РОССИЙСКАЯ АКАДЕМИЯ МЕДИЦИНСКИХ НАУК
СИБИРСКОЕ ОТДЕЛЕНИЕ
НАУЧНЫЙ ЦЕНТР ПРОБЛЕМ ЗДОРОВЬЯ СЕМЬИ И РЕПРОДУКЦИИ ЧЕЛОВЕКА
НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ИНСТИТУТ ЭПИДЕМИОЛОГИИ И МИКРОБИОЛОГИИ

ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ДОПОЛНИТЕЛЬНОГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ИРКУТСКАЯ ГОСУДАРСТВЕННАЯ МЕДИЦИНСКАЯ АКАДЕМИЯ ПОСЛЕДИПЛОМНОГО ОБРАЗОВАНИЯ»
МИНИСТЕРСТВА ЗДРАВООХРАНЕНИЯ И СОЦИАЛЬНОГО РАЗВИТИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Е.Д. Савилов • В.А. Астафьев
С.Н. Жданова • Е.А. Заруднев

ЭПИДЕМИОЛОГИЧЕСКИЙ АНАЛИЗ

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МАТЕРИАЛА

НОВОСИБИРСК
«НАУКА-ЦЕНТР»
2011

УДК 616.9–036.2:313.13
ББК 511(2)
С13

Савилов Е.Д. Эпидемиологический анализ: Методы статистической обработки материала / Е.Д. Савилов, В.А. Астафьев, С.Н. Жданова, Е.А. Заруднев. – Новосибирск: Наука-Центр, 2011. – 156 с.

ISBN 978-5-9554-0024-2.

В монографии рассмотрены возможность использования современных общепринятых методов медико-биологической статистики в эпидемиологии, а также способы и методические подходы, предложенные авторами применительно к анализу заболеваемости. Представленные материалы апробированы при проведении занятий с курсантами на кафедре эпидемиологии и микробиологии Иркутского института усовершенствования врачей и студентами Иркутского государственного медицинского университета.

Книга предназначена для врачей и научных работников, занимающихся вопросами инфекционной и неинфекционной эпидемиологии, а также студентов высших учебных заведений медицинского профиля.

Табл. 43. Ил. 9. Библиогр.: 39 назв.

Рецензент

доктор медицинских наук, профессор,
академик РАМН *В.И. Злобин*

Утверждено к печати Ученым советом
Института эпидемиологии и микробиологии
Научного центра проблем здоровья семьи и репродукции человека
СО РАМН

© Е.Д. Савилов, В.А. Астафьев,
С.Н. Жданова, Е.А. Заруднев, 2011
© Оформление. Издательство
Наука-Центр, 2011

ISBN 978-5-9554-0024-2

ПРЕДИСЛОВИЕ

Изучение и анализ эпидемиологической ситуации на различных территориях с целью выявления и устранения причин возникновения инфекционных заболеваний представляют важную научную и прикладную задачу. На основании знаний закономерностей эпидемического процесса и факторов, способствующих его проявлению, можно сделать заключение о дальнейшем ходе развития этого процесса во времени и пространстве. Важнейшим инструментом для реализации подобного заключения являются методы статистического анализа.

Необходимость использования этих методов в эпидемиологии обусловлена тем, что показатели проявления эпидемического процесса даже на одной и той же территории значительно варьируют во внутригодовом аспекте и испытывают межгодовые колебания. Статистический анализ позволяет объективно оценивать информацию об изменениях проявлений эпидемического процесса, о влиянии на него отдельных условий, в том числе антропогенного загрязнения, и позволяет убедительно доказать значимую роль конкретных факторов для данной территории.

К сожалению, эпидемиологи в своей практической работе далеко не всегда дают полноценную статистическую оценку полученных первичных материалов. Возможно, причиной этому является слабое знание биометрических методов или, что бывает чаще, исполнители теряются в выборе оперативных математических приемов. При этом следует также иметь в виду, что каждый из статистических методов имеет определенные границы применимости, а использование его вне этих границ может приводить к ошибочным выводам, которые особенно опасны своей видимостью математической точности. Это также затрудняет работу практического врача-эпидемиолога.

Таким образом, современный этап развития эпидемиологии предполагает знание различных математических приемов и границы их применения в эпидемиологической диагностике. Именно этим вопросам и

были посвящены первое (Новосибирск, «Наука», 1993 г.) и второе [31] издания настоящей монографии. Буквально с первых дней выхода эти книги оказались широко востребованными не только врачами-эпидемиологами (что, в общем-то и понятно), но и в не меньшей степени гигиенистами, педиатрами, гинекологами и многими другими специалистами, связанными по роду деятельности с вопросами неинфекционной эпидемиологии, что было для нас не только неожиданным, но и приятным. Одним из привлекательных для читателей моментов стало также то, что первое издание монографии было включено в учебную программу по эпидемиологии для студентов медико-профилактического и лечебного факультетов медицинских вузов, а также в соответствующие программы для постдипломной подготовки специалистов и для сдачи кандидатского минимума по этой специальности. Все это послужило для нас основанием доработать и расширить статистический материал, изложенный в указанных книгах, и представить на суд читателей практически новое издание.

В эту книгу вложена часть души всех ее авторов. Если отступить от лирических высказываний, то отмечу, что начиная с 1994 г. методам статистического анализа заболеваемости постоянно уделялось большое внимание на различных циклах усовершенствования на кафедре эпидемиологии и микробиологии Иркутского института усовершенствования врачей и при проведении занятий на кафедре эпидемиологии Иркутского медицинского университета. Это позволило нам скорректировать отдельные определения и примеры, расширить в некоторых случаях область применения статистических приемов, а также предложить новые подходы в такой многогранной области, как эпидемиологический анализ.

В последние годы в специализированных магазинах появилось большое количество книг, посвященных применению статистических методов в медицинских и биологических исследованиях, в том числе применительно к эпидемиологическому анализу. Тем не менее полагаю, что вынесенная на суд читателей книга имеет свою нишу.

В существующих на сегодняшний день изданиях представленный материал во многом излагается достаточно сложным языком для его адекватного восприятия практическим врачом-эпидемиологом. Нередко такие работы направлены в первую очередь на освоения методов обработки и оценки первичных материалов с использованием разнообразных прикладных статистических программ. И это очень хорошо! Однако ря-

довой читатель таких книг, проводя соответствующие вычисления, в итоге может остаться без понимания хотя бы азов и логики использованного статистического аппарата. Вот вам простейший (в самом сжатом виде) пример такого непонимания. Читая лекции в Институте усовершенствования врачей, я в соответствующем месте обычно спрашиваю курсантов: «О чем говорит (свидетельствует) коэффициент корреляции, равный 0,71?» При этом сразу предупреждаю, что вопрос провокационный, а в ряде случаев даются даже подсказки. Тем не менее за все годы преподавания лишь в единичных случаях был получен адекватный ответ, который сводится к тому, что в данном контексте этот коэффициент корреляции ни о чем не говорит, если мы не знаем достоверен ли он. Этот пример – лишь частный случай из целого ряда попыток приблизиться к пониманию того, насколько практические врачи осмысленно занимаются статистической обработкой первичных материалов.

Итак, мое отстраненное отступление сводится к тому, что одна из основных задач предлагаемой вашему вниманию книги – научить врача в достаточно понятных ему терминах логике статистического анализа. А это возможно, прежде всего, при «ручной» обработке данных. Только когда специалист освоит логику материала, только тогда (и не днем раньше!) он может воспользоваться различными статистическими программами и с помощью компьютера получать итоговые результаты. Обучение такого рода действиям – уже другая задача, и должно быть иное издание.

В этом я и вижу основу настоящей книги. Именно этим и объясняются многие отдельные частные примеры статистической обработки приведенных здесь данных. Я не призываю Вас делать все статистические вычисления в своей работе так, как изложено в настоящем пособии. Все частности направлены на одно – на понимание логики применяемого статистического аппарата.

Прекрасно понимаю, что предисловие в книге обычно мало кто читает, тем не менее не мог не упомянуть об особенностях предлагаемого вашему вниманию издания в связи с тем, что эти вопросы достаточно часто приходится обсуждать со своими коллегами (читателями, слушателями, аспирантами, соискателями).

И еще одно. Несмотря на то что большинство примеров в представленной книге дано для отдельных видов инфекционной патологии, нет принципиальных возражений для использования предлагаемых подходов

не только для анализа других инфекционных заболеваний, но и для описания различных состояний здоровья на популяционном уровне. Поскольку основа настоящей книги сложилась достаточно давно, многие первичные данные по отдельным инфекционным заболеваниям приведены для того времени. Мы сознательно не стали подыскивать более «свежие» примеры. Во-первых, чтобы не терять напрасно время (см. предыдущий абзац, а кто не понял, то это всего лишь шутка, а она не подвергается статистической обработке). Во-вторых, какая разница какие цифры предложены. Гораздо важнее последовательность действий и осознание анализируемого материала. И последнее. Любая система (биологическая, социальная и др., к которым, в частности, принадлежит и эпидемический процесс) обязательно находится в колебательном режиме. А когда многолетние циклические проявления, присущие указанной системе, вновь выведут ее на эпидемически значимый уровень, не дано предсказать никому. Наша с вами задача – постараться не допустить столь выраженного подъема инфекционной заболеваемости, чему, в частности, и будут способствовать материалы этой книги при внимательном ее изучении.

С учетом всего вышеизложенного, а также имеющегося значительного интереса к методам статистического анализа со стороны различных специалистов, в настоящее время назрела необходимость переиздания этой книги, тем более что просьбы такого рода постоянно поступают ее авторам.

Завершить свое предисловие я хочу словами благодарности своим коллегам и соавторам настоящей монографии, с которыми мы прошли долгий и, надеюсь, плодотворный путь. Искренне желаю, чтобы данная книга представила интерес для будущих ее читателей при решении вопросов в области как инфекционной, так и неинфекционной эпидемиологии.

Успехов вам и удачи!

Доктор медицинских наук, профессор,
заслуженный деятель науки РФ Е. Савилов

Глава 1

ОСНОВНЫЕ КРИТЕРИИ ЭПИДЕМИОЛОГИЧЕСКОГО АНАЛИЗА

1.1. ЭПИДЕМИОЛОГИЧЕСКИЕ ПОКАЗАТЕЛИ

Оценка эпидситуации и эффективности профилактических и противоэпидемических мероприятий проводится с использованием ряда показателей, которые являются общими для многих инфекций, а по методу получения относятся к статистическим. Под термином «эпидемиологические показатели» следует понимать качественную или количественную характеристику эпидемических явлений. Эпидемиологические показатели рассчитываются на определенную численность населения (на 1000, 10 000, 100 000 и т. д.), поэтому они являются относительными величинами, а именно **интенсивными показателями.**

Важнейшим критерием эпидемического процесса считается **заболеваемость** инфекционными болезнями за определенный период, которая определяется по формуле

$$\frac{\text{число заболеваний}}{\text{численность населения}} \cdot 100\,000.$$

При анализе заболеваемости дополнительно используют такие показатели, как инфицированность (число инфицированных лиц на 100, 1000 и т.д. обследованных), пораженность (число выявленных больных на 100 или 1000 обследованных), болезненность (число больных с активными и неактивными формами болезни на 10 000 населения).

В ходе анализа часто необходимо выяснить структуру заболеваемости, долю различного по форме или по степени тяжести течения болезни, оценить результаты лабораторного исследования и т.д. Этим задачам отвечает **экстенсивный показатель** – удельный вес, который характеризует распределение целого на составные части и выражается в процентах. При этом за 100 % принимается общее количество. Например, при изучении качества воды из 95 проб 60 оказалось без кишечной палочки, 30 – с

допустимым ее содержанием, 5 – с высоким содержанием кишечной палочки, в процентах это составляет 63,2; 31,6 и 5,2 соответственно.

Из экстенсивных показателей наиболее широко применяются следующие:

Показатель	Определение
Вспышечная заболеваемость	Число заболеваний во время вспышек / Общее число заболеваний · 100
Лабораторное обследование больных	Число больных, обследованных лабораторно / Число заболевших данной инфекцией · 100
Лабораторное подтверждение диагноза	Число диагнозов, подтвержденных лабораторно / Число поставленных диагнозов данной инфекции · 100
Подтверждение окончательного диагноза	Число подтвержденных первичных диагнозов / Число поставленных диагнозов · 100
Расхождение первоначального и окончательного диагнозов	Число первоначальных диагнозов – число окончательных диагнозов / Число поставленных диагнозов · 100
Охват мероприятиями в очаге (эпидемиологическое обследование, наблюдение в очаге, лабораторное обследование контактных лиц)	Число эпидемиологических обследований в очаге (число лабораторных обследований общавшихся и др.) / Общее число очагов данной инфекции · 100
Заключительная дезинфекция	Количество проведенной заключительной дезинфекции / Число очагов, подлежащих заключительной дезинфекции · 100
Камерная дезинфекция	Число очагов проведенной камерной дезинфекции / Число очагов, подлежащих такой обработке · 100
Степень выявления больных	Число активно выявленных больных / Число заболевших · 100

Показатель	Определение
Своевременность мероприятий в отношении больного (поступление экстренного извещения, госпитализация, лабораторные обследования)	$\frac{\text{Своевременное проведение в отношении больного мероприятия (поступление экстренного извещения, госпитализация, лабораторное обследование)}}{\text{Общее число зарегистрированных больных с данной инфекцией}} \cdot 100$

Для более полного представления об эпидситуации исследуют такой параметр, как индекс очаговости. Его вычисляют отношением всех случаев, зарегистрированных в очагах данной болезни, к числу очагов [37]. Например, если в 9 очагах зарегистрировано 27 случаев заболевания, то индекс очаговости равен 3,0 (27:9). В оптимальном варианте в каждом очаге должно быть по одному больному, в соответствии с этим индекс очаговости равен 1.

Показатель госпитализации выражают в процентах и определяют следующим образом:

$$\frac{\text{число госпитализированных больных}}{\text{число заболевших}} \cdot 100.$$

Под индексом эпидемиологической эффективности понимается практическая ценность профилактических мероприятий (вакцинирования, фагирования, гаммаглобулинопрофилактики). Иногда пользуются коэффициентом эпидемиологической эффективности. Рассчитывают эти критерии по формулам

$$\text{индекс эффективности} = \frac{a}{b},$$

$$\text{коэффициент эпидемиологической эффективности (\%)} = \frac{b - a}{b} \cdot 100,$$

где a – заболеваемость в испытуемой группе; b – в контрольной.

Индекс эффективности показывает во сколько раз заболеваемость в испытуемой группе ниже заболеваемости в контрольной группе.

Коэффициент эффективности отражает на сколько процентов заболеваемость в испытуемой группе ниже заболеваемости в контрольной группе.

Все рассмотренные эпидемиологические показатели обрабатываются с использованием методов статистического анализа. В связи с этим остановимся на определении таких понятий, как статистическая совокупность, варианта, динамический ряд и т.д.

1.2. СТАТИСТИЧЕСКАЯ ОЦЕНКА ЭПИДЕМИОЛОГИЧЕСКИХ ПОКАЗАТЕЛЕЙ

Развитие живого организма определяется очень многими и разнообразными факторами внутреннего и внешнего порядка, поэтому изучение качественного или количественного признака у особей одной и той же популяции всегда дает целый ряд значений, обычно не совпадающих между собой. Каждое значение в одинаковой степени может считаться истинным значением признака изучаемой популяции. Следовательно, анализируемую популяцию будут характеризовать все полученные значения. Вместе с тем очень неудобно иметь дело с таким множеством характеристик, желательно «уплотнить информацию», т.е. свести все многообразие показателей к немногим параметрам, которые достаточно емко и в компактной форме описывают свойства совокупности. Следует иметь в виду, что невозможно изучить все особи данной популяции, поэтому обычно исследуется лишь часть популяции, которую принято называть выборкой из генеральной совокупности, т.е. совокупности всех экземпляров или особей, которые вообще в принципе могут относиться к данной популяции [34].

Совокупность единиц, обладающих некоторыми общими свойствами, существенными для их характеристики (совокупность анализов, проб воды, показателей заболеваемости и др.), получила название **статистической совокупности.** Каждый член такой совокупности представляет отдельное значение изучаемого признака и именуется единицей совокупности. Число единиц в совокупности принято называть объемом совокупности.

Признак – свойство единиц (элементов) совокупности. Различают признаки: количественные, выраженные непосредственно числом (возраст больных, коли-индекс и т.д.), и качественные (атрибутивные), не поддающиеся непосредственному количественному выражению (пол больных, профессия и др.).

Варианта – отдельное конкретное значение варьирующего признака. Обозначается буквами латинского алфавита: x, y и т.д. с индексом в правом нижнем углу, указывающим номер, варианты:

$$x_i = 1, 2, 3 \ldots.$$

Вариация – колеблемость признака, изменчивость величины призна-

ка у единиц, входящих в состав совокупности. Вариация признака может быть выражена обобщающими величинами (средними и относительными). Различают дискретную вариацию, когда отдельные значения признака (варианты) отличаются друг от друга только на некоторое значение, обычно целое число (число членов семьи, число дней госпитализации), и непрерывную вариацию, когда значения признака могут отличаться сколь угодно мало (возраст заболевших, процент выполнения койко-дней).

Атрибутивная вариация – вариация качественного признака, т.е. признака, не имеющего количественного выражения (профессия заболевшего, нозологическая форма и др.).

Альтернативная вариация – вариация качественного признака, по которой признак принимает только два взаимоисключающих значения (болен – здоров, пробы стандартные – нестандартные).

Вариационный ряд – совокупность значений варьирующего признака и соответствующих им численных единиц совокупности. Вариационный ряд может быть прерывным (дискретным или интервальным) и непрерывным. В случае, если варианты расположены в возрастающем или убывающем порядке, вариационный ряд называется упорядоченным (ранжированным).

Динамический ряд (разновидность вариационного ряда) – совокупность значений признака, характеризующих изменения числовых значений данного признака, расположенных в хронологическом порядке [35]. Динамический ряд применяется обычно для ретроспективного анализа заболеваемости.

