственный медицинский университет

Ключевые слова: информатика, информационные технологии, статистика, статистическая обработка, обучение, учебный план, программы, шкала, методы.

Аннотация.

Некорректное применение статистических методов позволяет им посредством предоставления псевдонаучные заявления, которые могут нанести большой вред не только наука, но и вполне конкретным пациентам. Практика экспериментальных исследований разрабатывает некоторые идеи о том, как статистического анализа, которые являются стандартными в соответствующих областях.

В последние годы, медицина и биология вступили в новую фазу своего развития. Накопление больших объемов количественных данных и повсеместное распространение средств вычислительной техники облегчает применения сложных математических вычислительных методов в биологии и медицине.

История науки показывает следующие этапы разработки для создания доминирующей методологии и повышения уровень знаний предмета исследования, с применением компьютеров:

- эмпирическая стадия, которая характеризуется накоплением и описание фактов, их частичную систематизацию;
- теоретический этап - анализ и синтез накопленных фактических данных;
- количественный математический этап - на основе накопленных фактов определяют закономерности, создают математические модели изучаемых явлений и объектов.

Для авторов, которые не знакомы с математической теорией, но нуждаются в статистических расчетах, наиболее подходящим является использование готовых программных статистических пакетов. Решение задачи облегчается с распространением компьютерных сетей.

Resume.

Incorrect application of statistical methods makes them a means of giving pseudoscientific declarations that can cause great harm not only science, but also quite specific patients. The practice of experimental research develops some ideas about how to statistical analysis, which are standard in the respective areas.

In the last years, medicine and biology entered a new phase of its development. Accumulation of huge amounts of quantitative data and increased computing mathematization gives the availability of application the complex mathematical calculational methods in biology and medicine.

The history of science reveals the following development steps due to creating the dominant methodology and the High knowledge level of the research subject:

- empirical stage, characterized by the accumulation and description of facts, their partial systematization;
- theoretical stage - the analysis and synthesis of evidence accumulated in the form of individual concept that combines a relatively consistent theory;
- quantitative mathematical stage - based on accumulated facts investigated quantitative laws, mathematical models are studied phenomena and objects.

For authors who are not familiar with the mathematical theory, but require statistical calculations, the most suitable is to use ready-made software statistical packages. With the proliferation of computers and computer networks it becomes easy and convenient.

Доказательная медицина – это технология сбора, обобщения и анализа медицинской информации, которая позволяет формулировать научно обоснованные решения в медицине.

Доказательная медицина сегодня завоевывает твердые позиции в научном мире. Она отличается тем, что требует обязательно предоставлять доказательства правоты исследователя, полученные неопровержимыми и неподкупными методами, почти исключая возможность выполнения подгонки результатов под требуемый доказывающему ответ. На помощь исследователям приходят

статистические методы планирования экспериментов и обоснования выводов.

Статистическая обработка результатов медико-биологических экспериментов и данных повседневной медицинской практики сложна, многокомпонентна. Она тяжела для преподавания и понимания студентами медицинских вузов и многими авторами научных работ. Люди с гуманитарным образованием испытывают серьезные затруднения при освоении математической статистики, являющейся отраслью высшей математики, и основ теории вероятности, идущей рука об руку с математической статистикой.

Почти все медицинские колледжи и вузы СНГ преподают своим студентам ряд предметов, дающих представление о методах статистической обработки, - биостатистика, статистика, медстатистика, матстатистика и т.д. Это очень серьезные дисциплины, методический материал для которых готовят профессионалы с физико-математическим образованием. Для медиков он не всегда понятен, он лежит за пределами сформированного у них медицинского образа мышления [1,2]. Даже сдав на «хорошо» и «отлично» экзамены по соответствующим образовательным предметам, студенты не понимают, из каких соображений надо выбирать тот или иной метод статистической обработки данных [3]. Выучив последовательность действий при

выполнении статистической обработки результатов, они часто не могут грамотно интерпретировать полученные результаты статистических расчетов. Из всех методов сравнения во многих, в том числе диссертационных, работах исследователи выполняют только расчет коэффициента Стьюдента и, в лучшем случае, сравнивают полученный результат с табличным значением, не утруждаясь ни проверкой вида распределения имеющихся у них данных, ни правомерностью выбора метода сравнения данных. Нужно понимать, что каждый из методов математической статистики имеет свои возможности и ограниченную область применения.

