

*Российская академия медицинских наук
Волгоградский научный центр РАМН*

Д.А. Новиков, В.В. Новочадов

**СТАТИСТИЧЕСКИЕ МЕТОДЫ В
МЕДИКО-БИОЛОГИЧЕСКОМ
ЭКСПЕРИМЕНТЕ
(ТИПОВЫЕ СЛУЧАИ)**

**Волгоград
2005**

УДК 519.6
ББК 65, 74
Н 73

Новиков Д.А., Новочадов В.В. Статистические методы в медико-биологическом эксперименте (типовые случаи). Волгоград: Издательство ВолГМУ, 2005. – 84 с.

ISBN 5-9652-0011-0

Книга доктора технических наук, профессора Д.А. Новикова и доктора медицинских наук, профессора В.В. Новочадова содержит «рецепты» применения статистических методов в типовых случаях анализа экспериментальных данных в медико-биологических исследованиях.

Работа рассчитана, в первую очередь, на студентов медицинских вузов, биологических факультетов университетов, а также аспирантов и соискателей.

Рецензенты: А.И. Орлов – д.т.н., проф., президент Российской ассоциации статистических методов
А.И. Щеголев – д.м.н., проф., зав. патологоанатомическим отделом Института хирургии им. А.В. Вишневского РАМН

УДК 519.6
ББК 65, 74

ISBN 5-9652-0011-0

© Новиков Д.А., Новочадов В.В., 2005
© Волгоградский государственный
медицинский университет, 2005

СОДЕРЖАНИЕ

1. Введение.....	4
2. Структура медико-биологического эксперимента.....	6
3. Элементы теории измерений.....	13
3.1. Шкалы измерений.....	14
3.2. Допустимые преобразования.....	17
3.3. Применение шкал измерений в медико-биологическом эксперименте.....	20
3.4. Агрегированные оценки.....	24
3.5. Комплексные оценки.....	25
4. Анализ использования статистических методов в статьях и диссертационных исследованиях по биологии и медицине.....	27
5. Типовые задачи анализа данных в медико-биологическом эксперименте.....	30
5.1. Исходные данные: пример 1.....	30
5.2. Исходные данные: пример 2.....	35
5.3. Типовые задачи анализа данных.....	35
5.4. Использование компьютера.....	39
6. Методы обработки данных и примеры.....	41
6.1. Описательная статистика.....	41
6.2. Сравнение двух выборок.....	49
6.2.1. Общие подходы к определению достоверности совпадений и различий.....	49
6.2.2. Методика определения достоверности совпадений и различий для экспериментальных данных, измеренных в шкале отношений.....	51
6.2.3. Методика определения достоверности совпадений и различий для экспериментальных данных, измеренных в порядковой шкале.....	57
6.2.4. Алгоритм выбора статистического критерия.....	63
6.3. Исследование зависимостей.....	67
6.3.1. Корреляционный анализ.....	67
6.3.2. Дисперсионный анализ.....	71
6.3.3. Регрессионный анализ.....	73
6.4. Снижение размерности.....	77
6.5. Классификация и прогноз.....	79
7. Заключение.....	81
Литература.....	82

1. ВВЕДЕНИЕ

Экспериментальные исследования играют существенную роль во всех науках. Можно утверждать, что, чем менее строгой является наука, тем более значимую роль в ней играет *эксперимент*¹. Действительно, в науках сильной версии (см. [22]), использующих математический аппарат, многие результаты могут быть получены и обоснованы теоретически, на базе существующего эмпирического материала. В науках же слабой версии эксперимент зачастую является единственным способом подтверждения справедливости гипотезы и результатов теоретического исследования, так как отсутствие общепринятой аксиоматики и адекватного формального аппарата не позволяет привести должного обоснования, не прибегая к эксперименту. Например, можно ли априори сказать, что то или иное новое воздействие в медицине более эффективно, в сравнении с ранее известными и применяемыми до него? Вряд ли – пока оно не будет апробировано, и результаты его применения не будут сопоставлены с результатами применения традиционных воздействий, никаких выводов сделать нельзя.

При планировании и подведении результатов эксперимента существенную роль играют *статистические методы*, которые дают возможность²:

- компактно и информативно описывать результаты эксперимента;
- устанавливать степень достоверности сходства и различия исследуемых объектов на основании результатов измерений их показателей;
- анализировать наличие или отсутствие зависимости между различными показателями (явлениями);
- количественно описывать эти зависимости;
- выявлять информативные показатели;

¹ *Эксперимент – общий эмпирический метод исследования, суть которого заключается в том, что явления и процессы изучаются в строго контролируемых и управляемых условиях.*

² *Отметим, что в настоящей работе речь идет о применении статистических методов именно в медико-биологическом эксперименте. Другая обширная область применения статистики в медицине – так называемая медицинская статистика, описывающая деятельность медицинских учреждений и показатели здоровья населения – остается вне нашего рассмотрения.*

- классифицировать изучаемые объекты и прогнозировать значения их показателей и характеристик, и др.

Анализ статей и диссертационных исследований по биологическим и медицинским наукам (см. четвертый раздел настоящей работы) позволяет констатировать, что на сегодняшний день складывается следующая картина. С одной стороны, большинство исследователей четко представляет, что использование статистических методов необходимо (хотя бы потому, что это является общепринятым требованием в науке), и существует обширная литература по теоретической и прикладной статистике. С другой стороны, статистические методы при планировании и обработке результатов медико-биологического эксперимента либо не используются вообще, либо часто используются некорректно.

Объяснений этому несколько. Во-первых, необходимо признать, что существующая литература в большинстве своем ориентирована на людей, имеющих математическое или техническое образование, и практически недоступна специалистам с медицинским или биологическим образованием (немногочисленные книги по математической статистике для «не-математиков» подавляют своим объемом и, все-таки, наверное, слишком сложны). Во-вторых, класс *типовых* (наиболее распространенных, массовых) *задач* (случаев) анализа данных, возникающих в медико-биологических экспериментальных исследованиях, достаточно узок, и для эффективного решения этих задач вовсе не требуется знакомства со всем богатейшим арсеналом статистических методов. Все это приводит к тому, что исследователи боятся использовать статистические методы, а если и используют, то на уровне «шаманских заклинаний», особо не понимая, что и как надо делать, что они делают и какие результаты получают.

Поэтому основной целью настоящей работы является изложение «рецептов» применения статистических методов для решения типовых задач анализа данных в медико-биологическом экспериментальном исследовании³. Желаящим же получить более полное представление о том, как и в каких ситуациях, какие методы можно и нужно использовать, порекомендуем ознакомиться с перечисленными в списке литературы многочисленными учебниками и

³ Следует признать, что иногда мы были вынуждены немного жертвовать корректностью изложения в пользу его доступности.

книгами, содержащими методики и опыт применения статистических методов в различных областях научного знания, в том числе – в медицине и биологии [6, 8, 10, 21, 34, 35].

Дальнейшее изложение имеет следующую структуру. Во втором разделе описана модель медико-биологического эксперимента и алгоритм действий исследователя при организации эксперимента и обработке его результатов. Третий раздел содержит минимально необходимые сведения из теории измерений относительно того, какого рода данные существуют, и какие операции к ним применимы. В четвертом разделе проводится анализ использования статистических методов в статьях и диссертационных исследованиях по медицине и биологии, что позволяет сформулировать в пятом разделе типовые задачи анализа данных в медико-биологических экспериментальных исследованиях. Шестой раздел – основной – включает описание методов решения этих задач и соответствующие примеры.

Авторы признательны за ценные замечания и рекомендации рецензентам: доктору технических наук, профессору А.И. Орлову и доктору медицинских наук, профессору А.И. Щеголеву, а также кандидату технических наук С.А. Чижову и В.В. Засканову.

2. СТРУКТУРА МЕДИКО-БИОЛОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Объектом исследований в медико-биологических науках являются живые организмы, как правило – в срезе их взаимодействия с комплексом различных факторов внешней среды, часть из которых относится к болезнетворным. Другие факторы направлены на профилактику или лечение болезни. Основным подходом в этом направлении по-прежнему остается эксперимент, а основной его целью – установление закономерностей возникновения различных болезней, механизмов их развития, разработка и проверка эффективности новых методов профилактики и лечения.

В процессе познания в биологии и медицине научный эксперимент выполняет следующие основные функции:

- выступает средством получения новых научных данных;

- является способом выделения общего в серии сходных явлений, обоснования закономерностей, формирования гипотез;
- выступает средством проверки гипотез и теорий, критерием их истинности, основой для выдвижения новых гипотез;
- является относительным гарантом перед применением новых методов диагностики и лечения болезней в клинической практике.

По своей продолжительности различают острый эксперимент (например, воспроизведение острой кровопотери, ожоговой травмы, отравления угарным газом и т.д.) и хронический, который подразумевает длительное наблюдение за экспериментальным объектом (например, животным: введение электродов, операция изолированного желудка по И.П. Павлову, трансплантация органов и тканей и т.д.).

Эксперимент позволяет производить такие воздействия, например, на органы животного (травма, отравление, облучение), которые невозможны и недопустимы в клинике. В эксперименте на животных возможно испытание новых лекарственных препаратов. Некоторые болезни (например, психические заболевания, ряд опухолей, некоторые иммунопатологические состояния) пока в эксперименте воспроизвести не удалось. Оно и понятно. Сложность устройства человеческого организма несравнимо даже с организмом человекообразных обезьян. По образному выражению С. Хокинга один только человеческий мозг по сложности своего устройства превосходит всю Вселенную. Поэтому воспроизведение, например, психических нарушений возможно только на аналогичных структурах. Из всего многообразия животного мира наиболее адекватным экспериментальным субстратом является человеческий мозг. Но такие эксперименты запрещены Конвенцией о правах человека. Поэтому в данном случае приходится довольствоваться либо дорогостоящими точными методами клинического *наблюдения* (магнитно-резонансная томография, цветовая доплерография и т.п.), либо пытаться моделировать подобные состояния на компьютере. Другим важным фактором, обуславливающим невозможность многих патологических состояний, является формирование последних в условиях социального окружения и других факторов, специфичных для человеческого общества.

Этапами постановки и проведения эксперимента являются: формирование рабочей гипотезы, определение цели и задач исследования, выбор конкретных методик, непосредственное проведение эксперимента (серии опытов), фиксация и анализ данных эксперимента, обсуждение и выводы.

Рабочая гипотеза – первое, с чем сталкивается экспериментатор. Это – попытка объяснить какой-либо процесс на основе уровня знаний, имеющегося в настоящий момент. Нельзя просто ставить опыт, не зная, какие предположительно изменения в состоянии объекта в ту или другую сторону можно зафиксировать. Конечно, можно сказать, что в погоне за каким-либо массивом данных исследователь может пропустить более глубокие закономерности. Однако именно последовательное системное изучение патологического процесса может в конечном итоге привести к получению более-менее целостной картины его развития, сложенной из отдельных «кусочков» исследовательских работ. В противном случае произойдет простое распыление сил, средств и времени без получения сколь-нибудь существенного результата. Подобное «расчленение» исследования на несколько отдельных работ совершенно необходимо, чтобы в каждом конкретном случае абстрагироваться от множества посторонних влияний и оценить вклад каждого фактора в суммарный поток изменений.

Основу проведения эксперимента составляет получение максимально точной (адекватной) модели болезни. Моделирование – это воспроизведение структурно-функционального комплекса болезни в более упрощенной форме для выяснения причин, условий и механизмов ее развития, разработки методов лечения и профилактики.

Требования к модели:

- аналогия причин и условий, вызывающих болезнь и примененных при ее моделировании;

- сходство функциональных и морфологических изменений болезни и ее модели на системном, органном, клеточном и молекулярном уровнях;

- сходство в развитии типичных осложнений;

- эффективность лечения одними и теми же воздействиями.

Например, в патологии, в зависимости от степени упрощения и уровня изучения, можно моделировать: типовые патологические

процессы, отдельные нозологические формы болезни, нарушения отдельных органов и систем, нарушения деятельности отдельных клеток и их компонентов, молекулярные нарушения.

При моделировании важным вопросом является выбор животного. Например, язву желудка нецелесообразно вызывать у кроликов – у них часто наблюдаются спонтанные язвы, инфекционный процесс – у крыс – они резистентны к инфекциям. Наиболее часто для проведения экспериментов используются различные лабораторные животные: белые крысы и мыши, кролики, морские свинки, а также беспородные собаки и кошки, реже – овцы, свиньи и другие животные.

Конкретные экспериментальные методики, применяемые при моделировании, могут включать «выключение» функции какого-либо органа или ткани (удаление органа, перерезка нерва, пережатие сосуда, фармакологическая блокада и др.), раздражение (стимуляция) функции, моделирование изменений регуляции и баланса химических веществ-метаболитов. Возможно использование метода «включения» какого-либо вещества, чаще нейромедиаторной или гормональной природы, для анализа роли этой субстанции в патогенезе данного состояния. Например, таковым может быть моделирование эндотоксикоза компонентами бактериальных клеток для сравнения формирующегося состояния с патологией, вызванной бактериальной взвесью и доказательства роли именно липополисахаридного компонента в индукции эндогенной интоксикации. В последние десятилетия широкое распространение получил метод тканевых культур в оценке различных воздействий (в том числе и фармакологических препаратов) на изолированные клетки.

Большой вопрос, стоящий перед исследователем – это правомерность экстраполяции данных, полученных экспериментальным путем на животных на человеческий организм. Мы можем говорить только об условной адекватности используемых моделей, так как полная идентичность проявлений заболеваний у животных и у человека принципиально недостижима. Этому препятствуют иные законы, царящие в человеческом обществе. В человеческом обществе в определенной мере отсутствует естественный отбор, что предполагает наличие определенного числа вырождающихся функций и приобретения патологических реакций. Помимо этого,

чрезвычайно важную роль в формировании патологии у человека играет социальный строй, адекватное моделирование которого практически невозможно. Связанным с социальной функцией является и наличие второй сигнальной системы, играющей колоссальную роль в процессах компенсации и адаптации. Неслучайно еще со времен Гиппократата было замечено, что у врача есть три метода воздействия на больного, причем на первом месте, еще до «травы» и «ножа», стоит «слово». Кроме того, некоторые заболевания человека, такие, как корь и скарлатина, вообще невозможно воспроизвести на животных ввиду отсутствия у них того молекулярного субстрата, на котором данная болезнь может существовать.

Выполнение экспериментальных исследований строго регламентировано законодательством. Все работы проводятся только в рамках утвержденных учебных и научных программ, специалистами с медико-биологическим образованием, болезненные и травмирующие процедуры выполняются с необходимым уровнем обезболивания. Все эксперименты проводятся не из праздного любопытства, а для изучения механизмов развития заболеваний, актуальных для человека, это первый, необходимый, но не самодостаточный этап познания болезни.

Рассмотрим следующую *модель медико-биологического эксперимента* [24]. Пусть имеется некоторый *объект*, изменение *состояния* которого исследуется в ходе эксперимента. В качестве объекта может выступать отдельный организм, группа лабораторных животных (нередко в литературе – серия) и т.д. Состояние объекта измеряется⁴ теми или иными показателями⁵ (*характеристиками*) по *критериям*⁶, отражающим его существенные характеристики. Примерами критериев являются: выраженность интоксикации, выживаемость в группе животных на определенный период

⁴ Измерение – «процесс определения какой-либо мерой величины чего-либо». Величина – «то (предмет, явление и т.д.), что можно измерить, исчислить». Другими словами, величина – мера некоторого множества, относительно элементов которого имеют смысл утверждения – больше, меньше, равно. Мера – «единица измерения». Все определения здесь и далее взяты, если не оговорено особо, из словаря русского языка С.И. Ожегова.

⁵ Показатель – «то, по чему можно судить о развитии и ходе чего-либо».

⁶ Критерий – «1) средство для вынесения суждения; стандарт для сравнения; правило для оценки; 2) мера степени близости к цели».

после начала опыта и т.д., примерами характеристик – температура, активность тех или иных ферментов в биологических жидкостях, количественные показатели структуры внутренних органов и т.д.

Эксперимент заключается в целенаправленном *воздействии* на объект, призванном изменить его определенным образом. Собственно, это воздействие – его состав, структура, свойства и т.д. – и есть результат теоретического (теоретической части) исследования. Примерами воздействия являются любые методы воздействия на болезнь с целью ее излечения, включая потенциальные лекарственные средства, немедикаментозные воздействия, хирургические манипуляции.

Следовательно, при проведении медико-биологического эксперимента необходимо обосновать, что состояние объекта изменилось, причем в требуемую сторону. Но этого оказывается недостаточно. Ведь нужно доказать, что изменения произошли именно в результате произведенного воздействия.

Действительно, например, на утверждение о том, что температура тела у экспериментального животного снизилась в результате использования нового испытуемого вещества, можно всегда возразить, – а, может быть, она снизилась бы сама, без каких-либо нововведений, или в результате каких-либо других воздействий? Аналогично, на утверждение о том, что скорость и степень снижения температуры у животных, которым вводился новый препарат, отличаются от того, как это происходило у животных, леченных с применением традиционных препаратов, можно возразить, – а, может быть, сама группа имела до начала эксперимента внутренние отличия, позволившие ей продемонстрировать подобные «успешные» результаты.

Таким образом, для того, чтобы выделить в явном виде результат целенаправленного воздействия на исследуемый объект, необходимо взять аналогичный объект и посмотреть, что происходит с ним в отсутствии воздействий.

Традиционно эти два объекта⁷ называют соответственно *экспериментальной группой* (иногда основной) и *контрольной группой* (или группой сравнения).

На Рис. 1 представлена в общем виде структура эксперимента (двойными пунктирными стрелками отмечены процедуры сравнения⁸ характеристик объектов).

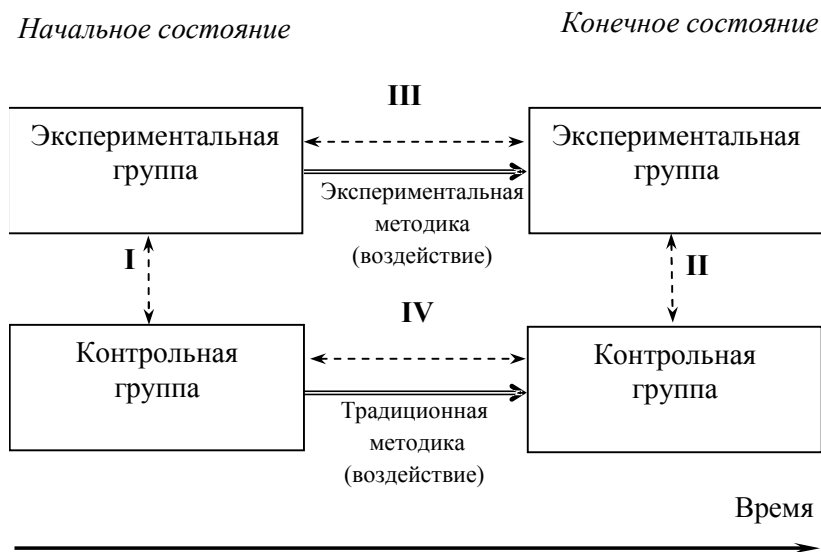


Рис. 1. Структура эксперимента

Констатации (в результате сравнения III – см. Рис. 1) различий начального и конечного состояний (динамики) экспериментальной группы недостаточно – быть может, аналогичные изменения происходят и с контрольной группой, что может быть установлено сравнением IV. Поэтому **алгоритм** действий исследователя заключается в следующем:

⁷ Некоторые изоцирненные статистические методы используют схемы с большим числом контрольных выборок, но малого объема (вплоть до единицы – при скользящем контроле) [29].

⁸ При этом мы по умолчанию подразумеваем, что методы измерения характеристик объектов одинаковы.

- 1) На основании сравнения I установить совпадение⁹ начальных состояний экспериментальной и контрольной группы;
- 2) Реализовать воздействие на экспериментальную группу¹⁰;
- 3) На основании сравнения II установить различие конечных состояний экспериментальной и контрольной группы.

После выполнения трех перечисленных шагов можно приступить к изучению зависимостей между различными характеристиками объектов (см. ниже).

Легко видеть, что, выполняя перечисленные шаги¹¹, мы, фактически, косвенным образом реализуем процедуру сравнения III, исключая влияние общих для экспериментальной и контрольной группы условий и воздействий.

Спрашивается, а где же место статистических методов? Роль их заключается в том, чтобы корректно и достоверно обосновать и объяснить совпадение или различие состояний контрольной и экспериментальной группы. Однако прежде чем описывать эти методы, давайте рассмотрим, что понимается под «состоянием объекта» и как это состояние измерять. Проблемами измерений занимается теория измерений, поэтому приведем минимально необходимые сведения из этой теории.

3. ЭЛЕМЕНТЫ ТЕОРИИ ИЗМЕРЕНИЙ

Информация, имеющаяся о начальных и конечных состояниях экспериментальной и контрольной группы, определяется проведенными измерениями. Любое измерение производится в той или иной *шкале*, и выбранная шкала определяет тип получающихся

⁹ Если говорить корректно, то с точки зрения математической статистики совпадение установить невозможно – можно установить различие или отсутствие статистически значимого различия. Точнее, при статистической проверке гипотез по статистическим данным выбирают одну из двух гипотез – есть различие или нет различия (точнее, различие не обнаружено).

¹⁰ При выполнении данного шага необходимо быть уверенным, что и экспериментальная, и контрольная группы находятся в одинаковых условиях, за исключением целенаправленно изменяемых исследователем.

¹¹ Эксперимент может следовать и более сложной, но укладывающейся в рамки описанной идеологии, схеме – например, характеристики контрольных и экспериментальных групп могут измеряться и сравниваться неоднократно, в различные моменты времени.

данных и множество операций, которые можно с этими данными осуществлять. Поэтому в настоящем разделе дается краткий обзор свойств основных шкал измерений, а затем описываются наиболее распространенные типы экспериментальных данных и методы их первоначальной обработки (до применения статистических методов).

3.1. ШКАЛЫ ИЗМЕРЕНИЙ

Состояние объекта оценивается по тем или иным критериям. В качестве критериев могут выступать: выживаемость животных, степень интоксикации, сохранение жизненно важных функций и т.д.

Оценки измеряются в той или иной шкале. *Шкала* (условно говоря, шкала – это множество возможных значений оценок по критериям) – числовая система, в которой отношения между различными свойствами изучаемых явлений, процессов переведены в свойства того или иного множества, как правило – множества чисел.

Различают несколько *типов шкал*. Во-первых, можно выделить *дискретные шкалы* (в которых множество возможных значений оцениваемой величины конечно – например, оценка в баллах – «1», «2», «3», «4», «5») и *непрерывные шкалы* (например, концентрация вещества в ммоль/л или активность фермента в сыворотке крови в мКат/л). Во-вторых, выделяют *шкалы отношений*, *интервальные шкалы*, *порядковые (ранговые) шкалы* и *номинальные шкалы* (шкалы наименований) – см. Рис. 2, на котором отражена также мощность шкал – то есть их «разрешающая способность» (иногда выделяют и иные шкалы, например, шкалу разностей, в которой измеряется (календарное) время).

Рассмотрим, следуя в основном [23, 32], свойства четырех основных типов шкал, перечисляя их в порядке убывания мощности.

Шкала отношений – самая мощная шкала. Она позволяет оценивать, во сколько раз один измеряемый объект больше (меньше) другого объекта, принимаемого за эталон, единицу. Для шкал отношений существует естественное начало отсчета (нуль), но нет естественной единицы измерений.

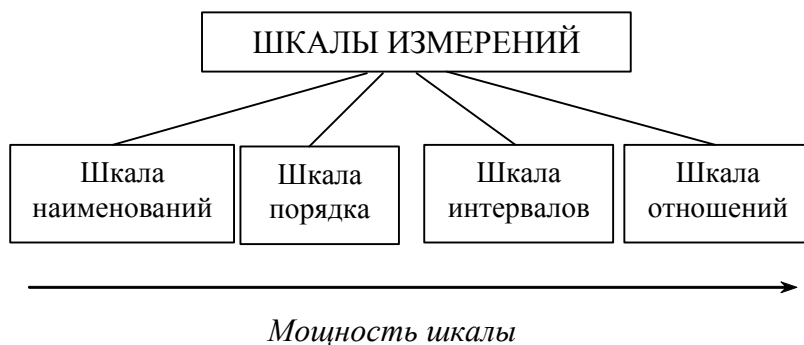


Рис. 2. Классификация шкал измерений

Шкалами отношений измеряются почти все физические величины – время, линейные размеры, площади, объемы, сила тока, мощность и т.д. В медико-биологических исследованиях шкала отношений будет иметь место, например, когда измеряется время появления того или иного признака после начало воздействия (порог времени, в секундах, минутах), интенсивность воздействия до появления какого-либо признака (порог силы воздействия в вольтах, рентгенах и т.п.). Естественно, к шкале отношений относятся все данные в биохимических и электрофизиологических исследованиях (концентрации веществ, вольтажи, временные показатели электрокардиограммы и т.п.). Сюда же, например, относятся и количество правильно или неправильно выполненных «заданий» в различных тестах по изучению высшей нервной деятельности у животных.

Шкала интервалов применяется достаточно редко и характеризуется тем, что для нее не существует ни естественного начала отсчета, ни естественной единицы измерения. Примером шкалы интервалов является шкала температур по Цельсию, Реомюру или Фаренгейту. Шкала Цельсия, как известно, была установлена следующим образом: за ноль была принята точка замерзания воды, за 100 градусов – точка ее кипения, и, соответственно, интервал температур между замерзанием и кипением воды поделен на 100 равных частей. Здесь уже утверждение, что температура 30⁰С в три раза больше, чем 10⁰С, будет неверным. В шкале интервалов сохраняется отношение длин интервалов. Можно сказать: температу-

ра в 30⁰С отличается от температуры в 20⁰С в два раза сильнее, чем температура в 15⁰С отличается от температуры в 10⁰С.

Порядковая шкала (шкала рангов) – шкала, относительно значений которой уже нельзя говорить ни о том, во сколько раз измеряемая величина больше (меньше) другой, ни на сколько она больше (меньше). Такая шкала только упорядочивает объекты, приписывая им те или иные баллы (результатом измерений является нестрогое упорядочение объектов).

Например, так построена шкала твердости минералов Мооса: взят набор 10 эталонных минералов для определения относительной твердости методом царапания. За 1 принят тальк, за 2 – гипс, за 3 – кальцит и так далее до 10 – алмаз. Любому минералу соответственно однозначно может быть приписана определенная твердость. Если исследуемый минерал, допустим, царапает кварц (7), но не царапает топаз (8), то соответственно его твердость будет равна 7. Аналогично построены шкалы силы ветра Бофорта и землетрясений Рихтера.

Шкалы порядка широко используются в педагогике, психологии, медицине и других науках, не столь точных, как, скажем, физика и химия. В частности, повсеместно распространенная шкала школьных отметок в баллах (пятибалльная, двенадцатибалльная и т.д.) может быть отнесена к шкале порядка. В медико-биологических исследованиях шкалы порядка встречаются сплошь и рядом и подчас весьма искусно замаскированы. Например, для анализа свертывания крови используется тромботест: 0 – отсутствия свертывания в течение времени теста (а через минуту?), 1 – «слабые нити», 2 – желеподобный сгусток, 3 – сгусток, легко деформируемый, 4 – плотный, упругий, 5 – плотный, занимающий весь объем и т.п. Понятно, что интервалы между этими плохо отличимыми и очень субъективными позициями произвольны. В этом случае фраза «Тромботест у исследуемых животных повышался в среднем с 3,3 до 3,7» выглядит абсурдной. Масса подобных шкал все еще встречается в экспериментальной токсикологии, экспериментальной хирургии, экспериментальной морфологии.

Частным случаем порядковой шкалы является *дихотомическая шкала*, в которой имеются всего две упорядоченные градации – например, «выжил после эксперимента», «не выжил».

Шкала наименований (номинальная шкала) фактически уже не связана с понятием «величина» и используется только с целью отличить один объект от другого: номер животного в группе или присвоенный ему уникальный шифр и т.п.

3.2. ДОПУСТИМЫЕ ПРЕОБРАЗОВАНИЯ

Результаты измерений необходимо анализировать, а для этого нередко приходится строить на их основании производные (вторичные) показатели, то есть, применять к экспериментальным данным то или иное преобразование. Используемая шкала определяет множество преобразований, которые допустимы для результатов измерений в этой шкале (подробнее см. публикации [20, 31, 32, 37, 41] по теории измерений).

Начнем с наиболее слабой шкалы – *шкалы наименований*, которая выделяет попарно различимые классы объектов. Например, в шкале наименований измеряются значения признака «пол»: «мужской» и «женский». Эти классы будут различимы независимо от того, какие различные термины или знаки для их обозначений будут использованы: «особи женского пола» и «особи мужского пола», или «female» и «male», или «А» и «Б», или «1» и «2», или «2» и «3» и т.д. Следовательно, для шкалы наименований применимы любые взаимно-однозначные преобразования, то есть сохраняющие четкую различимость объектов (таким образом, самая слабая шкала – шкала наименований – допускает самый широкий диапазон преобразований).

Отличие *порядковой шкалы* (шкалы рангов) от шкалы наименований заключается в том, что в шкале рангов классы (группы) объектов упорядочены. Поэтому произвольным образом изменять значения признаков нельзя – должна сохраняться упорядоченность объектов (порядок следования одних объектов за другими). Следовательно, для порядковой шкалы допустимым является любое монотонное преобразование. Например, если проба тромботеста у животного А – 5 баллов, а у животного Б – 4 балла, то их упорядочение не изменится, если мы число баллов умножим на одинаковое для всех животных положительное число, или сложим с некоторым одинаковым для всех числом, или возведем в квадрат и т.д. (например, вместо «1», «2», «3», «4», «5» используем соответст-

венно «3», «5», «9», «17», «102»). При этом изменятся разности и отношения «баллов», но упорядочение сохранится.

Для *шкалы интервалов* допустимо уже не любое монотонное преобразование, а только такое, которое сохраняет отношение разностей оценок, то есть линейное преобразование – умножение на положительное число и добавление постоянного числа. Например, если к значению температуры в градусах Цельсия добавить 273°C , то получим температуру по Кельвину, причем разности любых двух температур в обеих шкалах будут одинаковы. А если от шкалы Цельсия перейти к шкале Фаренгейта, то для любых четырех температур отношение разности первой и второй к разности третьей и четвертой будут одинаковы в обеих шкалах.

И, наконец, в наиболее мощной шкале – *шкале отношений* – возможны лишь преобразования подобия – умножения на положительное число. Содержательно это означает, что, например, отношение масс двух предметов не зависит от того, в каких единицах измерены массы – граммах, килограммах и т.д.

Суммируем сказанное в Табл. 1, которая отражает соответствие между шкалами и допустимыми преобразованиями.

Табл. 1. Шкалы и допустимые преобразования

Шкала	Допустимое преобразование
Наименований	Взаимно-однозначное
Порядковая	Строго возрастающее
Интервальная	Линейное
Отношений	Подобия

Как отмечалось выше, результаты любых измерений относятся, как правило¹², к одному из основных (перечисленных выше) типов шкал. Однако получение результатов измерений не является самоцелью – эти результаты необходимо анализировать, а для этого нередко приходится строить на их основании *производные показатели*. Эти производные показатели могут измеряться в других шкалах, нежели чем исходные. Например, можно для оцен-

¹² *Результатами измерений могут быть и более сложные данные – ранжировки, последовательности и т.д., встречающиеся в медико-биологических исследованиях чрезвычайно редко, поэтому их рассмотрение выходит за рамки настоящей работы – см. список литературы.*

ки степени регенерации суставного хряща после его повреждения применять 100-балльную шкалу (один из авторов участвовал в подобных разработках). Но она слишком детально, и ее можно перестроить в пятибалльную («1» – от «1» до «10»; «2» – от «10» до «30» и т.д.), или двухбалльную (например, положительная оценка – все, что выше 70 баллов, отрицательная – 70 и меньше). Следовательно, возникает проблема – какие преобразования можно применять к тем или иным типам исходных данных. Другими словами, переход от какой шкалы к какой является корректным. Эта проблема в теории измерений получила название *проблемы адекватности*.

Для решения проблемы адекватности можно воспользоваться свойствами взаимосвязи шкал и допустимых для них преобразований, так как отнюдь не любая операция при обработке исходных данных является допустимой. Так, например, такая распространенная операция, как вычисление среднего арифметического, не может быть использована, если измерения получены в порядковой шкале [20, 32]. Общий вывод таков – всегда возможен переход от более мощной шкалы к менее мощной, но не наоборот (например, на основании оценок, полученных в шкале отношений, можно строить балльные оценки в порядковой шкале, но не наоборот).

Завершая обсуждение шкал измерений, в качестве отступления отметим, что мы рассматриваем процесс обработки результатов измерений, но вовсе не затрагиваем проблемы, связанные, во-первых, с процедурой измерений (то есть с тем, каким образом получается информация об объекте), во-вторых, с тем, какого рода информация представляет интерес с точки зрения проводимого исследования, и, наконец, в-третьих, с тем, что понимать под «улучшением» или «ухудшением» состояния исследуемого объекта, то есть, каковы критерии эффективности (подобные содержательные аспекты находятся вне компетенции математики – статистические методы позволяют лишь установить и обосновать сходство или различие объектов, наличие или отсутствие связи между показателями и т.д., а как их интерпретировать – вопрос предметной области – соответствующих разделов биологии и медицины).

Не останавливаясь на том очевидном требовании, что для сравнения результатов измерений ко всем объектам должны при-

меняться одинаковые процедуры измерений (например, нельзя сравнивать результаты выполнения двумя различными группами крыс двух различных тестов двигательной активности), отметим, что отдельной и чрезвычайно интересной областью исследований является выбор показателей, которые наиболее адекватно, и, в то же время, емко отражают изучаемые свойства объекта. К этой содержательной области относятся задачи построения тестов, выбора методик оценки процессов и т.д. Кроме того, необходимо подчеркнуть, что проблема адекватности возникает не только при переходе от одной шкалы к другой, но и при выборе шкалы для получения первоначальных оценок – непосредственной информации об объекте. И здесь опять справедлив вывод о том, что шкала должна быть адекватна – если она слишком мощная, то возможен большой произвол (в том числе – при измерении качественных характеристик в шкале отношений), если слишком слабая, то происходят потери информации (в том числе – при измерении количественных показателей в номинальной шкале). Например, наверное, нецелесообразно, с одной стороны, оценивать результаты свертывания крови – тромботеста – в 100-балльной шкале, а с другой стороны, оценивать выраженность полиорганной патологии по трехбалльной шкале (по принципу 1 балл – да, 0 баллов – нет, в отношении общего суждения о наличии или отсутствии повреждения для каждого из трех органов: печень, почки и легкие).

Теперь, когда мы совершили небольшой экскурс в теорию измерений, рассмотрим вопрос о применении шкал измерений в медико-биологическом эксперименте.

3.3. ПРИМЕНЕНИЕ ШКАЛ ИЗМЕРЕНИЙ В МЕДИКО-БИОЛОГИЧЕСКОМ ЭКСПЕРИМЕНТЕ

В медико-биологических исследованиях, как уже говорилось, порядковые шкалы используются довольно часто. Вернемся к нашему примеру с регенерацией суставного хряща.

При оценке микроскопической картины учитывали: клеточный состав ткани регенерата, степень заполнения дефекта, структуру поверхностных и глубоких слоев, окрашивание матрикса, сращение образовавшейся ткани с окружающим гиалиновым хря-

щом, степень восстановления субхондрального слоя кости. Эти показатели объединяли с помощью разработанной нами морфологической шкалы оценки заживления поврежденного суставного хряща. Максимальным количеством баллов (до 20) оценивали такие критерии, как клеточный состав ткани, степень заполнения дефекта и восстановление субхондрального слоя кости (см. Табл. 2).

Табл. 2. Морфологические шкалы оценки заживления повреждений суставного хряща

	Критерии	Баллы
1	Клеточный состав ткани в области дефекта	
А	Соответствует нормальному гиалиновому хрящу	20
Б	Более 50% клеток имеют строение хондроцитов	15
В	Менее 50% клеток имеют строение хондроцитов, а большинство – строение фибробластов	10
Г	Наибольшее количество клеток имеют нехондроцитарное строение, встречается единичное число хондроцитов	5
Д	Значительное количество бесклеточных зон. Встречающиеся клетки имеют нехондроцитарное строение	0
2	Степень заполнения дефекта	
А	90-110%	20
Б	75%-89% или 111%-125%	15
В	50%-74% или 126%-150%	10
Г	25%-49% или более 150%	5
Д	менее 25%	0
3	Структура поверхностных слоев дефекта	
А	Соответствует нормальному гиалиновому хрящу	10
Б	Имеется небольшое разволокнение	6
В	Разволокнение средней степени и беспорядочное расположение волокон	2
Г	Значительное разволокнение, непрерывность поверхности нарушена	0
4	Структура глубоких слоев дефекта	
А	Соответствует нормальному гиалиновому хрящу	10
Б	Небольшое разволокнение и образование нескольких маленьких полостей. Клетки в средней зоне расположены столбиками	6
В	Разволокнение средней степени и образование больших и малых полостей. В средней зоне – редкие столбики клеток	2
Г	Значительное разволокнение и образование многочисленных больших и маленьких полостей. Клетки располагаются беспорядочно во всех зонах	0

	Критерии	Баллы
5	Окрашивание матрикса	
А	Соответствует нормальному гиалиновому хрящу	5
Б	Интенсивность снижена незначительно	4
В	Интенсивность снижена значительно	2
Г	Не окрашивается	0
6	Сращение образовавшейся в области дефекта ткани с окружающим гиалиновым хрящом	
А	Полное сращение и переход одной ткани в другую по всей глубине края дефекта	15
Б	Перерыв или неполное сращение на границе между тканями менее чем на 1/2 глубины края дефекта	10
В	Перерыв или неполное сращение на границе между тканями более чем на 1/2 глубины края дефекта	5
Г	Несращение между тканями по всей глубине края дефекта	0
7	Степень восстановления субхондрального слоя кости в области дефекта	
А	90%-100%	20
Б	75%-89%	15
В	50%-74%	10
Г	25%-49%	5
Д	менее 25%	0

Понятно, что, выделив какие-либо уровни по каждому признаку, исследователь приписывает этим уровням соответствующие значения баллов: «5», «10», «15» и т.д., или «0», «2», «4», что, в принципе, безразлично. Использование порядковой шкалы в современных исследованиях нежелательно, хотя и не исключено. И дело здесь не только в известной необъективности, о чем уже говорилось, но и в свойствах самой шкалы порядка. В этой шкале ничего нельзя сказать о равномерности или неравномерности интервалов между соседними значениями оценок. Мы не вправе, к примеру, сказать о том, что регенерация хряща, оцененная на «50», настолько же отличается от регенерации, оцененной на «40», как и процессы, оцененные на «90» и «80». И поэтому совершенно некорректно использование подобных шкал даже в качестве дополнительных аргументов для обоснования эффективности новых методов воздействия на болезнь, поскольку усреднение предполагает сложение значений величины, а операция суммы для порядковых шкал не может быть корректно определена. Соответственно не

могут быть определены и все остальные арифметические и алгебраические действия.

Поэтому, например, утверждение о том, что регенерация хряща в группе животных с применением нового высокоточного малоинвазивного метода хирургического лечения в среднем на 5,5 баллов выше, чем в группе контрольных животных, будет неправомочным, некорректным. Тем более при использовании балльных оценок некорректны (даже абсурдны) утверждения типа: «эффективность экспериментальной методики в 1,6 раза выше контрольной».

Таким образом, **операция вычисления среднего арифметического не является корректной в порядковой шкале**. В принципе, шкалу балльных оценок, также как и другие шкалы порядка, можно использовать в экспериментальных исследованиях, но в этом случае необходимо применять адекватные методы обработки данных, не вычисляя «среднего балла». Корректной характеристикой набора балльных оценок является *медиана* (такое значение оценки, справа и слева от которого расположено одинаковое число оценок в их упорядоченной совокупности). Более подробно методы обработки результатов измерений в порядковой шкале рассмотрены ниже – в шестом разделе.

По приведенным выше соображениям целесообразно использовать такие способы оценки, которые позволяют применить шкалу отношений или шкалу интервалов, а не шкалу порядка.

К слову сказать, в работе по изучению регенерации суставного хряща исследователи вполне отдавали себе отчет в ограниченности балльных оценок. Поэтому основными данными, которые легли в основу доказательной части работы, стали результаты измерений в шкале отношений. Такими измерениями, например, являлись: определение толщины регенерата (мм), числа клеток в объеме ткани регенерата ($1/\text{мкм}^3$), процента заполнения объема дефекта и процента клеток хондрального ряда среди всех клеток в регенерате, времени формирования остеохондральной линии в регенерате.

В общем же случае можно выделить следующие характеристики, измеряемые в шкале отношений:

- временные (время изменения показателя после начала воздействия, время реакции, и т.д.);

- скоростные (частоты различных событий, изменений, скорость реакции, движения и т.д. – величины, отнесенные ко времени);

- структурные (массы, объемы, моли и соотнесенные между собой величины – концентрации и т.д.);

- информационные (объем информации).

Методы обработки величин, измеренных в шкале отношений, рассмотрены ниже – в шестом разделе.

3.4. АГРЕГИРОВАННЫЕ ОЦЕНКИ

Как правило, в любом медико-биологическом эксперименте имеется значительное число (десятки, сотни, а иногда и тысячи) субъектов/объектов – от целого организма до его клеток или субклеточных структур и т.д. В результате измерения показателей этих субъектов получается набор их *индивидуальных оценок*. Понятно, что сравнивать между собой и анализировать одновременно все индивидуальные оценки невозможно, да и нецелесообразно, так как всегда существует их разброс, обусловленный неконтролируемым различием объектов эксперимента (каждый субъект неповторим).

Поэтому для того, чтобы, во-первых, получить обозримое число характеристик и, во-вторых, для того, чтобы сгладить индивидуальные колебания, используют так называемые *агрегированные* (коллективные, групповые, производные) *оценки*. Например, если имелись индивидуальные оценки порога болевой реакции отдельных животных, то агрегированной оценкой будет средний порог болевой реакции в группе.

Получение агрегированных оценок на основании индивидуальных является их преобразованием, и преобразование это следует выполнять корректно. Приведем некоторые корректные процедуры агрегирования для наиболее распространенных в экспериментальных исследованиях показателей (см. также раздел 6.1 «Описательная статистика»).

Для абсолютных величин, измеренных в шкале отношений (см. их перечисление выше), наиболее типичным является вычисление среднего арифметического по группе. Эта процедура вполне корректна, и обычно ее реализация не вызывает затруднений.

Наибольшее число ошибок возникает при агрегировании показателей, измеренных в порядковых шкалах. Еще раз повторим – **не следует складывать, вычитать, умножать или делить баллы друг на друга, да и на чтобы то ни было – все это абсолютно бессмысленные операции.** В порядковой шкале для усреднения используют медиану [30, 31].

Если имеется набор индивидуальных баллов, то характеристикой группы будет число ее членов, получивших тот или иной балл¹³ (например, 4 животных получили балл «40»). Аналогичным образом агрегируется и информация о выделении уровней – если введены три уровня (например, уровни регенерации: низкий, средний и высокий) и имеется информация о распределении всех членов нескольких групп (контрольных или экспериментальных) по этим уровням, то агрегированной информацией об объединенной группе будет число ее членов, обладающих тем или иным уровнем (вычисляемое как сумма по всем группам числа их членов, обладающих данным уровнем регенерации) – соответствующие примеры приводятся ниже.

Если в настоящем разделе речь шла об агрегировании индивидуальных оценок по группе с целью получения характеристик группы в целом, то в следующем разделе рассматривается проблема агрегирования показателей, характеризующих один и тот же объект.

3.5. КОМПЛЕКСНЫЕ ОЦЕНКИ

Нередко встречаются случаи, когда какое-либо изучаемое явление, процесс характеризуется несколькими показателями – *вектором показателей*. При этом часто возникает вопрос о возможности однозначной оценки этого явления, процесса или изучаемых их свойств одной величиной – *комплексной оценкой*. Так, во многих спортивных состязаниях победитель выявляется по сумме очков, баллов, набранных на отдельных этапах состязания или в отдельных играх, в многоборье – в отдельных видах спорта. Или же другой пример из нашей практики – общие результаты

¹³ Отметим, что такая агрегированная характеристика группы как число ее членов, получивших данный балл, является величиной, измеренной в шкале отношений.

операции оцениваются по степени регенерации хряща, качеству восстановления движений в суставах и времени, когда возможна полноценная функция конечности.

На практике комплексные оценки встречаются довольно часто и, очевидно, без них не обойтись, хотя способы их определения нередко и вызывают множество недоуменных вопросов. Но в любом случае такие комплексные оценки, применяемые в повседневной жизни, являются либо результатом определенных общественных соглашений, которые признаются всеми участниками, либо установлены каким-либо нормативным актом определенного директивного органа – министерства, ведомства и т.д. и в силу этого также признаются всеми заинтересованными лицами.

Другое дело – применение комплексных оценок в научном исследовании. Здесь сразу на первое место встает вопрос о научной, в том числе математической, строгости применяемой оценки. В частности, не вызывает сомнений возможность использования такой векторной оценки, как суммарные затраты времени на выполнение той или иной манипуляции. Здесь суммируются однородные величины, заданные шкалами отношений.

Между тем суммирование баллов довольно часто встречается в статьях и диссертациях по медицине и биологии. В некоторое оправдание подобным построениям комплексных оценок следует отметить, что проблема агрегирования векторных оценок на сегодняшний день исследована не полностью, а существующие результаты, даже для их применения на практике, зачастую требуют хорошего знания высшей математики. Достаточно простым и интуитивно понятным (но, в то же время, корректным) методом агрегирования балльных оценок является использование так называемых матриц свертки [25], элементы которых содержат значения агрегированного показателя, а агрегируемые баллы задают номер строки и столбца.

Для тех, кто глубже заинтересуется проблемой комплексных оценок и принятия решений при многих критериях, можно рекомендовать ознакомиться с соответствующими публикациями [27, 36, 43]. Но в любом случае при построении комплексных оценок нужно быть предельно внимательным и осторожным. Кстати, нередко можно обойтись и без них. Если получены количественные результаты по отдельным показателям, то можно ограничиться

их качественной интерпретацией, не «загоняя под общий знаменатель», проанализировать и сравнить исследуемые объекты отдельно по каждому из показателей. И пусть по каким-то показателям результаты экспериментальных групп будут лучше контрольных, а по каким-то хуже – от этого исследование только обогатится, станет достовернее.

4. АНАЛИЗ ИСПОЛЬЗОВАНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ В СТАТЬЯХ И ДИССЕРТАЦИОННЫХ ИССЛЕДОВАНИЯХ ПО БИОЛОГИИ И МЕДИЦИНЕ

Авторами настоящей работы были исследованы 225 диссертаций (22 докторских и 203 – кандидатских), успешно защищенных в различных диссертационных советах за последние годы и без замечаний аттестованных ВАК РФ. Исследования охватывали большинство медицинских и биологических наук (16 специальностей из перечня), и в качестве основной методологии содержали экспериментальное исследование как инструмент получения новых знаний и установления новых закономерностей в соответствующей отрасли знаний.

Главной особенностью работ медико-биологического профиля является практически 100%-ная декларация статистической обработки экспериментальных данных (221 работа). Лишь четыре диссертации не содержали сколько-нибудь приемлемых указаний на статистические методы, а являлись практически описательными полуколичественными работами без убедительного анализа данных, полученных в параллельных исследованиях номинально сравниваемых групп.

Анализируемые диссертационные работы распределились следующим образом на пять неравнозначных групп.

1. Работы с декларированным, но некорректно примененным набором простейших статистических методов («неправильно обсчитанные»). Эти работы не были многочисленными и составили 28 из 225, или 12,4%.

2. Диссертации, содержащие правильно примененные простейшие статистические методы, но не содержащие необходимого набора экспериментальных групп/состояний («неправильно спла-

нированные»). К сожалению, в этой группе оказалось 75 работ (33,3%), в том числе 6 докторских диссертаций. Никакая статистика не поможет в формировании научно обоснованных положений при отсутствии адекватно подобранных условий эксперимента, подбора групп и измеряемых параметров (см. выше раздел 2).

3. Диссертации, в которых рутинные статистические исследования соответствуют простой схеме попарно сравниваемых групп (состояний) и т.п. – см. второй раздел выше. Это группа была максимально представительной – 84 работы, или 37,3%. Именно подобная практика в настоящее время составляет «золотой стандарт» техники обработки экспериментальных данных при выполнении диссертационных работ по медицине и биологии.

4. Диссертации, в которых присутствует расширенный статистический анализ выборок, на основании которого не подтверждаются предварительные качественные (прогнозируемые заранее) выводы, а строятся самостоятельные выводы. Таких работ оказалось 35 (15,6%), они выгодно отличались более продвинутым математическим аппаратом и содержали обычно более глубокие положения и результаты. Преимущественно, подобные работы соответствовали специальностям, близким к цитологии, биохимии, генетике, молекулярной биологии.

5. Работы, содержащие сложные продвинутое математические построения с прицелом на создание или реальную разработку математических моделей. Таких работ оказалось всего три (все – докторские), или 1,3%. Арсенал применяемых в таких работах методов выходит далеко за рамки настоящей книги.

Таким образом, при анализе результатов применения статистических методов в экспериментальной биологии и медицине, на примере диссертационных работ по медико-биологическим специальностям вырисовываются два интересных момента.

Во-первых, только чуть более половины (52,5%) всех диссертаций выполнялось с применением статистических методов, адекватных задачам, характеру выбранных параметров и схеме эксперимента. Типичной ошибкой в этом случае являлось недостаточно корректное планирование эксперимента, выводящее статистический анализ за рамки достаточности для получения научно обоснованных выводов и положений.

Во-вторых, среди работ с адекватным набором методик преобладали диссертации, выполненные с явно минимальным набором статистических процедур. Возникает закономерный вопрос: является ли это данью традициям соответствующих научных школ или недостаточностью информированностью о возможностях статистического анализа? Складывалось также впечатление, что в части работ целеполагание и задачи были явно заужены именно в результате исходной ущербности статистического аппарата у исследователя. Если это так, то незнание современных статистических методов становится проблемой, оказывающей влияние на научную сторону исследований, что уж вовсе недопустимо.

Конечно, авторы отдают себе отчет в том, что выполнение диссертационных работ вовсе не исчерпывает интересы отечественных исследований в области экспериментальной биологии и медицины, но по математическому аппарату, на наш взгляд – вполне отражает их спектр. Оно, к тому же, достаточно открыто анализу, в отличие, например, от столь модных теперь коротких сообщений со ссылками на стандартный протокол, который на деле в этой области знаний нередко отсутствует.

Помимо неутешительных выводов, анализ диссертационных работ по экспериментальной биологии и медицине позволил выделить *типовые* (для рассматриваемой предметной области) *задачи анализа данных*.

1. Описание данных используется в подавляющем большинстве работ. Даже в части работ, не использующих статистические методы, в иллюстративных целях применялись некоторые производные показатели – среднее, дисперсия и т.д. Краткое рассмотрение *описательной статистики*, то есть описания результатов эксперимента с помощью различных агрегированных показателей и графиков, приведено в разделе 6.1 ниже.

2. Анализ сходства/различий объектов в рамках схемы эксперимента, описанной в разделе 2, встречается примерно в половине работ. Соответствующие статистические методы подробно рассматриваются в разделе 6.2 настоящей работы.

3. Анализ зависимостей – установление наличия/отсутствия зависимостей между различными характеристиками объектов, а также нахождение этих зависимостей – используется примерно в

20-30% работ. Соответствующие статистические методы описаны в разделе 6.3 настоящей работы.

4. Задачи, требующие использования «продвинутых» статистических методов (в первую очередь – многомерного анализа [2, 32, 40]). К задачам данного класса (обработки «сложных» данных) можно отнести задачи выявления существенных признаков, задачи классификации, прогнозирования и т.д. (см. краткое описание подобных задач, встречающихся не более чем в 10% статей и диссертационных работ, в разделах 6.4 и 6.5 ниже).

Завершив анализ использования статистических методов в статьях и диссертационных исследованиях по медицине и экспериментальной биологии, и выделив типовые задачи, перейдем к формализации последних.

5. ТИПОВЫЕ ЗАДАЧИ АНАЛИЗА ДАННЫХ В МЕДИКО-БИОЛОГИЧЕСКОМ ЭКСПЕРИМЕНТЕ

В настоящем разделе приводятся два числовых примера исходных данных (на которых в дальнейшем – см. шестой раздел – иллюстрируется применение статистических методов), полученных в результате эксперимента (разделы 5.1 и 5.2), перечисляются типовые задачи анализа данных (раздел 5.3) и приводятся рекомендации по использованию компьютерных программ для анализа экспериментальных данных (раздел 5.4).

5.1. ИСХОДНЫЕ ДАННЫЕ: ПРИМЕР 1

Предположим, что имеется экспериментальная группа, состоящая из N объектов – например, людей (пациентов) или экспериментальных животных и т.д., и контрольная группа, состоящая из M объектов (где N и M – целые положительные числа, например, $N = 25$, $M = 30$). Допустим, что в результате измерения одного и того же показателя с помощью одной и той же процедуры измерений были получены следующие данные:

$x = (x_1, x_2, \dots, x_N)$ – выборка¹⁴ для экспериментальной группы
и

$y = (y_1, y_2, \dots, y_M)$ – выборка для контрольной группы,

где x_i – элемент выборки – значение исследуемого показателя (признака¹⁵) у i -го члена экспериментальной группы, $i = 1, 2, \dots, N$, а y_j – значение исследуемого показателя у j -го члена контрольной группы, $j = 1, 2, \dots, M$. Число элементов выборки называется ее объемом – например, объем выборки x равен N , а объем выборки y равен M .

В зависимости от того, в какой шкале – шкале отношений или порядковой шкале – производились измерения, получаем следующие два случая.

Шкала отношений. Если измерения производились в шкале отношений, то $\{x_i\}$ и $\{y_j\}$ – числа, для которых имеют смысл все арифметические операции.

Рассмотрим пример¹⁶. Пусть имеется экспериментальная группа, состоящая из 25 человек ($N = 25$), и контрольная группа, состоящая из 30 человек ($M = 30$), и измерение заключается в определении числа хондроцитов в 1 мкм^3 ткани регенерата на седьмые сутки после травмы. Это число принимает в реальной ситуации натуральные значения от 0 до 20, то есть, измерения производятся в шкале отношений. Поскольку по результатам работы число хондроцитов оказалось критериальным, то в дальнейшем оно будет обозначено как *степень регенерации суставного хряща* (СРСХ).

Экспериментальным воздействием в рассматриваемом примере выступала стимуляция заживления (регенерации), в то время как в контрольной группе имела место естественная регенерация.

Отметим, что аналогичные рассматриваемому примеру ситуации возникают чрезвычайно часто – когда результатом эксперимента является набор значений числовых характеристик для двух групп. Например, масса экспериментальных животных в граммах, количество органов, в которых обнаружены изменения, и т.д.

¹⁴ Выборка – совокупность значений одного и того же признака у наблюдаемых объектов.

¹⁵ Признак – свойство (характеристика) наблюдаемого объекта.

¹⁶ Данный пример рассматривается на протяжении всего настоящего и ряда последующих разделов. Все таблицы, диаграммы и графики экспортированы из компьютерной программы Microsoft Excel для Windows.

Результаты измерений СРСХ в контрольной и экспериментальной группах до и после эксперимента приведены в Табл. 3, строки которой соответствуют членам групп (отдельным пациентам).

Табл. 3. Результаты измерений СРСХ в контрольной и экспериментальной группах до и после эксперимента

Контрольная группа до начала эксперимента	Экспериментальная группа до начала эксперимента	Контрольная группа после окончания эксперимента	Экспериментальная группа после окончания эксперимента
15	12	16	15
13	11	12	18
11	15	14	12
18	17	17	20
10	18	11	16
8	6	9	11
20	8	15	13
7	10	8	7
8	16	6	14
12	12	13	17
15	15	17	19
16	14	19	16
13	19	15	12
14	13	11	15
14	19	9	19
19	12	19	18
7	11	8	14
8	16	6	13
11	12	9	18
12	8	12	13
15	13	11	13
16	7	17	15
13	15	10	18
5	8	8	9
11	9	8	14
19	–	20	–
18	–	19	–
9	–	6	–
6	–	14	–
15	–	10	–

Например, первый пациент из контрольной группы до начала эксперимента характеризуется СРСХ, равной 15 (см. Табл. 3), а второй пациент из экспериментальной группы характеризуется СРСХ, равной 11, и т.д.

Результаты эксперимента могут быть получены и в порядковой шкале (или переведены из шкалы отношений в порядковую), поэтому рассмотрим представление данных в порядковой шкале.

Порядковая шкала. Если использовалась порядковая шкала (шкала рангов) с L градациям (например, в трехбалльной шкале $L = 3$), то будем считать, что $\{x_i\}$ и $\{y_j\}$ – натуральные числа, принимающие одно из L значений. Для простоты можно считать, что множество значений (баллов) есть множество чисел от единицы до L . Тогда характеристикой группы будет число ее членов, получивших тот или иной балл (см. раздел «Агрегированные оценки» выше). То есть, для экспериментальной группы вектор баллов есть

$$n = (n_1, n_2, \dots, n_L),$$

где n_k – число членов экспериментальной группы, получивших k -ый балл, $k = 1, 2, \dots, L$. Для контрольной группы вектор баллов есть

$$m = (m_1, m_2, \dots, m_L),$$

где m_k – число членов контрольной группы, получивших k -ый балл, $k = 1, 2, \dots, L$. Очевидно, что

$$n_1 + n_2 + \dots + n_L = N, m_1 + m_2 + \dots + m_L = M.$$

Пусть в рассматриваемом примере (в котором $N = 25$, $M = 30$) выделены три уровня СРСХ ($L = 3$): низкий (СРСХ меньше либо равна 10), средний (СРСХ строго больше 10, но меньше либо равна 15) и высокий (СРСХ строго больше 15). Сформируем в компьютерной программе Microsoft Excel для Windows Табл. 4, в которой указаны верхние границы диапазонов.

Табл. 4. Переход от шкалы отношений к порядковой шкале

Уровень СРСХ	Максимальная оценка
Низкий	10
Средний	15
Высокий	20

Поставим в соответствие уровням СРСХ (низкому, среднему и высокому) баллы – 1, 2 и 3 (эта операция является корректной для порядковой шкалы – см. раздел «Допустимые преобразования» выше). Вычислим на основании данных Табл. 3, например, сначала для контрольной группы до начала эксперимента, число ее членов, получивших балл, принадлежащий тому или иному диапазону: $m_1 = 9$ (то есть, 9 членов контрольной группы до начала эксперимента характеризовались низким уровнем СРСХ), $m_2 = 14$, $m_3 = 7$. Результаты занесем в Табл. 5 (в компьютерной программе Microsoft Excel для Windows таблица Табл. 5 получается из Табл. 3 и Табл. 4 применением инструмента анализа данных «Гистограмма» (Меню\Сервис\Анализ данных\Гистограмма).

Табл. 5. Уровни СРСХ членов контрольной группы до эксперимента

Уровень СРСХ	Частота (число человек)
Низкий (1 балл)	9
Средний (2 балла)	14
Высокий (3 балла)	7

Для каждого из столбцов Табл. 3 по аналогии с Табл. 5 определяем распределение членов экспериментальной и контрольной групп по уровням СРСХ и получаем Табл. 6.

Табл. 6. Результаты измерений уровня СРСХ в контрольной и экспериментальной группах до и после эксперимента

Уровень СРСХ	Контрольная группа до начала эксперимента (чел.)	Экспериментальная группа до начала эксперимента (чел.)	Контрольная группа после окончания эксперимента (чел.)	Экспериментальная группа после окончания эксперимента (чел.)
Низкий	9	7	12	2
Средний	14	12	10	13
Высокий	7	6	8	10

Табл. 6 построена по Табл. 3 введением диапазонов значений оценок, попадание в которые считалось соответствующим уровням СРСХ. Отметим, что при подобном переходе от шкалы отношений к порядковой шкале часть информации теряется – в рассматриваемом примере одному и тому же уровню СРСХ соответствуют

несколько различных оценок. Следовательно, труднее становится устанавливать совпадения и различия характеристик исследуемых объектов. Поэтому, рекомендуется использовать всю имеющуюся информацию, то есть, если при измерениях использовалась шкала отношений, то и обрабатывать данные следует в этой шкале.

5.2. ИСХОДНЫЕ ДАННЫЕ: ПРИМЕР 2

Рассмотрим следующие первичные данные другого эксперимента: для 27 экспериментальных животных (крыс) измерялись следующие параметры: пол (0 – мужской, 1 – женский), тип – порог болевой чувствительности (три градации), M_0 – масса тела до начала эксперимента¹⁷ (мг), M – масса тела после окончания эксперимента (мг), $M1$ – коэффициент массы печени, то есть отношение ее массы к массе тела (мг/г), $M2$ – коэффициент массы почек (мг/г), $M3$ – коэффициент массы легких (мг/г), а также содержание ряда эндогенных токсических веществ в плазме крови: $ОП_0$ – олигопептиды до начала эксперимента (мг/л), $ВСММ_0$ – вещества средней молекулярной массы до начала эксперимента (ед. экстинкции), $ОП$ – олигопептиды после окончания эксперимента (мг/л), $ВСММ$ – вещества средней молекулярной массы после окончания эксперимента (ед. экстинкции). Числовые данные приведены в Табл. 7. Пол – номинальный признак, тип измерен в порядковой шкале, остальные характеристики – в шкале отношений.

То есть в данном случае имеется 11 характеристик для каждого из 27 объектов. Задача статистического анализа заключается в установлении взаимосвязи между переменными и выделении существенных характеристик. Результаты анализа приведены в шестом разделе ниже.

5.3. ТИПОВЫЕ ЗАДАЧИ АНАЛИЗА ДАННЫХ

Перечислим кратко типовые задачи анализа результатов медико-биологического эксперимента – см. также четвертый раздел выше. Можно выделить следующие этапы (см. Табл. 8).

¹⁷ Экспериментальное воздействие заключалось в введении токсинов.

Табл. 7. Результаты эксперимента

№	Пол	Тип	M ₀	M	M1	M2	M3	ОП ₀	BCMM ₀	ОП	BCMM
1	0	1	240	250	32,91	6,25	6,04	150	0,24	283	0,43
2	0	2	205	220	29,02	7,02	7,56	97	0,12	220	0,38
3	1	1	250	245	25,20	5,20	7,40	125	0,18	184	0,51
4	0	1	285	300	24,74	6,14	6,61	115	0,14	252	0,42
5	1	2	230	235	26,84	6,75	6,67	136	0,22	215	0,51
6	1	2	235	220	32,50	6,85	5,68	120	0,16	196	0,28
7	1	3	220	225	31,29	8,13	7,71	100	0,16	174	0,39
8	1	1	210	225	27,86	7,00	8,14	125	0,18	201	0,43
9	0	2	240	255	28,75	7,30	7,90	75	0,12	132	0,32
10	1	2	210	220	29,60	7,55	7,67	90	0,14	144	0,31
11	0	2	225	230	31,50	7,48	8,15	120	0,20	260	0,39
12	0	2	220	245	30,00	7,30	8,20	110	0,16	165	0,45
13	1	2	230	235	33,61	7,48	8,07	85	0,10	154	0,44
14	1	3	170	180	32,18	6,94	7,88	85	0,15	140	0,56
15	0	3	240	245	35,75	7,60	8,00	60	0,09	133	0,26
16	1	1	240	235	28,24	6,45	6,84	85	0,12	170	0,28
17	1	1	280	305	28,21	7,86	7,03	80	0,14	124	0,37
18	1	2	270	275	30,27	8,02	7,34	110	0,15	215	0,40
19	1	2	215	215	29,77	7,67	7,20	125	0,20	260	0,52
20	1	3	260	280	30,37	6,90	7,20	65	0,11	104	0,29
21	1	3	270	275	33,81	7,49	6,82	95	0,14	177	0,32
22	0	1	250	250	29,70	6,35	6,35	70	0,13	150	0,43
23	0	1	270	260	32,16	7,12	6,32	100	0,15	222	0,28
24	0	2	280	285	33,40	7,33	7,00	72	0,09	134	0,31
25	0	2	300	300	33,46	7,88	6,77	83	0,12	155	0,34
26	0	3	290	295	35,05	7,32	6,54	70	0,20	183	0,46
27	0	3	350	365	34,25	8,46	8,00	60	0,23	131	0,50

Табл. 8. Этапы анализа экспериментальных данных

Э Т А П	<i>Описание данных</i>	<i>Изучение сходства/различий</i>	<i>Исследование зависимостей</i>	<i>Снижение размерности</i>	<i>Классификация и прогноз</i>
М Е Т О Д Ы	- описательная статистика, - определение необходимого объема выборки. (раздел 6.1)	Статистические критерии: Крамера-Уэлча, Вилкоксона-Манна-Уитни, χ^2 , Фишера. (раздел 6.2)	- корреляционный анализ, - дисперсионный анализ, - регрессионный анализ. (раздел 6.3)	- факторный анализ, - метод главных компонент. (раздел 6.4)	- дискриминантный анализ, - кластерный анализ, - группировка. (раздел 6.5)

1. Описание данных – компактное описание имеющихся данных с помощью различных агрегированных показателей и графиков. К этому классу можно отнести также задачу определения необходимого объема выборки. Методы описания данных рассмотрены ниже в разделе 6.1.

2. Изучение сходства/различий (сравнение двух выборок). Задача заключается в установлении совпадений или различий характеристик двух выборок¹⁸. Подробное описание решения этой задачи и методы ее решения приведены в разделе 6.2.

3. Исследование зависимостей. Следующим шагом после изучения сходства/различий является установление факта наличия/отсутствия зависимости между показателями (переменными – будем употреблять эти термины как синонимы) и количественное описание этих зависимостей. Для этих целей используются, соответственно, корреляционный и дисперсионный анализ, а также – регрессионный анализ (см. раздел 6.3).

4. Снижение размерности заключается в уменьшении числа анализируемых переменных либо посредством выделения существенных переменных, либо/и построения новых показателей (на основании полученных в результате эксперимента). Соответствующие статистические методы (факторный анализ, метод главных компонент) и примеры их использования рассмотрены в разделе 6.4.

¹⁸ В соответствии с описанной выше схемой эксперимента (см. Рис. 1) используются две группы – экспериментальная и контрольная. При этом считается, что каждая выборка является результатом независимых наблюдений реализации некоторой случайной величины с соответствующей функцией распределения, неизвестной исследователю. Предполагается также, что наблюдения в одной выборке не зависят от наблюдений в другой, поэтому выборки называются независимыми (см. п. 4.4 в [32]). Возможен другой вариант – когда выборки связаны, например, рассматриваются одни и те же пациенты до и после лечения. Предположим, что для каждого пациента имелся небольшой положительный сдвиг. Если рассматривать результаты измерений в начале и в конце лечения как независимые выборки и применять соответствующие методы проверки однородности, то эффект лечения может не быть обнаружен, поскольку будет скрыт большим размахом данных. Методы, основанные на анализе связанных выборок, обнаружат эффект (например, критерий знаков в п. 4.7 в [32]).

5. Классификация и прогноз. Задачи классификации и прогноза выходят за рамки настоящей работы. Качественное их обсуждение и ссылки на литературу приведены в разделе 6.5.

Статистические методы используются с целью описания наблюдаемых процессов и явлений и объяснения. Поэтому следующим шагом (не отраженным в Табл. 8) в исследовании является переход к математическим моделям биологических систем (см. краткое обсуждение в заключении к настоящей работы).

5.4. ИСПОЛЬЗОВАНИЕ КОМПЬЮТЕРА

Использование компьютера при анализе результатов эксперимента, несомненно, целесообразно¹⁹. С одной стороны, ряд статистических методов реализован в такой популярной программе, как Microsoft Excel для Window, входящей в стандартный комплект Microsoft Office, и установленной, наверное, на любом современном компьютере. С другой стороны, на сегодняшний день существует множество специальных профессиональных программ, позволяющих осуществлять статистический анализ данных.

Среди последних можно выделить и рекомендовать к использованию такие наиболее распространенные пакеты статистического анализа как: Statistica, StatGraphics и SPSS. Однако, упомянутые программы достаточно сложны и требуют значительных временных затрат для их освоения.

Поэтому можно рекомендовать следующее: если для решения задач исследования хватает возможностей Excel, то можно ограничиться использованием этой программы (недостатком ее, правда, является практически полное отсутствие вразумительных объяснений, которые помогли бы неподготовленному пользователю понять, что получилось в результате расчетов). Если возможностей Excel недостаточно, то нужно обращаться к профессиональным статистическим программам. Каждая из них обладает своими достоинствами и недостатками (в одной более полно реализованы одни методы, в другой другие и т.д.). С методической точки зрения

¹⁹ Все обсуждаемые в настоящем разделе достоинства и недостатки программного обеспечения отражают лишь субъективное мнение авторов и не могут рассматриваться как попытка рекламы того или иного продукта.

можно рекомендовать использовать программу *StatGraphics* (версии 5.0 и выше). Ее достоинством с позиций непрофессионального статистика является наличие «советчика», который разъясняет, что означает та или иная вычисленная величина, и что исследователю следует делать дальше. Поэтому ниже расчет примеров производится в программе *StatGraphics*, и приводятся инструкции, как в этой программе вычислить интересующие исследователя показатели.

Приведем рекомендации по использованию компьютера для решения перечисленных выше типовых задач анализа данных.

Описательная статистика корректно и в достаточно полном объеме реализована в электронных таблицах *Microsoft Excel*.

Статистические критерии сравнения двух выборок (Вилкоксона-Манна-Уитни, χ^2 и Фишера) корректно реализованы в профессиональных статистических пакетах²⁰. К сожалению, ни один из рекомендуемых ниже в разделе 6.2 статистических критериев не реализован в *Excel*²¹, поэтому можно посоветовать производить расчет эмпирических значений критериев либо вручную²² (все необходимые формулы приведены ниже), используя компьютер или калькулятор для получения описательной статистики и автоматизации расчетов, либо скачав в Интернете (www.mtas.ru) бесплатную программу «Педагогическая статистика», в которой реализованы рассмотренные в настоящей работе статистические критерии сравнения двух выборок.

Коэффициенты линейной корреляции и простая линейная регрессия могут быть вычислены в *Excel*. Для более сложных задач исследования зависимостей и снижения размерности (а также для

²⁰ Относительно критерия Крамера-Уэлча следует отметить, что в стандартных компьютерных статистических пакетах реализован двухвыборочный критерий Стьюдента, который совпадает с критерием Крамера-Уэлча лишь при равенстве объемов выборок.

²¹ В компьютерной программе *Microsoft Excel* для *Windows* имеется критерий согласия χ^2 , отличающийся от описанного выше критерия однородности χ^2 , поэтому применение первого может привести к неверным результатам.

²² Альтернативой является использование дополнительных статистических надстроек к *Excel* – *Megastat*, *XLStat*, которые можно найти в свободном доступе в Интернете. В этих пакетах хорошо представлены непараметрические методы – критерий Вилкоксона-Манна-Уитни и другие.

решения задач классификации и прогноза) необходимо использование профессиональных статистических пакетов.

Завершив перечисление типовых для медико-биологических исследований задач анализа данных, рассмотрим методы их решения.

6. МЕТОДЫ ОБРАБОТКИ ДАННЫХ И ПРИМЕРЫ

Настоящий раздел содержит изложение методов анализа данных для выделенных выше типовых задач:

- описательная статистика (раздел 6.1);
- анализ совпадений и различий характеристик экспериментальной и контрольной групп на основании измерений, проведенных в порядковой шкале или шкале отношений (раздел 6.2);
- исследование зависимостей (раздел 6.3, в том числе – корреляционный анализ (раздел 6.3.1), дисперсионный анализ (раздел 6.3.2), регрессионный анализ (раздел 6.3.3.));
- снижение размерности (раздел 6.4),

а также качественное обсуждение задач классификации и прогноза (раздел 6.5).

В качестве иллюстрации рассматривается реализация перечисленных выше методов для числовых примеров 1 и 2 (см. соответственно разделы 5.1 и 5.2).

6.1. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

В практических задачах обычно имеется совокупность наблюдений (десятки, сотни, а иногда – тысячи результатов измерений индивидуальных характеристик), поэтому возникает задача компактного описания имеющихся данных. Для этого используют методы *описательной статистики* – описания результатов с помощью различных агрегированных показателей и графиков. Кроме того, некоторые показатели описательной статистики используются в других статистических методах (см., например, раздел 6.2).

Для результатов измерений в шкале отношений показатели описательной статистики можно разбить на несколько групп [44]

(для первоначального ознакомления с теорией вероятностей и математической статистикой можно порекомендовать учебник [9]):

- *показатели положения* описывают положение экспериментальных данных на числовой оси. Примеры таких данных – *максимальный и минимальный элементы выборки, среднее значение*²³, *медиана*²⁴, *мода*²⁵ и др.;

- *показатели разброса* описывают степень разброса данных относительно своего центра (среднего значения). К ним относятся: *выборочная дисперсия*²⁶, *разность между минимальным и максимальным элементами (размах, интервал выборки)* и др.

- *показатели асимметрии*: положение медианы относительно среднего и др.

- *гистограмма*²⁷ и др.

Данные показатели используются для наглядного представления и первичного («визуального») анализа результатов измерений характеристик экспериментальной и контрольной группы.

Приведем формулы расчета основных показателей. *Среднее арифметическое (Mean) \bar{x}* выборки $\{x_i\}_{i=1 \dots N}$ (*выборочное среднее*) рассчитывается следующим образом²⁸:

$$(1) \bar{x} = \frac{1}{N} (x_1 + x_2 + x_3 + \dots + x_{n-1} + x_n) = \frac{1}{N} \sum_{i=1}^N x_i,$$

²³ Имеется в виду среднее арифметическое значение.

²⁴ Медианой называется значение исследуемого признака, справа и слева от которого находится одинаковое число упорядоченных элементов выборки. Если объем выборки – четное число, то медианой является среднее арифметическое двух центральных членов.

²⁵ Модой называется такое значение измеренного признака, которым обладает максимальное число элементов выборки, то есть значение, которое встречается в выборке наиболее часто.

²⁶ Выборочная дисперсия рассчитывается как средняя сумма квадратов разностей между элементами выборки и средним значением. Дисперсия характеризует разброс элементов выборки вокруг среднего значения.

²⁷ Гистограммой называется графическое изображение зависимости частоты попадания элементов выборки от соответствующего интервала группировки (диапазона значений показателя).

²⁸ Символ $\sum_{i=1}^N x_i$ здесь и далее обозначает сумму элементов $\{x_i\}$ по индексу i ,

пробегаящему последовательно все значения от единицы до N : $x_1 + x_2 + \dots + x_N$.

а выборочная дисперсия (Variance) s^2 :

$$(2) s^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2 .$$

В компьютерной программе Microsoft Excel для Windows описательная статистика получается применением инструмента анализа данных «Описательная статистика» (Сервис\Анализ данных\Описательная статистика). Описательная статистика для первого столбца Табл. 3 (степень регенерации суставного хряща (СРСХ) в контрольной группе до начала эксперимента) приведена в Табл. 9.

Табл. 9. Описательная статистика СРСХ в контрольной группе до начала эксперимента (см. первый столбец Табл. 3)

Среднее	12,6
Стандартная ошибка	0,76
Медиана	13
Мода	15
Стандартное отклонение	4,16
Выборочная дисперсия	17,28
Эксцесс ²⁹	-0,89
Асимметричность	-0,03
Интервал (размах)	15
Минимум	5
Максимум	20
Сумма	378
Счет (объем выборки)	30

В описаниях результатов медико-биологических экспериментов часто используют одно из двух представлений результатов.

²⁹ Коэффициент эксцесса является характеристикой того, насколько кучно основная масса данных группируется около центра; коэффициент асимметрии характеризует асимметричность распределения. Значения эксцесса и асимметричности, лежащие вне интервала от -2 до +2 свидетельствуют об отклонениях от нормального распределения (могут служить грубым тестом на нормальность – см. ниже).

Первое – в виде « $\bar{x} \pm s$ », где \bar{x} – среднее, а s – стандартное отклонение (Standard Deviation, равное по определению корню квадратному из дисперсии).

Второе представление результатов – в виде « $\bar{x} \pm \Delta\bar{x}$ », где $\Delta\bar{x}$ – стандартная ошибка среднего (Standard Error of Mean) – вторая строка Табл. 9, определяемая следующим образом: $\Delta\bar{x} = s / \sqrt{n}$, то есть стандартная ошибка среднего в \sqrt{n} раз меньше стандартного отклонения.

В каждом конкретном случае, необходимо оговаривать, какое из представлений результатов используется, так как запись «одно число плюс/минус другое» может толковаться неоднозначно.

Гистограмма в Excel получается применением инструмента анализа данных «Гистограмма» (Сервис\Анализ данных\Гистограмма). Гистограмма СРСХ в контрольной группе до начала эксперимента (первый столбец Табл. 3) представлена на Рис. 3.

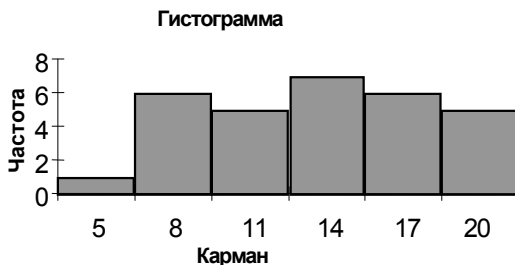


Рис. 3. Гистограмма числа СРСХ в контрольной группе до начала эксперимента («частота» в Excel³⁰ – число элементов выборки, попавших в заданный диапазон, называемый в Excel «карманом»)

Рассмотрим теперь показатели описательной статистики для данных, измеренных в порядковой шкале.

Для результатов измерений в порядковой шкале при небольшом числе градаций единственным информативным показателем описательной статистики является гистограмма³¹.

³⁰ Обычно частотой называют не «число элементов выборки», а «долю элементов выборки».

Для визуального (качественного) сравнения экспериментальной и контрольной групп удобно строить для них совместные гистограммы. Например, по результатам Табл. 6 (см. выше) можно построить несколько парных гистограмм, на которых отложены одновременно частоты для двух групп (например, контрольной и экспериментальной). На Рис. 4 и Рис. 5 приведены две из них – позволяющие сравнивать контрольную и экспериментальную группу до начала и после окончания эксперимента (на самом деле визуальный анализ не дает возможности сказать, значительно ли различаются данные выборки – для этого необходимо использовать статистические методы – см. ниже раздел 6.2). Для их построения сначала перейдем от Табл. 6 к Табл. 10, отличающейся от первой тем, что в ее ячейках стоят не абсолютное число членов той или иной группы, набравших соответствующий балл (уровень СРСХ), а доля³² (в процентах) членов группы, получивших данный балл, так как подобное преобразование (деление на одно и то же число – количество членов в данной группе) позволяет качественно сравнивать группы разных размеров (например, разного количества пациентов). Затем строим гистограммы в компьютерной программе Microsoft Excel для Windows (Меню\Вставка\Диаграмма) – см. Рис. 4 и Рис. 5, на которых по вертикали отложен процент членов той или иной группы, набравших соответствующий балл.

Табл. 10. Результаты измерений уровня СРСХ в контрольной и экспериментальной группах до и после эксперимента

Уровень СРСХ	Контрольная группа до начала эксперимента (%)	Экспериментальная группа до начала эксперимента (%)	Контрольная группа после окончания эксперимента (%)	Экспериментальная группа после окончания эксперимента (%)
Низкий	30,00	28,00	40,00	8,00
Средний	46,67	48,00	33,33	52,00
Высокий	23,33	24,00	26,67	40,00

³¹ Если число градаций (различных значений) велико, то информативными также являются мода и медиана.

³² Доля принимает значения от нуля до единицы. Для перехода к процентам следует долю умножить на 100%.



Рис. 4. Гистограммы контрольной и экспериментальной групп до начала эксперимента

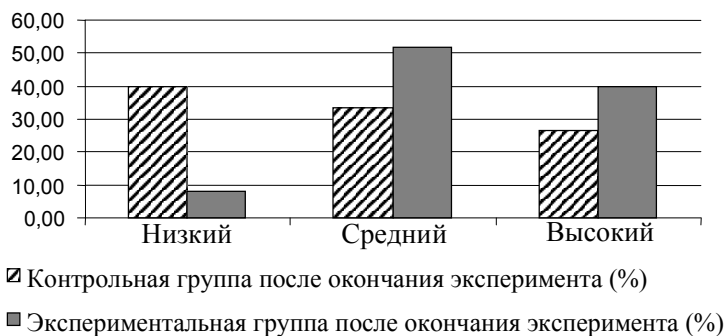


Рис. 5. Гистограммы контрольной и экспериментальной групп после окончания эксперимента

Таким образом, описательная статистика позволяет представить результаты эксперимента в компактном и информативном виде, что дает возможность проводить качественный анализ исследуемых объектов³³.

³³ Показатели описательной статистики (объем выборки, среднее, гистограммы и т.д.) обычно приводятся в тексте диссертационных работ и авторефератов по биологии и медицине.

Отметим, что к описательной статистике можно отнести и **задачи качественного анализа «однородности» выборки.**

Например, при исследовании средних размеров ядер клеток (мкм²) на срезах миокарда в контроле и опыте были получены следующие показатели – см. Табл. 11. Выборки, естественно, не связаны между собой. Количество измерений произвольно ограничено 48 в каждой ситуации (в реальных условиях компьютерной морфометрии оно достигает 100-150 на поле зрения ×5 полей зрения ×2 препарата/случай ×5-7 случаев в группе, то есть около 6000 измерений в каждой группе).

Табл. 11. Результаты эксперимента

№	Контроль	Эндотоксикоз	№	Контроль	Эндотоксикоз
1	4,274	4,498	25	4,295	4,117
2	4,377	4,519	26	4,233	4,054
3	4,336	1,954	27	4,109	3,845
4	4,192	4,314	28	4,439	1,773
5	4,254	1,338	29	4,439	4,66
6	4,274	4,171	30	4,398	4,514
7	4,274	4,171	31	4,192	4,201
8	4,274	4,11	32	4,006	1,349
9	3,923	3,926	33	4,088	4,222
10	3,882	4,191	34	4,212	4,201
11	4,109	4,089	35	3,964	1,5
12	3,923	3,864	36	3,799	3,824
13	4,316	1,238	37	3,841	3,845
14	3,964	3,803	38	4,006	1,42
15	4,026	3,844	39	3,799	3,908
16	4,047	1,777	40	3,964	3,95
17	4,192	4,089	41	4,047	4,096
18	4,192	1,293	42	4,047	4,201
19	4,192	4,15	43	4,171	4,222
20	4,192	4,253	44	4,171	4,284
21	4,13	4,232	45	4,377	4,368
22	4,13	4,253	46	4,501	4,514
23	4,212	1,898	47	4,501	4,577
24	4,316	4,478	48	4,398	4,389

Простейший статистический анализ (см. методы сравнения двух выборок в разделе 6.2) приведет к заключению, что при эндотоксикозе (ЭТ) размеры ядер клеток в миокарде «в среднем» уменьшаются. Но, взглянув на гистограмму (Рис. 6), придется изменить формулировку: «при эндотоксикозе среди кардиомиоцитов появляется до 15% клеток с крайне малыми размерами ядер, размер же ядер остальных клеток – велик». Согласитесь, что это не одно и то же!

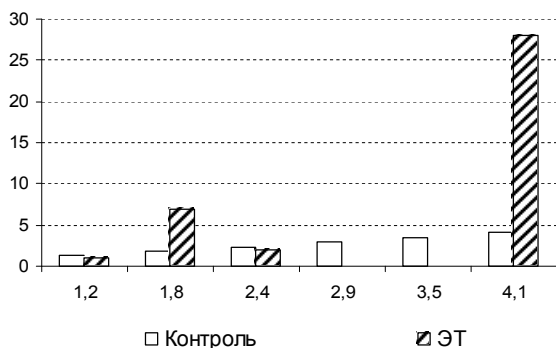


Рис. 6. Гистограммы средних размеров ядер кардиомиоцитов в норме и при эндотоксикозе (компьютерная морфометрия на аппарате Видеостет-Морфо 4.0)

Планирование медико-биологического эксперимента. В заключение настоящего раздела отметим, что, несмотря на то, что в настоящей работе рассматривается применение статистических методов к уже полученным в результате проведения эксперимента данным, знание этих методов позволяет планировать эксперимент на стадии его подготовки. Например, формулы, определяющие эмпирические значения критериев сравнения двух выборок (см. раздел 6.2), совместно с фиксированными критическими их значениями, позволяют заранее (до проведения эксперимента) оценить необходимый объем выборки и другие важные параметры³⁴.

³⁴ Конечно, чем больше объемы выборок, тем в некотором смысле лучше, то есть тем проще будет обосновать различия, если они есть. Но, с другой стороны, расширение выборки требует от исследователя определенных усилий, поэтому целесообразно заранее примерно определить требуемый объем выборки.

В большинстве профессиональных компьютерных статистических программ имеется специальный блок – «Определение объема выборки» (Sample size determination), в котором требуется указать, например, какую величину различия средних значений двух выборок исследователь планирует обосновать. Результатом будет минимально необходимый для обоснования с заданным уровнем значимости (степенью уверенности) объем выборок. В программе StatGraphics эта функция реализована в «Describe\Sample Size Determination» или «Compare\Two Samples\Samples Size Determination». Рассматривать подробно и приводить примеры определения объема выборки мы не будем.

Завершив рассмотрение показателей описательной статистики, перейдем к статистическим методам сравнения двух выборок – определения степени достоверности совпадений и различий.

6.2. СРАВНЕНИЕ ДВУХ ВЫБОРОК

6.2.1. Общие подходы к определению достоверности совпадений и различий

В настоящем разделе рассмотрены общие подходы к определению достоверности совпадений и различий характеристик исследуемых объектов [24]. Правило принятия решений относительно того, какой конкретный статистический критерий (метод обработки экспериментальных данных) следует использовать в том или ином случае, описано ниже в разделе 6.2.4 «Алгоритм выбора статистического критерия».

Как отмечалось выше, типовой задачей анализа данных в медико-биологических исследованиях является установление совпадений или различий характеристик экспериментальной и контрольной группы. Для этого формулируются *статистические гипотезы* [9, 32, 47]:

- гипотеза об отсутствии различий (так называемая *нулевая гипотеза*);
- гипотеза о значимости различий (так называемая *альтернативная гипотеза*).

Для принятия решений о том, какую из гипотез (нулевую или альтернативную) следует принять, используют решающие правила – *статистические критерии*³⁵. То есть, на основании информации о результатах наблюдений (характеристиках членов экспериментальной и контрольной группы) вычисляется число, называемое *эмпирическим значением* критерия. Это число сравнивается с известным (например, заданным таблично) эталонным числом, называемым *критическим значением* критерия.

Критические значения приводятся, как правило, для нескольких уровней значимости. *Уровнем значимости* называется вероятность ошибки, заключающейся в отклонении (не принятии) нулевой гипотезы, когда она верна, то есть вероятность того, что различия сочтены существенными, а они на самом деле случайны.

Обычно используют уровни значимости (обозначаемые α), равные 0,05, 0,01 и 0,001. В медико-биологических экспериментальных исследованиях обычно ограничиваются значением 0,05, то есть, грубо говоря, допускается не более чем 5%-ая возможность ошибки.

Если полученное исследователем эмпирическое значение критерия оказывается меньше или равно критическому, то принимается нулевая гипотеза – считается, что на заданном уровне значимости (то есть при том значении α , для которого рассчитано критическое значение критерия) характеристики экспериментальной и контрольной групп совпадают. В противном случае, если эмпирическое значение критерия оказывается строго больше критического, то нулевая гипотеза отвергается и принимается альтернативная гипотеза – характеристики экспериментальной и контрольной группы считаются различными с достоверностью различий $1 - \alpha$. Например, если $\alpha = 0,05$ и принята альтернативная гипотеза, то достоверность различий равна 0,95 или 95%. То есть, ***достоверность различия*** характеристик – это дополнение до

³⁵ Заметим, что в математической статистике исторически сложилось называть статистическими критериями не только решающие правила, но и методы расчета определенного числа (используемого в решающих правилах), а также само это число.

единицы уровня значимости при проверке гипотезы о совпадении характеристик двух независимых выборок³⁶.

Другими словами, чем меньше эмпирическое значение критерия (чем левее оно находится от критического значения), тем больше степень совпадения характеристик сравниваемых объектов. И наоборот, чем больше эмпирическое значение критерия (чем правее оно находится от критического значения), тем сильнее различаются характеристики сравниваемых объектов.

В дальнейшем мы ограничимся уровнем значимости $\alpha = 0,05$, поэтому, **если эмпирическое значение критерия оказывается меньше или равно критическому, то можно сделать вывод, что «характеристики экспериментальной и контрольной групп совпадают на уровне значимости 0,05».** Если эмпирическое значение критерия оказывается строго больше критического, **то можно сделать вывод, что «достоверность различий характеристик экспериментальной и контрольной групп равна 95%».**

Опишем методики расчета эмпирических значений критериев для двух типовых задач анализа данных – сравнения выборок, содержащих данные, измеренные в шкале отношений (раздел 6.2.2) и порядковой шкале (раздел 6.2.3). Численно задача сравнения двух выборок может быть решена с помощью программы «Педагогическая статистика» (скачать которую бесплатно можно с сайта www.mtas.ru), в которую «заложены» все рассматриваемые в разделе 6.2 статистические критерии.

6.2.2. Методика определения достоверности совпадений и различий для экспериментальных данных, измеренных в шкале отношений

Рассмотрим случай (см. описание исходных данных выше, в разделе 5.1), когда для измерений используется шкала отношений. Предположим, что имеется экспериментальная группа, состоящая

³⁶ Отметим, что термин «достоверность различий» принят в медико-биологических исследованиях, в то время как в математической статистике [9, 11, 17, 32] он практически не используется. Тем не менее, в настоящей работе мы будем применять его как словосочетание, устоявшееся в рассматриваемой предметной области.

из N пациентов, и контрольная группа, состоящая из M пациентов. Допустим, что в результате измерения одного и того же показателя с помощью одной и той же процедуры измерений были получены следующие данные: $x = (x_1, x_2, \dots, x_N)$ – выборка для экспериментальной группы и $y = (y_1, y_2, \dots, y_M)$ – выборка для контрольной группы, где x_i – элемент выборки – значение исследуемого показателя у i -го члена (пациента) экспериментальной группы, $i = 1, 2, \dots, N$, а y_j – значение исследуемого показателя у j -го члена контрольной группы, $j = 1, 2, \dots, M$. Так как измерения производились в шкале отношений, то $\{x_i\}$ и $\{y_j\}$ – положительные, в том числе, возможно – целые, числа, для которых имеют смысл все арифметические операции. В качестве примера будем рассматривать результаты измерений степени регенерации суставного хряща (СРСХ) в контрольной и экспериментальной группах до и после эксперимента (см. Табл. 3).

Для данных, измеренных в шкале отношений, для проверки гипотезы о совпадении характеристик двух групп целесообразно³⁷ использование либо критерия Крамера-Уэлча³⁸ [17, 32], либо критерия Вилкоксона-Манна-Уитни³⁹ [32, 44, 47].

Критерий Крамера-Уэлча. Эмпирическое значение данного критерия рассчитывается на основании информации об объемах N и M выборок x и y , выборочных средних \bar{x} и \bar{y} и выборочных дисперсиях s_x^2 и s_y^2 сравниваемых выборок (эти значения могут быть вычислены вручную⁴⁰ по формулам (1)-(2) или с помощью инструмента «Описательная статистика» в компьютерной программе Microsoft Excel для Windows – см. раздел 6.1) по следующей формуле:

³⁷ Выбор критериев достаточно широк, в чем можно убедиться, ознакомившись с приведенными в списке литературы публикациями. Однако нашей целью является описание статистических критериев, адекватных типовым для медико-биологических исследований задачам анализа данных.

³⁸ Критерий Крамера-Уэлча является более эффективным «заменителем» такого известного в физике и технике критерия как t -критерий (критерий Стьюдента) [32].

³⁹ Критерий Вилкоксона-Манна-Уитни плохо применим в условиях, когда число отличающихся друг от друга значений в выборках мало.

⁴⁰ Отметим, что мы даем описание методики расчетов с тем, чтобы читатель, даже пользуясь компьютерными статистическими программами, понимал, что эти программы делают и как интерпретировать выдаваемые ими результаты.

$$(3) T_{эмт} = \frac{\sqrt{M \cdot N} |\bar{x} - \bar{y}|}{\sqrt{M \cdot s_x^2 + N \cdot s_y^2}}.$$

Алгоритм определения достоверности совпадений и различий характеристик сравниваемых выборок для экспериментальных данных, измеренных в шкале отношений, с помощью критерия Крамера-Уэлча заключается в следующем:

1. Вычислить для сравниваемых выборок $T_{эмт}$ – эмпирическое значение критерия Крамера-Уэлча по формуле (3).
2. Сравнить это значение с критическим значением $T_{0,05} = 1,96$: если $T_{эмт} \leq 1,96$, то сделать вывод: «характеристики сравниваемых выборок совпадают на уровне значимости 0,05»; если $T_{эмт} > 1,96$, то сделать вывод «достоверность различий характеристик сравниваемых выборок составляет⁴¹ 95%».

В качестве примера применим алгоритм для данных из Табл.

3.

Для этого сравним сначала оценки СРСХ в контрольной и экспериментальной группе до начала эксперимента. Вычисляем⁴² по формуле (3) значение $T_{эмт} = 0,04 \leq 1,96$. Следовательно, гипотеза о совпадении характеристик контрольной и экспериментальной групп до начала эксперимента принимается на уровне значимости 0,05.

Теперь сравним характеристики контрольной и экспериментальной групп после окончания эксперимента. Вычисляем по формуле (3) значение $T_{эмт} = 2,42 > 1,96$. Следовательно, достоверность различий характеристик контрольной и экспериментальной групп после окончания эксперимента составляет 95%.

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают, а конечные (после окончания эксперимента) – различаются. Следовательно, можно

⁴¹ *Корректнее говорить, что достоверность различий составляет не менее 95%, однако, так мы условились считать достаточной 95%-ую достоверность различий, то будем говорить, что достоверность различий составляет 95%.*

⁴² *Для сокращения ручных расчетов средние и дисперсии могут быть вычислены в рамках описательной статистики в компьютерной программе Microsoft Excel для Windows – см. выше.*

сделать вывод, что эффект изменений обусловлен именно применением экспериментального воздействия.

Отметим, что мы не рассматриваем вопрос о том, «в какую сторону» экспериментальная группа отличается от контрольной, то есть, улучшились или ухудшились (с содержательной точки зрения, не имеющей отношения к статистическим методам и являющейся прерогативой биологии и медицины) исследуемые характеристики.

Критерий Вилкоксона-Манна-Уитни⁴³. Данный критерий оперирует не с абсолютными значениями элементов двух выборок, а с результатами их парных сравнений. Например, существенно, что у пациента А оценка степени регенерации суставного хряща (СРСХ) больше, чем у пациента В, а на сколько больше – не важно.

Возьмем две выборки⁴⁴: $\{x_i\}_{i=1\dots N}$ и $\{y_j\}_{j=1\dots M}$ и для каждого элемента первой⁴⁵ выборки x_i , $i = 1\dots N$, определим число a_i элементов второй выборки, которые превосходят его по своему значению (то есть, число таких y_j , что $y_j > x_i$), а также число b_i элементов второй выборки, которые по своему значению равны ему (то есть число таких y_j , что $y_j = x_i$). Сумма

$$a_1 + a_2 + \dots + a_N + \frac{1}{2} (b_1 + b_2 + \dots + b_N) = \sum_{i=1}^N a_i + \frac{1}{2} \sum_{i=1}^N b_i$$

по всем N членам первой выборки называется *эмпирическим* значением критерия Манна-Уитни и обозначается U .

⁴³ Существуют два критерия – Вилкоксона и Манна-Уитни, однако, так как они однозначно связаны между собой, будем говорить об одном критерии Вилкоксона-Манна-Уитни.

⁴⁴ Минимальное ограничение на использование критерия Вилкоксона-Манна-Уитни следующее: каждая выборка должна содержать не менее трех элементов, если же в одной из выборок всего два элемента, то во второй их должно быть не менее пяти. Однако, так как правило принятия решений основано на асимптотической нормальности статистики критерия, то объемы выборок должны быть не менее 10. Иначе надо пользоваться специальными таблицами.

⁴⁵ Какую выборку считать первой, а какую второй, не имеет значения, хотя при вычислениях вручную удобнее первой считать ту выборку, в которой меньше членов.

Определим эмпирическое значение критерия Вилкоксона⁴⁶:

$$(4) W_{эмп} = \frac{|\frac{N \cdot M}{2} - U|}{\sqrt{\frac{N \cdot M \cdot (N + M + 1)}{12}}}.$$

Алгоритм определения достоверности совпадений и различий для экспериментальных данных, измеренных в шкале отношений, с помощью критерия Вилкоксона-Манна-Уитни заключается в следующем:

1. Вычислить для сравниваемых выборок $W_{эмп}$ – эмпирическое значение критерия Вилкоксона по формуле (4).
2. Сравнить это значение с критическим значением $W_{0,05} = 1,96$: если $W_{эмп} \leq 1,96$, то сделать вывод: «характеристики сравниваемых выборок совпадают на уровне значимости 0,05»; если $W_{эмп} > 1,96$, то сделать вывод «достоверность различий характеристик сравниваемых выборок составляет 95%».

В качестве примера применим алгоритм для данных из Табл.

3.

Для этого сравним сначала оценки СРСХ в контрольной и экспериментальной группе до начала эксперимента. В Табл. 12 приведены результаты экспериментальной группы (второй столбец), и контрольной группы (пятый столбец), а также для каждого члена экспериментальной группы подсчитано число членов контрольной группы, характеризуемых строго большей (чем он) оценкой СРСХ, плюс полусумма числа членов контрольной группы, характеризуемых такой же (что и он) оценкой СРСХ (третий столбец).

Например, в Табл. 12 серым цветом в пятом столбце помечены члены контрольной группы, имеющие строго большие оценки СРСХ, чем первый член (то есть $i = 1$) экспериментальной группы (его оценка СРСХ равна 12).

⁴⁶ Иногда при ручных расчетах удобнее использовать другой алгоритм расчета эмпирического значения критерия Вилкоксона – просуммировать сумму рангов элементов первой выборки (номеров мест, на которых находятся соответствующие элементы в упорядочении в порядке возрастания элементов объединенной выборки, то есть полученной объединением исходных выборок) – см. п. 4.5. в [32].

Табл. 12. Пример вычисления эмпирического значения критерия Манна-Уитни

Номер члена экспериментальной группы	Оценка СРСХ i -го члена экспериментальной группы до начала эксперимента	Число членов контрольной группы, имеющих строго большую оценку СРСХ, чем у i -го члена экспериментальной группы $a_i + b_i / 2$	Номер члена контрольной группы	Оценка СРСХ j -го члена контрольной группы до начала эксперимента
i	x_i		j	y_j
1	12	17	1	15
2	11	19,5	2	13
3	15	9	3	11
4	17	5	4	18
5	18	4	5	10
6	6	28,5	6	8
7	8	24,5	7	20
8	10	21,5	8	7
9	16	6	9	8
10	12	17	10	12
11	15	9	11	15
12	14	12	12	16
13	19	2	13	13
14	13	14,5	14	14
15	19	2	15	14
16	12	17	16	19
17	11	19,5	17	7
18	16	6	18	8
19	12	17	19	11
20	8	24,5	20	12
21	13	14,5	21	15
22	7	27	22	16
23	15	9	23	13
24	8	24,5	24	5
25	9	22,5	25	11
–	–	–	26	19
–	–	–	27	18
–	–	–	28	9
–	–	–	29	6
–	–	–	30	15

Значит $x_1 = 12$, и число таких y_j , что $y_j > x_1$ равно 16. Следовательно, $a_1 = 16$. Число таких y_j , что $y_j = x_1$ равно 2. Следовательно, $b_1 = 2$. Итак, $a_1 + b_1/2 = 17$, то есть число затененных ячеек равно 17. Записываем это число во вторую строку третьего столбца. Аналогично заполняются остальные строки третьего столбца.

Сумма всех 25 чисел в третьем столбце Табл. 12 дает эмпирическое значение критерия Манна-Уитни $U = 373$. Вычисляем по формуле (4) значение $W_{эм} = 0,0338 \leq 1,96$. Следовательно, гипотеза о том, что сравниваемые выборки совпадают, принимается на уровне значимости 0,05.

Теперь аналогичным образом (построив таблицу, аналогичную Табл. 12, и вычислив эмпирическое значение критерия Вилкоксона) сравним оценки СРСХ в контрольной и экспериментальной группе после окончания эксперимента. Вычисляем по формуле (4) значение $W_{эм} = 2,1974 > 1,96$. Следовательно, достоверность различий сравниваемых выборок составляет 95%.

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают, а конечные (после окончания эксперимента) – различаются. Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментального воздействия.

6.2.3. Методика определения достоверности совпадений и различий для экспериментальных данных, измеренных в порядковой шкале

Рассмотрим случай, когда используется порядковая шкала с L различными баллами. Характеристикой группы будет число ее членов, набравших тот или иной балл. Для экспериментальной группы вектор баллов есть $n = (n_1, n_2, \dots, n_L)$, где n_k – число членов экспериментальной группы, получивших k -ый балл, $k = 1, 2, \dots, L$. Для контрольной группы вектор баллов есть $m = (m_1, m_2, \dots, m_L)$, где m_k – число членов контрольной группы, получивших k -ый балл, $k = 1, 2, \dots, L$. Для рассматриваемого нами числового примера ($L = 3$ – «низкий», «средний» или «высокий» уровень СРСХ) данные приведены в Табл. 6.

Для данных, измеренных в порядковой шкале (см., например, таблицу 5), целесообразно использование критерия однородности χ^2 («хи» – буква греческого алфавита, название критерия читается: «хи-квадрат») [39], эмпирическое значение $\chi_{эм}^2$ которого вычисляется по следующей формуле⁴⁷ (пример расчета приведен ниже):

$$(5) \chi_{эм}^2 = N \cdot M \cdot \sum_{i=1}^L \frac{\left(\frac{n_i}{N} - \frac{m_i}{M}\right)^2}{n_i + m_i}.$$

Критические значения $\chi_{0.05}^2$ критерия χ^2 для уровня значимости 0,05 приведены в Табл. 13 (статистические таблицы критических значений статистических критериев для различных уровней значимости и различных – в том числе больших 10 – градаций шкалы отношений можно найти, практически, в любом учебнике по статистическим методам, или в специальных статистических таблицах [5]).

Табл. 13. Критические значения критерия χ^2 для уровня значимости $\alpha = 0.05$

$L-1$	1	2	3	4	5	6	7	8	9
$\chi_{0.05}^2$	3,84	5,99	7,82	9,49	11,07	12,59	14,07	15,52	16,92

Алгоритм определения достоверности совпадений и различий для экспериментальных данных, измеренных в порядковой шкале, заключается в следующем:

1. Вычислить для сравниваемых выборок $\chi_{эм}^2$ – эмпирическое значение критерия χ^2 по формуле (5).
2. Сравнить это значение с критическим значением $\chi_{0.05}^2$, взятым из Табл. 13: если $\chi_{эм}^2 \leq \chi_{0.05}^2$, то сделать вывод:

⁴⁷ Обычно считают, что критерий хи-квадрат применим при условии, что для любого значения балла в любой из сравниваемых выборок не менее пяти ее членов получили данный балл, то есть: $n_i \geq 5, m_i \geq 5, i = 1, 2, \dots, L$. Кроме того, желательно, чтобы число градаций L было не менее трех. Если $L = 2$, то есть используется дихотомическая шкала, то можно применять критерий Фишера – см. ниже настоящий раздел.

«характеристики сравниваемых выборок совпадают на уровне значимости 0,05»; если $\chi_{эмп}^2 > \chi_{0,05}^2$, то сделать вывод «достоверность различий характеристик сравниваемых выборок составляет 95%».

Применим алгоритм для данных из Табл. 6. Сначала вычисляем по формуле (5) эмпирические значения критерия χ^2 . Для примера приведем расчет. Параметры экспериментальной группы ($N = 25$) после окончания эксперимента: $n_1 = 2$, $n_2 = 13$, $n_3 = 10$ (то есть 2 пациента характеризуются «низким» уровнем СРСХ, 13 – «средним» и 10 – «высоким» – см. выше таблицу Табл. 6), контрольной группы ($M = 30$): $m_1 = 12$, $m_2 = 10$, $m_3 = 8$. Подставляя в формулу (5), получаем:

$$\begin{aligned} \chi_{эмп}^2 = 25 \cdot 30 \cdot [& \left(\frac{2}{25} - \frac{12}{30} \right)^2 / (2 + 12) + \left(\frac{13}{25} - \frac{10}{30} \right)^2 / (13 + 10) + \\ & + \left(\frac{10}{25} - \frac{8}{30} \right)^2 / (10 + 8)] = 7,36. \end{aligned}$$

В рассматриваемом примере $L = 3$ (выделены три уровня СРСХ – «низкий», «средний» и «высокий»), то есть $L - 1 = 2$. Из Табл. 13 получаем для $L - 1 = 2$: $\chi_{0,05}^2 = 5,99$. Итак, эмпирическое значение критерия χ^2 , равное $\chi_{эмп} = 7,36$, строго больше критического $\chi_{0,05}^2 = 5,99$, поэтому «достоверность различий характеристик экспериментальной и контрольной групп после окончания эксперимента составляет 95%».

Аналогичным образом, сравнивая экспериментальную и контрольную группу до начала эксперимента, получаем $\chi_{эмп}^2 = 0,03$. Это значение строго меньше критического $\chi_{0,05}^2 = 5,99$, поэтому можно заключить, что «характеристики экспериментальной и контрольной групп до начала эксперимента, совпадают с уровнем значимости 0,05».

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают, а конечные (после окончания эксперимента) – различаются. Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментального воздействия.

Отметим, что использование критерия χ^2 для сравнения экспериментальной группы после окончания эксперимента с какой-либо другой не совсем корректно, так как только два ее члена имеют низкий уровень СРСС – см. Табл. 6 (а критерий χ^2 , как отмечалось выше, применим в случае, если для любого значения балла в любой из сравниваемых выборок не менее пяти ее членов получили данный балл). Выходом является снижение числа градаций. Однако в рассматриваемом случае градаций всего три, что также является ограничением снизу на использование критерия χ^2 . Поэтому альтернативой может служить либо изменение границ между классами (границы диапазонов – см. Табл. 4 – определяют к какому классу порядковой шкалы относится то или иное измерение в шкале отношений), либо переход к дихотомической шкале.

Дихотомическая шкала. Отдельно рассмотрим случай, когда используется дихотомическая шкала – порядковая шкала с двумя различными упорядоченными баллами – «высокий»-«низкий», «прошел тест»-«не прошел» и т.д. Характеристикой группы, помимо общего числа ее членов, будет число членов (или доля, процент от общего числа), набравших заданный, например – максимальный, балл (в общем случае – число членов, обладающих заданным признаком).

Для экспериментальной группы, описываемой двумя числами (n_1, n_2) , где n_1 – число членов рассматриваемой группы, набравших низкий балл, n_2 – набравших высокий балл, $n_1 + n_2 = N$, доля p ее членов, набравших максимальный балл, равна: $p = n_2 / N$. Для контрольной группы, описываемой двумя числами (m_1, m_2) , где $m_1 + m_2 = M$, доля q ее членов, набравших максимальный балл, равна: $q = m_2 / M$.

Рассмотрим пример: для каждого из столбцов Табл. 3, считая, что возможны два уровня СРСХ – «не удовлетворительный» (оценка СРСХ меньше либо равна 10) и «удовлетворительный» (оценка СРСХ строго больше 10) определяем распределение членов экспериментальной и контрольной группы по двум уровням СРСХ и получаем Табл. 14 (для экспериментальной группы до начала эксперимента $p = 0,72$ (или 72%), после окончания эксперимента $p = 0,92$; для контрольной группы до начала эксперимента $q = 0,70$, после окончания эксперимента $q = 0,60$).

Табл. 14. Результаты дихотомических измерений уровня СРСХ в контрольной и экспериментальной группах до и после эксперимента

	Контрольная группа до начала эксперимента	Экспериментальная группа до начала эксперимента	Контрольная группа после окончания эксперимента	Экспериментальная группа после окончания эксперимента
Доля, которую составляют пациенты с неудовлетворительным уровнем СРСХ	0,30	0,28	0,40	0,08
Доля, которую составляют пациенты с удовлетворительным уровнем СРСХ	0,70	0,72	0,60	0,92

Для данных, измеренных в дихотомической шкале целесообразно⁴⁸ использование критерия Фишера⁴⁹, для которого эмпирическое значение $\varphi_{эмп}$ вычисляется по следующей формуле (арксинус может быть вычислен в Excel):

$$(6) \varphi_{эмп} = |2 \arcsin(\sqrt{p}) - 2 \arcsin(\sqrt{q})| \sqrt{\frac{M \cdot N}{M + N}}.$$

Критическое значение $\varphi_{0,05}$ критерия Фишера для уровня значимости 0,05 равно 1,64.

Алгоритм определения достоверности совпадений и различий для экспериментальных данных, измеренных в дихотомической шкале, заключается в следующем:

1. Вычислить для сравниваемых выборок $\varphi_{эмп}$ – эмпирическое значение критерия Фишера по формуле (6).

⁴⁸ Можно использовать и критерий, соответствующий ранее рассмотренному критерию «хи-квадрат»; этот критерий подробно описан в главе 2 [32].

⁴⁹ В математической статистике существует несколько критериев Фишера. Мы используем один из них – так называемое угловое преобразование, поэтому далее под критерием Фишера будем понимать именно угловое преобразование Фишера.

2. Сравнить это значение с критическим значением $\varphi_{0,05} = 1,64$: если $\varphi_{эм} \leq 1,64$, то сделать вывод: «характеристики сравниваемых выборок совпадают на уровне значимости 0,05»; если $\varphi_{эм} > 1,64$, то сделать вывод: «достоверность различий характеристик сравниваемых выборок составляет 95%».

Применим алгоритм для экспериментальных данных из Табл. 14. Сначала вычисляем по формуле (2) эмпирические значения критерия Фишера. Для примера приведем расчет. Параметры экспериментальной группы ($N = 25$) после окончания эксперимента: $p = 0,92$, контрольной группы ($M = 30$): $q = 0,60$ (см. Табл. 14). Подставляя в формулу (6), получаем:

$$\varphi_{эм} = |2 \arcsin(\sqrt{0,92}) - 2 \arcsin(\sqrt{0,6})| \sqrt{\frac{25 \cdot 30}{25 + 30}} = 2,94.$$

Так как $\varphi_{эм} = 2,94 > 1,64 = \varphi_{кр}$, то «достоверность различий состояний экспериментальной и контрольной групп после окончания эксперимента составляет 95%».

Аналогичным образом определяем, что эмпирическое значение критерия Фишера, получаемое при сравнении характеристик контрольной и экспериментальной групп до начала эксперимента, равно 0,16. Следовательно «состояния экспериментальной и контрольной групп до начала эксперимента совпадают с уровнем значимости 0,05».

Итак, начальные (до начала эксперимента) состояния экспериментальной и контрольной групп совпадают, а конечные (после окончания эксперимента) – различаются. Следовательно, можно сделать вывод, что эффект изменений обусловлен именно применением экспериментального воздействия.

Подчеркнем, что данный вывод (один и тот же) был получен при применении к соответствующим экспериментальным данным

всех четырех критериев – Крамера-Уэлча, Вилкоксона-Манна-Уитни, χ^2 и Фишера⁵⁰.

Отметим также, что, применяя критерий Крамера-Уэлча или Вилкоксона-Манна-Уитни к данным второго примера (см. Табл. 7), можно убедиться, что достоверность различий таких переменных как ВСММ (вещества средней молярной массы) и ОП (олигопептиды) до и после окончания эксперимента составляет 95%.

6.2.4. Алгоритм выбора статистического критерия

Рассмотренные выше методы анализа сходства/различий являются базовыми по следующим причинам. Во-первых, они включают большинство (см. четвертый раздел) задач анализа данных, встречающихся в медико-биологическом эксперименте. Во-вторых, они сформулированы для простейшей схемы организации эксперимента (см. второй раздел) – когда состояние исследуемых объектов описывается одним показателем и измеряется два раза – до начала и после завершения воздействия. Сделаем пояснение для других случаев.

Если возникает многокритериальность (объекты описываются одновременно по нескольким критериям – см. раздел «Комплексные оценки» выше), то описание и сравнение экспериментальной и контрольной групп⁵¹ по каждому из критериев может производиться независимо в рамках одной из базовых задач.

⁵⁰ Перечисленные четыре критерия обладают различной «мощностью» – возможны случаи, когда, например, применение критерия Крамера-Уэлча к данным, измеренным в шкале отношений, свидетельствует о наличии статистически значимых различий, а применение критерия χ^2 к тем же результатам эксперимента, переведенным в порядковую шкалу, свидетельствует о совпадении характеристик (см. также обсуждение потерь информации при переходе от шкалы отношений к порядковой шкале выше в пятом разделе). Поэтому можно рекомендовать максимально использовать всю полученную в результате эксперимента информацию – если измерения проводились в шкале отношений, то и обрабатывать данные следует в этой шкале, переходя к порядковой шкале только в случае крайней необходимости.

⁵¹ Встречаются случаи, когда имеется несколько экспериментальных или несколько контрольных групп. При этом попарное их сравнение все равно является одной из базовых задач.

Аналогично, если возникает динамика (то есть, состояния объектов измеряются более чем два раза), то описание и сравнение групп может производиться несколько раз независимо (в каждый момент времени) в рамках одной из базовых задач. Если же у исследователя имеется желание сразу анализировать одновременно несколько групп (в динамике) и/или несколько показателей, то необходимо применение соответствующих статистических методов. Их описание выходит за рамки настоящей работы, ознакомиться с ними можно в публикациях [32, 40, 44].

Завершив описание методик анализа данных, поясним, как следует выбирать статистические критерии, то есть приведем алгоритм выбора статистического критерия – процедуру принятия решения относительно того, какой статистический критерий (из четырех описанных выше – Крамера-Уэлча, Вилкоксона-Манна-Уитни, χ^2 и Фишера) использовать в той или иной ситуации сравнения двух выборок.

Алгоритм выбора статистического критерия (см. Табл. 15). Во-первых, необходимо определить какая шкала измерений используется – отношений, порядковая или номинальная (см. раздел 3).

Для шкалы отношений целесообразно использовать критерий Крамера-Уэлча (напомним, что условия применимости всех критериев приведены выше при описании этих критериев). Если число различающихся между собой значений в сравниваемых выборках велико (более десяти)⁵², то возможно использование критерия Вилкоксона-Манна-Уитни⁵³.

Для порядковой шкалы целесообразно использовать критерий Вилкоксона-Манна-Уитни, возможно также использование критерия χ^2 .

⁵² Например, выборка (1, 2, 2, 2, 1, 1, 2, 1, 1, 1) содержит всего два различных значения – единицу и двойку. В то же время, например, выборка (2, 0, 1, 5, 8, 4, 2, 7, 3, 9) того же объема (десять элементов) содержит десять различных значений. Понятно, что приводимые границы числа различающихся между собой значений – 10 – примерны, приближительны.

⁵³ Критерий Вилкоксона-Манна-Уитни не является самостоятельным; самостоятельными являются критерии Смирнова и Лемана-Розенблатта, именно с их помощью можно проверять гипотезу о том, что две выборки «одинаковы» (в том числе, что совпадают их средние, дисперсии и все другие показатели – см. пп. 4.5, 4.6. [32]).

Для номинальной шкалы следует использовать критерий χ^2 .

Для дихотомической шкалы следует использовать критерий Фишера.

Табл. 15. Алгоритм выбора статистического критерия

Шкала измерений	Статистический критерий
Отношений	Крамера-Уэлча, Вилкоксона-Манна-Уитни
Порядковая	Вилкоксона-Манна-Уитни, χ^2
Номинальная	χ^2
Дихотомическая	Фишера

Как отмечалось выше (см. раздел 2), целью любого эксперимента является эмпирическое подтверждение или опровержение гипотезы исследования и/или справедливости теоретических результатов, например, обоснование того, что предлагаемое воздействие:

а) направлено на лечение данного патологического процесса (а не способствует его развитию) и

б) более эффективно (или, возможно, наоборот – менее эффективно) в сравнении с воздействиями, ранее рекомендованными к использованию для этих целей.

Для этого, как минимум, необходимо показать, что, будучи примененным к тому же объекту (например – к группе животных с рассматриваемым заболеванием), предлагаемое воздействие дает другие результаты, чем применение традиционного воздействия.

Для этого выделяется экспериментальная группа, которая сравнивается с контрольной группой. Различие эффектов тестируемых воздействий будет обосновано, если две эти группы, первоначально совпадающие по своим характеристикам, различаются после реализации воздействий. Следовательно, требуется провести два сравнения и показать, что при первом сравнении (до начала эксперимента) характеристики экспериментальной и контрольной группы совпадают, а при втором (после окончания эксперимента) – различаются.

Так как объектом медико-биологического эксперимента, являются живые организмы, каждый из которых индивидуален, то говорить о совпадении или различии характеристик эксперимен-

тальной и контрольной групп можно лишь в чисто формальном, статистическом смысле. Эти различия, естественно, распространяются только на критерии, выбранные к конкретному анализу – в миллионах прочих характеристик объекты в группах могут и совпадать. Для того чтобы выяснить, являются ли совпадения или различия случайными, используются статистические методы, которые позволяют на основании данных, полученных в результате эксперимента, принять обоснованное решение о совпадениях или различиях.

Общий алгоритм использования статистических критериев прост: до начала и после окончания эксперимента на основании информации о результатах наблюдений (характеристиках членов экспериментальной и контрольной группы) вычисляется эмпирическое значение критерия (алгоритм выбора статистического критерия приведен выше, формулы для вычислений – в разделах 6.2.2 и 6.2.3). Это число сравнивается с известным (табличным) числом – критическим значением критерия (критические значения⁵⁴ для всех рекомендуемых нами критериев приведены выше в разделах 6.2.2 и 6.2.3). Если эмпирическое значение критерия оказывается меньше или равно критическому, то можно утверждать, что **«характеристики экспериментальной и контрольной групп совпадают на уровне значимости 0,05 по статистическому критерию ...** (далее следует название использованного критерия: Крамера-Уэлча, Вилкоксона-Манна-Уитни, хи-квадрат, Фишера)». В противном случае (если эмпирическое значение критерия оказывается строго больше критического) можно утверждать, что **«достоверность различий характеристик экспериментальной и контрольной групп по статистическому критерию ... равна 95%»**.

Следовательно, если характеристики экспериментальной и контрольной групп до начала эксперимента совпадают с уровнем значимости 0,05, и, одновременно с этим, достоверность различий характеристик экспериментальной и контрольной групп после эксперимента равна⁵⁵ 95%, то можно сделать вывод, что⁵⁶ **«приме-**

⁵⁴ Напомним, что выше мы решили ограничиться 0,05 уровнем значимости и, соответственно, 95%-ым уровнем достоверности различий.

⁵⁵ Еще раз напомним, что уровень значимости и достоверность различий в сумме равны единице (или 100 %) – см. также выше раздел 6.2.1.

нение рассматриваемого экспериментального воздействия (например, нового лекарственного препарата) приводит к статистически значимым (на уровне 95% по критерию ...) отличиям результатов».

6.3. ИССЛЕДОВАНИЕ ЗАВИСИМОСТЕЙ

Если рассмотренные в предыдущих разделах описательная статистика и статистические критерии позволяли, соответственно, компактно представлять полученные результаты и определять сходства и различия, то следующим этапом анализа данных обычно является исследование зависимостей. Для этих целей применяются корреляционный анализ и дисперсионный анализ (для установления факта наличия/отсутствия зависимости между переменными), а также регрессионный анализ (для нахождения количественной зависимости между переменными). Эти три группы статистических методов кратко рассматриваются ниже в настоящем разделе: корреляционный анализ – раздел 6.3.1, дисперсионный анализ – раздел 6.3.2, регрессионный анализ – раздел 6.3.3.

6.3.1. Корреляционный анализ

Корреляция (Correlation) – связь между двумя или более переменными (в последнем случае корреляция называется множественной). Цель корреляционного анализа – установление наличия или отсутствия этой связи.

В случае, когда имеются две переменных, значения которых измерены в шкале отношений⁵⁷, используется *коэффициент линейной корреляции Пирсона r* , который принимает значения от -1 до +1 (нулевое его значение свидетельствует об отсутствии корреля-

⁵⁶ Понятно, что в каждом конкретном случае общие термины «характеристика группы», «экспериментальное воздействие», «результат» заменяются на конкретные характеристики, воздействия и результаты.

⁵⁷ Единицы измерений при этом не важны – например, масса тела может быть измерена в граммах, килограммах, тоннах – они не влияют на значение коэффициента корреляции.

ции⁵⁸) – см. Рис. 7. Термин «линейный» свидетельствует о том, что исследуется наличие линейной связи между переменными – если $r(x, y) = 1$, то одна переменная линейно зависит от другой (и, естественно, наоборот), то есть существуют константы a и b , причем $a > 0$, такие что $y = a x + b$.

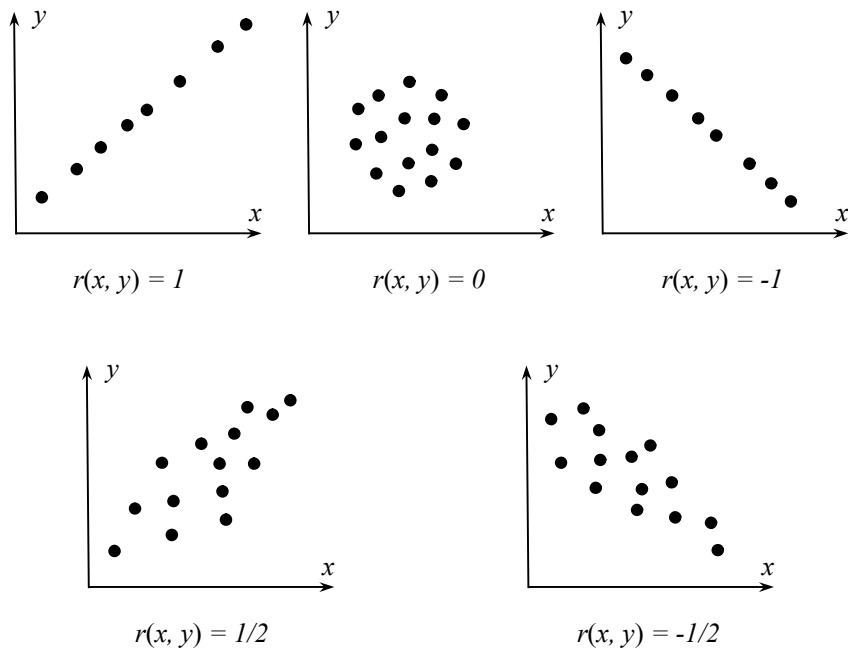


Рис. 7. Величины коэффициента линейной корреляции в различных ситуациях

Рассмотрим первый пример (см. раздел 5.1). Обозначим x_0 – вектор характеристик контрольной группы до начала эксперимента, x – после окончания эксперимента, y_0 – вектор характеристик экспериментальной группы до начала эксперимента, y – после окончания эксперимента.

Сделаем следующее важное замечание. Если решение задачи о выявлении сходства/различия двух выборок не зависит от перестав-

⁵⁸ Корректно говоря, этот факт справедлив в случае, если анализируемая пара переменных описывается двумерным нормальным распределением.

новок значений переменных внутри одной и той же выборки, то при изучении связи между переменными важно сравнивать значения переменных, относящихся к одному и тому же объекту. Так, например, при анализе связи между оценкой степени регенерации суставного хряща (СРСХ) до начала эксперимента и после его окончания, необходимо сравнивать пары оценок, характеризующие одного и того же пациента.

Коэффициент корреляции Пирсона может быть вычислен в программе StatGraphics следующим образом: «Describe\Numeric Data\Multiple Variable Analysis». Для данных, приведенных в Табл. 3, получаем: $r(x_0, x) = 0,76$; $r(y_0, y) = 0,42$. Коэффициент корреляции Пирсона также может быть вычислен в программе Excel: «Сервис\Анализ данных\Корреляция». Отметим, что коэффициент корреляции Пирсона симметричен, то есть не зависит от перестановки переменных: $r(x_0, x) = r(x, x_0)$.

Достаточно высокий коэффициент корреляции $r(x_0, x)$ в рассматриваемом примере означает, что те из членов контрольной группы, которые имели достаточно высокие оценки СРСХ до эксперимента, имеют высокие оценки СРСХ и после эксперимента. Невысокое значение коэффициента корреляции $r(y_0, y)$ свидетельствует, что у членов экспериментальной группы оценка СРСХ до начала эксперимента слабо линейно связана с оценкой СРСХ после окончания эксперимента.

Для данных, измеренных в порядковой шкале, следует использовать *коэффициент ранговой корреляции Спирмена* (он может применяться и для данных, измеренных в интервальной шкале, так как является непараметрическим и улавливает тенденцию – изменения переменных в одном направлении), который обозначается s и определяется сравнением *рангов* – номеров значений сравниваемых переменных в их упорядочении. Коэффициент корреляции Спирмена может быть вычислен в программе StatGraphics следующим образом: «Describe\Numeric Data\Multiple Variable Analysis». Для данных, приведенных в Табл. 3, получаем: $s(x_0, x) = 0,78$; $s(y_0, y) = 0,39$.

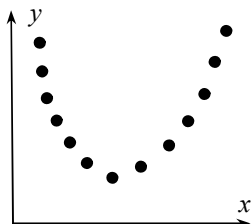


Рис. 8. Пример: коэффициент линейной корреляции (Пирсона) равен нулю для функционально (нелинейно и немонотонно) связанных переменных

Коэффициент корреляции Спирмена является менее чувствительным, чем коэффициент корреляции Пирсона⁵⁹ (так как первый в случае измерений в шкале отношений учитывает лишь упорядочение элементов выборки). В то же время, он позволяет выявлять корреляцию между монотонно *нелинейно* связанными переменными (для которых коэффициент Пирсона может показывать незначительную корреляцию – см. Рис. 8).

Универсальных рецептов установления корреляции между немонотонно и нелинейно связанными переменными на сегодняшний день не существует.

Отметим, что большое (близкое к плюс единице или к минус единице) значение коэффициента корреляции говорит о связи переменных, но ничего не говорит о причинно-следственных отношениях между ними. Так, например, из высокой корреляции температуры воздуха за окном и времени суток нельзя делать вывод о том, что движение солнца обусловлено изменениями температуры воздуха.

Рассмотрим второй пример (см. Табл. 7 в разделе 5.2). Вычислим в программе StatGraphics коэффициент корреляции Пирсона («Describe\Numeric Data\Multiple Variable Analysis»). Результаты представлены в Табл. 16.

Существенная корреляция (коэффициент корреляции более 0,8) наблюдается только между двумя парами переменных (соответствующие ячейки выделены в Табл. 16 жирным шрифтом). Значение коэффициента корреляции в диапазоне от 0,5 до 0,8 свидетельствует о возможной взаимосвязи между соответствующими переменными (см. выделенные курсивом значения в Табл. 16).

⁵⁹ Для линейного коэффициента корреляции (Пирсона) требуется выполнение предположения о нормальности результатов измерений, что редко имеет место в медико-биологическом эксперименте. Ранговые коэффициенты корреляции Спирмена и Кендалла [1, 32, 44] свободны от этого недостатка.

Табл. 16. Корреляция между переменными во втором примере

	M_0	M	$M1$	$M2$	$M3$	BC MM_0	OP_0	$ВСММ$	OP
M_0	1								
M	0,97	1							
$M1$	0,23	0,18	1						
$M2$	0,24	0,29	<i>0,54</i>	1					
$M3$	-0,30	-0,16	-0,04	0,38	1				
$BCMM_0$	-0,38	-0,41	-0,43	-0,39	-0,21	1			
OP_0	0,06	0,05	-0,07	-0,06	-0,12	<i>0,59</i>	1		
$ВСММ$	-0,19	-0,25	-0,23	-0,27	-0,30	0,83	<i>0,56</i>	1	
OP	-0,15	-0,09	-0,22	-0,13	0,21	0,34	<i>0,62</i>	0,26	1

Итак, корреляционный анализ позволяет устанавливать наличие или отсутствие зависимости между переменными. Другим инструментом, дающим ответ на этот вопрос, является дисперсионный анализ.

6.3.2. Дисперсионный анализ

Изучение наличия или отсутствия зависимости между переменными можно проводить и с помощью *дисперсионного анализа* (Analysis of Variance – ANOVA). Его суть заключается в следующем. Дисперсия характеризует «разброс» значений переменной. Переменные связаны, если для объектов, отличающихся значениями одной переменной, отличаются и значения другой переменной. Значит, нужно для всех объектов, имеющих одно и то же значение одной переменной (*называемой независимой переменной*), посмотреть, насколько различаются (насколько велика дисперсия) значения другой (или других) переменной, называемой *зависимой переменной*. Дисперсионный анализ как раз и дает возможность сравнить отношение дисперсии зависимой переменной (*межгрупповой дисперсии*) с дисперсией внутри групп объектов, характеризуемых одними и теми же значениями независимой переменной (*внутригрупповой дисперсией*).

Другими словами, дисперсионный анализ «работает» следующим образом. Выдвигается гипотеза о наличии зависимости между переменными. Выделяются группы элементов выборки с одинаковыми значениями независимой переменной (число таких групп равно числу попарно различных значений независимой переменной).

ной). Если гипотеза о зависимости верна, то значения зависимой переменной внутри каждой группы должны не очень различаться (внутригрупповая дисперсия должна быть мала). Напротив, значения зависимой переменной для различных групп должны различаться сильно (межгрупповая дисперсия должна быть велика). То есть, переменные зависимы, если отношение межгрупповой дисперсии к внутригрупповой (обычно обозначаемое буквой F) велико. Если же гипотеза не верна, то это отношение должно быть мало.

Рассмотрим первый пример (см. Табл. 3 в разделе 5.1). Например, в контрольной группе до начала эксперимента оценку степени регенерации суставного хряща (СРСХ), равную 15, получили 4 пациента (с номерами 1, 11, 21 и 30). Но, после окончания эксперимента они получили, соответственно, оценки: 16, 17, 11 и 10. В то же время до начала эксперимента оценку СРСХ, равную 18, получили два пациента (с номерами 4 и 27). А после окончания эксперимента они получили, соответственно, оценки 17 и 19. И так далее. Спрашивается, связаны ли между собой оценки СРСХ членов контрольной группы до и после окончания эксперимента.

Большое значение F отношения межгрупповой дисперсии к внутригрупповой (вычисляемое в программе StatGraphics следующим образом: «Compare\Analysis of Variance\One-way ANOVA»), равное для рассматриваемого случая 5,82, подтверждает гипотезу, что оценки СРСХ после эксперимента у пациентов, принадлежащих контрольной группе, зависят от их оценок до эксперимента⁶⁰. Отметим, что этот вывод согласован с относительно высоким коэффициентом корреляции между анализируемыми характеристиками, равным 0,76 (см. выше).

Для экспериментальной группы отношение межгрупповой дисперсии к внутригрупповой сравнительно мало – равно 1,79. То есть, не существует устойчивой связи между оценками СРСХ у членов экспериментальной группы до и после эксперимента. Этот вывод согласован с относительно низким коэффициентом корреляции.

⁶⁰ Модуль дисперсионного анализа в программе StatGraphics вычисляет не только отношение дисперсий, но и определяет уровень значимости, на котором можно утверждать наличие или отсутствие зависимости между переменными. Отметим, что в дисперсионном анализе принимается нереалистическое предположение многомерной нормальности результатов измерений.

ляции между анализируемыми характеристиками, равным 0,42 (см. выше).

Подчеркнем, что, во-первых, результаты дисперсионного анализа зависят от того, какая выборка взята в качестве независимой переменной, а какая в качестве зависимой (что не удивительно – ведь исследуются «причинно-следственные» связи). Во-вторых, применять дисперсионный анализ имеет смысл только когда первая выборка содержит повторяющиеся значения – если в ней все значения различны, то внутригрупповая дисперсия во второй выборке будет равна нулю. И, наконец, в третьих, дисперсионный анализ возможен и для случая нескольких независимых переменных (так называемый многофакторный дисперсионный анализ – MANOVA – Multifactor Analysis of Variance). «Идеология» его такая же, что и однофакторного дисперсионного анализа, поэтому рассмотрим его кратко применительно к данным второго примера (см. раздел 5.2).

Исследуем, зависит ли масса печени (M1) от пола и типа животного (см. исходные данные в Табл. 7). Для этого проведем многофакторный дисперсионный анализ в котором масса печени является зависимой переменной, а пол и тип – факторами. Определяем в программе StatGraphics («Compare\Analyses of variance\Multifactor ANOVA»), что F -отношение равно 4,41 для переменной «пол» и 8,49 для переменной «тип», причем оба фактора влияют на переменную «масса печени» с 95%-ым уровнем значимости. Поскольку разделение на группы по порогу болевой чувствительности является весьма чутким тестом на выделение общей неспецифической реактивности организма, этот факт вполне соответствует теоретическим предположениям, что особи с высоким порогом болевой чувствительности имеют более крупные внутренние органы.

6.3.3. Регрессионный анализ

Если корреляционный и дисперсионный анализ, качественно говоря, дают ответ на вопрос, существует ли взаимосвязь между переменными, то регрессионный анализ предназначен для того, чтобы найти «явный вид» этой зависимости.

Цель регрессионного анализа – найти функциональную зависимость между переменными. Для этого предполагается, что зави-

симая переменная (иногда называемая *откликом*) определяется известной функцией (иногда говорят – *моделью*), зависящей от зависимой переменной или переменных (иногда называемых *факторами*) и некоторого параметра. Требуется найти такие значения этого параметра, чтобы полученная зависимость (модель) наилучшим образом описывала имеющиеся экспериментальные данные. Например, в *простой*⁶¹ *линейной регрессии* предполагается, что зависимая переменная y является линейной функцией $y = ax + b$ от независимой переменной x . Требуется найти значения параметров a и b , при которых прямая $ax + b$ будет наилучшим образом описывать (аппроксимировать) экспериментальные точки (x_1, y_1) , (x_2, y_2) , ..., (x_n, y_n) .

Фраза «описывать наилучшим образом» требует пояснений⁶². В качестве характеристики «качества» аппроксимации в регрессионном анализе используется *коэффициент детерминации* R^2 (читается – «эр квадрат»), который измеряется в процентах (соответственно, может принимать значения в интервале от 0 до 100 процентов) и характеризует ту долю изменений зависимой переменной, которая объясняется моделью (выбранной зависимостью между переменными при условии, что коэффициенты подобраны «наилучшим» образом). Понятно, что чем выше значение коэффициента детерминации⁶³ (чем ближе оно к 100%), тем лучше модель.

Рассмотрим первый пример (см. раздел 5.1). Для данных, приведенных в Табл. 3, найдем регрессию оценок x степени регенерации суставного хряща (СРСХ) членов контрольной группы после окончания эксперимента (зависимая переменная), по оценкам x_0 СРСХ членов контрольной группы до начала эксперимента. Для

⁶¹ *Простой* называется регрессия одной переменной по другой.

⁶² Для поиска параметров обычно используют метод наименьших квадратов – ищут такие значения параметров, чтобы сумма по всем экспериментальным точкам квадратов расстояний (по вертикали) от них до построенной зависимости была минимальной.

⁶³ Коэффициент детерминации, то есть квадрат коэффициента корреляции, имеет ясный смысл только тогда, когда обе переменные – случайные. Точнее, когда результат измерения – двумерный (или многомерный) вектор, причем имеющий двумерное (многомерное) распределение. Если одна из переменных, например – время, то это не так.

этого в программе StatGraphics⁶⁴ определим наилучшую линейную модель («Relate\Simple Regression»):

$$x = 2,35475 + 0,789306 x_0.$$

Коэффициент детерминации при этом равен 58,2%, то есть в рамках линейной модели только чуть более половины изменений переменной x может быть объяснено изменением переменной x_0 .

Можно использовать *полиномиальную регрессию*, в которой предполагается, что зависимая переменная является полиномом некоторой степени от независимой переменной (напомним, что линейная зависимость является полиномом первой степени). Например, полиномом второй степени будет зависимость вида $y = a x^2 + b x + c$ и задачей регрессии будет нахождение коэффициентов a , b и c .

Выбирая в программе StatGraphics («Relate\Polynomial regression») шестую степень полинома, получаем, что этот полином описывает зависимость y от y_0 с коэффициентом детерминации 33%. Увеличивая степень полинома, при помощи полиномиальной регрессии, можно увеличить «объясняющие» свойства модели (в нашем примере коэффициент детерминации вырос с 20,6% для линейной модели до 33% для полинома шестой степени⁶⁵). Но, необходимо помнить, что, во-первых, если число оцениваемых коэффициентов сравнивается с объемом данных, то коэффициент детерминации достигнет единицы⁶⁶. Во-вторых, желательно, чтобы найденная зависимость между переменными имела содержательную интерпретацию (найти такую интерпретацию для полинома шестой степени, наверное, будет затруднительно).

Выше мы рассмотрели простую регрессию (по одной независимой переменной) – линейную и нелинейную (в том числе – полиномиальную). Рассмотрим второй пример (см. раздел 5.2), иллюстрирующий *множественную регрессию* – определение зависимости одной переменной от нескольких факторов (независимых переменных). Предположим, что нужно построить модель

⁶⁴ Простая линейная регрессия может быть построена и в программе Excel для Windows («Сервис\Анализ данных\Регрессия»).

⁶⁵ Отметим, что низкие значения коэффициентов детерминации – 20,6% и 33% – показывают, что связь между переменными y и y_0 практически отсутствует.

⁶⁶ Более подробно с проблемой подбора модели заинтересованный читатель может ознакомиться в [15, 15, 32].

изменения ВСММ (вещств средней молярной массы) после окончания эксперимента в зависимости от других переменных, измеряемых до начала эксперимента (см. исходные данные в Табл. 7).

Для этого сначала определим, какие комбинации переменных M0, M1, M2, M3, ВСММ0 и ОП0 наилучшим образом объясняют изменения ВСММ. Оказывается («Special\Advanced Regression\Regression Model Selection»), что это – три переменных: M0, M1 и ВСММ0 – они объясняют 71,7% изменений ВСММ.

Далее, строим регрессионную модель («Relate\Multiple regression»), в которой зависимой переменной является ВСММ а независимыми переменными (факторами) – M0, M1 и ВСММ0:

$$\text{ВСММ} = -111,17 + 0,17 \text{ M0} + 2,42 \text{ M1} + 1,81 \text{ ВСММ0}.$$

Отметим, что более простая (однофакторная) регрессионная модель $\text{ВСММ} = 26,79 + 1,59 \text{ ВСММ0}$ не намного хуже – она объясняет 68,1% изменений ВСММ. Найденные зависимости позволяют «прогнозировать» значения зависимых переменных. То есть, регрессионный анализ, помимо того, что он позволяет количественно описывать зависимость между переменными, дает возможность прогнозировать значения зависимых переменных – подставляя в найденную формулу значения независимых переменных, можно получать прогноз значений зависимых.

При этом следует помнить, что построенная модель «локальна», то есть, получена для некоторых вполне конкретных значений переменных. Экстраполяция результатов модели на более широкие области значений переменных может привести к ошибочным выводам. Например, если моделировать эндотоксикоз путем повреждения ткани поджелудочной железы, то к токсическому компоненту (в формировании пула ВСММ) присоединится банальный протеолитический компонент от ферментов разрушающейся железы. Соответствующие крайне высокие значения ВСММ прогнозироваться данной моделью на основе регрессионного анализа, естественно, не будут. При хроническом токсическом процессе необратимые изменения в печени и почках по-иному будут воздействовать на формирование пула ВСММ, занижая значения показателя, предсказанные формулой регрессии.

Необходимо подчеркнуть, что при описании методов исследования зависимостей выше (а также ниже – при описании методов снижения размерности, классификации и прогноза) нами даны

лишь вводные соображения. Заинтересованному читателю можно порекомендовать обратиться к более добротной и профессиональной литературе по этой тематике [2, 11-13, 15, 28, 32, 40, 44, 47].

6.4. СНИЖЕНИЕ РАЗМЕРНОСТИ

Часто в результате экспериментальных исследований возникают большие массивы информации. Например, каждый из исследуемых объектов описывается по нескольким критериям (измеряются значения нескольких переменных – признаков). Тогда результатом измерений будет таблица с числом ячеек, равным произведению числа объектов на число признаков. Возникает вопрос, а все ли переменные являются информативными, например, отражают изменения, произошедшие в результате изучаемого воздействия. Конечно, исследователю желательно было бы выявить эти существенные переменные (это важно с содержательной точки зрения) и сконцентрировать внимание на них. Кроме того, всегда желательно сокращать объемы обрабатываемой информации (не теряя при этом сути). Чем тут могут помочь статистические методы?

Существует целый класс задач статистического анализа – методы *снижения размерности* – цель которых как раз и заключается в уменьшении числа анализируемых переменных либо посредством выделения существенных переменных, либо/и построения новых показателей (на основании полученных в результате эксперимента). Но, за все (в том числе за агрегирование информации) надо платить – такой платой в задачах снижения размерности является та часть вариации (изменений, дисперсии) исходных показателей, которая объясняется изменениями тех показателей, которые не «остаются» в результате снижения размерности (наименее изменчивые показатели или их комбинации).

Для снижения размерности используется *факторный анализ* [28, 46], а основными методами являются кратко рассматриваемый ниже метод главных компонент и многомерное шкалирование [42, 46].

Метод главных компонент заключается в получении нескольких новых показателей – главных компонент, являющихся линей-

ными комбинациями исходных показателей (напомним, что линейной комбинацией называется взвешенная сумма), полученных в результате эксперимента. Главные компоненты упорядочиваются в порядке убывания той дисперсии, которую они «объясняют». Первая главная компонента объясняет большую часть дисперсии, чем вторая, вторая – большую, чем третья и т.д. Понятно, что чем больше главных компонент будет учитываться, тем большую часть изменений можно будет объяснить.

Преимущество метода главных компонент заключается в том, что зачастую первые несколько главных компонент (одна-две-три) объясняют большую часть (например, 80-90%) изменений большого числа (десятков, а иногда и сотен) параметров. Кроме того, может оказаться, что в первые несколько главных компонент входят не все исходные параметры. Тогда можно сделать вывод о том, какие параметры являются существенными, и на них следует обратить внимание в первую очередь.

Приведем пример. Для этого возьмем числовые данные из Табл. 7 (см. раздел 5.2). Применяя метод главных компонент («Special\Multivariate methods\Principal components») к шести переменным, описывающим каждый объект до начала эксперимента (M0, M1, M2, M3, ВСММ0, ОП0), получаем, что более 80% дисперсии исходных данных можно объяснить всего тремя линейными комбинациями этих шести переменных – размерность можно снизить с 6 до 3.

То, что в методе главных компонент используются линейные комбинации факторов, иногда затрудняет содержательные интерпретации главных компонент. Например, если главной компонентой (объясняющей 95% дисперсии всех переменных) оказалось среднее арифметическое массы тела, температуры и артериального давления (или, как в рассмотренном выше примере – линейная комбинация масс и концентраций), то оперировать с данным показателем (основывая на нем какие-либо дальнейшие рассуждения или выводы), не имеющим физического смысла, надо очень осторожно. Тем не менее, если первоначально, помимо трех перечисленных, измерялись еще два десятка показателей, то на последние можно почти не обращать внимания – их значения будут почти всецело определяться значениями массы, температуры и давления.

6.5. КЛАССИФИКАЦИЯ И ПРОГНОЗ

Решив задачи описания данных, установления сходства/отличий, проанализировав качественно и количественно зависимости между переменными и выявив существенные переменные, можно анализировать соотношение групп переменных и пытаться прогнозировать значения одних переменных в зависимости от значений других переменных или времени развития того или иного процесса.

Классификация. Обширную группу задач анализа данных, основывающихся на применении статистических методов, составляют так называемые задачи классификации. В близких смыслах (в зависимости от предметной области) используются также термины: «группировка», «систематизация», «таксономия», «диагностика», «прогноз», «принятие решений», «распознавание образов». Обсудим некоторые различия между этими терминами. В [13] предложено выделить три подобласти теории классификации: дискриминация (дискриминантный анализ), кластеризация (кластерный анализ) и группировка.

В *дискриминантном анализе*⁶⁷ классы предполагаются заданными (например, обучающими выборками, для элементов которых известно, каким классам они принадлежат: больной-здоровый, легкая степень заболевания – средняя – тяжелая и т.д.). Задача заключается в том, чтобы вновь появляющийся объект отнести к одному из этих классов. У термина «дискриминация» имеется множество синонимов: *диагностика* (требуется поставить диагноз из конечного списка возможных диагнозов, если известны определенные характеристики пациента и известно, какие диагнозы ставились пациентам, вошедшим в обучающую выборку), распознавание образов с учителем, автоматическая (или статистическая) классификация с учителем и т.д.

Если в дискриминантном анализе классы заданы, то кластеризация и группировка предназначены для выявления и выделения классов. Синонимами являются: построение классификации, таксономия, распознавание образов без учителя, автоматическая классификация без учителя и т.д.

⁶⁷ В пакете StatGraphics функции дискриминантного анализа соответствуют «Special\Multivariate Methods\Discriminant Analysis».

Задача *кластерного анализа*⁶⁸ заключается в выделении по эмпирическим данным резко различающихся групп (кластеров) объектов, которые схожи между собой внутри каждой из групп.

При *группировке*, когда резких границ между кластерами не существует, исследователю приходится самому вводить границы между группами объектов.

Задачам классификации (как теоретическим их аспектам, так и опыту успешного решения конкретных прикладных задач, в том числе – в области медицины и биологии [16, 19]) посвящена многочисленная литература [1, 32, 40]. Рассматривать их подробно в настоящей работе мы не будем.

К задачам прогноза обычно относят и задачи анализа временных рядов.

Анализ временных рядов. *Временным рядом* называется последовательность чисел – значений некоторого показателя, измененного в различные моменты времени. Временные ряды используются для описания динамики процессов, например, изменения температуры тела, концентрации определенного вещества в крови, и т.д. На основании конечного отрезка временного ряда исследователь должен сделать выводы о свойствах рассматриваемого процесса и тех механизмах (в рамках статистики – вероятностных механизмах), которые порождают этот ряд.

При изучении временных рядов ставятся следующие цели [44]: агрегированное описание характерных особенностей ряда; подбор статистических моделей, описывающих временной ряд; предсказание будущих значений на основании прошлых наблюдений (*прогноз динамики*); выработка рекомендаций по управлению процессом, порождающим временной ряд.

На сегодняшний день существует множество моделей и методов, позволяющих достигать перечисленных выше целей с учетом специфики исследуемого процесса. Эти методы подробно описаны в литературе [1, 32, 40, 44] и реализованы в компьютерных статистических пакетах⁶⁹. Рассматривать их подробно мы не будем.

⁶⁸ В пакете *StatGraphics* функции анализа временных рядов соответствует «*Special/Time Series Analysis*».

⁶⁹ В пакете *StatGraphics* функции анализа временных рядов соответствует «*Special/Time-Series Analysis*».

7. ЗАКЛЮЧЕНИЕ

Настоящая работа посвящена использованию статистических методов при обработке результатов медико-биологического эксперимента. Цель этих методов – получить описание изучаемой системы. Однако, помимо описательного – *дескриптивного* (феноменологического) – аспекта, существуют еще два аспекта – *прогностический* и *нормативный*.

Прогностическое свойство любой теории заключается в том, что она в состоянии предсказать поведение изучаемой системы в будущем. Нормативный подход заключается в поиске таких воздействий на систему, которые привели бы ее в желаемое состояние.

Статистические методы отчасти дают возможность построить модель исследуемой системы и в результате решения задач регрессионного анализа, классификационного анализа и т.д. в той или иной степени предсказывать реакцию системы на различные воздействия. Но этого мало для нормативного подхода, так как последний требует наличия математических (количественных аналитических или имитационных) моделей, отражающих все существенные свойства исследуемой системы.

На сегодняшний день степень проникновения математики в биологию и медицину, к сожалению, достаточно мала – см. [3, 4, 7, 14, 16, 26, 33, 38, 48]. Объясняется это, в первую очередь колоссальной сложностью биологических систем по сравнению с физическими объектами и техническими системами, на успешном описании которых математика заработала сегодняшнюю свою высокую репутацию. Поэтому разработка и исследование математических моделей биологических систем представляется перспективным направлением дальнейших исследований, требующих совместных усилий биологов, медиков и математиков.

Итак, в настоящей работе мы попытались изложить на доступном уровне «рецепты» применения статистических методов при решении типовых задач анализа данных экспериментальных медицинских и биологических исследованиях. В то же время, не следует забывать, что рассмотрены лишь несколько, хотя и наиболее распространенных, но все-таки достаточно простых ситуаций. Арсенал же современных статистических методов гораздо богаче.

Быть может, освоение и применение этого арсенала подтолкнет исследователей в области медицины и биологии как к расширению соответствующих предметных областей, так и к повышению уровня обоснованности научных результатов.

ЛИТЕРАТУРА

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. М.: ЮНИТИ, 2001. – 270 с.
2. Анализ нечисловой информации в социологических исследованиях. М.: Наука, 1985. – 220 с.
3. Антомонов Ю.Г. Моделирование биологических систем. Киев: Наукова Думка, 1977. – 260 с.
4. Беллман Р. Математические проблемы в биологии М.: Мир, 1966. – 280 с.
5. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. М.: Наука, 1983. – 416 с.
6. Власов В.В. Введение в доказательную медицину. М.: Медиа Сфера, 2001. – 242 с.
7. Герасимов А. Математические модели в биологии, экологии и медицине: Учебное пособие. М.: МИФИ, 1998. – 40 с.
8. Гланц С. Медико-биологическая статистика. М: Практика, 1999. – 334 с.
9. Гнеденко Б.В., Хинчин А.Я. Элементарное введение в теорию вероятностей. М.: Едиториал УРСС, 2003. – 208 с.
10. Гринхальх Т. Основы доказательной медицины. М.: ГЭОТАР-МЕД, 2004. – 240с.
11. Ивченко Г.И., Медведев Ю.И. Математическая статистика. М.: Высшая школа, 1984. – 248 с.
12. Кендалл М. Ранговые корреляции. М.: Статистика, 1975. – 212 с.
13. Кендалл М., Стьюарт А. Многомерный статистический анализ и временные ряды. М.: Наука, 1976. – 736 с.
14. Коган А.Б., Наумов Н.П., Режабек В.Г., Чораян О.Г. Биологическая кибернетика. М.: Высшая школа, 1972. – 384 с.
15. Колмогоров А.Н. Теория вероятностей и математическая статистика. М.: Наука, 1986. – 535 с.

16. Котов Ю.Б. Новые математические подходы к задачам медицинской диагностики. М.: Едиториал УРСС, 2004. – 328 с.
17. Крамер Г. Математические методы статистики. М.: Мир, 1975. – 648 с.
18. Кудлаев Э.М., Орлов А.И. Вероятностно-статистические методы исследования в работах А.Н.Колмогорова // Заводская лаборатория. 2003. Т. 69. № 5. С. 55 – 61.
19. Ларичев О.И. Выявление экспертных знаний. М.: Наука, 1989. – 128 с.
20. Литвак Б.Г. Экспертная информация: методы получения и анализа. М.: Радио и связь, 1982. – 184 с.
21. Малета Ю.С., Тарасов В.В. Математические методы статистического анализа в биологии и медицине. М.: МГУ, 1982. – 179 с.
22. Новиков А.М. Докторская диссертация? М.: Эгвес, 2003. – 120 с.
23. Новиков А.М. Как работать над диссертацией. М.: Эгвес, 2003. – 104 с.
24. Новиков Д.А. Статистические методы в педагогических исследованиях (типовые случаи). М.: МЗ-Пресс, 2004. – 67 с.
25. Новиков Д.А. Теория управления организационными системами. М.: ИПУ РАН, 2005. – 472 с.
26. Новосельцев В.Н. Теория управления и биосистемы. М.: Наука, 1978. – 319 с.
27. Ногин В.Д. Принятие решений в многокритериальной среде: количественный подход. М.: Физматлит, 2002. – 176 с.
28. Окунь Я. Факторный анализ. М.: Статистика, 1974. – 200 с.
29. Орлов А.И. Надо ли разбивать выборку? // Заводская лаборатория. 1997. Т. 63. № 1. С. 54 – 54.
30. Орлов А.И. О применении статистических методов в медико-биологических исследованиях // Вестник Академии медицинских наук СССР. 1987. № 2. С. 88 – 94.
31. Орлов А.И. Устойчивость в социально-экономических моделях. М.: Наука, 1979. – 296 с.
32. Орлов А.И. Эконометрика. М.: Экзамен, 2004 (3-е изд.). – 576 с.
33. Парин В.В., Баевский Р.М. Введение в медицинскую кибернетику. М.: Медицина, 1966. – 298 с.

34. Петри А., Сэбин К. Наглядная статистика в медицине. М.: ГЭОТАР-МЕД, 2003. – 144 с.
35. Платонов А.Е. Статистический анализ в медицине и биологии: задачи, терминология, логика, компьютерные методы. М: Изд-во РАМН, 2001. – 52 с.
36. Подиновский В.В., Ногин В.Д. Парето-оптимальные решения многокритериальных задач. М.: Наука, 1982. – 386 с.
37. Пфанцгль И. Теория измерений. М.: Мир, 1976. – 248 с.
38. Ризниченко Г. Ю. Лекции по математическим моделям в биологии М.: Регулярная и хаотическая динамика, 2002. – 232 с.
39. Смирнов Н.В., Дунин-Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969.
40. Справочник по прикладной статистике. М.: Финансы и статистика. Том 1, 1989. – 510 с., Том 2, 1990. – 526 с.
41. Суппес П., Зинес Д. Основы теории измерений / Психологические измерения. М.: Мир, 1967. С. 9 – 110.
42. Терехина А.Ю. Анализ данных методами многомерного шкалирования. М.: Наука, 1986. – 220 с.
43. Трахтенгерц Э.А. Компьютерная поддержка принятия решений. М.: Синтег, 1998. – 376 с.
44. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. М.: ИНФРА-М, 1998. – 528 с.
45. Тюрин Ю.Н., Литвак Б.Г., Орлов А.И., Сатаров Г.А., Шмерлинг Д.С. Анализ нечисловой информации. М.: Научный совет АН СССР по комплексной проблеме «Кибернетика», 1981. – 80 с.
46. Харман Г. Современный факторный анализ. М.: Статистика, 1972. – 477 с.
47. Холлендер М., Вулф Д.А. Непараметрические методы статистики. М.: Финансы и статистика, 1983. – 518 с.
48. Цетлин М.Л. Исследования по теории автоматов и моделированию биологических систем. М.: Наука, 1969. – 316 с.