Интервал – значения варьирующего признака, лежащие в определенных границах. Например, госпитализация в первые 2–3 дня, 4–5 дней и т.д. Различают интервалы закрытые (замкнутые), с конкретными численными значениями обеих границ интервалов, и открытые, в которых нижняя или верхняя граница не фиксирована. Например, эффективность профилактических мероприятий составила до 500 млн р. или 501 млн р. и более. Обычно интервалы выбирают так, чтобы на каждый из них приходилось не более 15–20 % всех значений признака.

Величина интервала (интервальная разность) есть разность между верхней (x_n) и нижней (x_1) границами интервала.

Если вариационный ряд содержит открытые интервалы, границы интервала устанавливают по аналогии с соседними. Для расчета оптимальной величины интервала (z) можно использовать формулу Стерджесса [35]:

$$z = \frac{x_n - x_1}{1 + 3{,}322 \lg n} \cdot 100,$$

где n – число единиц в совокупности; x_n и x_1 – наибольшая и наименьшая варианты соответственно.

Интервальные вариационные ряды – ряды со сгруппированными вариантами. Различают интервальные вариационные ряды с равными и неравными интервалами.

Варьирующий признак – признак, значения которого (варианты) отличаются друг от друга. Обычно признак, принятый за основу группировки, т. е. группировочный, называют варьирующим.

Амплитуда – вариация значений признака в пределах от минимальной до максимальной варианты.

Доля – относительная численность (в процентах) единиц совокупности, обладающих данным значением признака, например доля детей в совокупности всех больных дизентерией.

Частота (p) – абсолютное число, показывающее сколько раз та или иная варианта встречается в совокупности. Например, если в совокупности госпитализированных 3 поступили в первый день, 10 – во второй, то число 3 будет частотой первой варианты, а 10 – второй. Иногда частоты называют статистическими весами.

Частость – относительная величина, определяющая долю частот отдельных вариант, выраженная в процентах.

Критерии – показатели, вычисляемые на основании фактических наблюдений, которые позволяют судить о приемлемости некоторых гипотез.

В области эпидемиологических исследований широко применяется проверка гипотезы о статистической значимости различий сравниваемых выборочных совокупностей. Для этого формулируется **нулевая гипотеза** – предположение о том, что разница между генеральными параметрами сравниваемых совокупностей равна нулю, и различия, наблюдаемые между выборочными показателями, носят случайный характер [16]. В дальнейшем с помощью соответствующих критериев проверяется справедливость принятия этой гипотезы или она отвергается.

Процедура проверки гипотезы связана с объемом выборки, соответствующим числом степеней свободы (подробнее в гл. 2) и уровнем

значимости. Уровень статистической значимости – вероятность получения ошибки, допускаемой при оценке принятой гипотезы, может различаться (5, 1, 0,1 %, или соответственно в долях единицы 0,05; 0,01; 0,001), но в медико-биологических приложениях, если специально не оговорено другое значение, он обычно принимается равным 5 %. Если результаты значимы на уровне 1–5 %, обычно говорят о наличии статистической значимости, на уровне меньше 1 % – о высокой статистической значимости.

С уровнем значимости обратно связан **доверительный уровень**, который выражается в процентах случаев или в долях единицы. Значение этого показателя, близкое к 100 % (или к 1), интерпретируется как сильный довод против нулевой гипотезы. В биологических исследованиях чаще всего применяется 95 %-й доверительный уровень. В более ответственных случаях, когда выводы должны быть особенно строгими, требуется достижение 99 %-го или более высокого доверительного уровня.

Для оценки справедливости нулевой гипотезы рассчитывается также ***P*-значение**, которое оценивает вероятность получения такого же значения критерия при многократном повторении исследования и отсутствии различий между сравниваемыми совокупностями. Если в результате проверки нулевая гипотеза была отвергнута на уровне значимости α, то для отражения наличия статистически значимых различий результат сравнения может быть записан $P < \alpha$. Это означает, что, например, при $\alpha = 5$ % (это соответствует $\alpha = 0{,}05$) ошибка сравнения возможна не более чем в $\alpha = 5$ % случаев, а значит маловероятна. Более корректной и информативной считается запись о степени значимости полученных значений критериев в виде $0{,}01 < P < 0{,}05$ или его точное значение, например $P = 0{,}026$, которое получают при работе с пакетами статистических программ на персональных компьютерах.

Для проверки гипотез возможны два вида критериев: **параметрические** (построенные на основании специальных параметров совокупности) и **непараметрические** (построенные непосредственно по вариантам данной совокупности и их частотам). Первые критерии служат для проверки гипотез о параметрах совокупностей, распределенных по нормальному закону, а вторые – для проверки гипотез, не зависимых от формы распределения совокупностей. Так, при нормальном распределении признака параметрические критерии обладают большей мощностью, чем непараметрические, поэтому, если известно, что сравнивае-

мые выборки извлечены из нормально распределенных совокупностей, предпочтение следует отдавать параметрическим критериям. В случае распределения, отличного от нормального, при малых объемах выборок, а также для анализа вариационных рядов с атрибутивным или альтернативным выражением признаков применение непараметрических критериев является единственно возможным. Проверить, извлечена ли рассматриваемая выборка из нормально распределенной совокупности, можно с помощью специальных тестов, рассмотренных в разд. 1.3.

Нормальное распределение является одним из основных в медико-биологической статистике, так как многие эмпирические изменения биологических признаков, характеризующиеся непрерывной вариацией, приближаются к нормальному, следуют ему [20]. Кроме нормального в вариационной статистике применяется биноминальное и пуассоново распределения. Биноминальное распределение относится к признакам, варьирующим дискретно, прерывисто, пуассоново также характеризуется дискретной и прерывистой изменчивостью, оно в принципе является тем же биноминальным, но относится к явлениям, обладающим очень малой вероятностью, поэтому это распределение асимметрично. Теоретическая основа вариации – результат совместного действия многих разнонаправленных и независимых друг от друга факторов. Согласно теореме Ляпунова, если случайная величина является суммой большого числа независимых слагаемых, то она с достаточной степенью точности будет распределяться по нормальному закону [21]. Вот почему закон нормального распределения широко применяется в биологии. Наряду с нормальным распределением Н.П. Ашмарин и А.А. Воробьев [6] указывают, что здесь нередко приходится встречаться с данными, варьирование которых подчиняется законам распределения, отличным от нормального (**от авторов** – не надо использовать термин «ненормальное распределение», который иногда встречается в литературе). Так, довольно часто встречаются в разной степени асимметричные распределения, характеризующиеся наличием более чем одного максимума: лог-нормальное, которое приобретает вид нормального, если при его графическом изображении откладывать не просто результат, а его логарифмы [6].

Нормальным распределением называется такое распределение, которое следует закону, выраженному определенной математической формулой, предложенной Гауссом и независимо от него Лапласом. Нормальное распределение еще называется распределением Гаусса (или гауссовым) и

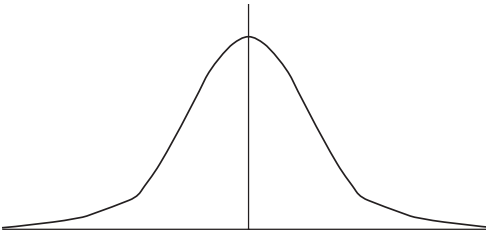

Рис. 1. Вид кривой нормального распределения.

широко используется в самых различных областях науки и практики, в том числе в биологии и медицине. Основными статистическими параметрами, характеризующими нормальное распределение, являются среднее значение (среднее арифметическое), дисперсия (σ^2) и ее производное – среднее квадратическое отклонение (σ), описание которых изложено в следующем разделе. При нормальном распределении крайние значения – наименьшее и наибольшее – встречаются редко, чем ближе значение признака к средней арифметической, тем оно чаще встречается; в центре распределения находятся такие значения, которые встречаются наиболее часто.

Таким образом, при нормальном распределении подавляющее большинство вариант изучаемого признака находится в пределах $M \pm 3\sigma$, причем 68,26 % из них приходится на середину вариационного ряда ($M \pm 1\sigma$). Графическое изображение нормального распределения имеет форму колокола (рис. 1).

1.3. СРЕДНИЕ ВЕЛИЧИНЫ, ИХ РАЗМЕРЫ И КОЛИЧЕСТВЕННЫЕ ОТНОШЕНИЯ ПРИЗНАКОВ

Средняя величина есть обобщающая количественная характеристика совокупности по одному варьирующему признаку, например продолжительность инкубационного периода, число заболевших в очагах и т.д. Эти величины имеют огромное познавательное значение, так как теснейшим образом связаны с существом рассматриваемых процессов и являются основным средством статистических расчетов [5]. Отличительная особенность средних величин в том, что в них взаимно погашаются индивидуальные отклонения признаков единиц совокупности, тем самым устраняются случайные колебания и проявляются значения признаков, объективно присущие тому или иному массовому явлению.

Из средних величин наиболее широкое распространение имеет средняя арифметическая. Следует иметь в виду, что если в расчетах не указан вид средней, то определялась средняя арифметическая.

Средняя арифметическая – одна из распространенных форм средней величины. Рассчитывается как частное от деления суммы индивидуальных значений (вариантов) признака на их число:

$$M = \frac{\Sigma x}{n} \text{ или } M_{взв} = \frac{\Sigma xp}{\Sigma p},$$

где M – средняя арифметическая; $M_{взв}$ – средняя арифметическая взвешенная; x – варианты признака; n – число вариантов; p – веса.

При вычислении средней арифметической необходимо проверить совокупность на присутствие так называемых «выскакивающих» вариант. Для этого существует несколько методов. Быстрый способ, позволяющий решить эту задачу с достаточной строгостью для эпидемиологии, приведен в руководстве Н.П. Ашмарина с соавторами [5]. Метод основан на оценке различий крупных вариант в упорядоченном ряду данной совокупности. Для этого пользуются предложенной формулой, а результат сравнивают с табличными значениями (см. приложение, табл. 1):

$$\frac{x_n - x_{n-1}}{x_n - x_1},$$

где в числителе – разность между подозреваемой «выскакивающей» вариантой (x_n) и вариантой, которая ей предшествует в упорядоченном ряду (x_{n-1}); в знаменателе – разность между наибольшей (x_n) и наименьшей (x_1) вариантами ряда.

Допустим, что мы наблюдали 13 случаев инфекционного заболевания, при которых известны даты заражения и начала заболевания. Интересующий эпидемиологов инкубационный период составлял 6; 5; 7; 5; 8; 9; 5; 8; 10; 14; 5; 7; 3 дня (пример 1).

В этом вариационном ряду после расположения вариант в порядке возрастания (3; 5; 5; 5; 5; 6; 7; 7; 8; 8; 9; 10; 14) подозревается на «выскакивающую» варианта 14. После вычисления отношения по указанной формуле получаем

$$\frac{14-10}{14-3} = \frac{4}{11} = 0{,}363.$$

Результат сравниваем с табличным для 13 вариант (см. приложение, табл. 1). Оказалось, что пограничное значение при 95%-м уровне достоверности составляет 0,338, а при 99%-м – 0,438. Следовательно, наибольшая варианта (14) не может считаться «выскакивающей» при 99%-м уровне достоверности и данная величина может быть использована для дальнейших расчетов. Следует заметить, что выбор критерия достоверности исследователь должен логически обосновать, исходя из поставленных целей и задач.

Можно таким же образом проверить и наименьшую варианту. Для этого следует применить формулу

$$\frac{x_2 - x_1}{x_n - x_1},$$

где в числителе – разность между значением, предшествующим минимальному (x_2), и наименьшим (x_1); в знаменателе – разность между наибольшим (x_n) и наименьшим (x_1) вариантами. Рассматривая также пример 1, находим $x_2 = 5$, $x_1 = 3$, $x_n = 14$ и рассчитаем отношение:

$$\frac{5-3}{14-3} = 0{,}182.$$

Полученное значение (0,182) меньше табличных критериев как при 95%-м (0,338), так и при 99%-м (0,438) уровнях достоверности. Следовательно, наименьшее значение (3) не является «выскакивающим». После проверки вариант «выскакивающие» значения можно приступить к получению средней арифметической. Это сумма значений варьирующего признака, разделенная на их число:

$$M = \frac{\Sigma x}{n} = \frac{x_1 + x_2 + \ldots x_n}{n}.$$

В рассматриваемом нами вариационном ряду (пример 1) средняя арифметическая равна

$$M = \frac{6+5+7+5+8+9+5+8+10+14+5+7+3}{13} = 7{,}1.$$

Среднюю арифметическую можно рассчитывать и с учетом статистических весов (частот):

$$M_{взв} = \frac{\Sigma xp}{\Sigma p} = \frac{x_1 p_1 + x_2 p_2 + \ldots x_n p_n}{p_1 + p_2 \ldots p_n},$$

в нашем примере

$$M_{взв} = \frac{(3 \cdot 1)+(5 \cdot 4)+(6 \cdot 1)+(7 \cdot 2)+(8 \cdot 2)+(9 \cdot 1)+(10 \cdot 1)+(14 \cdot 1)}{1+4+1+2+2+1+1+1} = 7{,}1.$$

Это так называемая «взвешенная средняя», т.е. суммирование произведения вариант на их частоту и деление полученного значения на сумму статистических весов.

Особенно удобно пользоваться средней арифметической взвешенной в тех случаях, когда эпидемиологу необходимо проанализировать большой вариационный ряд, в котором многие варианты повторяются часто.

Классическим же методом вычисления средней арифметической является метод произведений [21]. Для этого составляется таблица, в которой указываются варианты, частота, отклонение вариант от средней, выбираемой по моде или медиане, так называемая условная средняя, и произведения варианты на отклонения (табл. 1). Вычисления проводятся по формуле

$$M = A + \frac{\Sigma pa}{n},$$

где M – средняя арифметическая; A – условная средняя (варианта, находящаяся в середине ряда или наиболее часто встречающаяся); p – частота, a – отклонения вариант от условной средней ($x - A$), n – число наблюдений.

Рассмотрим вычисление средней арифметической с учетом статистических весов методом произведений. В качестве примера воспользуемся данными по количеству нестандартных проб питьевой воды децентрализованного водопровода (см. табл. 1).

Таблица 1

Расчет средней арифметической методом произведений (по данным нестандартных проб в воде децентрализованного водопровода)

Варианта (x)	Частота (p)	$\frac{a}{(x-A)}$	pa
25	7	– 5	– 35
26	9	– 4	– 36
27	15	– 3	– 45
28	17	– 2	– 34
29	18	–1	– 18
30	21	0	0
31	17	1	17
32	11	1	22
33	6	3	18
34	5	4	20
35	3	5	15
Σ	129		– 76

Примечание. A =30.

Варианты: 25; 26; 27; 28; 29; 30; 31; 32; 33; 34; 35.
Частоты: 7; 9; 15; 17; 18; 21; 17; 11; 6; 5; 3.

$$M = (25 \cdot 7 + 26 \cdot 9 + 27 \cdot 15 + 28 \cdot 17 + 29 \cdot 18 + 30 \cdot 21 + 31 \cdot 17 + 32 \cdot 11 + 33 \cdot 6 + 34 \cdot 5 + 35 \cdot 3) / (7 + 9 + 15 + 17 + 18 + 21 + 17 + 11 + 6 + 5 + 3) = 29{,}4.$$

В интервальном вариационном ряду при расчете средней арифметической методом произведений учитывается величина интервала, на которую делится каждое отклонение вариант от условной средней, т.е.

$$a = \frac{x - A}{i}, \text{ где } i \text{ – величина интервала.}$$

Тогда формула принимает вид $M = A + i \frac{\Sigma pa}{n}$.

$$M = 151 + 4\ (7/167) = 151{,}17.$$

Допустим, что дан вариационный интервальный ряд с вариантами от 135 до 167, с частотой их встречаемости от 3 до 52 и интервалом – 4 (табл. 2).

Таблица 2

Расчет средней арифметической интервального вариационного ряда методом произведений (данные заболеваемости дизентерией в городе Н.)

Варианта (x)	Частота (p)	$x - A$	$a = \frac{x - A}{i}$	pa
135	3	– 16	– 4	– 12
139	15	– 12	– 3	– 45
143	17	– 8	+ 2	– 34
147	11	– 4	+ 1	– 11
151	52	0	0	0
155	42	+ 4	1	42
159	18	+ 8	2	36
163	5	+12	3	15
167	4	+16	4	16
Σ	167			7

Найдем среднюю арифметическую с учетом статистических весов

$$M = \frac{135 \cdot 3 + 139 \cdot 15 + 143 \cdot 17 + 147 \cdot 11 + 151 \cdot 52 + 155 \cdot 42}{3 + 15 + 17 + 11 + 52 + 42 + 18 + 5 + 4} = 151{,}17.$$

Теперь найдем среднюю арифметическую методом произведений (см. табл. 2). Как видно, расчет средней арифметической методом произведений проще.

Применение метода произведений удобно тем, что по его схеме после вычисления средней арифметической, проведя некоторые дополнительные расчеты, можно найти среднее квадратическое отклонение и ошибку средней.

Средняя геометрическая применяется главным образом при изучении динамики, например при расчете роста или снижения в вариационных рядах. Средняя геометрическая вычисляется по формуле

$$M_{\text{геом}} = \sqrt[n]{x_1 x_2 x_3 \ldots x_n},$$

где x_1, x_2... – показатели 1-го, 2-го и последующих лет исследований.

Более подробно этот вопрос рассматривается в разделе «Статистический анализ многолетней динамики заболеваемости».

Из средних величин широкое распространение также имеют медиана и мода.

Медиана – числовая характеристика вариационного ряда, находящаяся посередине и делящая его пополам. Например, мы имеем вариационный ряд, состоящий из 13 значений, описывающих различную длительность инкубационного периода при дизентерии: 6; 5; 7; 5; 3; 9; 5; 8; 10; 14; 5; 7; 8. Располагаем варианты в порядке возрастания: 3; 5; 5; 5; 5, 6; 7; 7; 8; 8; 9; 10; 14. В середине ряда находится варианта, имеющая числовое значение 7, она и будет медианой. Если совокупное число единиц вариационного ряда четное, медиана равна средней арифметической двух соседних вариант, находящихся посередине. Рассмотрим предыдущий пример, где вариационный ряд состоит из 12 вариант: 3; 5; 5; 5; 5; 6; 7; 7; 8; 8; 9; 10. В этом случае медиана равна

$$\frac{6+5}{2} = 6{,}5.$$

Медиану, являющуюся описательной характеристикой вариационного ряда, иногда называют непараметрической средней (понятие о параметрических и непараметрических критериях см. в гл. 2). Применение медианы особенно целесообразно, когда мы не знаем о характере закона распределения.