Безграмотность в статистике приводит к ошибочным результатам, на которых базируются выводы об итогах проведенных исследований. То есть, вся проведенная работа разрушается из-за неверно выполненной статистической обработки данных.

Исправить это положение можно несколькими способами:

- увеличить количество часов, отводимых на углубленное изучение особенностей применения многих методов математической статистики;
- старательно разработать пошаговые алгоритмы и применять везде один-два метода для всех результатов всех исследований;

- предложить инструмент, позволяющий даже не очень грамотному в математике человеку верно выполнить статистические расчеты.

Каждый из перечисленных способов по-своему верен. Но для врача, специалиста совсем в нематематической области знаний, с нашей точки зрения, наилучшим будет предложение ознакомиться с инструментом, который возьмет на себя большую часть вычислений.

Под инструментом мы понимаем:

- метод обработки данных;
- программный пакет(программу) для проведения статистических расчетов.

На кафедре информатики Запорожского государственного медицинского университета для студентов проводятся занятия по различным аспектам медицинской информатики. В тоже время, учитывая сложности, связанные с введением дополнительных часов в учебный план подготовки специалистов, дополнительные знания в области основ математической статистики решили давать студентами в процессе выполнения лабораторных работ, в рамках, действующих в настоящее время учебных программ по дисциплине «Информатика»[4].

Цель лабораторных работ - ознакомление с широко распространенными программными пакетами, позволяющими

провести статистические расчеты. Методической основой были определены существующие программы статистической обработки данных – MS Excel и STATISTICA. MS Excel находится на каждом компьютере, в пакете MS Office. Программа STATISTICA используется во всем мире, в ней есть встроенные блоки с методами, рассчитанные на обработку именно медицинских данных [5,6].

При таком подходе, не требуется увеличения учебных часов, отводимых на изучение непрофильного предмета. Уменьшается объем материала, который студент должен освоить часто в ущерб основным, наиболее важным профессионально предметам медицинского вуза [7].

Для выбора метода обработки данных студент должен сосредоточиться на нескольких моментах. Так как выбираемый метод статистической обработки напрямую зависит от вида имеющихся у исследователя данных, то, прежде всего надо разобраться с терминами. Некоторые из этих понятий, например, популяция или генеральная совокупность, даются медикам ещё в начальный период их обучения в вузах [5]. Следующий шаг - это объяснение понятия выборки, которая представляет собой часть объектов популяции, генеральной совокупности, извлечённых из неё случайным образом. Важнейшим понятием статистики является понятие шкалы[8,9].

«Шкала» - инструмент, которым были проведены исследования; это набор алгоритмов, по которым результатам наблюдений присваиваются некоторые числовые значения.

В медицине и биологии часто результатом проведенного исследования являются количественные признаки. Большинство из них имеют эталоны для используемых шкал: массы, длины, времени и т.д. Значения других количественных признаков получаются путём вычислений, с использованием величин, имеющих эталоны. Для всех таких величин допустимы не только отношения "Равно - Не равно" и "Больше - Меньше", но также и арифметические операции. Для них допустимо вычисление таких статистик, как среднее арифметическое, ошибка среднего, дисперсия, стандартное отклонение и т.п. Эталон измеряемых величин необходим для выполнения условия эквидистантности, равноудалённости между соседними делениями шкалы. Он существует не для всех шкал.

В медицине наиболее часто используются такие шкалы:

- номинальная
- порядковая
- интервальная
- шкала отношений

Шкала наименований, номинальная шкала – это качественная, а не количественная шкала, она не содержит эталона для измерения, и

единиц измерения. Такими шкалами пользуются для обозначения пола, цвета, названия заболевания и т.п. Для значений этой шкалы можно использовать сравнения с заключениями «равно» и «не равно». Но для этой шкалы нет соотношений «больше» или «меньше», как нет и процедур использующих арифметические действия (сложение, вычитание и т.д.)

Для порядковой, или ранговой, или балльной, шкалы, в дополнение к отношениям «равно - не равно» вводится соотношение «больше» или «меньше». Здесь также недопустимо использование арифметических действий. Пример порядковой шкалы - шкала Апгар. Она содержит суммарные значения пяти критериев, каждый из которых оценивается субъективно, а не измеряется прибором.

