

**М. Ш. Мухаматзанова, М. А. Захарова, В. А. Вельш**

Новосибирский государственный медицинский университет

## **О ВЫБОРЕ МЕТОДА СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ ДЛЯ МЕДИКО-СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЙ**

УДК 614.1:614.2

В настоящее время значительно повышаются требования к математической культуре научных сотрудников и врачей-специалистов, требуется применение адекватного математического аппарата для доказательства полученных результатов. От качества используемых данных, от их достоверности и точности зависит достоверность результатов анализа. Поэтому изучение прикладной статистики является неотъемлемой частью формирования культуры научных сотрудников, студентов, аспирантов, планирующих и проводящих исследования в социологии медицины.

*Ключевые слова:* дизайн исследования, метод описательных статистик, доказательная статистика, критериальный анализ, база данных.

**M. Sh. Mukhamatzanova, M. A. Zakharova, A. A. Velsh**

## **ON THE CHOICE OF METHOD OF STATISTICAL DATA PROCESSING IN MEDICOSOCIOLOGICAL RESEARCH**

The type of data distribution in the study should be evaluated with the help of mathematic-statistical criterion. Parametric tests are used when the obtained data distribution is considered as normal. The type of distribution for nonparametric tests is of no importance. The suggested algorithm will help the investigator to classify the facts, develop a study design and choose the best technique of processing obtained information. Therefore, mathematic statistics as applied to medico-sociological study may help to define the purpose, develop the study design, choose the statistic method for revelation of statistically significant difference during an analysis of obtained results and drawing the conclusion.

*Key words:* study design, descriptive statistics method, evidential statistics, criteria analysis, data base.

В последние годы широкое распространение программных средств для статистического анализа данных в различных прикладных областях, включая медицинские приложения, тем не менее не снимает необходимости владения хотя бы основами математической статистики. От пользователя требуется умение грамотно выбрать подходящие статистические процедуры, знание их возможностей и ограничений, корректная и осмысленная интерпретация полученных результатов. Произвольное применение статистических методов может привести к ложным выводам [3].

Перед статистической наукой встают актуальные проблемы дальнейшего совершенствования системы показателей, приемов и методов сбора, обработки, хранения и анализа статистической информации. Статистическая работа состоит в том, чтобы собрать числовые данные о массовых явлениях, обработать их, представить в форме, удобной для анализа, проанализировать и интерпретировать полученные результаты. Сбор данных лежит в основе всего исследования.

Недостаточное внимание к планированию исследований влечет за собой нехватку данных для формирования статистически значимого вывода

после окончания этапа сбора информации. В этом случае даже самые сложные математические методы анализа полученных результатов не смогут дать необходимой исследователю информации [2].

### **ЦЕЛЬ РАБОТЫ**

Разработать алгоритм, демонстрирующий наглядно ту последовательность действий, которую следует выполнять исследователю при описании и анализе результатов научного исследования.

Для реализации обозначенной цели потребуются решение следующих задач:

1. Проанализировать наиболее часто применяемые критерии в медико-социологических исследованиях на предмет получения значимой информации с позиции доказательной медицины.

2. Разработать схему дизайна медико-социологического исследования на основании выбранных методов обработки получаемой в ходе исследования информации.

Предлагаемый алгоритм (рис. 1) поможет исследователю систематизировать знания, разработать дизайн собственного исследования и выбрать оптимальное сочетание методов обработки полученной информации.



Рис. 1. Алгоритм анализа данных научного исследования

Описание данных научно-практического медицинского исследования включает в себя ряд единиц наблюдения (больных, лабораторных подопытных животных и т.п.), характеризующихся определенными признаками. Количественные признаки выражаются числовыми значениями, например возраст, рост, вес, давление. Порядковые признаки могут быть измерены в шкалах (например, школьные оценки, степень тяжести заболевания: легкая, средняя, тяжелая). Качественные признаки характеризуют некоторое состояние объекта, но не могут быть измерены количественно (например, пол, профессия, диагноз). Включая признак в описание данных, исследователь должен достаточно четко представить, для чего этот признак понадобится ему в дальнейшем. Это необходимо, чтобы избежать перегруженности информации, тем не менее база данных должна быть достаточно полной и информативной.

В зависимости от типа данных выбираем метод описательных статистик:

- для количественных данных, подчиняющихся нормальному закону распределения, рассчитывают среднее и стандартное отклонение. Представляются в виде  $M \pm StD$  (например, концентрация вещества «Х» составила  $(14,5 \pm 0,5)$  мг/л);

- для количественных данных, не подчиняющихся нормальному закону распределения рассчитывают медиану и квартили (перцентили). Представляются в виде  $Me (Q1-Q3)$  (например, возраст участников исследования составил  $10 (5-11)$  лет);

- для качественных данных рассчитываются частоты или проценты. Представляются либо только относительные величины (встречаемость заболевания «Х» в исследуемой совокупности составила 25%), либо абсолютные и относительные вместе (выявлено 16 случаев заболеваний «Х», что составило 25%).

После описательных статистик следует этап доказательной статистики. На этом этапе исследователь выбирает метод критериального анализа, в зависимости от изучаемого признака и вида исследования [1].

Одним из главных вопросов перед исследователем является метод получения им информации. Зачастую выбор происходит неосознанно, по аналогии с другими исследователями, коллегами, что порождает ряд ложных результатов. Для того чтобы ориентироваться в выборе необходимого критерия существует несколько алгоритмов (С. Гланц, В. В. Безляк, Н. М. Жилина и др.). Алгоритм выбора критерия представлен на рис. 2, 3 и зависит от типа данных.

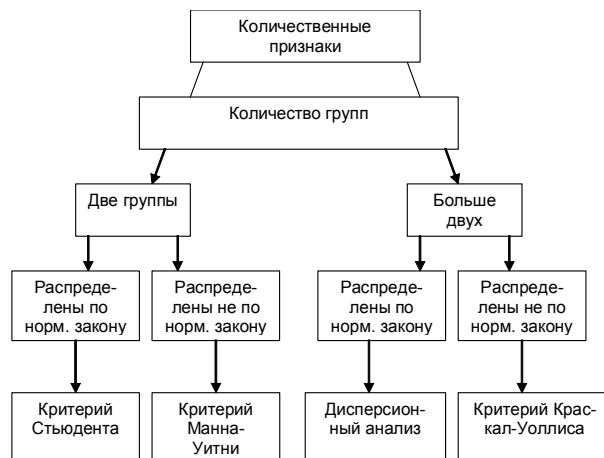


Рис. 2. Алгоритм сравнения групп по количественному признаку

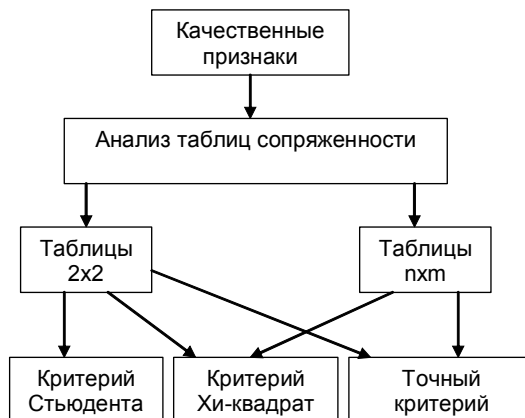


Рис. 3. Алгоритм сравнения групп по качественному признаку

Множество критериев, приводимых обычно в учебниках по математической статистике, и сложное описание процедур их вычисления часто смущают исследователя. Многие из них используются достаточно редко. Каждый исследователь (и научный руководитель в том числе) предпочитает статистические критерии, исходя из своих знаний, опыта, типа задачи и вида данных, которые подлежат обработке. При выборе математико-статистического критерия нужно ориентироваться также на тип распределения данных, который получен в исследовании. Параметрические критерии используются в том случае, когда распределение полученных данных рассматривается как нормальное. Для непараметрических критериев тип распределения данных не имеет значения.

При осуществлении поиска различий в медико-социологических показателях у испытуемых, имеющих те или иные особенности, могут использоваться критерии для выявления различий в уровне исследуемого признака или в его распределении. Для определения значимости различий в проявлении признака в психологических исследованиях часто используются такие показатели, как парный критерий Вилкоксона, U-критерий Манна-Уитни, критерий хи-квадрат ( $\chi^2$ ), точный критерий Фишера, биномиальный критерий.

Осуществляя поиск взаимосвязи показателей у одних и тех же испытуемых, следует использовать коэффициенты корреляции. Связь величин друг с другом и их зависимость часто характеризуется коэффициентом линейной корреляции Пирсона и коэффициентом ранговой корреляции Спирмена.

Структура данных, а также их взаимосвязь выявляются факторным анализом. Во многих медико-социологических исследованиях центральное место занимает анализ изменчивости признака под влиянием каких-либо контролируемых факторов, или, другими словами, оценка влияния разных факторов на изучаемый признак. Для математической обработки данных в таких задачах может использоваться U-критерий Манна-Уитни, критерий Краскела-Уоллиса, T-критерий Вилкоксона, критерий Фридмана.

Следует заметить, что для исследования влияния, а тем более взаимовлияния нескольких факторов на изучаемый параметр полезнее может ока-

заться дисперсионный анализ. Исследователь исходит из предположения, что одни переменные могут рассматриваться как причины, а другие — как следствия. Переменные первого рода считаются факторами, а переменные второго рода — результативными признаками. В этом отличие дисперсионного анализа от корреляционного, в котором предполагается, что изменения одного признака просто связаны с определенными изменениями другого [3].

Важно обратить внимание на ограничения, которые имеет каждый критерий. Если один критерий не подходит для анализа имеющихся данных, всегда можно найти какой-либо другой, возможно, изменив тип представления самих данных. Прежде чем обрабатывать данные, полезно проверить, существуют ли в пособии, которым вы пользуетесь, критические значения, соответствующие количеству и типу ваших данных. В противном случае вас может ждать разочарование, когда ваши подсчеты окажутся напрасными по причине отсутствия в таблице критических значений при объеме выборки, которая у вас была.

Не зная статистических методов, исследователь вычисляет только самые элементарные статистические показатели, может неправильно применить критерий, не полностью использует возможности статистической обработки.

Таким образом, математическая статистика применительно к проведению медико-социологических исследований может помочь в формулировании цели, разработке дизайна, выборе статистического метода для выявления статистически значимого различия, непосредственно в проведении анализа полученных результатов, формировании заключения.

## ЛИТЕРАТУРА

1. Гланц С. Медико-биологическая статистика — М.: Практика, 1999. — 461 с.
2. Жилина Н. М. Приложения математической статистики к медицинским научным исследованиям: учебное пособие — Новокузнецк: Изд-во МОУ ДПО ИПК, 2005. — 41 с.
3. Сергиенко В. И., Бондарева И. Б. Математическая статистика в клинических исследованиях. — М.: ГЭОТАР МЕДИЦИНА, 2000. — 256 с.