Мода – представляет собой наиболее типичное значение случайной величины, т.е. это наиболее часто встречающаяся варианта в вариационном ряду. Мода также является непараметрической средней. В нашем условном вариационном ряду: 6; 5; 7; 5; 8; 9; 5; 8; 10; 14; 5; 7; 3 – наиболее часто встречается варианта 5 – она и будет модой.

Изучаемые эпидемиологией явления чаще носят характер нормального распределения. При этом из всех вариант подавляющее большинство составляют варианты среднего размера, и чем дальше они отклоняются от среднего значения признака, тем реже встречаются в данной совокупности. В симметричном вариационном ряду, соответствующем нормальному распределению, значения средней арифметической, моды и медианы совпадают.

В случаях, когда какие-нибудь причины благоприятствуют появлению значений признака (например, аварии водопроводной сети могут повысить численность заболевших инфекциями, передающи-

мися водным путем; правильно выполненные профилактические мероприятия уменьшают количество заболевших и т. д.), отличающихся от среднего значения в сторону уменьшения или увеличения, образуются асимметричные распределения. В таких рядах средняя арифметическая, мода и медиана не совпадают. Асимметрия может быть учтена при помощи коэффициента асимметрии (*As*), который вычисляется по формуле [21]

$$As = \frac{\sum D^3}{n\sigma^3},$$

где *As* – показатель асимметрии; $\sum D^3 = \sum (W–M)^3$ – сумма кубов отклонений вариант от средней арифметической; σ^3 – среднее квадратическое отклонение, возведенное в третью степень; *n* – общее количество наблюдений.

Этот показатель колеблется от – 3 до +3. Если $As > 0$, то симметрия правосторонняя, положительная (т.е. большинство вариант расположены справа от середины ряда). Если $As < 0$, то асимметрия левосторонняя, отрицательная (большинство вариант расположены слева от середины ряда). В том случае, когда $As = 0$, вариационный ряд принято считать симметричным [10]. Таким образом, коэффициент асимметрии позволяет проверить гипотезу о законе распределения генеральной совокупности.

Кроме того, необходимым для оценки формы распределения является также коэффициент эксцесса [20]

$$E = \frac{\sum D^4}{n\sigma^4} - 3,$$

где *E* – показатель эксцесса; $\sum D^4 = \sum (W - M)^4$ – сумма отклонений вариант от средней арифметической в четвертой степени; σ^4 – среднее квадратическое отклонение в четвертой степени; *n* – общее число наблюдений.

Для нормального распределения эксцесс равен 0, если значение этого коэффициента больше нуля, говорят о положительном эксцессе, если меньше нуля – отрицательном. При нормальном распределении показатели асимметрии и эксцесса должны быть близки к нулю. Однако на практике равенство нулю почти не встречается, так как

показатели оцениваются по выборочным совокупностям; они также являются случайными величинами, имеющими ошибки. Для проверки нормальности распределения рекомендуется использовать соответствующие таблицы, в которых указаны критические точки для этих коэффициентов при различных уровнях значимости и объемах выборки n [34]. Если рассчитанные значения для асимметрии и эксцесса превосходят эти критические точки, гипотеза о нормальности распределения отвергается.

Воспользуемся примером, приведенным в руководстве по математической статистике В.Ю. Урбаха [34]. При изучении распределения эритроцитов крови человека по размерам (в выборке из $n = 1000$) обнаружена некоторая асимметрия ($As = – 0,12$) и небольшой положительный эксцесс ($E = + 0,17$). Можно ли считать распределение эритроцитов в организме нормальным?

По таблице граничных значений асимметрии и эксцесса (см. приложение, табл. 14) находим критические значения для n =1000: $As_{0,05}(1000) = 0,15$ и $E_{0,05}(1000) = 0,30$.

Так как фактические значения As и E меньше критических значений, то нулевая гипотеза не отвергается и можно считать, что исследуемая выборка получена из нормально распределенной генеральной совокупности.

Еще одним методом оценки принадлежности выборки к совокупности с нормальным распределением является расчет критериев достоверности асимметрии и эксцесса. Для этого вычисляются ошибки коэффициентов асимметрии и эксцесса по формулам [21]

$$m_{As} = \sqrt{\frac{6}{n}}, \; m_E = \sqrt[2]{\frac{6}{n}},$$

где m_E, m_{As} – ошибки показателей асимметрии и эксцесса; n – общее количество наблюдений.

Затем рассчитываются значения критериев достоверности асимметрии (t_{As}) и эксцесса (t_E) [21]:

$$t_{As} = \frac{A}{m_{As}}, \; t_E = \frac{E}{m_E}.$$

Показатели асимметрии и эксцесса свидетельствуют о достовер-

ном отличии распределения испытуемой выборки от нормального в том случае, если они превышают свою ошибку в 3 и более раза, т.е. если t_{As} и t_E больше или равно 3, то исследуемое выборочное распределение вариант признака не принадлежит нормальному.

Вывод о нормальном распределении генеральной совокупности может быть получен и на основании использования «правила трех сигм», согласно которому считается, что нормально распределенная случайная величина X попадает в $M \pm 3\,\sigma$-интервал с вероятностью события 0,9973. Эта вероятность достаточно велика при $n < 20$, например $P_{10}(10) \approx 0{,}973$; $P_{19}(19) \approx 0{,}95$; $P_{20}(20) \approx 0{,}947$ [23]. Таким образом, если выборка объемом $n = 19$ принадлежит к совокупности с нормальным распределением, то 95 % ее вариант содержатся в интервале $M \pm 3\,\sigma$. Наличие в выборке объема $n < 20$ значений вне интервала имеет веский повод усомниться в нормальности распределения генеральной совокупности.

При изучении массовых явлений нельзя ограничиться только средними значениями их признаков, необходимо широко подвергать всестороннему анализу отклонения от средней, поскольку без этого нельзя увидеть весь процесс в его динамическом развитии. Для этого в статистике используются показатели вариации или колеблемости.

В эпидемиологических исследованиях широко применяется **среднее квадратическое отклонение** (σ), которое является вторым по значимости параметром характеристики вариационного ряда. По средней арифметической и среднему квадратическому отклонению можно установить с определенной статистической значимостью принадлежность той или иной варианты к рассматриваемому вариационному ряду. Этот показатель представляет собой меру колеблемости и определяется по формулам

$$\sigma_{\text{невзв}} = \sqrt{\frac{\sum (x - M)^2}{n}},$$

$$\sigma_{\text{взв}} = \sqrt{\frac{\sum (x - M)^2 \cdot p}{n}}.$$

Более правильно в случаях $n \leq 30$ в знаменателе подкоренного

выражения ставить не n, а $n - 1$. Это связано с тем, что при достаточно большом ($n > 30$) числе наблюдений уменьшение знаменателя на 1 не сказывается сколько-нибудь существенно на результате, так как после их вычислений изменяется значение второго или даже третьего десятичного знака.

В нашем примере 1 инкубационных периодов дизентерии

$$\sigma_{\text{невзв}} = \sqrt{[(6-7{,}1)^2+(5-7{,}1)^2+(7-7{,}1)+(5-7{,}1)^2+(3-7{,}1)^2+(9-7{,}1)^2+(5-7{,}1)^2+}$$

$$\overline{+(8-7{,}1)^2+(10-7{,}1)^2+(14-7{,}1)^2+(5-7{,}1)^2+(7-7{,}1)^2+(8-7{,}1)^2]/(13-1)} = 2{,}8,$$

$$\sigma_{\text{взв}} = \sqrt{[(6-7{,}1)^2 \cdot 1 + (5-7{,}1)^2 \cdot 4 + (7-7{,}1) \cdot 2 + (9-7{,}1)^2 \cdot 1 + (8-7{,}1)^2 \cdot 2 +}$$

$$\overline{+(10-7{,}1)^2 \cdot 1 + (14-7{,}1)^2 \cdot 1]/(1+4+2+1+1+2+1+1)} = 2{,}8.$$

Если средняя арифметическая вычисляется методом произведений, то удобнее всего среднее квадратическое отклонение находить этим же методом по формуле

$$\sigma = \sqrt{\frac{\sum pa^2}{n} - \left(\frac{\sum pa}{n}\right)^2}$$ (для несгруппированных рядов),

$$\sigma = i\sqrt{\frac{\sum pa^2}{n} - \left(\frac{\sum pa}{n}\right)^2}$$ (для сгруппированных рядов),

где p – частота вариант; a – отклонения вариант от условной средней; n – число наблюдений; i – величина интервала.Пример расчета среднего квадратического отклонения представлен в табл. 3.

В большинстве случаев $\left(\frac{\sum pa}{n}\right)^2$ очень мала и не может существенно влиять на среднее квадратическое отклонение, чаще всего пользуются более простой формулой:

для сгруппированных рядов – $\sigma = \sqrt{\frac{pa^2}{n}}$,

для несгруппированных рядов – $\sigma = i\sqrt{\frac{pa^2}{n}}$, откуда $\sigma = \sqrt[4]{\frac{485}{167}} = 6{,}82$.

Таблица 3

Расчет среднего квадратического отклонения

x	*p*	*a*	*pa*	pa^2
135	3	– 4	– 12	48
139	15	– 3	– 45	135
143	17	– 2	– 34	68
147	11	– 1	+11	11
151	52	0	0	0
155	42	+1	+42	42
159	18	+2	+36	72
163	5	+3	+15	45
167	4	+4	+16	64
$\sum$	167		7	485

Среднее квадратическое отклонение можно вычислить и экспресс-методом, который рекомендуется применять как при большом, так и при малом числе наблюдений. Формула расчета:

$$\sigma = \frac{(x_n - x_1)}{A}.$$

При этом кроме амплитуды $(x_n - x_1)$ используется коэффициент A, который зависит от числа наблюдений (см. приложение, табл. 3). Для числа наблюдений 167 коэффициент A равен 5,43.

$$\sigma = \frac{(x_n - x_1)}{A} = \frac{(167 - 135)}{5{,}43} = 5{,}89.$$

Значения, полученные классическим (6,82) и экспрессным (5,89) методами, достаточно близки, но техника расчета экспресс-методом проще и быстрее.

Дисперсия (σ^2) – средний квадрат отклонений вариант (x) от средней арифметической (M). Она является мерой вариации, т. е. колеблемости признака. Этот показатель – основной при проведении дисперсионного факторного анализа и определяется по формулам

$$\sigma_{\text{невзв}} = \frac{\sum (x - M)^3}{n},$$

$$\sigma_{взв} = \frac{\sum (x - M)^2 \cdot p}{n}.$$

Для того же примера 1 дисперсия будет равняться

$$\sigma^2_{невзв} = \sqrt{[(6 - 7{,}1)^2 + (5 - 7{,}1)^2 + (7 - 7{,}1) + (5 - 7{,}1)^2 + (3 - 7{,}1)^2 + (9 - 7{,}1)^2 + (5 - 7{,}1)^2 + (8 - 7{,}1)^2 + (10 - 7{,}1)^2 + (14 - 7{,}1)^2 + (5 - 7{,}1)^2 + (7 - 7{,}1)^2 + (8 - 7{,}1)^2] / 13} = 8{,}1.$$

Глава 2

СРАВНЕНИЕ ДВУХ ВЫБОРОК

При эпидемиологическом анализе, как правило, большое значение имеет достоверность различий или сходства между статистическими характеристиками, полученными при исследовании сравниваемых выборок. Особенно часто приходится анализировать показатели заболеваемости, связанные с различными факторами риска, данные по эффективности вакцинации, фагирования в основной и контрольной группах и т.д.

Расхождение между выборками может быть отражением какого-либо реального различия или же является результатом случайности, сопровождающей попадание вариант в выборку. Гипотезу об отсутствии реального различия называют нулевой. Критерии, проверяющие нулевую гипотезу о значениях каких-либо параметров распределения, называют параметрическими. Критерии, не требующие вычисления параметров распределения (средних значений, дисперсий и т.д.), называют непараметрическими.

2.1. ПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ДВУХ ВЫБОРОК

Параметрические методы применяются только при сравнении выборок, относящихся к нормальному распределению (см. гл. 1). Как указывалось выше, сравнение двух выборок из нормально распределенной совокупности изучаемых признаков основано на количественной оценке различий основных параметров: средних значений и их дисперсий. Для вычисления статистически значимых различий между средними значениями нужен еще один показатель – стандартная ошибка (m). Стандартная ошибка – это мера вариабельности средней арифметической, которая показывает, какие можно было бы ожидать

случайные вариации этого показателя для групп, взятых из одной и той же совокупности. Этот показатель вычисляется по формуле

$$m = \pm \frac{\sigma}{\sqrt{n-1}},$$

где m – стандартная ошибка; σ – стандартное отклонение; n – число наблюдений. В случае, если $n \geq 30$, указанная формула имеет вид

$$m = \frac{\sigma}{\sqrt{n}}.$$

Рассчитаем ошибку средней, используя пример, приведенный в табл. 3:

$$m = \pm \frac{6{,}82}{\sqrt{167}} = \pm 0{,}53.$$

Для вычисления стандартной ошибки существует экспресс-метод [6], при котором этот показатель определяется по разнице между наибольшей и наименьшей вариантами (т. е. по амплитуде) и по коэффициенту (см. приложение, табл. 4):

$$m = \pm \frac{(x_{\max} - x_{\min})}{B},$$

где $x_{\max}$ – наибольшая варианта; $x_{\min}$ – наименьшая; B – коэффициент.

Найдем ошибку средней арифметической, используя пример 1 из гл. 1, в котором рассматривается различная длительность инкубационного периода при дизентерии. Максимальная длительность (x_n) – 14 дней, минимальная (x_1) – 3 дня. Табличный коэффициент для 13 наблюдений равен 12,00:

$$m = \pm \frac{x_n - x_1}{B} = \frac{14-3}{12{,}0} = 0{,}92.$$

Статистическая значимость различий двух средних арифметических оценивается по критерию Стьюдента (t).

Формула вычисления критерия Стьюдента имеет вид

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}},$$

где t – критерий Стьюдента; M – средняя арифметическая; m – стандартная ошибка. Значения t могут колебаться от 0 до ∞.

Рассмотрим вычисление t на примере заболеваемости гепатитом А (табл. 4).

Находим среднюю арифметическую обычным способом и стандартную ошибку для каждой группы наблюдений экспресс-методом:

$$M_1 = \frac{x_1 + x_2 + \ldots . x_n}{n} = \frac{3712}{11} = 337,$$

$$m_1 = \frac{x_{\max} - x_{\min}}{B} = \frac{877 - 120}{10{,}05} = 75{,}3,$$

$$M_2 = \frac{2043}{11} = 185{,}7,$$

$$m_2 = \frac{325 - 107}{10{,}05} = 21{,}7.$$

На основании полученных данных находим критерий Стьюдента:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}} = \frac{337{,}4 - 185{,}7}{\sqrt{75{,}3^2 + 21{,}7^2}} = 1{,}9.$$

Значения критерия для различных уровней значимости приводят-

Таблица 4

Число детей, заболевших гепатитом А в различных районах области

Район	Число заболевших		Район	Число заболевших	
	2001 г.	2002 г.		2001 г.	2002 г.
1	365	325	7	205	127
2	234	157	8	363	107
3	877	216	9	205	230
4	197	153	10	618	155
5	120	151	11	165	233
6	363	189	Σ	3712	2043

ся в приложении (табл. 5). При числе наблюдений 11 (число степеней свободы K равно 11 + 11 – 2 = 20, т.е. $K = (n_1 + n_2) – 2$ t по таблице равен 2,09 при уровне достоверности 95 % . Значение коэффициента t (Стьюдента) зависит не только от уровня достоверности, но и от объема выборки. Если оценивается разность средних из выборочных совокупностей с одинаковым объемом n, то число степеней свободы (K) равно $2(n – 1)$, а если объемы сравниваемых выборок различны, то число степеней свободы рассчитывается как $n_1 + n_2 – 2$ [16].

В нашем примере t равен 1,9. Таким образом, различие средних не имеет статистически значимых различий, т.е. не доказано. Следовательно, по приведенным данным мы не можем сказать, что в 2001 г. степень риска заражения гепатитом у детей была больше или меньше, чем в 2002 г. Для доказательства возможных различий требуется большее число наблюдений.

Существует односторонний вариант критерия t, наиболее часто излагаемый в статистических справочниках, когда большая средняя ставится на первое место в разности и значения колеблются от 0 до ∞, и двусторонний вариант критерия t, когда значения могут принимать от $– \infty$ до 0 до $+\infty$.

Следует также помнить, что вышеприведенная формула расчета критерия Стьюдента применима для сравнения выборок по средним значениям только в случаях, когда характеристики исследуемого признака не являются связанными, что не так часто встречается в эпидемиологических исследованиях.

В приведенном выше примере сравнивали две выборки, характеризующие проявления эпидемического процесса. О действии какого-либо выделенного (изучаемого) фактора (благоприятного или неблагоприятного) судили по выявленным различиям между двумя группами показателей. Если разброс в группах велик, сила действия изучаемого фактора «тонет» среди действия других случайных факторов, и мы не можем реально выявить существующие различия. Для оценки такой ситуации возможно применение другого подхода сравнения. Вместо двух групп показателей создается одна, и сравнению подлежит состояние каждого показателя при воздействии изучаемого фактора и без него. Методически такой подход достаточно трудно выполним, ведь нужно быть уверенным, что изменение параметров не обусловлено естественным течением эпидемического процесса. Тем не менее учет изменения

состояния каждого показателя выборки в отдельности, нивелируя влияние разброса данных, значительно повышает чувствительность статистических критериев.