Таблица 1. Шкала Апгар:

	0 баллов	1 балл	2 балла
Окраска кожного покрова	Генерализованная бледность или генерализованный цианоз	Розовая окраска тела и синюшная окраска конечностей (акроцианоз)	Розовая окраска всего тела и конечностей
Частота сердечных сокращений	Отсутствует	<100	>100
Рефлекторная возбудимость	Не реагирует	Реакция слабо выражена (grimаса, движение)	Реакция в виде движения, кашля, чиханья, громкого крика
Мышечный тонус	Отсутствует, конечности свисают	Снижен, некоторое сгибание конечностей	Выражены активные движения
Дыхание	Отсутствует	Нерегулярное, крик слабый (гиповентиляция)	Нормальное, крик громкий

На этом примере видно, что разность в 1 балл между 3 и 2 баллами не равна разности между 9 и 8 баллами, потому что в порядковой шкале отсутствует эквидистантность между ее делениями. Для измеренных в порядковой шкале величин допустимы только операции сравнения "Равно - Не равно" и "Больше - Меньше". Но не допустимы такие операции, как сложение, деление и т.д. Поэтому не имеет смысла понятие «Средний балл», «ошибка среднего» и т.п.

В шкале интервалов нет физически аргументированного нулевого значения. Пример такой шкалы – шкала времени. За точку отсчёта можно выбирать удобный в каждом конкретном случае момент. Например, в христианских странах начало летоисчисления ведут от рождества Христова, а в мусульманских – от переезда пророка Мухаммеда в Медину.

В шкале отношений имеется естественное нулевое значение, которое устанавливается в зависимости от точности измерений. В такой шкале мы измеряем вес. Для нее характерно свойство эквидистантности: между весами в 43 кг и 44, 1018 и 1019, 1 и 2 кг разница одинаковая: 1 кг.

Определившись с видом шкал для измерения исследуемых признаков и получив результаты эксперимента, можно перейти к их первичной статистической обработке. На этом этапе требуется

знакомство с некоторыми понятиями математической статистики и теории вероятности. Это трудная задача для человека с медицинским или гуманитарным образованием. Для помощи в ее решении существуют стандартные пакеты прикладных программ. К ним относятся выше упомянутые пакеты STATISTICA, StatSoft, Inc, и MS Excel из пакета MicroSoft Office. Но для осмысленного выбора метода статистической обработки имеющихся данных нужно сделать несколько стандартных шагов:

- определение числовых характеристик переменных и оценка их точности и надежности;
- определение статистических рядов распределения переменных и оценка их соответствия теоретическим законам распределения;
- оценка значимости различия показателей в зависимых и несвязанных выборках.

Это необходимо, чтобы сделать правильный выбор статистического метода обработки результатов исследований.

Существует две группы методов обработки числовых данных: параметрические и непараметрические. Параметрические методы, к числу которых относится широко известный, но имеющий очень строгие ограничения на применение критерий Стьюдента, могут использоваться только для нормально распределенных данных со

сходными дисперсиями. Непараметрические методы применяются для оценки различий ненормально распределенных данных. Непараметрические методы часто являются аналогами соответствующих параметрических, но могут применяться для оценки данных без учета вида их распределения, для качественных и количественных признаков, если переменные являются порядковыми или содержат повторяющиеся значения. Все перечисленные в Таблице 2 параметрические и непараметрические методы есть в пакете Статистика на соответствующих ветках меню, открывающегося при выборе элемента Главного меню «Анализ» (в русифицированном варианте) или «Statistica».

Таблица 2. Соответствие между параметрическими и непараметрическими статистическими методами.

№	Задание	Параметрический Метод	Непараметрический аналог
1	Оценка двух независимых выборок	T-критерий Стьюдента для двух независимых выборок	Критерий серий Вальда-Вольфовица U критерий Манна-Уитни Двухвыборочный критерий Колмогорова-Смирнова Критерий Вилкоксона парных сравнений критерий знаков
2	Сравнение нескольких независимых выборок	Дисперсионный анализ	Ранговый дисперсионный анализ Краскела-Уоллиса Медианный тест

3	Оценка двух зависимых выборок	T-критерий для зависимых выборок	Критерий знаков Критерий Вилкоксона парных сравнений
4	Две зависимые категориальные переменные (представлены в виде частот попавших в определенные категории)		Хи-квадрат Макнемара
5	Более двух категориальных зависимых переменных	Дисперсионный анализ (ANOVA) с повторными измерениями.	Ранговый дисперсионный анализ Фридмана Q критерий Кохрена (используется и для оценки изменений частот (долей))
6	Оценка зависимости (связь) между двумя переменными	Корреляционный анализ: стандартный коэффициент корреляции Пирсона	Статистики Спирмена R Тау Кендалла Коэффициент Гамма Для категориальных: Хи-квадрат Фи коэффициент Точный критерий Фишера. Критерий зависимости между несколькими переменными Коэффициент конкордации Кендалла.
7	Оценка частоты очень редких событий	-	распределение Пуассона

Вычисление выборочных характеристик, применение спектра статистических методов можно легко и быстро проводить с помощью этих программ. Они обладают удобными инструментами визуализации

данных – построение двух- и трехмерных графиков, создание многих видов диаграмм делает статистический анализ данных очень удобным и понятным.

Но окончательный выбор метода для обработки результатов каждого эксперимента все равно принадлежит исследователю.

Конечно, чтобы получить навыки работы с этими программами, студент должен потратить определенное время, но объем материала, который он должен усвоить, при этом значительно уменьшается и упорядочивается. Информационные технологии очень способствуют оптимизации учебного процесса по непрофильным предметам в медицинских вузах.

Данный материал представляет собой первый этап обучения студентов осознанному применению статистических методов для анализа данных, построения модели и прогнозирования возможных сценариев развития любых процессов. Это позволит перевести статистическую обработку данных для медицины и фармации из разряда непонятных манипуляций в разряд осмысленного научного этапа, обязательного при переходе к доказательной медицине.

Такой способ изложения и закрепления учебного материала позволит студенту после недлительной и несложной подготовки самостоятельно ориентироваться в разнообразии существующих статистических методов обработки данных, грамотно применять

статистические методы оценивания в эксперименте или при проведении маркетингового исследования.

Целесообразно развить курс статистической обработки данных в направлении представлений о классификационных методах и кластерном анализе, применимых к неколичественным данным, полученным с использованием номинальных или балльных шкал. Также ввести в курсы «Медицинская информатика» и «Компьютерные технологии в фармации» несколько занятий с практическим использованием существующих программ статистической обработки данных, освещающих разные аспекты применения статистических методов для анализа результатов экспериментов, анализа и прогнозирования бизнес-процессов и процессов медицинского администрирования.

ЛИТЕРАТУРА

1. Савельев А.Я. Модель формирования специалиста с высшим образованием на современном этапе. - М., 2005. – 72 с.
2. Геращенко Л.А., Воронов В.К., Ким Де Чан. Проблема познавательных затруднений студентов вузов в изучении цикла математических и естественнонаучных дисциплин //Успехи современного естествознания.- 2007.- №2 .
3. Коровина В.А. , Пащенко В.П.. Некоторые особенности преподавания математической статистики в медицинском вузе// Международный журнал медицинской практики .- 2006.- №2.
4. Бубнов В.А. О методах применения информационных технологий в учебном процессе на гуманитарных факультетах // Вестник МГПУ. — М., 2001. - №1.
5. Халафян А. STATISTICA 6.0. Статистический анализ данных: Учебник./А.Халафян– М.:ООО "Бином-Пресс", 2007.–512 с.
6. Орлов А.И. Прикладная статистика. М.: Издательство «Экзамен», 2004.
7. Шахов Б.Е. Подготовка медицинских кадров в свете новых ФГОС // Аккредитация в образовании: Электронный журнал об образовании [Электронный ресурс].
URL: http://www.akvobr.ru/fgos_podgotovka_medicinskih_kadrov.html .
8. Типы статистических шкал.
http://www.datuapstrade.lv/rus/spss/section_5/2/.
9. Общая теория статистики: Учебник/ Под ред. Башиной О.Э., Спирина А.А. - 5-е изд., доп. и перераб. - М.: Финансы и статистика, 2001. - 440 с