



Колдаев Владимир Михайлович
Доктор биологических наук, профессор. Более 40 лет заведовал кафедрой медицинской физики ТГМУ, опубликовал два учебника по физике для студентов лечебного и стоматологического факультетов. Ныне ведущий научный сотрудник лаборатории лекарственных растений Федерального научного центра ДВО РАН Министерства науки и высшего образования, имеет более 200 научных печатных работ в отечественных и зарубежных изданиях.



Кропотов Александр Валентинович
Доктор медицинских наук, профессор. Почти 30 лет заведовал кафедрой фармакологии ТГМУ, в настоящее время профессор этой кафедры, имеет более 150 научных публикаций.



Издательство «Медицина ДВ»
690950 г. Владивосток, пр-т Острякова, 4
Тел.: (423) 245-56-49. E-mail: medicinaDV@mail.ru



ОСНОВНЫЕ ПРИЕМЫ СТАТИСТИКИ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Учебная литература ТГМУ

В.М. Колдаев, А.В. Кропотов

ОСНОВНЫЕ ПРИЕМЫ СТАТИСТИКИ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Учебное пособие



Владивосток
Медицина ДВ
2019

Министерство науки и высшего образования РФ
Федеральный научный центр биоразнообразия наземной биоты Восточной Азии
Дальневосточного отделения российской академии наук
Министерство здравоохранения Российской Федерации
Тихоокеанский государственный медицинский университет

В.М. Колдаев, А.В. Кропотов

ОСНОВНЫЕ ПРИЕМЫ СТАТИСТИКИ В МЕДИКО-БИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Учебное пособие



Владивосток
Медицина ДВ
2019

УДК [61+57]:311(0.75.8)
ББК 51.1(РОС),02я73
К 601

*Издано по рекомендации редакционно-издательского совета
Тихоокеанского государственного медицинского университета*

Рецензенты:

Игнатюк В.А. – доктор физико-математических наук, профессор,
кафедра Информационных технологий и систем ВГУЭС

Хотимченко Ю.С. – доктор биологических наук, профессор,
директор Школы биомедицины ДВФУ

Маняхин А.Ю. – кандидат биологических наук, старший научный
сотрудник ФНЦ биоразнообразия ДВО РАН

Колдаев, В.М.

К 601 Основные приемы статистики в медико-биологических исследо-
ваниях: учебное пособие / В.М. Колдаев, А.В. Кропотов. – Владивосток:
Медицина ДВ, 2019. – 104 с. Илл. 13. Библ. 4.

ISBN 978-5-98301-181-6

В учебном пособии просто и доступно описаны основные приемы современной статистики, наиболее широко используемые в анализе результатов медико-биологических исследований. По каждому методу приводятся наглядные конкретные примеры статистической обработки данных из экспериментальной практики, а также даются рекомендации по их применению, что, несомненно, способствует повышению грамотности, расширению приложений статистики в биологии и медицине. Пособие предназначено для студентов, аспирантов и соискателей медико-биологических специальностей. Безусловно, книгой может пользоваться и более широкая аудитория.

УДК [61+57]:311(0.75.8)
ББК 51.1(РОС),02я73

ISBN 978-5-98301-181-6

© В.М. Колдаев, А.В. Кропотов, 2019
© «Медицина ДВ», 2019

СПИСОК СОКРАЩЕНИЙ

$НГ_K, ВГ_K$	– нижняя и верхняя границы классов
a, b	– границы доверительного интервала
D	– дисперсия
K	– номер класса
m	– абсолютная частота, численность группы
Mo, Me	– мода и медиана
n	– объем выборки, количество объектов в выборке
n_T	– количество групп
n_K	– количество классов
N	– число степеней свободы
Q_1, Q_2, Q_3	– 1-й, 2-й, 3-й квартиль, процентиль
p	– относительная частота, частость, вероятность
P_0	– доверительная вероятность
R	– коэффициент парной линейной корреляции
R_s	– коэффициент корреляции Спирмена
i	– номер измерения, номер значения признака
T -критерий	– критерий Манна-Уитни
t -критерий	– критерий Стьюдента
$s_{\bar{x}}$	– ошибка среднего
s_x^2	– выборочная дисперсия

s_x	– выборочное среднее квадратичное отклонение
W -критерий	– критерий Уилкоксона
X, Y	– обозначения, названия, имена признаков
x, y	– значения признаков X, Y
\bar{x}	– среднее значение
x_{max}, x_{min}	– максимальное и минимальное значения признака, лимиты
α	– уровень значимости нулевой гипотезы
η	– корреляционное отношение
μ	– математическое ожидание
ν	– число ограничений
Σ	– сумма
σ	– среднее квадратичное отклонение, стандартное отклонение
χ^2	– критерий хи-квадрат

Светлой памяти
Киры Александровны Мещерской
посвящается

ПРЕДИСЛОВИЕ

Статистика (лат. *status* – состояние) – вид деятельности по сбору, систематизации и анализу информации об окружающей действительности. Статистика включает несколько направлений: демографическое, экономическое и др. Настоящее пособие посвящено *математической статистике*, занимающейся описанием, обработкой и анализом результатов научных исследований.

По математической статистике имеется немало руководств, однако 40 лет общения авторов со студентами и соискателями ученых степеней, а также более чем 25-летний опыт работы авторов в диссертационных советах, показывают, что статистические методы зачастую используются или недостаточно грамотно, не по назначению, или ошибочно. Недооценка важности применения статистических методов, наверное, связана с тем, что биологи и медики сами редко занимаются статистическим обоснованием своих исследований, поскольку их познания в этой области, как правило, весьма скромны. Обычно вся статистическая часть в научном исследовании перепоручается «консультантам», которые нередко действительно разбираются в вопросах статистики, но имеют довольно смутное представление о медико-биологической проблематике. По-видимому, выход состоит в том, чтобы медики и биологи сами занялись статистическим анализом своих данных, поскольку именно они знают суть и цель исследования. Повышение статистической грамотности важно еще и потому, что на самом деле математическая статистика – единственный путь обоснования достоверности полученных в исследовании результатов. Никакая наука без статистики совершенно невозможна. Все это вместе взятое побудило нас к составлению настоящего пособия, цель которого показать молодым начинающим исследователям, как должным образом обставить эксперимент, оценить достоверность результатов, провести убедительные сравнения и сделать на их основе соответствующие правильные выводы.

Пособие состоит из десяти глав. В 1-й, вводной, главе приводятся некоторые необходимые понятия теории вероятностей – основы матема-

тической статистики. Вторая глава посвящена приемам статистического описания и представления результатов исследований. В главах с 3-й по 8-ю показаны разные современные методы решения основных статистических задач – сравнения однородных групп данных. В 9-й главе описан статистический анализ зависимости типа «эффект-доза», довольно часто встречающийся при изучении закономерностей реакций организма на воздействие разнообразных факторов (радиации, химических веществ и др.), и анализ выживаемости, имеющий значение в первую очередь для онкологии. Главы (кроме первой) включают наглядные примеры реализации того или иного статистического метода и контрольные вопросы с типовыми задачами для самостоятельной практической отработки навыков статистики. В вычислениях конкретных примеров авторы ссылаются на функции приложений к электронным таблицам без разъяснений, т.к. надеются, что читатели имеют достаточно высокую компьютерную грамотность.

В приложениях приводятся справочные таблицы, в конце пособия даны ответы на задачи, а также указана рекомендуемая дополнительная литература по медико-биологической статистике.

Пособие предназначено для практических занятий студентов по специальностям «Клиническая психология» (37.05.01), «Лечебное дело» (35.05.01), «Медико-профилактическое дело» (32.05.01), «Медицинская биохимия» (30.05.01) и других, а также для аспирантов, соискателей ученых степеней медико-биологического профиля при обработке материалов экспериментальных исследований. Безусловно, пользоваться книгой может и более широкая аудитория.

Авторы

ГЛАВА 1

НЕКОТОРЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ

В природе, в окружающей действительности многие явления многократно повторяются в определенных условиях. При этом, несмотря на постоянство условий, явления не совсем совпадают друг с другом, как говорится, *рассеиваются*. Это обусловлено или свойствами самих явлений, или какими-либо неучитываемыми, или нередко неизвестными факторами. Рассеяние не дает возможности заранее *точно предсказать* конкретное явление. Однако закономерности многократно повторяющихся рассеивающихся явлений позволяет *изучать*, описывать и анализировать специальный раздел математики *теория вероятностей*, на которой основывается статистика.

В теории вероятностей сложилась определенная терминология. Так, реализацию комплекса условий, воспроизводимого сколь угодно много число раз, принято называть *испытанием*. Для однотипных испытаний условия должны быть строго одинаковыми, хотя, конечно, на практике осуществить это почти никогда не удастся. Тождественность условий испытаний принимается *приближенно*. Комплекс условий может создавать человек по своему желанию, например, в исследовательской деятельности, тогда испытания называют опытами или *экспериментами*.

В настоящем разделе приводятся базовые понятия теории вероятностей, которые потребуются в дальнейшем.

1.1. Случайное событие и его вероятность

Явления, происходящие при испытаниях, обозначают как *события*. Если о событии *точно заранее нельзя* сказать произойдет оно или нет, то такое событие называют *случайным*.

Количество всех событий в данной системе испытаний обозначают обычно буквой *n* (англ. *number* – количество). Однотипные события в испытаниях происходят с разной частотой – одни чаще, другие реже. Количество появлений некоторого события в испытании обозначают буквой *m* (англ. *modicum* – малое количество) и называют *абсолютной*

частотой события. Отношение абсолютной частоты к общему числу испытаний (m/n) называют *относительной частотой* или *частостью*. Например, в процессе испытаний по определению группы крови у $n = 100000$ людей такое событие, как первая группа крови появилось $m = 40000$ раз, значит, в этом случае для этого события относительная частота $m/n = 40000/100000 = 0,4$. Относительную частоту можно выразить и в процентах: $m \times 100 = 0,4 \times 100 = 40\%$. Всегда m не больше n , поэтому относительная частота не превышает единицу.

Если событие появляется довольно редко, а его относительная частота мала в данной системе испытаний, то говорят, что такое событие *маловероятно*, а о событиях, появляющихся очень *часто*, говорят, что они имеют *высокую вероятность*. При увеличении числа испытаний относительная частота стремится к некоторому пределу – числу, обусловленному самой природой исследуемых событий.

Предел относительной частоты, когда число испытаний стремится к бесконечности, называется *вероятностью* P (англ. *probability* – возможность осуществления) события. Вероятность всегда положительна и принимает значения от нуля до единицы: $0 \leq P \leq 1$.

При больших сериях испытаний, когда n велико, относительная частота приближается к вероятности и может служить ее оценкой тем более точной, чем больше число испытаний.

1.2. Случайные величины

В эксперименте обычно исследуют признаки объектов. Признаки обозначают прописными буквами, например, температура T , а значения признаков строчными буквами, например, $t = 36,8^\circ\text{C}$. В общем случае для признака X в эксперименте в результате нескольких измерений получают набор величин, так называемый *эмпирический ряд*:

$$x_1, x_2, \dots, x_p, \dots, x_n,$$

где i – номер измерения, n – количество измерений.

При этом каждое наблюдаемое значение признака считается *событием*. Если точное значение x заранее предсказать невозможно, то такое значение по аналогии считается случайным. Величина, принимающая в результате испытаний то или иное значение в зависимости от *случайного исхода* испытания, называется *случайной величиной*. В этом смысле все измеряемые величины являются *случайными*, поскольку все измерительные инструменты имеют погрешность, и результат измерения, зависящий от многих (порой неизвестных) случайных обстоятельств, заранее точно предсказать невозможно. Случайные величины принято

подразделять на *дискретные* (счетные), принимающие целые значения, например, количество объектов, и *непрерывные*, принимающие как целые, так и дробные значения, например, масса тела. Случайные величины (как и случайные события) можно характеризовать относительной частотой появления в данной системе измерений.

1.3. Распределение случайных величин, основные параметры

Обычно в эксперименте одни значения величины признака появляются чаще, а другие реже, другими словами – одни значения *более вероятны*, а другие *менее вероятны*, т.е. полученные в эксперименте случайные величины *распределяются* по вероятностям. Появление каждого значения случайной величины в данной системе измерений имеет свою вероятность.

Совокупность значений x_i случайной величины и соответствующих вероятностей p_i их появления называется *законом распределения*.

Распределение характеризуется несколькими параметрами (числами). Для решения практических задач наибольшее значение имеют следующие два параметра.

Во-первых, такое значение случайной величины, которое встречается в испытании наиболее часто, с *наибольшей вероятностью*, и получило название *математического ожидания*. Математическое ожидание (μ) равно сумме произведений вероятностей (p_i) появления тех или иных значений случайной величины на сами значения (x_i):

$$\mu = p_1x_1 + p_2x_2 + \dots + p_ix_i + \dots + p_nx_n = \sum_{i=1}^n p_ix_i,$$

где i – номер величины, n – количество величин, Σ – знак суммирования.

Оценкой математического ожидания служит *среднее значение* \bar{x} случайной величины, равное сумме всех полученных значений, деленной на их количество:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_i + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$

Во-вторых, одни из полученных в эксперименте величин всегда находятся близко от математического ожидания, другие подальше или совпадают с ним, т. е. величины имеют тот или иной разброс или *рассеяние*. Поэтому вторым важнейшим параметром распределения яв-