Выявить изменение, располагая парами наблюдений, позволяет парный критерий Стьюдента и критерий Мак-Нимара. Изменим условия примера с заболеваемостью вирусным гепатитом А, чтобы оценить эффект воздействия изучаемого фактора (мероприятия). Ранее мы выбирали две группы показателей. Затем вычисляли среднее по каждой выборке и определяли статистическую значимость различия этих средних. Теперь мы набираем одну группу, оцениваем разницу каждой пары показателей для оценки эффективности мероприятия, осуществленному с целью снижения заболеваемости в 2002 г. Затем находим среднее изменение и проверяем статистическую значимость его отличия от нуля (табл. 5).

Такой подход более точно улавливает различия, вызванные действием фактора.

Парный критерий Стьюдента рассчитывается как отношение среднего изменений к его ошибке. Без подробного рассмотрения расчетов средней,

Таблица 5

Пример расчета парного критерия Стьюдента по данным числа заболевших гепатитом А в различных районах области

Район	Число заболевших		Изменение (d)
	2001 г.	2002 г.	
1	365	325	40
2	234	157	77
3	877	216	461
4	197	153	44
5	120	151	– 31
6	363	189	174
7	205	127	78
8	363	107	156
9	205	230	– 45
10	618	155	453
11	165	233	– 68
Среднее ($\overline{d}$)	337,5	165,7	151,7
Среднеквадратичное отклонение			227,1
Ошибка средней (m_d)			71,8

среднеквадратичного отклонения и ошибки средней, проведем вычисления значения парного критерия Стьюдента для нашего примера:

$$t = \frac{\bar{d}}{m_d} = \frac{151{,}7}{71{,}8} = 2{,}11.$$

При числе степеней свободы для уровня значимости α = 0,05 $v = n - 1 = 11 - 1 = 10$ критическое значение равно 2,23, что свидетельствует об отсутствии статистически достоверных различий показателей в сравниваемые годы, несмотря на проведенное мероприятие по снижению заболеваемости.

Следует учитывать, что если непарный критерий Стьюдента требует нормального распределения самих данных выборки, то парный требует нормально распределенных его изменений. Кроме того, он применим, если только зависимый признак является числовым. Для количественного выражения признака, т.е. когда значениями параметра являются не числа, а качественные характеристики (название нозологий, присутствие или отсутствие изменений), применяют критерий Мак-Нимара. Его вычисление сходно с принципом определения χ^2.

Критерий Стьюдента предназначен для сравнения двух групп. Однако на практике он широко (и неправильно) используется для оценки различия большего числа групп посредством их попарного сравнения. При этом вступает в силу эффект множественных сравнений, что приводит к увеличению вероятности принятия неправильной гипотезы. Если исследуемых групп больше двух, то следует воспользоваться дисперсионным анализом и (или) специальными методами множественного сравнения. Дисперсионный анализ нужен для проверки гипотезы о равенстве всех сравниваемых средних, а затем, если найдено, что они различаются, выделяют среди них отличные от остальных, используя методы множественного сравнения. Простейший из них – введение поправки Бонферрони. Если k раз применить критерий с уровнем значимости α, то в каждом из сравнений мы должны принять уровень значимости α/k. Например, при трехкратном сравнении уровень значимости должен быть $0{,}05/3 = 0{,}017$.

Поправка Бонферрони работает, когда число сравнений не превышает 8. В ином случае уровень значимости принимает очень малые значения, и даже весьма большие различия становятся статистически незначимыми с учетом применения этой поправки. Для большого числа групп

следует применять критерий Ньюмена – Кейлса. При этом методе средние упорядочиваются по возрастанию и сравниваются попарно, каждый раз вычисляется значение критерия согласно двум правилам:

1. В упорядоченном ряду (от 1 до m) сначала сравниваются наибольшее с меньшим, т.е. m-е с 1-м, затем m-е со 2-м, с 3-м, и так далее, до m – 1-го. Предпоследнее (m – 1-е) тем же порядком сравниваем с 1-м, затем со 2-м и т.д. до m – 2-го. Перебираем таким образом все пары. Например, в случае 5 групп выполняется следующий порядок сравнений: 5 – 1, 5 – 2, 5 – 3, 5 – 4, 4 – 1, 4 – 2, 4 – 3, 3 – 1, 3 – 2, 2 – 1.

2. Если какие-либо средние не различаются, то все средние, лежащие между ними, тоже не различаются. Например, если не выявлено различий между 4-м и 1-м средними, то сравнения 4 – 2, 4 – 3, 3 – 1, 3 – 2, 2 – 1 проводить не нужно.

Расчет значения критерия проводится по формуле

$$q = \frac{M_1 - M_2}{\frac{\delta^2_{\text{внутр}}}{2}(\frac{1}{n_1} + \frac{1}{n_2})},$$

где q – критерий Ньюмена – Кейлса, M_1 и M_2 – средние выборок, n_1 и n_2 – объемы выборок, $\delta^2_{\text{внутр}}$ – внутригрупповая дисперсия, которая вычисляется по формуле

$$\delta^2_{\text{внутр}} = \frac{1}{m}(\delta_1^2 + \delta_2^2 + ...\delta_m^2),$$

где $\delta_1^2, \delta_2^2, \delta_m^2$ – выборочные дисперсии всех сравниваемых групп, m – число групп.

Для определения критического значения критерия Ньюмена – Кейлса (см. приложение, табл. 16, 17) необходимо рассчитать число степеней свободы: $\nu = N - m$, где N – сумма численностей всех групп, m – число групп и величины интервала сравнения l, равного разнице мест сравниваемых пар в упорядоченном ряду. Если сравниваются пары, находящиеся на j-м и i-м местах, то интервал сравнения $l = j - i + 1$. Например при сравнении 8-го и 1-го членов упорядоченного ряда $l = 8 - 1 + 1 = 8$.

Иногда задача заключается в том, чтобы сравнить несколько групп с контрольной. Вышеперечисленные критерии будут слишком строги

для таких расчетов, поэтому следует использовать специальный метод множественного сравнения с контрольной группой с помощью критерия Даннета:

$$q' = \frac{M_{\text{контр}} - M_m}{\sqrt{\delta^2_{\text{внутр}}\left(\frac{1}{n_{\text{контр}}} + \frac{1}{n_m}\right)}},$$

где q' – критерий Даннета, $M_{\text{контр}}$ – средняя контрольной выборки, M_m – средняя выборки, с которой проводится сравнение, $n_{\text{контр}}$ и n_m – объемы выборок, $\delta^2_{\text{внутр}}$ – внутригрупповая дисперсия, которая вычисляется по формуле

$$\delta^2_{\text{внутр}} = \frac{1}{m}(\delta_1^2 + \delta_2^2 + \ldots \delta_m^2),$$

где $\delta_1^2, \delta_2^2, \delta_m^2$ – выборочные дисперсии всех сравниваемых групп, m – число групп.

Число сравнений равно числу групп, не считая контрольной, и существенно меньше, чем при применении критерия Ньюмена – Кейлса. Средние упорядочиваются по абсолютной величине отличия от контрольной группы. Затем контрольную группу сравнивают с остальными, начиная с наиболее отличной от контрольной. Если различия с очередной группой не найдены, вычисления прекращают. Параметр l постоянен и равен числу групп, включая контрольную. Число степеней свободы вычисляют как в критерии: $v = N - m$, где N – сумма численности всех групп, m – число групп. Критические значения критерия Ньюмена – Кейлса приведены в приложении (табл. 18 и 19) для двух уровней значимости.

В тех случаях, когда сравниваемые выборки отличаются не по средним значениям, а по степени варьирования (σ^2) , определить статистически значимые различия можно при помощи критерия Фишера (F). Он вычисляется по формуле

$$F = \frac{\sigma_1^2}{\sigma_2^2},$$

где σ_1^2 и σ_2^2 – выборочные дисперсии.

В нашем примере (см. табл. 4) находим σ_1 и σ_2 экспресс-методом:

$$\sigma_1^2 = \frac{x_{max} - x_{min}}{A} = \frac{877 - 120}{3{,}17} = 238{,}8,$$

$$\sigma_2^2 = \frac{x_{max} - x_{min}}{A} = \frac{325 - 107}{3{,}17} = 68{,}8.$$

Большие значения σ всегда ставят в числитель.

Затем находим $F = 238{,}8^2 : 68{,}82^2 = 12{,}1$, значения которого могут колебаться от 0 до ∞. Статистически значимые различия между квадратами отклонений определяются по критическим значениям (см. приложение, табл. 6). При F больше табличного значения различие между дисперсиями признается достоверным, при меньшем или равном табличному – случайным. По таблице критических значений (F) уровню значимости 0,05 (при $K_1 = n - 1 = 10$ и $K_2 = n - 1 = 10$) соответствует значение 2,97. Значит в рассмотренном примере мы можем говорить о статистически значимых различиях по размаху сравниваемых выборок.

Приведенные вычисления относятся к случаям количественной изменчивости. Когда различия между вариантами выражаются в каких-то качествах, такую изменчивость называют качественной, или альтернативной. Для оценки различий между показателями при альтернативном распределении изменчивости используют более простую формулу вычисления стандартной ошибки

$$m = \pm\sqrt{\frac{pq}{n}},$$

где m – стандартная ошибка; p – частота с наличием признака; q – частота с отсутствием признака ($100 - p$ для процентов, $1000 - p$ и т.д. – для других интенсивных показателей).

Например, нам нужно сравнить показатели заболеваемости дизентерией в двух населенных пунктах с разным уровнем коммунального благоустройства (т.е. с учетом фактора риска). В первом населенном пункте (А) проживает 225 056 чел., из них заболело 1904 чел.; во втором пункте (В) из 332 123 жителей заболело 1633 чел. (пример 5). Вычисляем заболеваемость на 100 000 населения и стандартные ошибки.

$$p_A = \frac{1904 \cdot 100\,000}{225\,056} = 846{,}0,$$

$$m_A = \pm\sqrt{\frac{846{,}0\,(100\,000 - 846{,}0)}{225\,056}} = 19{,}3,$$

$$p_B = \frac{1633 \cdot 100\,000}{332\,123} = 491{,}7,$$

$$m_B = \pm\sqrt{\frac{491{,}7\,(100\,000 - 491{,}7)}{332\,123}} = 12{,}1,$$

$$t = \frac{491{,}7 \cdot 846{,}0}{\sqrt{19{,}3^2 + 12{,}1^2}} = 15{,}5.$$

По таблице критических значений t (см. приложение, табл. 5) находим, что t при уровне значимости 0,05 должен быть 1,96, а в нашем случае он равен 15,5. Значит, в данном примере показатели заболеваемости в изучаемых населенных пунктах высокодостоверно различны ($P < 0{,}01$) и, следовательно, в населенном пункте А с низким уровнем коммунального благоустройства (фактор риска) вероятность заболеть дизентерией статистически значимо выше, чем в населенном пункте В с более высоким уровнем коммунального благоустройства.

Аналогичным образом определяется статистическая значимость различий для экстенсивных величин (относительных долей). Например, нам нужно оценить статистически значимые различия по следующим данным: удельный вес нестандартных проб воды в разводящей сети водопровода в июле был 24,6 % (M_1 из 56 наблюдений), а в январе – 11,7 % (M_2 из 40 наблюдений). Вычисляем ошибки и критерий Стьюдента (пример 6):

$$m_1 = \pm\sqrt{\frac{24{,}6\,(100 - 24{,}6)}{56}} = 5{,}76,$$

$$m_2 = \pm\sqrt{\frac{11{,}7\,(100 - 11{,}7)}{40}} = 5{,}08,$$

$$t = \frac{24{,}6 - 11{,}7}{\sqrt{5{,}76^2 + 5{,}08^2}} = \frac{12{,}9}{7{,}68} = 1{,}68.$$

Полученное значение t = 1,68 оцениваем по таблице критических значений (см. приложение, табл. 5) при $K = (n_1 + n_2) - 2 = 94$. В таблице нет значений t для K = 94, воспользуемся близким к нему значением K = 100, при этом t окажется равным 1,98 (уровень значимости 0,05). Как видно, в нашем случае мы не можем говорить о статистически значимом различии удельных весов нестандартных проб воды в разводящей сети водопровода, а можем отметить лишь тенденцию.

Для быстрой оценки значимости различий между сравниваемыми показателями в относительных величинах можно воспользоваться критерием Боярского (см. приложение, табл. 7). В таблице указана минимальная разность между относительными величинами, достижение которой означает, что превышение одного показателя над другим неслучайно.

Допустим, требуется установить статистически значимые различия между средними M_1 и M_2. Известно, что M_1 = 27 % (n = 70) и M_2 = 12 % (n = 70). Пользуясь таблицей Боярского, находим в первой графе по вертикали значение 10, близкое нашей средней – 12 %, а по горизонтали число наблюдений, равное 70, на пересечении получаем 12, это минимальная разность между средними, при которой превышение одной средней над другой считается статистически значимым. Следовательно, в рассматриваемом случае различия достоверны.

При разном числе наблюдений в сопоставимых группах пользуются формулой расчета усредненного числа наблюдений [22]

$$\frac{2 \cdot n_1 \cdot n_2}{n_1 + n_2},$$

где n_1 – число наблюдений в первой выборке; n_2 – число наблюдений во второй выборке.

Например, в двух детских дошкольных учреждениях в течение года зарегистрировано по 25 случаев острой кишечной инфекции. При этом в одном из них число детей 130 (n_1), в другом 250 (n_2), удельный вес переболевших 19,2 (M_1) и 10 % (M_2) соответственно.

Необходимо установить, случайны или нет различия по удельному весу переболевших в этих детских дошкольных учреждениях.

По вышеприведенной формуле находим усредненное число детей в обоих учреждениях:

$$\frac{2 \cdot 130 \cdot 250}{130 + 250} = 171.$$

Далее определяем, что минимальная разность, достижение которой свидетельствует о неслучайности различий, должна составлять 6,8 % (см. приложение, табл. 7). В нашем примере это различие составляет 9,2 %, следовательно, имеющиеся различия по доле переболевших неслучайны.

2.2. НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ СРАВНЕНИЯ ДВУХ ВЫБОРОК

Выше были рассмотрены оценки сравнения двух совокупностей с помощью критериев Стьюдента (t) и Фишера (F), которые связаны с нормальной формой распределения событий. Применение указанных критериев основано на расхождении или сходстве между двумя выборками по таким параметрам, как средние значения или дисперсии соответственно. В связи с этим они называются параметрическими критериями.

В медико-биологических исследованиях для оценки значимости различий между двумя эмпирическими совокупностями, когда сравниваются не сами варианты, а их порядковые номера (ранги) в ранжированном ряду, следует использовать критерии, не требующие вычисления параметров нормального распределения (среднее и дисперсия), поэтому указанные критерии получили название непараметрические, или порядковые (ранговые).

Однако при эпидемиологических исследованиях методы непараметрической статистики все еще используются недостаточно, за исключением критерия соответствия и коэффициента корреляции рангов Спирмена. В настоящем разделе рассматривается возможность использования ряда непараметрических критериев применительно к основным задачам эпидемиологического анализа.

В исследованиях, где оценивается близость двух распределений,

т.е. там, где явления представлены в виде динамических рядов, целесообразно применять критерий соответствия, обозначаемый χ^2. Он представляет собой сумму квадратов отклонений эмпирических частот от частот теоретических, отнесенную к теоретическим частотам динамического ряда, т.е.

$$\chi^2 = \sum \frac{(p - p')^2}{p'},$$

где p – эмпирическая частота; p' – соответствующая теоретическая, т.е. вычисленная или ожидаемая, частота.

Рассмотрим использование критерия соответствия на примере, заимствованном из книги Б.С. Бессмертного, М.Н. Ткачевой [11]. На материалах помесячной динамики заболеваемости (p) скарлатиной (в условных единицах) (табл. 6) в качестве типового (теоретического) ряда принимаются средние данные о заболеваемости в каждом месяце той же области за 10 лет (p'). В этом случае критерий соответствия равен

$$\chi^2 = 0{,}42 + 0{,}50 + 0{,}17 + 0{,}22 \ldots\ldots = 23{,}95.$$

Если бы $p' = p$, т.е. «ожидаемые» числа соответствовали фактическим, то χ^2 был бы равен нулю. Такое положение свидетельствовало бы об отсутствии различий в сравниваемых рядах распределений и подтверждала бы справедливость нулевой гипотезы ($P > 0{,}05$). Чем больше разность ($p - p'$), тем больше и размеры χ^2. Чтобы ус-

Таблица 6

Критерий соответствия χ^2 между фактической (p) и теоретической (p') заболеваемостью скарлатиной по месяцам

Показатель	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
p	34	28	22	16	14	18	19	35	58	54	52	28
p'	38	32	24	18	20	24	40	52	34	40	40	50
$p - p'$	– 4	– 4	– 2	– 2	– 6	– 6	– 9	– 5	6	– 10	12	– 22
$(p - p')^2$	16	16	4	4	36	36	81	25	36	100	144	484
$\frac{(p - p')^2}{p'}$	0,42	0,50	0,17	0,22	1,80	1,50	2,03	0,58	1,06	2,50	3,60	9,68

тановить существенность или случайность разности ($p - p'$), значения χ^2 оценивают по специальной таблице, в которой содержатся критические (стандартные) значения критерия y^2 для двух уровней значимости.

В нашем примере каждая пара $p - p'$ является фиксированной и $K = n - 1$=11. По табл. 8 приложения находим, что значение, равное 23,70 при 11 степенях свободы, превышает критический показатель уровня статистической значимости 95 %.

Следовательно, отличия сезонной динамики заболеваемости скарлатиной в году *N* от обычного типового характера ее были неслучайны и вызваны существенными обстоятельствами (в данном случае резким изменением условий жизни эвакуированных детей).

Эпидемиологи часто решают задачи с альтернативными выражениями (заболел или не заболел, привит или не привит, водопользование централизованное или местное и т.д.). В таких случаях статистическую значимость различий также можно оценивать с помощью χ^2.

В качестве примера рассмотрим данные А.М. Меркова, Л.Е, Полякова [21] о болевших и оставшихся здоровыми среди привитых в непривитых (табл. 7).

В качестве нулевой гипотезы предполагаем, что прививка не оказывает влияния на размеры заболеваемости привитых (нулевой эффект). В этом случае заболеваемость привитых и непривитых должна быть одинаковой и соответствовать заболеваемости всего населения, т.е. 328 · 100 : 184 83 = 1,77 %.

Среди привитых тогда будет 6815 · 1,77 : 100 = 121 заболевших и 6815 – 121 = 6694 незаболевших, а среди непривитых 328 – 121 = 207 заболевших и 11 668 – 207 = 11 461 незаболевших.

Сопоставим фактические числа заболевших и незаболевших с «ожидаемыми», вычисленными согласно нулевой гипотезе, предполагающей, что прививки неэффективны (табл. 8).

Таблица 7

Количество заболевших среди привитых и непривитых детей

Дети	Заболело	Не заболело	Итого
Привитые	56	6759	6815
Непривитые	272	11396	11668
Всего …	328	18155	18483

Таблица 8

Фактическое и ожидаемое число заболевших среди привитых и непривитых детей

Показатели	Привитые		Непривитые	
	Заболело	Не заболело	Заболело	Не заболело
Фактическое число	56	6759	272	11396
Ожидаемое число	121	6694	207	11461

Используя приведенную выше формулу, рассчитаем показатель соответствия:

$$\chi^2 = \frac{(26-121)^2}{121} + \frac{(6759-6694)^2}{6694} + \frac{(272-207)^2}{207} + \frac{(11396-11461)^2}{1164} = 56{,}2.$$

Число степеней свободы (K) определяется по формуле

$$K = (c-1)-(r-1),$$

где c – число граф первоначальной таблицы (кроме графы «итого»); r – число строк в таблице (кроме строки «всего»).

Таким образом, по данным табл. 7, число степеней свободы

$$K = (2-1)-(2-1) = 1.$$

Полученное значение $\chi^2 = 56{,}2$ значительно превышает критический показатель для уровня статистической значимости 99 % (см. приложение, табл. 8), т. е. вероятность подтверждения «нулевой гипотезы» – предположения об отсутствии связи между прививками и уменьшением заболеваемости привитых – очень мала, значительно меньше 1 %.

В данном примере (альтернативная вариация) можно упростить вычисление показателя (табл. 9). Вычисления в этом случае можно вести по формуле

$$\chi^2 = \frac{(ad-bc)^2 \cdot (a+b+c+d)}{(a+b)\cdot(c+d)\cdot(a+c)\cdot(b+d)}.$$

Подставим в эту формулу числа взятого примера (см. табл. 7):

$$\chi^2 = \frac{(56\cdot 11396 - 6759\cdot 272)^2 \cdot 18483}{6815\cdot 11668\cdot 328\cdot 18155} = 56{,}2.$$

Таблица 9

Макет таблицы для расчетов χ^2

Дети	Заболело	Не заболело	Всего
Привитые	a	b	$a + b$
Непривитые	c	d	$c + d$
Итого …	$a + c$	$b + d$	$a + b + c + d = n$

В этом случае, как и в предыдущем, $\chi^2 = 56{,}2$. Необходимо подчеркнуть, что критерий должен всегда определяться на основании абсолютных чисел, а не относительных. При этом все «ожидаемые» числа должны быть не меньше 5. В противном случае необходимо изменить группировку исследуемого материала, уменьшить число групп и получить большие числа наблюдений в каждой клетке первоначальной таблицы. Кроме того, объемы выборок должны составлять не менее 20 – 30 вариант.

Помимо изложенного выше критерия χ^2 существует большое количество других непараметрических критериев. На простых примерах рассмотрим применение этих методов.

В нашей работе [29] показано, что при сравнении попарно связанных выборок очень удобно применять критерий знаков и парный критерий Вилкоксона (критерий *T*). В эпидемиологических исследованиях связанные выборки обусловлены одновременностью происходящих изменений в каждом динамическом ряду и показатели заболеваемости за год, месяц, неделю и т. д. в этих рядах строго соответствуют друг другу.

Применение критерия знаков может найти широкое распространение при ретроспективном анализе (табл. 10). Принцип метода критерия знаков заключается в подсчете числа однонаправленных эффектов в попарно сопряженных вариантах.

При рассмотрении годовых показателей заболеваемости дизентерией в двух городах Иркутской области оказалось, что в городе А в 2 случаях из 9 заболеваемость была выше, чем в городе Б. Нулевая гипотеза заключается в том, что города А и Б по уровню многолетней заболеваемости не различаются. Чтобы ее отвергнуть, максимальное число менее часто встречающихся знаков при данном числе вариант должно быть равно 1 (см. приложение, табл. 9), а в нашем случае – 2. Следовательно, нулевая гипотеза не отвергается, т.е. по критерию знаков различия в двух рассмотренных выборках несущественны.

Таблица 10

Вычисление парного критерия Вилкоксона на примере заболеваемости дизентерией в двух городах Иркутской области (в показателях наглядности)

Год наблюдения	Город		Разность	Ранг	
	А	Б		абсолютного значения разности	числа однозначных разностей
1–й	100	146,4	– 46,4	7	
2–й	105,2	222,7	– 117,5	9	
3–й	109,5	193,6	– 84,1	8	
4–й	117,7	108,4	9,3	3	3
5–й	71,0	84,9	– 13,9	4	
6–й	43,2	74,4	– 31,2	6	
7–й	53,3	51,8	1,5	1	1
8–й	54,5	56,1	– 1,6	2	
9–й	27,7	49,9	– 22,2	5	

Однако логический анализ материала, приведенного в табл. 10, свидетельствует, что различия в уровнях заболеваемости в тех случаях, когда они были выше в городе А, крайне незначительны. Когда заболеваемость была выше в городе Б, эти расхождения были значительными. Можно предположить, что здесь имело место не отсутствие различий в двух выборках, а недостаточная мощность критерия знаков. В таком случае целесообразно или увеличить число наблюдений (анализируемых лет), или применить другой, более мощный, критерий, например критерий *T*, который учитывает не только направленность изменения, но и их размеры.

Расчет критерия *T* проводится следующим образом. Установленным разностям в попарно сопряженных показателях присваивают номера (ранги) в порядке возрастания абсолютных значений разности (без учета знака). Далее определяется значение *T*, равное сумме ранговых номеров менее часто (по знаку) встречающихся разностей (табл. 10, см. приложение, табл. 10). В нашем случае менее часто встречающиеся разности (по знаку) имели ранговые номера 1 и 3. Таким образом, $T = 1 + 3 = 4$. Это меньше 8 (критическое значение *T* при $n = 9$ и $P_T = 0{,}05$) и равно критическому значению при $P_T = 0{,}01$. Следовательно, ретроспективный анализ заболеваемости двух динамических рядов с помощью критерия *T* позволил доказать, что ее уровень в городах А и Б существенно различается, причем более высок он в городе Б.

Таблица 11

Даты начала и окончания сезонных подъемов заболеваемости дизентерией у детского и взрослого населения городе И. в 1990 – 2001 гг.

Год	Дети (0 – 14 лет)	Взрослые (15 лет и старше)	Направленность знака относительно взрослых
1990	18.07	23.06	+
1991	19.07	17.06	+
1992	22.07	17.07	+
1993	22.07	3.07	+
1994	30.06	11.07	–
1995	21.07	7.07	+
1996	27.07	4.07	+
1997	28.06	22.06	+
1998	29.06	23.06	+
1999	3.07	27.06	+
2000	24.07	14.07	+
2001	10.08	30.06	+

Непараметрические критерии могут использоваться в эпидемиологическом анализе не только для оценки рядов заболеваемости. В качестве примера приведем данные наших исследований, в которых с помощью критерия знаков и парного критерия Вилкоксона показано, что более раннее начало и окончание сезонного подъема дизентерии отмечается у взрослого населения, затем последовательно у детей старших (7 – 14 лет) и младших возрастных групп (табл. 11).

За 12 лет лишь в одном случае (знак минус) отмечено более раннее начало сезонного подъема заболеваемости дизентерией у детей. По таблице критериев знаков (см. приложение, табл. 9) находим, что при числе вариант 12 максимальное число менее часто встречающихся знаков для $P = 0{,}05$ и $P = 0{,}01$ равно соответственно 2 и 1. Следовательно, критерий знаков позволил выявить статистически значимое ($P = 0{,}01$) различие в сроках начала сезонного подъема у взрослых и детей.

Следует отличать *T*-критерий Виклоксона для связанных выборок от *U*-критерия Вилкоксона, который применяется для проверки гипотезы о принадлежности сравниваемых независимых (несвязанных) выборок к одной и той же генеральной совокупности. Часто этот критерий называют *U*-критерий (Манна – Уитни). Гипотезу проверяют,

расположив в обобщенный ряд значения сравниваемых выборок в возрастающем порядке. Всем значениям полученного обобщенного ряда присваиваются ранги от 1 до $N = n_1+n_2$, т.е. последний ранг должен быть равен сумме вариант выборок n_1 и n_2. Для меньшей выборки находят сумму рангов U. Если численность групп одинаковая, то сумму рангов U можно вычислять для любой из них.

Если выборки не различаются, то мы не должны ожидать преобладания значений одной выборки на одном из концов объединенного ранжированного ряда, их значения должны быть достаточно равномерно рассеяны по всему обобщенному ряду. Таким образом, слишком большие и слишком малые значения U должны заставить нас усомниться в справедливости нулевой гипотезы. Полученное значение U сравнивают с критическими значениями. Если U меньше или равно первому либо больше или равно второму, то нулевая гипотеза отвергается (различия статистически значимы).

Критические значения Манна – Уитни приведены в приложении (табл. 13). Столбец критических значений содержит пары чисел. Различия статистически значимы, если U не больше первого и не меньше второго. Например, когда в одной группе 3 варианты, а в другой 6, различия статистически значимы, если $U \leq 7$ или $U \geq 23$.

Если нужной численности групп в таблице критических значений не оказалось, то можно самому построить распределение U. Учитывая, что распределение U с увеличением объема выборки приближается к нормальному, то для него можно рассчитать среднее и стандартное отклонения, необходимые для вычисления искомого критического значения (z_U).

Для этого по следующим формулам рассчитывают среднее (μ_U) и стандартное отклонение σ_U [16]:

$$\mu_U = \frac{n_м(n_м + n_б + 1)}{2},$$

$$\sigma_U = \sqrt{\frac{n_м n_б(n_м + n_б + 1)}{12}},$$

где μ_U – среднее; σ_U – стандартное отклонение; $n_м$ – объем меньшей по количеству вариант выборки; $n_б$ – объем большей по количеству вариант выборки.

Затем вычисляют z_U и сравнивают его с критическим значением t нормального распределения (последняя строка табл. 5 приложения) для соответствующего уровня статистической значимости, т.е. со значениями 1,96 для 95%-го уровня статистической значимости или 2,58 – для 99 %-го:

$$z_U = \frac{\left|U - \mu_U\right| - \frac{1}{2}}{\sigma_U}.$$

Нулевая гипотеза об отсутствии различий может быть отклонена, если $z_U < t_{0,05}$.

Таблица 12

Расчет рангов для объединенной выборки показателей длительности стационарного лечения острой дизентерии Флекснера при легком и среднетяжелом ее течении

Длительность стационарного лечения (в днях) при течении болезни		Объединенная выборка	Ранг для объединенной выборки
легком	среднетяжелом		
1	2	3	4
8	8	8	5
8	9	8	5
9	9	9	10
10	9	10	14
7	11	7	2,5
7	12	7	2,5
9	12	9	10
9	13	9	10
11	14	11	16
6	12	6	1
	11	8	5
	9	9	10
		9	10
		9	10
		11	16
		12	19
		12	19
		13	21
		14	22
		12	19
		11	16
		9	10

Рассмотрим пример расчета U-критерия Манна – Уитни для оценки различий длительности стационарного лечения острой дизентерии Флекснера при легкой и средней степени развития болезни. Для расчета критерия U объединим варианты сравниваемых выборок и присвоим ранги для объединенного ряда в порядке возрастания (табл. 12). Надо обратить внимание на то, что если имеются одинаковые варианты, им присваивается средний ранг. Как оговаривалось выше, значение последнего ранга должно быть равно сумме вариант выборок n_1 и n_2. Это правило используется для проверки правильности ранжирования.

Для меньшей выборки рассчитываем суммы рангов U:

$$U = 5+5+10+14+2{,}5+2{,}5+10+10+16+1=76.$$

Для проверки значимости полученного критерия необходимо сравнить его с критическим значением (см. приложение, табл. 13) для $n_1 = 10$ и $n_2 = 12$. В нашем случае в табл. 13 приложения нет соответствующей строки. Так как размер групп достаточен, чтобы воспользоваться нормальным приближением для получения критического значения z_U, рассчитаем необходимые средние (μ_U) и стандартное отклонение σ_U:

$$\mu_U = \frac{10(10+12+1)}{2} = 115,$$

$$\sigma_U = \sqrt{\frac{10 \cdot 12\,(10+12+1)}{12}} \approx 15{,}2,$$

$$z_U = \frac{\left|76-115\right| - \frac{1}{2}}{15{,}2} = 2{,}53.$$

Полученное значение z_U сравниваем с критическим t, равным 1,96 для бесконечного числа степеней свободы при уровне статистической значимости 95 %. Найденное критическое значение больше полученного. Тем самым нулевая гипотеза отклоняется, и это, в свою очередь, позволяет говорить о статистически значимом различии показателей длительности стационарного лечения дизентерии Флекснера при легком и среднетяжелом течении болезни.

Непараметрическим критерием для сравнения нескольких групп является показатель Краскела – Уоллиса, который представляет собой обобщение критерия Манна – Уитни. Принцип основан на том, что все значения, независимо от того, какой выборке они принадлежат, упорядочивают по возрастанию. Каждому присваивают ранг – номер его в упорядоченном ряду (совпадающим величинам присваивают общий ранг, равный среднему тех мест, которые эти величины делят между собой в общем упорядоченном ряду). Затем вычисляют сумму рангов, относящихся к каждой группе, и для каждой группы определяют средний ранг. При отсутствии межгрупповых различий средние ранги групп должны оказаться близки. Напротив, если существует значительное расхождение средних рангов, то гипотезу об отсутствии различий следует отвергнуть. Значения критерия Краскела – Уоллиса (H) является мерой такого расхождения средних рангов.

Для простоты положим, что групп всего три. Имеются результаты некоторого признака в трех группах. Численность группы n_1 n_2 n_3. значения объединим, упорядочим и каждому присвоим ранг. Вычислим сумму рангов для каждой группы R_1 R_2 R_3, найдем средний ранг $R_{1/n}$ $R_{2/n}$ $R_{3/n}$. Общее число наблюдений $N = n_1 + n_2 + n_3$. Для объединенной группы рангов общая сумма равна

$$\frac{N(N+1)}{2},$$

тогда средний ранг равен

$$R = \frac{N+1}{2}.$$

Критерий Крускала – Уоллиса (H) рассчитывается по следующей формуле:

$$H = \frac{D}{N(N+1)12} = \frac{n_1(R_1 - R) + n_2(R_2 - R) + n_3(R_3 - R)}{N(N+1)12}.$$

Суммирование проводится по всем группам.

Рассчитанный показатель необходимо соотнести с критическим значением. Если группы не слишком малы, распределение H хорошо приближается к распределению χ^2 с числом степеней свободы, где

Таблица 13

Заболеваемость гепатитом В среди различных возрастных групп мужского населения

Город	40 лет и старше		30 – 39 лет		20 – 29 лет	
	Заболеваемость	Ранг	Заболеваемость	Ранг	Заболеваемость	Ранг
1	2,0	1	5,3	12	10,4	25
2	2,2	10	7,3	19	13,3	29
3	6,1	15	9,0	21	11,8	28
4	5,8	14	6,6	16	4,5	6
5	5,4	13	4,6	8	11,0	26
6	3,5	4	5,2	11	10,1	24
7	3,2	2	7,3	18	14,5	31
8	4,6	7	3,5	3	9,4	23
9	4,8	9	7,6	20	13,4	30
10	11,3	27				
11	3,8	5				
12	9,0	22				
13	7,2	17				
Сумма рангов …		146		128		222
Средний ранг …		11,2		14,2		24,7

k – число групп. Тогда для проверки нулевой гипотезы нужно просто вычислить по имеющимся наблюдениям значение P и сравнить его с критическим значением χ^2. В случае трех групп приближение с помощью χ^2 пригодно, если численность каждой группы не меньше 5. Для четырех груп, если общее число наблюдений не менее 10, но группы совсем малы, значение проверяют по таблице распределения Манна – Уитни (см. приложение, табл.13).

Разберем пример использования критерия Краскела – Уоллиса для оценки различий уровней заболеваемости в трех возрастных группах риска (табл. 13).

Общий средний ранг $\overline{R} = \dfrac{1+2+3...+31}{31} = \dfrac{N+1}{2} = \dfrac{31+1}{2} = 16.$

Вычисление суммы квадратов отклонений от средней по группам от общего среднего:

$$D = 13\,(11{,}2 - 16)^2 + 9\,(14{,}2 - 16)^2 + 9\,(24{,}7 - 16)^2 = 1000{,}8;$$

вычисление значения критерия Краскела – Уоллиса:

$$H = \frac{12}{N(N+1)} \cdot D = \frac{12}{31\,(31+1)} \cdot 1008{,}82 = 12{,}107.$$

По таблице критических значений находим критическое значение с числом степеней свободы $V = k - 1 = 3 - 1 = 2$, оно равно 9,210, т.е. меньше полученного нами. Таким образом, различия групп статистически достоверно.

В литературе [20, 21, 32 и др.] приводятся и другие непараметрические критерии статистики, которые в ряде случаев могут успешно заменить описанные выше методы оценки статистически значимых различий сравниваемых рядов. При этом задача нахождения различий может быть решена не только по средним тенденциям, что показано выше, но и по ширине распределений (серийные критерии, точный метод Фишера и др.).

Глава 3

АНАЛИЗ МНОГОЛЕТНЕЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ В ЭПИДЕМИОЛОГИЧЕСКОЙ ДИАГНОСТИКЕ

По современным представлениям факторы, влияющие на уровень инфекционной заболеваемости в многолетней динамике, можно разделить на три группы. Для краткой характеристики каждой группы воспользуемся классификацией, предложенной В.Д. Беляковым с соавт. [9].

Первая группа включает постоянно действующие факторы, которые определяют основное направление изменений параметров интенсивности эпидемического процесса, т.е. его многолетнюю эпидемическую тенденцию. Из природно-климатических факторов к постоянно действующим относят географическое положение, высоту над уровнем моря, рельеф местности. Установлено, что начало сезонного подъема заболеваемости во многом определяет географическое положение данной территории.

Во вторую группу объединены факторы, действующие периодически. Изменения интенсивности эпидемического процесса, связанные с влиянием факторов этой группы, называют периодическими, или циклическими, колебаниями (иногда просто цикличностью). Особенность данной формы проявления процесса – относительно равномерно появляющиеся изменения уровня регистрируемой заболеваемости. Особенно большое значение имеют периодические природно-климатические (метеорологические) факторы для внутригодовой динамики заболеваемости.

Третью группу составляют факторы случайные, непрерывно меняющиеся, которые вызывают нерегулярные колебания (вариации) интенсивности эпидемического процесса. Эти факторы отклоняют обычно регистрируемую заболеваемость от нормативного уровня, т.е. того уровня, который является результирующим между периодическими колебаниями и многолетней тенденцией.

К нерегулярным факторам можно отнести крупные аварии в системе централизованного водоснабжения, нарушение технологического процесса приготовления пищевых продуктов, стихийные бедствия (наводнение, землетрясение) и т.д.

Роль каждого фактора из этих групп вычленить довольно сложно, так как они, как правило, взаимосвязаны. В разных регионах в определенные периоды времени в качестве ведущих причин могут выступать различные сочетанные действия множества факторов. Причем по силе воздействия на уровень заболеваемости они приобретают первостепенное значение, т.е. выступают в роли пускового механизма. К сожалению, при обобщении результатов исследователи не всегда учитывают, что эти факторы значимы только для определенного типа регионов.

Сложность рассматриваемой проблемы связана с отсутствием четких критериев, оценивающих роль отдельных причин, обусловливающих интенсификацию заболеваемости. Как видно из вышеизложенного, в настоящее время многие причины, формирующие многолетние тенденции заболеваемости и конкретные механизмы, посредством которых эти причины реализуют свое влияние, для большинства инфекций остаются все еще предметом дискуссии [9]. Кроме того, проблема изучения многолетней динамики заболеваемости связана с отсутствием исходных материалов, собранных по единой методике, и стандартных приемов ретроспективного анализа.

Таким образом, основу изучения многолетней динамики заболеваемости составляет анализ многолетней эпидемической тенденции и многолетней периодичности. Многолетние тенденции заболеваемости в зависимости от характера процесса могут быть прямо- или криволинейными. Прямолинейная тенденция указывает на равномерное изменение интенсивности эпидемического процесса на протяжении анализируемого периода. Равномерное изменение обычно связано с пропорциональным изменением роли ведущих причин заболеваемости в распространении инфекции. Криволинейная тенденция обусловлена такими причинами, которые на протяжении анализируемого периода действуют неравномерно, и эффект их действия постепенно (или одномоментно) уменьшается или нарастает. Колебания интенсивности эпидемического процесса под действием периодических причин накладываются на многолетнюю тенденцию и образуют многолетнюю периодичность.

Исследование многолетней динамики инфекционной заболеваемости позволяет более углубленно изучить сущность эпидемического процесса, вскрыть причинно-следственные связи в проявлении его интенсивности. Задача статистического анализа – основного инструмента эпидемиологической диагностики – оценить количественно многочисленные и разнообразные эпидемические явления в пределах данной группы населения района, города и т. д.

Динамические ряды многолетней заболеваемости состоят из следующих компонентов: 1) многолетняя эпидемическая тенденция, или тренд; 2) долговременные периодические, или циклические, колебания; 3) нерегулярные (случайные) колебания.

Следовательно, цель анализа многолетней динамики заболеваемости – оценка возможных ее причин по характеру многолетней тенденции (рост, снижение, стабилизация), по многолетней периодичности и по соотношению уровня заболеваемости в анализируемый период со среднемноголетним или прогнозируемым уровнем с учетом тенденции и периодичности [9].

3.1. ОБОБЩАЮЩАЯ ХАРАКТЕРИСТИКА МНОГОЛЕТНЕЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ И ИЗМЕРЕНИЕ ЕЕ ИНТЕНСИВНОСТИ

Исследования многолетней динамики заболеваемости начинают с построения так называемых динамических (временных, хронологических) рядов. В связи с тем, что показатели заболеваемости получены за определенный интервал времени (месяц, год и т. д.), ряд, составленный из них, является интервальным.

После составления динамических рядов рассчитывают **средний уровень заболеваемости**. Он характеризует периоды, представленные в динамическом ряду. Средний уровень целесообразно вычислять в тех случаях, если общая тенденция динамики характеризуется монотонным изменением показателей во времени. При выраженной тенденции к росту (снижению) эта величина значительно менее информативна, поскольку не отражает характера изменения значений ряда во времени. Для решения вопроса о возможности использования одного показателя среднего уровня для характеристики всего ряда можно воспользоваться простым тестом, предложенным И.Л. Шаханиной с соавт. [36]: если максимальное и минимальное значения ана-

лизируемого динамического ряда различаются более чем в 2,5 раза, то расчет средней многолетней нецелесообразен. В таком случае (наличие выраженной тенденции) рекомендуется расчленить динамический ряд на несколько близких по уровням отрезков и оценивать каждый из них в отдель-ности.

Средний арифметический уровень заболеваемости динамического (интервального) ряда (y) вычисляется по формуле

$$y = \frac{\sum y}{n},$$

где y – отдельные уровни; n – число членов ряда.

Интенсивность изменений в динамике эпидемического процесса оценивают количественно с помощью ряда параметров:

1. **Абсолютный прирост** – показатель, характеризующий насколько изменился (стал больше или меньше) уровень за данный период по сравнению с предыдущим уровнем, т.е. это разность между абсолютными значениями уровней в данный и предыдущий периоды. При возрастании уровня в данный период прирост выражается положительными значениями, а при убывании – отрицательными.

2. **Темп роста** – это критерий, позволяющий проследить процесс изменения изучаемых явлений, т.е. на сколько процентов увеличился (или уменьшился) исследуемый уровень. Рассчитывается как отношение данного уровня к предшествующему и выражается, как правило, в процентах. Его называют темпом роста при цепном основании. Иногда пользуются вычислением темпа роста при постоянном основании, т. е. любой период условно считают исходным и принимают за 100 %, по отношению к нему рассчитывают значение всех остальных членов ряда. Убывающие динамические ряды следует преобразовывать по максимуму, а возрастающие – по минимуму.

3. **Темп прироста** – это отношение между абсолютным приростом данного периода и абсолютным уровнем предшествующего периода, выраженное в процентах. Темп прироста может быть представлен как отрицательными, так и положительными значениями. Легко видеть, что абсолютный прирост можно получить путем вычитания 100 % от темпов роста при цепном основании (темп прироста = темп роста – 100 %). Информативное значение прироста такое же, как и у темпа роста.

Таблица 14

Многолетняя динамика заболеваемости дизентерией, абсолютный прирост и темпы роста и прироста

Год наблюдения	Фактический уровень	Абсолютный прирост	Темп роста, %		Темп прироста, %
			при постоянном основании	при цепном основании	
1-й	872,1	-	-	-	-
2-й	908,6	36,5	104,2	104,2	4,2
3-й	926,6	18,0	106,2	102,0	-2,0
4-й	589,9	– 336,7	67,6	63,7	36,3
5-й	358,4	– 231,5	41,2	60,8	39,2
6-й	442,7	84,3	50,8	123,5	23,3
7-й	286,3	– 156,4	32,8	64,7	– 35,3
8-й	230,4	– 55,9	26,4	80,5	– 19,5
9-й	275,6	45,2	31,6	119,6	19,6
10-й	192,1	-83,5	22,0	69,7	– 30,3
11-й	336,0	153,9	38,5	174,9	74,9

В качестве примера рассмотрим показатели заболеваемости дизентерией за 11 лет в городе А (табл. 14). Приведенные показатели оценивают изменения заболеваемости в конкретный период времени по отношению к предыдущему или какому-нибудь исходному периоду. Для общей характеристики динамики изучаемого явления за весь анализируемый период используют средние показатели, т.е. средний абсолютный прирост, а также средний темп роста и прироста.

Средний абсолютный прирост характеризует средний ежегодный рост или снижение уровня ряда. Вычисляют его по формуле

$$y = \frac{y_n - y_1}{n-1},$$

где y_n и y_1 – показатели конечного и начального периодов соответственно; n – число периодов.

Средний темп роста показывает среднюю скорость изменения уровня в динамическом ряду. Для определения среднего темпа роста следует вычислять среднюю геометрическую (см. гл. 1).

Средний темп роста, как правило, выражают в процентах и определяют следующим образом:

$$\mathrm{T_p} = \sqrt[n-1]{\frac{x_n}{x_1}} \cdot 100\,\%,$$

где $T_р$ – средний темп роста; x_1 и x_n – показатели начального и конечного периодов соответственно, n – число лет исследуемого периода.

Для определения среднегодового темпа прироста $T_{пр}$ используется следующая формула: $T_{пр} = T_р - 100\ \%$.

Рассмотрим применение указанных формул на фактических данных по заболеваемости дизентерией (см. табл. 14).

Средняя абсолютного прироста (ежегодно) равна

$$\frac{336{,}0 - 872{,}1}{11-1} = \frac{-536{,}1}{10} = -53{,}6\ {}^{0}/_{0000}.$$

Средний темп роста при использовании предлагаемой формулы равен

$$\sqrt[11-1]{\frac{336{,}0}{872{,}1}} \cdot 100\,\% = 0{,}909 \cdot 100\,\% = 90{,}9\,\%.$$

Темп прироста в рассматриваемом динамическом ряду равен
$T_{пр} = 90{,}9 - 100\ \% = -9{,}1\ \%$.

3.2. МЕТОДЫ АНАЛИЗА МНОГОЛЕТНЕЙ ТЕНДЕНЦИИ

Следующим этапом анализа является вычисление многолетней тенденции с помощью выравнивания по плавной линии, чтобы освободиться от периодических и случайных колебаний. Динамические ряды можно выравнивать с помощью графического изображения ряда. В этом случае на график наносят фактические данные, а затем на основе расположения фактических точек проводят линию, отражающую тенденцию. Для снижения субъективной ошибки определений теоретической линии пользуются различными математическими приемами. Наиболее простые – методы сглаживания по скользящей средней и укрупнения периодов.

Сущность метода скользящей средней состоит в том, что фактические показатели динамического ряда заменяют на теоретические ожидаемые ($y_т$) путем осреднения двух или трех рядом стоящих фактических показателей. Например, если принять 2-летние периоды для осреднения (шаг в 2 года), то первая средняя будет представлять сглаженные показатели заболеваемости из взятых нами первых 2 лет:

$$y_{T_1} = \frac{y_1 + y_2}{2}.$$

Вторая средняя получается путем осреднения фактических показателей за 2-й и 3-й годы:

$$y_{T_2} = \frac{y_2 + y_3}{2} \text{ и т.д.}$$

Если же нами для сглаживания приняты 3-летние циклы, то показатели теоретически ожидаемых значений (сглаженных) будут соответственно равны:

$$y_{T_1} = \frac{y_1 + y_2 + y_3}{3};\ y_{T_2} = \frac{y_2 + y_3 + y_4}{3} \text{ и т.д.}$$

На основе данных по заболеваемости (см. табл. 14) рассчитаем скользящие средние (табл. 15).

Легко заметить, что выровненные показатели более четко выявляют тенденцию снижения заболеваемости по сравнению с фактическими материалами. Это особенно отчетливо проявилось при выравнивании с шагом в 3 года. Однако при таком подходе теряется часть сведений (год при двухшаговом выравнивании и 2 года – при трехшаговом). Кроме того, в этом случае не полностью устраняется влияние

Таблица 15

Динамические ряды фактической и выравненной (методом скользящей средней) заболеваемости дизентерией, %

Год наблюдения	Фактический уровень	Заболеваемость после сглаживания	
		с шагом в 2 года	с шагом в 3 года
1-й	872,1	-	-
2-й	908,6	890,4	902,4
3-й	926,6	917,6	808,4
4-й	589,9	758,2	625,0
5-й	358,4	474,2	463,7
6-й	442,7	400,6	362,5
7-й	286,3	364,5	319,8
8-й	230,4	258,4	264,1
9-й	275,6	233,8	232,7
10-й	192,1	264,0	267,9
11-й	336,0	264,0	-

случайных колебаний и имеет место субъективный подход исследователя при определении периодов для осреднения.

Более простым, чем сглаживание по скользящей средней, хотя и менее точным, является метод укрупнения периодов. В вышерассмотренном примере можно укрупнить периоды динамического ряда и вместо годовых показателей оценивать средние по периодам (см. табл. 15). В этом случае, с учетом движения заболеваемости, динамический ряд будет выглядеть следующим образом:

$$x_{1-3}=(872{,}1+908{,}6+926{,}6)/3=902{,}4,$$
$$x_{4-6}=(589{,}9+358{,}4+442{,}7)/3=463{,}7,$$
$$x_{7-9}=(286{,}3+230{,}4+275{,}6)/3=264{,}1,$$
$$x_{10-11}=(192{,}1-336{,}0)/2=264{,}0.$$

Полученный динамический ряд показывает четкую тенденцию к снижению заболеваемости дизентерией. Однако методу укрупнения периодов свойственны те же недостатки, что и методу скользящих средних.

Указанных недостатков лишен метод наименьших квадратов. Название его связано с тем, что сумма квадратов разностей между фактическим уровнем и теоретическим наименьшая. Следовательно, для каждого конкретного случая линия тенденции наиболее близка к фактическим результатам. В связи с этим метод наименьших квадратов является наиболее объективным способом определения тенденции эпидемического процесса.

Рассмотрим применение метода наименьших квадратов на примере прямолинейной тенденции. Уравнение прямой линий, иначе называемое параболой первой степени (первого порядка), имеет вид

$$y_т = bx + a,$$

где $y_т$ – теоретические показатели; a – начальный уровень ряда; b – коэффициент регрессии прирост показателя); x – время порядковый номер года.

Фактические показатели ряда возьмем из представленного выше примера (см. табл. 15). В этом случае порядок анализа будет следующим:

1. Составляется матрица таблицы

x	y	xy	x^2	$y_т$

2. В графе x приводится условная нумерация лет. Для удобства расчетов за начальный год ($x = 0$) принимается не первый год, а центральный. Нумерация последующих лет будет в порядке возрастания чисел, т. е. 1, 2, 3, 4 и т. д. Более ранние годы (до центрального) располагаются в таком же порядке, но номера имеют отрицательный знак.

3. В графу y вносят фактические показатели динамического ряда.

4. Показатели x и y перемножают и результаты записывают в графу xy.

5. Нумерацию исследуемых лет возводят в квадрат и полученные значения вносят в графу x^2.

6. Суммируют показатели граф: y, xy, x^2.

7. Вычисляют значения a и b по следующим формулам:

$$a = \frac{\sum y}{n};\ b = \frac{\sum xy}{\sum x^2},$$

где n – число членов ряда (число анализируемых лет).

8. Находим теоретические уровни линии тенденции по формуле

$$y_т = bx + a.$$

9. Полученные величины вносим в графу $y_т$.

Ход всех вычислений приведен в табл. 16. Получив все суммы, определяем значения a и b.

$$a = \frac{5418,7}{11} = 492,6,\ b = \frac{-8290,6}{110} = -75,4.$$

Вводя полученные значения a и b в уравнение $y_т = bx + a$, при этом придавая x значения чисел своего столбца, рассчитываем выравненные показатели ряда, которые будут равны

$y_т$ для 1-го года = $(-75,4) \times (-5) + 492,6 = 869,6$,
$y_т$ для 2-го года = $(-75,4) \times (-4) + 492,6 = 794,2$,
$y_т$ для 3-го года = $(-75,4) \times (-3) + 492,6 = 718,8$ и т.д.

Показатели выравненного ряда наносим на график и точки со-

Таблица 16

Динамический ряд заболеваемости, выравненный методом наименьших квадратов

Год наблюдения	x	y	xy	x^2	$y_т$
1-й	– 5	872,1	– 4360,5	25	869,6
2-й	– 4	908,6	– 3634,4	16	794,2
3-й	– 3	926,6	– 2779,8	9	718,8
4-й	– 2	589,9	– 1179,8	4	643,4
5-й	– 1	358,4	– 358,4	1	568,0
6-й	0	442,7	0	0	492,6
7-й	1	286,3	286,3	1	417,2
8-й	2	230,4	230,4	4	341,8
9-й	3	275,6	275,6	9	266,4
10-й	4	192,1	192,1	16	191,6
11-й	5	336,0	336,0	25	115,6
$\sum$		5418,7	– 8290,6	110	

единяем прямой. В целях экономии времени можно вычислять $y_т$ первого и последнего членов ряда и только по двум точкам вычертить линию тренда.

Разобранный выше пример выравнивания методом наименьших квадратов показан для нечетного числа членов динамического ряда. При четном числе анализируемых лет нумерацию их ведут так же от середины ряда, который при этом состоит из 2 лет, а значения x приминают равными – 1 и +1, далее идут значения – 3, – 5, – 7 и т.д., вниз +3, +5, +7 и т.д. Таким образом, в нечетном ряду от середины вверх и вниз значения увеличиваются на одну единицу, в четном – на две.

По результатам расчета многолетней динамики заболеваемости (см. табл. 14) рассчитаем темпы ее роста и прироста:

$$T_р = \sqrt[n-1]{\frac{x_n}{x_1}} \cdot 100\,\% = \sqrt[11-1]{\frac{115,6}{869,6}} \cdot 100\,\% = 0,817 \cdot 100\,\% = 81,7\,\%,$$

$$T_{пр} = T_р - 100\,\% = 81,7 - 100 = -18,3\,\%.$$

Как видим, показатели, характеризующие многолетнюю динамику заболеваемости, рассчитанные на основе данных, выравненных методом наименьших квадратов, существенно различаются с данными, полученными на основе первичного (статистически необработанного) материала.

Итак, мы рассмотрели пример прямолинейной тенденции. Однако нередко тенденция явления (заболеваемости) может быть криволинейной и прямая не подойдет для ее оценки. В таких случаях применяют выравнивание с помощью различных кривых, т. е. параболами, гиперболами, экспоненциальными и степенными кривыми и т.д. Примеры выравнивания динамических рядов подробно освещены в ряде руководств [20, 23]. При выравнивании по кривым значительно возрастают сложность и объем вычислений. Следует отметить, что в преобладающем большинстве случаев кривые высоких порядков можно условно делить на два (реже на три) периода и каждый из них выровнять по прямой линии. Такой подход целесообразно применять не только для уменьшения объема вычислений, но и для более четкого проявления закономерности эпидемического процесса. Например, предварительным анализом выявлено наличие качественных изменений во внешних условиях (это может быть введение в эксплуатацию очистных сооружений на водозаборе, проведение вакцинопрофилактики и др.), влияющих на многолетнюю заболеваемость. В этом случае кривую сложного порядка можно разложить на составляющие прямые отрезки, учитывающие влияние факторов. Такой прием будет значительно лучше отражать тенденцию.

Оценивая изменения динамического ряда, необходимо помнить, что снижение или рост заболеваемости за многолетний период не всегда бывают строго прямолинейными и часто представлены отдельными скачкообразными изменениями. В этом случае применение для характеристики ряда среднегодовых темпов по фактическим данным с использованием среднегеометрических показателей становятся не совсем корректным, так как в указанной формуле учитываются показатели только крайних членов динамического ряда, в связи с чем полученные знания не отвечают истинному движению заболеваемости. Объективные показатели темпов динамики при использовании средней геометрической имели бы место при двух условиях: во-первых, должно быть устранено действие случайных факторов, влияющих на многолетнее течение эпидемического процесса; во-вторых, расчет необходимо вести по теоретической линии, которая должна быть максимально приближена ко всем фактическим данным. Этим условиям наиболее полно отвечает построение линии тенденции по методу наименьших квадратов с использованием общепринятого уравнения прямой линии.

В качестве примера можно привести использование метода опреде-

ления среднегодового темпа роста (прироста) с учетом фактических и выравненных показателей (табл. 17). В приведенных материалах среднегодовое снижение заболеваемости по выравненным показателям оказалось на 1,4 % больше, чем вычисление по фактическим данным. Несмотря на кажущееся небольшое различие, значение становится довольно существенным и составляет 25,4 %, если принять темп прироста по фактическим значениям за 100 %.

Колеблемость годовых показателей приводит к тем большему отклонению фактических данных от выравненных, чем больше эти величины и чем ближе приходятся они к началу или концу динамического ряда. Лишь в тех случаях, когда динамический ряд представлен в виде однонаправленного непрерывно меняющегося уровня, темпы динамики, вычисленные по фактическим и выравненным данным, различаются незначительно.

При анализе динамического ряда иногда может возникнуть необходимость оценить динамику заболеваемости по внутригодовым эпидемическим периодам. Такой анализ обязателен при инфекциях с выраженными сезонными подъемами, так как каждый внутригодовой период имеет свои особенности проявления эпидемического процесса.

Так, нашими исследованиями [3, 27] показано, что многолетнее снижение заболеваемости дизентерией Зонне и Флекснера определяется прежде всего снижением заболеваемости в межсезонный и предсезонный периоды соответственно. При этом заболеваемость, заре-

Таблица 17

Заболеваемость дизентерией (в показателях наглядности) в г. И., %

Год наблюдения	Данные	
	фактические	выравненные
1-й	100,0	100,0
2-й	60,8	94,6
3-й	75,0	89,1
4-й	48,5	83,7
5-й	38,1	78,2
6-й	46,7	72,8
7-й	32,6	67,3
8-й	57,0	61,9
9-й	63,7	56,4
$T_р$	94,5	93,1
$T_{пр}$	– 5,5	– 6,9

гистрированная в период сезонного подъема, снижалась значительно менее интенсивно, а в предсезонный период для дизентерии Зонне имелась даже тенденция к росту (табл. 18).

Необходимо отметить, что, получив в результате преобразований выравненный ряд, исследователь должен проверить, является ли обнаруженное снижение (рост) заболеваемости случайным или же оно носит закономерный характер. Это особенно важно, когда имеющиеся тенденции затушевываются многолетними циклами или нерегулярными колебаниями.

Такая проверка может проводиться несколькими способами.

1. Показатели динамического ряда разбиваются на несколько групп, и проверяется нулевая гипотеза о случайности различий их средних. При опровержении нулевой гипотезы полагают, что различие средних не случайно, т.е. тенденция достоверна.

Из рассмотренных нами примеров можно воспользоваться данными табл. 13, на основании которых получены четыре укрупненных периода: 902,4; 463,7; 264,1; 264,0. Последовательное сравнение указанных средних при помощи критерия Стьюдента дает высокодостоверную разницу между 1-м и 2-м, а также 2-м и 3-м периодами заболеваемости (t = 21 и 17 соответственно). И лишь между 3-м и 4-м периодами разница оказалась статистически незначимой (t = 0,02). Однако анализ динамического ряда свидетельствует, что уровень 4-го периода связан преимущественно с последней вариантой (336,0). При

Таблица 18

Многолетняя динамика заболеваемости дизентерией в г. Иркутске по эпидемическим периодам

Этиология	Годы	Темпы прироста заболеваемости, %				
		годовой	сезонной	внесезонной	предсезонной	межсезонной
Дизентерия Зонне	1977 – 1986	– 2,8	– 1,6	– 8,0	+ 2,6	– 11,4
	1978 – 1986	– 2,0	– 0,2	– 9,3	+ 1,5	– 14,1
	1977 – 1982	– 15,7	– 15,2	– 20,2	+ 6,1	– 27,1
	1978 – 1982	– 20,5	– 17,8	– 31,8	+ 8,5	– 49,6
Дизентерия Флекснера	1977 – 1986	– 11,7	– 11,4	– 12,7	– 15,1	– 10,8
	1978 – 1986	– 9,6	– 8,6	– 11,4	– 12,2	– 10,0
	1977 – 1982	– 16,8	– 17,9	– 15,0	– 21,6	– 12,6
	1978 – 1982	– 15,3	– 15,3	– 15,3	– 21,3	– 2,3

ее исключении можно сравнить заболеваемость десятого по порядку члена ряда (192,1) и предшествующего периода (264,1). В этом случае разница между показателями будет также высокодостоверна ($t = 8$). Таким образом, с учетом выбранных нами укрупненных периодов и исключением последнего члена ряда, нулевая гипотеза отвергается со статистической значимостью более чем 99,9 %.

2. Фазочастотный критерий знаков разностей. Находим разности уровней ряда: нулевая гипотеза состоит в утверждении, что знаки этих разностей образуют случайную последовательность. Назовем последовательность одинаковых знаков разностей фазой и будем рассчитывать число фаз (без первой и последней). Фактические значения критерия $t_{ф}$ определяем по следующей формуле:

$$t_{ф} = \frac{[h - (2n - 7)/3] - 0{,}5}{(16n - 29)/90},$$

где h – фаза; n – число членов ряда.

Рассмотрим применение фазочастотного критерия на примере заимствованного из монографии И.Г. Венецкого, В.И. Венецкой [14]. Допустим, имеется динамический ряд с числом, равным 22 (табл. 19).

Находим знаки отклонения ($y_{i+1} - y_i$) и проставляем нумерацию фаз. Получаем $h = 7$. Используя значения вероятностей $t_{т}$:

p	$t_{т}$	p	$t_{т}$
0,001	3,29	0,03	2,17
0,005	2,81	0,04	2,05
0,01	2,58	0,05	1,96
0,02	2,33	0,1	1,64

находим, что при $p = 0{,}05$, т.е. 5 % уровня, теоретическое значение критерия $t_{т} = 1{,}96$.

Таблица 19

Уровни и фазы динамического ряда

Уровень ряда	10	9	8	7	6	4	8	5	4	6	8	10	9	8	4	5	7	6	4	3	7	6
Знак отклонения ($y_{i+1} - y_i$)		–	–	–	–	–	+	–	–	+	+	+	–	–	–	+	+	–	–	–	+	–
Нумерация фаз							1	2		3			4			5					6	7

Фактическое значение критерия t_{Φ} равно:

$$t_{\phi} = \frac{[7-(2\cdot 22-7)/3]-0{,}5}{(16\cdot 22-29)/90} = \frac{4{,}83}{1{,}89} = 2{,}56.$$

Таким образом 2,56 > 1,96, т.е. $t_{\phi} > t_{т}$, следовательно, нулевая гипотеза отвергается. Уровни динамического ряда не образуют случайную последовательность.

3. Критерий Кокса и Стьюарта. Весь динамический ряд делим на три группы уровней. Численности первой и третьей групп уровней должны быть равны между собой и составлять треть уровней каждая (при n, не делящемся на 3, добавляем единицу или два). Фиксируем знаки отклонений каждого уровня третьей группы от соответствующего уровня первой группы. Суммируем знаки «плюс» (при возрастании тенденции) или «минус» (при убывающей тенденции) и получаем число S. Фактическое значение критерия определяем по формуле

$$t_{\phi} = \frac{(S-n/6)-0{,}5}{\sqrt{n/12}},$$

где S – сумма знаков «плюс» (при возрастании тенденции) или «минус» (при убывании тенденции).

Используя данные предыдущего примера, добавляем число членов ряда до 24 и делим на три части. Получаем уровни групп:

Уровень группы		Знак отклонения
последней	первой	
4	10	–
5	9	–
7	8	–
6	7	–
4	6	–
3	4	–
7	8	–
6	5	+

Получено 7 знаков «минус» из 8. Проверка на убывающую тенденции дает:

$$t_{ф} = \frac{(7 - 22/6) - 0{,}5}{\sqrt{22/12}} = \frac{2{,}83}{1{,}35} = 2{,}10.$$

На основе значения вероятностей $t_{т}$ находим, что соответствующая вероятность меньше 0,04. Таким образом, на 5%-м уровне установлено наличие убывающей тенденции.

Из приведенных выше примеров статистической оценки динамического ряда на наличие тенденции наиболее мощным критерием является сравнение средних величин в нескольких группах. Два последних метода пригодны только при условии большого количества вариант в оцениваемом ряду. Приемлемые результаты по этим непараметрическим методам можно получить вариант в ряду, равном 20 и выше.

Помимо рассмотренных выше способов оценки статистической значимости различий тенденций можно применить предложенный нами [28] метод с использованием коэффициента корреляции рангов Спирмена. Необходимо подчеркнуть, что в этом случае коэффициент корреляции рангов Спирмена применяется в новой для себя функции, т.е. его нельзя трактовать как меру измерения связи заболеваемости со временем. Рассмотрим использование этого метода на динамических рядах заболеваемости дизентерии Зонне и Флекснера в г. И. за 10 лет. С этой целью ранжируем годы наблюдений в порядке их возрастания и соответствующие им показатели заболеваемости (табл. 20). Коэффициент ранговой корреляции Спирмена (ρ) равен

$$\rho = 1 - \frac{6\sum d^2}{n(n^2 - 1)},$$

где d – разность рангов, n – число пар наблюдений.

Не приводя здесь хода вычислений, укажем лишь конечный результат, который для дизентерии Зонне равен 0,15, для дизентерии Зонне и для дизентерии Флекснера – 0,90.

Статистическую значимость вычисленного коэффициента определяем по достоверности коэффициента корреляции рангов изучаемой величины (заболеваемости) со временем (см. приложение, табл. 11).

В нашем случае (число вариант 10) критические значения показа-

Таблица 20

Заболеваемость различными этиологическими видами дизентерии (в показателях наглядности в городе И.)

Год	Ранг года	Дизентерия Зонне		Дизентерия Флекснера	
		Заболеваемость	Ранг заболеваемости	Заболеваемость	Ранг заболеваемости
1977	1	100,0	7	100	10
1978	2	152,2	10	87,8	9
1979	3	40,0	2	47,0	6
1980	4	46,5	3	60,4	8
1981	5	102,7	8	48,3	7
1982	6	39,0	1	44,0	4
1983	7	126,9	9	46,2	5
1984	8	71,3	5	33,6	1
1985	9	57,8	4	34,8	2
1986	10	95,3	6	43,0	3

теля корреляции рангов при $P = 0,05$ и $P = 0,01$ составляют 0,64 и 0,79 соответственно. Таким образом, за изученный 10-летний период установлено статистически значимое ($P < 0,01$) снижение заболеваемости только для дизентерии Флекснера.

Проведя аналогичные вычисления по представленным ранее материалам (см. табл. 14), получаем те же выводы, как и при использовании критерия Стьюдента: наличие статистически значимой тенденции ($P < 0,01$) при исключении последнего члена ряда и отсутствие ее при оставлении этой варианты. Однако ход вычислений при использовании коэффициента корреляции рангов был значительно проще. Следует также отметить, что непараметрический критерий оценивает совокупность всех вариант, а не отдельные точки (часто субъективные) динамического ряда. Полное совпадение выводов о достоверности тенденции ($P < 0,05$) динамического ряда по приведенным выше данным (см. табл. 16) отмечалось также для критериев фазочастотного, Кокса и Стюарта. При этом для предлагаемого нами метода оценки достоверности тенденций динамический ряд ограничен лишь числом пар наблюдений, который должен состоять не меньше чем из пяти вариант.

В рассматриваемом случае коэффициент корреляции рангов был использован в качестве меры направленности процесса, и в этой роли его можно назвать коэффициентом монотонности. При таком

подходе указанный коэффициент будет выполнять в ретроспективном эпидемиологическом анализе еще одну важную количественную функцию – оценивать колеблемость динамических рядов. Подобный подход применен ранее при оценке урожайности зерновых культур М. Юзбашевым, Ф. Афанасьевым [38].

Коэффициент монотонности принимает значения от 0 до ±1. Чем ближе этот показатель к 1, тем выше монотонность движения заболеваемости. Оценивать монотонность как низкую, среднюю и высокую следует по общепринятым мерам тесноты связи. В нашем примере (см. табл. 20) многолетняя динамика заболеваемости дизентерией Зонне имеет низкий (– 0,15), а дизентерией Флекснера – высокий коэффициент монотонности (– 0,90). Это связано с тем, что для дизентерии Зонне характерны многолетние циклы с шагом в 2–3 года, которые и обусловили низкий коэффициент монотонности (см. табл. 18).

Следовательно, использование коэффициента монотонности позволяет выявить достоверные тенденции и оценить отклонения в циклическом течении и (или) нерегулярных колебаниях эпидемического процесса.

3.3. ОЦЕНКА МНОГОЛЕТНЕЙ ПЕРИОДИЧНОСТИ

Одной из закономерностей эпидемического процесса, присущей всем инфекциям, является многолетняя периодичность (цикличность). Динамика многолетней цикличности выражается, как правило, в синусоидоподобном колебании относительно линии многолетней тенденции (тренда). Продолжительность полного эпидемического цикла (длина волны) зависит как от особенностей инфекции, так и от условий, в которых протекает эпидемический процесс. Кроме этого, на длину волны цикла может оказывать влияние и субъективный подход исследователя, т.е. продолжительность выявленных циклов зависит также от использованного статистического метода. Так, В.Н. Ягодинский [39] указывает, что длительность периодов для скарлатины, согласно разным авторам, колеблется от 2 – 3 до 10 – 15 лет. В связи с этим для выделения многолетних циклов необходимо располагать данными по заболеваемости за длительные периоды времени. Для надежного заключения о наличии цикличес-

ких колебаний желательно опираться на несколько периодов, различающихся между собой.

Оптимальные требования к методам выявления циклов эпидемического процесса включают [39]: 1) установление реальности и продолжительности циклов; 2) выявление ведущих и вторичных циклических составляющих; 3) разграничение циклов и «случайных» колебаний; 4) статистическую оценку цикличности; 5) учет вероятного рассеивания частот колебаний; 6) возможность исследования циклов по кратким рядам наблюдений.

Для решения указанных задач применяется несколько методов, дополняющих и контролирующих друг друга (автокорреляционный, спектральный, когерентный анализы и др.).

Наиболее простым методом выявления эпидемических циклов является визуальная оценка графического изображения ряда. Несмотря на всю свою простоту и малую надежность, иногда при выраженной цикличности эпидемического процесса можно вычленить имеющиеся периодические колебания. Для погашения случайных колебаний следует применять сглаживание рядов по скользящей средней (по показателям 2 или 3 смежных лет).

Более надежным, хотя также ориентировочным методом оценки многолетней периодичности заболеваемости является анализ отклонений фактических показателей заболеваемости от теоретической линии тенденции. Выявленные отклонения отражают влияние на эпидемический процесс только двух групп причин: периодических и нерегулярных. Последняя группа причин элиминируется методами сглаживания кривой, а влияние длительно действующих факторов устраняется многолетней линией тенденции.

На основе фактических показателей заболеваемости вирусным гепатитом А в одном из городов Восточной Сибири нами составлена кривая, имеющая четко выраженные циклы с периодом в 5 лет, за исключением последнего периода, когда подъем приходится на 5-й год (рис. 2). Кривая настолько упорядочена, что в данном случае можно не проводить ее сглаживания. И лишь для чисто учебных целей на этом рисунке приведена выровненная кривая с шагом осреднения в 3 года.

Выявление периодических колебаний с помощью анализа отклонения фактических данных от теоретических представлено в табл.

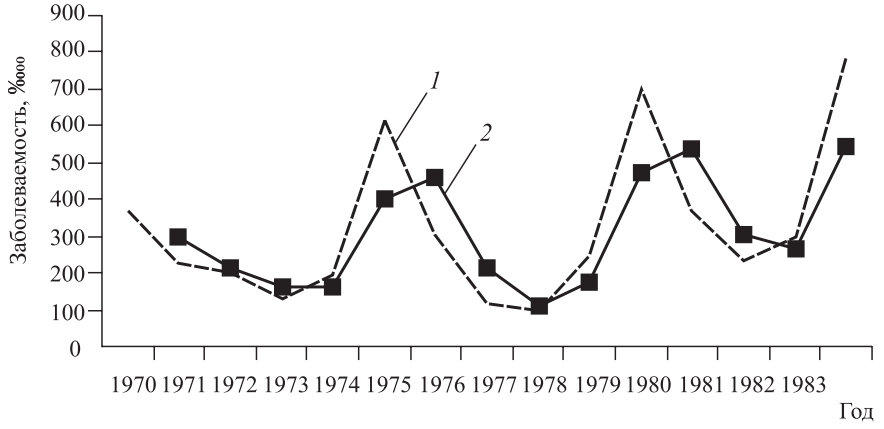

Рис. 2. Фактические показатели заболеваемости гепатитом А (*1*) и сглаженная (по методу скользящих средних) кривая с шагом 3 года (*2*).

21 и на рис. 3. В таблице приведены лишь фактические и выровненные данные, так как порядок расчетов подробно рассмотрен в разделе 3.1. Выравнивание эмпирического ряда показывает выраженную тенденцию к росту заболеваемости в анализируемый период (коэффициент регрессии равен +18,0). Тенденция ряда в нашем случае выражается уравнением $y_т = 325{,}8 + 18{,}0x$. После вычисления теоретических показателей находим отклонения фактических

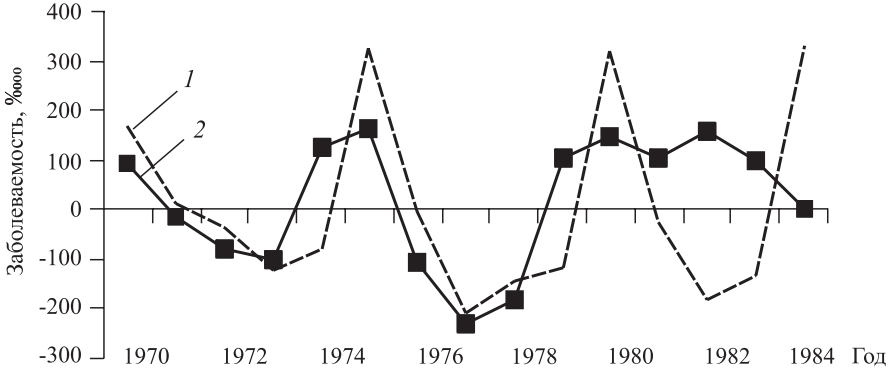

Рис. 3. Многолетняя периодичность гепатита А по фактическим данным (*1*) и сглаженным с шагом в 2 года (*2*).

Таблица 21

Оценка многолетней периодичности гепатита А (на 100 тыс. населения)

Год наблюдения	Заболеваемость		Отклонения фактических данных от теоретических	Скользящая средняя отклонений с шагом в 2 года
	фактическая	теоретическая		
1-й	370,1	199,8	170,3	90,7
2-й	228,9	217,8	11,1	– 13,0
3-й	198,8	235,8	– 37,0	– 80,4
4-й	130,1	253,8	– 123,7	– 101,6
5-й	192,3	271,8	– 79,5	123,0
6-й	615,2	289,8	325,4	162,0
7-й	306,5	307,8	– 1,3	– 106,2
8-й	114,7	325,8	– 211,1	– 227,4
9-й	100,2	243,8	– 143,6	– 180,2
10-й	244,9	361,8	– 116,9	101,2
11-й	699,0	379,8	319,2	146,2
12-й	371,1	397,8	– 26,7	105,4
13-й	231,6	415,8	– 184,2	159,0
14-й	300,0	433,8	– 133,8	99,4
15-й	784,4	451,8	332,6	-

показателей от теоретических и полученные значения наносим на график, соединив точки ломаной линией (см. рис. 3). Полученная кривая подтверждает наше предварительное заключение о наличии цикличности с периодами в 5 лет. Сглаживание кривой по скользящей средней (шаг в 2 года) еще более подчеркивает выявленную закономерность.

Предполагаемое несовпадение длины волны последнего периода (4 года вместо 5) объясняется следующим образом. Высокие показатели заболеваемости гепатитом А в последний год наблюдения обусловлены массивным фекальным загрязнением воды в городском водопроводе. При этом увеличение количества проб воды, в которых показатели качества отклонялись от стандарта, предшествовало подъему заболеваемости гепатитом А со сроком, укладывающимся в инкубационный период. Этот факт при наличии выраженной тенденции к росту заболеваемости на фоне периодического подъема и обусловил высокие показатели заболеваемости, а также укоротил длину волны многолетнего цикла до 4 лет.

Проведенный анализ многолетней периодичности с элиминацией длительно действующих причин позволяет нам определить долю циклических колебаний в отдельные годы эпидемического цикла в общей динамике заболеваемости. Например, для первого года наблюдения (см. табл. 21) подъем заболеваемости превысил теоретически ожидаемый уровень на 170,3 в расчете на 100 тыс. населения (370,1 – – 199,8 = 170,3). Если показатель фактической заболеваемости (370,1) принять за 100 %, то доля циклической надбавки составит 46,0 %. Примерно такая же доля заболеваемости связана с другими максимальными показателями циклических подъемов вирусного гепатита А (от 42,4 % в последнем году наблюдения до 52,9 % на 6-й год наблюдения), что подчеркивает стабильность циклического процесса.

3.4. АНАЛИЗ НЕРЕГУЛЯРНЫХ КОЛЕБАНИЙ ЗАБОЛЕВАЕМОСТИ

Наряду с причинами, действующими на динамику заболеваемости как постоянно, так и периодически на протяжении многих лет, широкое влияние на этот показатель оказывают множество природных и социальных факторов, которые действуют без определенной закономерности.

Природные факторы, вызывающие нерегулярные подъемы или спады заболеваемости, связаны, как правило, с необычными климатическими условиями для данной местности, а именно наводнениями, обильными выпадениями осадков или длительной засухой, необычно высокой или низкой температурой воздуха и т.п. Социальные причины нерегулярных колебаний заболеваемости более многообразны и могут включать в себя такие явления, как миграция населения и сменяемость организованных коллективов, нахождение населения на территориях риска и т.д.

Вспышечная заболеваемость анализируется с целью выявления наиболее частых причин их развития на конкретной территории за многолетний период. Для этого можно воспользоваться разработанными классификациями типов эпидемических вспышек в пределах группы инфекций одного механизма передачи. В качестве примера воспользуемся классификацией признаков эпидемического процесса при дизентерии конкретных факторов передачи (табл. 22).

Таблица 22

Дифференциально-диагностические признаки для предварительного определения характера вспышки дизентерии в населенном пункте

Дифференциально-диагностические признаки эпидемического процесса	Характер вспышки	
	пищевой	водный
1	2	3
Кратность прироста заболеваемости в первые 3 дня:		
1 – 1,5 раза	–	–
2 – 5 раз	++	+
более 5 раз	+	++
У 35 % и более больных наблюдается:		
клиническое проявление начала заболеваемости по типу пищевой токсикоинфекции (гастроэнтероколитические формы)	–	++
клиническое течение заболевания:		
тяжелое и среднетяжелое	+	++
легкое	++	+
Одновременный рост заболеваемости на разных территориях с обособленным водоснабжением	–	++
Наибольшая кратность прироста заболеваемости в 1-ю неделю вспышки среди больных в возрасте, лет:		
до 1	–	+
1 – 6	+	++
7 – 14	++	+
15 и старше	++	+
В 5 раз и более чаще заболевают работники:		
пищевых объектов	–	++
коммунального хозяйства	++	–
Дети, посещающие детские учреждения	–	+
Дети, не посещающие детские учреждения	++	++
Удельный вес бактериологически подтвержденных диагнозов, %:		
до 30	+	–
30 – 50	++	+
Более 50	–	++
Отмечается монотипаж выделяемых возбудителей	–/+/	++
Выделяются разные виды, типы шигелл и других возбудителей кишечных инфекций	++	+
Увеличение числа множественных семейных очагов с двумя и более одновременно возникшими заболеваниями	+	++

Окончание табл. 22

1	2	3
Результаты опроса больных о заподозренных факторах передачи:		
пили сырую воду до 60 %	+	+
пили только сырую воду до 30 % и более	–	++
ели молочные продукты до 70 %	+	+
ели только определенные виды молочных или других продуктов 35 % и более	++	–

Примечание. «–» признаки не проявляются; «+» признак выражен слабо; «++» признак выражен отчетливо.

Следует отметить, что подробный анализ может проводиться, как правило, лишь для расследования текущих эпизодических вспышек. При ретроспективном анализе такой подход весьма затруднителен.

В настоящее время нет строгих количественных методов выделения доли заболеваемости, связанной со случайными причинами. Вычленение нерегулярных колебаний в многолетней динамике чаще всего проводится визуальным способом при анализе графических материалов.

После завершения всех этапов анализа многолетней динамики заболеваемости необходимо обосновать заключение об особенностях отдельных форм проявления эпидемического процесса. Выводы должны касаться прежде всего наличия или отсутствия многолетней тенденции, ее вида (прямолинейная, криволинейная), а также скорости изменения динамического ряда. При наличии криволинейной тенденции могут быть выдвинуты гипотезы об эффективности проводимых мероприятий или влияния определенных факторов, которые привели к криволинейной зависимости. Необходимы выводы о наличии многолетней периодичности и основных ее характеристик (начала, окончания и продолжительности циклов), а также предположение о возможных их причинах. Следует выявить наличие нерегулярных колебаний заболеваемости, возникающих в определенные годы, с указанием возможных причин этого явления. Все это в целом будет составлять эпидемиологический диагноз.

3.5. ОЦЕНКА ЭПИДЕМИОЛОГИЧЕСКОЙ СИТУАЦИИ ПО ОБОБЩЕННОМУ КРИТЕРИЮ

При комплексной оценке степени эпидемиологического благополучия или неблагополучия той или иной территории и (или) здоровья

какой-либо группы населения основываются преимущественно на сопоставлении отдельных статистических показателей (заболеваемость, болезненность, смертность, инвалидность и др.). Не отрицая необходимости и важности подобных подходов, следует признать, что для объективного и всестороннего обобщения эпидемиологической ситуации в динамике за несколько лет или на разных территориях целесообразно дополнительно опираться и на обобщенный показатель, без которого бывает весьма проблематично адекватно оценить сходные объекты как «во времени», так и «в пространстве». Понятно, что окончательные выводы проще делать в тех ситуациях, когда значения всех анализируемых показателей состояния здоровья изменяются однонаправлено (либо наблюдается рост показателей, либо их снижение). В этом случае надежный вывод о различиях в эпидемиологической ситуации на разных территориях или на одной в динамике может быть сделан и без интегральной количественной оценки. Однако подобная оценка была бы полезна, так как позволяет устанавливать обобщенную количественную меру различия. При разнонаправленных изменениях изучаемых показателей надежный вывод, помимо логического анализа, поможет сделать единый интегральный показатель.

В настоящем разделе монографии приведено представление обоснования использования единого обобщающего статистического критерия для дополнительной количественной оценки эпидемиологической ситуации и (или) состояния популяционного здоровья [26].

Анализ представленного материала основан на фактических статистических показателях туберкулезной инфекции на примере Республики Саха (Якутия), которая по своим климатическим, социально-экономическим и жилищно-бытовым условиям большинства населения является территорией высокого риска распространения туберкулеза.

Для эпидемиологической характеристики туберкулеза приведен анализ четырех показателей: заболеваемость, болезненность, бациллярность и смертность. Движение заболеваемости оценено за 39 лет (с 1970 по 2008 г.), а по остальным показателям, в том числе и заболеваемости, – за 19 лет (с 1990 по 2008 г.). При оценке статистических показателей использованы стандартные методы эпидемиологического анализа, а именно сравнение средних уровней и темпов прироста, динамика которых рассчитана по выравненным методом наименьших квадратов данным. Кроме этого использован метод оценки достоверности тенденций

динамических рядов и корреляционный анализ с изучением парных и множественных связей (о корреляционном анализе см. гл. 5).

Сущность предлагаемого нами метода заключается в приведении разнородных показателей в единообразную систему с использованием универсального количественного критерия – доля признака (в процентах) в оцениваемой системе параметров и представляет собой отклонение от среднего состояния в относительных величинах.

Первоначально для оценки каждого отдельного взятого в анализ статистического показателя рассчитываем индивидуальный индекс (ИИ) эпидемиологической ситуации по следующей формуле:

$$\text{ИИ} = \frac{a_i}{M},$$

где a_i – значение отдельно i-го годового (месячного или другого) показателя; M – среднее значение совокупности показателей в целом за соответствующий период наблюдения.

В дальнейшем получаем основной показатель – суммарный индекс эпидемиологической ситуации (ИЭС), который вычисляется как средняя величина из всех взятых в разработку индивидуальных индексов

$$\text{ИЭС} = (\text{ИИ}_1 + \text{ИИ}_2 + \ldots. + \text{ИИ}_n) / n.$$

Рассматриваемый подход нашел достаточно широкое использование в эпидемиологических исследованиях при описании внутригодового движения какого-либо показателя (так называемый «индекс сезонности», использование которого в эпидемиологическом анализе будет представлено в разд. 4.1). Однако такое использование индекса основано на оценке лишь одной совокупности данных (например, заболеваемости). В предлагаемом нами варианте все индивидуальные индексы эпидемиологической ситуации приводятся к единому обобщенному показателю.

Следующий этап анализа сводится к ранжированию показателей ИЭС и построению оценочной шкалы. Ранжирование проводится путем упорядочивания индексов по их нарастанию. Полученный непрерывный ряд распределяется на дискретные классы (группы), для каждого из которых рассчитывается диапазон значений либо в пределах доверительного интервала, либо выше или ниже его границ.

Многолетний анализ заболеваемости туберкулезом на терри-

Таблица 23

Эпидемиологические показатели туберкулеза в Республике Саха (Якутия) за 1990 – 2008 гг.

Статистический критерий	Заболеваемость	Болезненность	Бациллярность	Смертность
Интенсивный показатель, $^{0}/_{0000}$	64,5±1,5	242,9±5,4	86,9±1,3	8,7±0,2
Темп прироста, %	2,8	– 1,3	1,0	1,1

тории Республики Саха (Якутия) показал, что исследуемый период (1970 – 2008 г.) характеризовался ее выраженным достоверным снижением (с 138,1 до 69,1 $^{0}/_{0000}$), с ежегодными отрицательными темпами прироста 1,4 %. В то же время установлено, что общая благоприятная тенденция движения заболеваемости имела место за счет первых 20 лет наблюдения (1970 – 1990 г.), когда отрицательные темпы прироста составляли – 4,2 %, с минимальным показателем заболеваемости в 1990 г. (45,4 $^{0}/_{0000}$). В последующий период отмечался стабильный достоверный ($P < 0,0001$) рост заболеваемости туберкулезом с положительным среднегодовым темпом прироста в 2,8 %. Выявленная неблагоприятная тенденция в движении заболеваемости за 1990 – 2008 г. послужила основанием провести комплексную оценку эпидемиологической ситуации в этот период. Основные статистические показатели эпидситуации на территории Республики Саха (Якутия) представлены в табл. 23.

Проведенный анализ показал, что за анализируемый период в Республике Саха (Якутия) имеет место неблагоприятная тенденция в развитии заболеваемости, смертности и бациллярности (рост в 1,5; 1,3 и 1,1 раза соответственно) и лишь для болезненности имеет место отрицательный темп прироста. Интегрирующим показателем эпидемиологической ситуации является ИЭС с ростом в 1,2 раза.

Корреляционный анализ (табл. 24) продемонстрировал достоверную связь между парами таких эпидемиологических показателей, как заболеваемость и бациллярность. Выраженная тенденция ($P < 0,1$) имеет место между заболеваемостью и смертностью.

Правомерность использования обобщенного критерия в оценке эпидемиологической ситуации в целом указывает на наличие корреляционной связи между каждым показателем в отдельности и ИЭС (за

Таблица 24

Коэффициенты корреляции (КК) между отдельными эпидемиологическими показателями

Показатель	Заболеваемость		Болезненность		Бациллярность		Смертность	
	r	*P*	*r*	*P*	*r*	*P*	*r*	*P*
Болезненность	– 0,24	0,3316	-	-	-	-	-	-
Бациллярность	0,66	**0,0023**	– 0,21	0,3976	-	-	-	-
Смертность	0,41	0,0789	– 0,02	0,9308	0,19	0,4346	-	-
ИЭС	0,81	**< 0,0001**	0,17	0,4621	0,62	**0,0044**	0,73	**< 0,0001**

Примечание. Здесь и далее: жирным шрифтом выделены достоверные КК.

исключением показателя болезненности). Еще более наглядно об этом свидетельствуют и множественные корреляционные связи различных вариант стандартных эпидемиологических показателей с ИЭС, которые во всех случаях носили высокодостоверный характер с наличием, как правило, сильной связи (табл. 25).

Дальнейший этап исследований связан с ранжированием многолетних показателей ИЭС и расчетом доверительного интервала для анализируемого периода. Показатели суммарного индекса эпидемиологической ситуации, входящие в пределы доверительного интервала, соответствуют «нормативному» уровню эпидемиологической обстановки на данной территории за рассматриваемый период времени. Все значения этого показателя, лежащие ниже или выше границ доверительного интервала, свидетельствуют соответственно о низком (благоприятном) или высоком (неблагоприятном) уровне этой обстановки.

Пример оценки эпидемиологической ситуации с использованием предложенного нами индекса представлен в табл. 26. Помимо приве-

Таблица 25

Множественные корреляционные связи между ИЭС и различными вариантами стандартных эпидемиологических показателей

Показатель	КК	*P*
Заболеваемость – болезненность	0,90	< 0,0001
Заболеваемость – бациллярность	0,82	0,0001
Заболеваемость – смертность	0,92	< 0,0001
Болезненность – бациллярность	0,69	0,0050
Болезненность – смертность	0,75	0,0012
Бациллярность – смертность	0,88	< 0,0001

Таблица 26

Оценка эпидемиологической ситуации по туберкулезу в Республике Саха (Якутия) по ИЭС с использованием доверительного интервала

Ранжированный ИЭС	Год	Шкала ранжирования	Уровень ИