

## НЕКОТОРЫЕ ВОПРОСЫ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ В МОРФОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

С. Н. Семенов, О. С. Семенова

ГБОУ ВПО «Воронежская государственная медицинская академия им. Н.Н.Бурденко»  
Минздрава России, г. Воронеж, Россия

В работе рассматриваются вопросы статистического анализа результатов морфологических исследований. Обсуждаются вопросы вариабельности параметров медико-биологических объектов и форма представления статистических данных.

*Ключевые слова:* статистика, морфологическое исследование, вариабельность параметров биологических объектов.

© S. N. Semenov, O. S. Semenova, 2012

On Statistical Analysis of the Data in Morphological Studies

The article covers the questions of statistical analysis of data in morphological studies. The variability of parameters of medico-biological objects and the form of their presentation are discussed.

*Keywords:* statistics, morphological study, the variability of the parameters of biological objects

Настоящей краткой публикацией мы предлагаем начать обсуждение важных, на наш взгляд, вопросов статистического анализа результатов изучения морфологических объектов.

Исследования, особенно выполняемые на микроскопическом и субмикроскопическом уровне, являются в методическом плане весьма трудоемкими. Нередко их конечная результативность оказывается ограниченной вследствие того, что при анализе первичных данных, полученных в результате морфометрии, количественной гистохимии, не используются возможности полноценного статистического анализа. Не обсуждая вопросы применения расширенных статистических исследований к данным количественной морфологии, рассмотрим некоторые аспекты анализа первичных данных на уровне так называемой «описательной статистики» и представления результатов этого анализа в научных публикациях.

В последние 40–50 лет накопление огромных массивов количественных показателей усилило математизацию биологии и медицины. Это ознаменовало реальный переход в развитии биологии, как и большинства наук, к конечному этапу — количественному, математическому, когда на базе накопленных фактов исследуются количественные закономерности, создаются математические модели исследуемых явлений и объектов. Для биологии и медицины данный этап начался

еще в XIX веке работами Ф. Гальтона, а широкое распространение различных статистических компьютерных программ вооружило современных исследователей мощными инструментами статистического анализа различной степени сложности.

Одной из важнейших составляющих качества научных работ является корректность использования методов статистического анализа результатов экспериментальных и клинических исследований, полнота описания как использованных методик анализа, так и его результатов.

Сам по себе термин «вариационная статистика», часто используемый авторами, содержит элементы тавтологии, поскольку без вариации нет и статистики, а статистический метод обязательно предусматривает и изучение вариации.

Исходя из идей Н. Винера и разработок К. Шеннона, У. Р. Эшби открыл закон, названный законом необходимого разнообразия, ставший одним из основных законов кибернетики — науки об общих закономерностях процессов управления и передачи информации в различных системах, будь то машины, живые организмы или общество.

Для управления состоянием кибернетической системы необходим регулятор, ограничивающий разнообразие возмущений, которые могут разрушить систему. При этом регулятор допускает такое их разнообразие, которое необходимо и

полезно для системы. При допустимом разнообразии состояний кибернетической системы  $P_c$  и разнообразии возмущений  $P_e$  количество разнообразия регулятора соответствует

$$P_p = P_e / P_c.$$

Эта формула является одной из количественных форм выражения закона необходимого разнообразия. В логарифмической форме этот закон имеет вид

$$\log P_p = \log P_e / P_c \text{ или } \log P_p = \\ = \log P_e - \log P_c.$$

Обозначив соответствующие логарифмы разнообразия как информационные содержания систем, получим

$$I_e = I_p + I_c.$$

Из формулы следует, что сумма информационных содержаний системы и регулятора равна информационному содержанию внешних возмущений.

Согласно закону необходимого разнообразия, информационное содержание системы должно быть не меньше, чем разнообразие внутренних и внешних негативных воздействий. Наиболее эффективно обеспечение безопасности (самосохранение) происходит через прогрессивное либо устойчивое развитие, которое реализуется в ходе самоорганизации материальных систем.

Переходя от изложенного понятийного аппарата кибернетики к биологическим терминам, мы можем обсуждать “допустимое состояние системы” как “норму”, “регулятор состояния системы” как “физиологические системы обеспечения гомеостаза”, “разнообразие элементов системы” как “полиморфизм”. В этом случае становится очевидным, что разнообразие, мерой которого может выступать информация, является обязательным атрибутом биологической системы, а характеристики разнообразия – показателями ее состояния.

Таким образом, показатели, характеризующие вариабельность изучаемых признаков, являются не менее важными и содержательными, чем средние величины, представлением которых ограничиваются авторы большинства отечественных научных публикаций.

Наиболее распространенной формой представления результатов описательной статистики в отечественных публикациях являются характеристики выборок в виде

$M \pm t$ . При этом не учитывается, что информация о величине стандартной ошибки средней арифметической содержит лишь промежуточные сведения о доверительных границах выборочной средней, которые определяются величиной доверительного интервала

$$\Delta = \pm tm,$$

где  $t$  – критерий Стьюдента, выбранный для заданной доверительной вероятности в соответствии с объемом выборки (числом степеней свободы). Но и в этом случае доверительные границы выборочной средней не дают представления об особенностях распределения признака и его вариабельности.

Очевидно, что развернутому статистическому анализу должна предшествовать диагностика типа распределения, как минимум – оценка его соответствия нормальному распределению, которая чаще всего проводится с использованием критерия  $\chi^2$  (Пирсона). Важную информацию о типе распределения представляют моменты третьего и четвертого порядка – коэффициенты асимметрии и эксцесса, величины которых можно получить при элементарном использовании стандартных электронных таблиц MS Excel. Хотя, нередко, бывает достаточно построить график распределения вариант выборки, чтобы обнаружить его отклонения от нормального распределения. Необходимо учитывать, что нормальное распределение признаков, вопреки распространенному заблуждению, встречается в медико-биологических исследованиях не так уж часто.

Диагностика типа распределения, с одной стороны, необходима для определения возможности использования параметрической статистики, с другой – позволяет получить принципиально важную информацию об изучаемых объектах и признаках и определить дальнейшие направления их изучения, например, при анализе результатов цитометрии обнаружить в исследуемой клеточной популяции несколько типов клеток в случае би- или полимодального распределения.

Полную оценку вариабельности признака дает величина стандартного (среднеквадратичного) отклонения ( $\sigma$  или  $SD$  – *Standard Deviation*). Не случайно, в большинстве зарубежных научных публикаций базовые результаты приводятся в виде

$$M \pm 1,96\sigma,$$

поскольку диапазон  $\pm 1,96\sigma$  в случае нормального распределения охватывает ~95,5% всех возможных значений признака. В этой связи уместно напомнить “правило трех сигм”: ~99,7 % всех единиц наблюдения при нормальном распределении находятся в пределах  $M \pm 3\sigma$ . Один из выводов этого правила: разность  $(M-3\sigma)$  не может лежать в области отрицательных значений, с чем нередко приходится встречаться при детальном анализе научных публикаций.

Возвращаясь к вопросу о вариативности как важном атрибуте биологических объектов и явлений, хочется напомнить о таком элементарном показателе как коэффициент вариации, который, элементарно рассчитываясь,

$$Cv(\%) = \frac{\sigma}{M} \times 100\%,$$

дает наглядное представление о соотношении стандартного отклонения и выборочной средней и об изменениях этих соотношений в процессе динамического исследования.

Наряду с затронутыми вопросами, связанными с вариативностью, авторы рассчитывают на обсуждение таких важных проблем, как оценка необходимого числа наблюдений, выбор оптимальных критериев значимости, статистические подходы к классификации в морфологии и др.

#### Информация об авторах

**Семенов Сергей Николаевич** – д.м.н., профессор, зав. кафедрой нормальной анатомии человека ГБОУ ВПО “Воронежская государственная медицинская академия им. Н. Н. Бурденко” Минздрава России. 394036, г. Воронеж, ул. Студенческая, д. 10. E-mail: sns7250@mail.ru

**Семенова Ольга Сергеевна** – к.м.н., ассистент кафедры эндокринологии ГБОУ ВПО “Воронежская государственная медицинская академия им. Н. Н. Бурденко” Минздрава России. 394036, г. Воронеж, ул. Студенческая, д. 10.

Поступила в редакцию 2.11.2012 г.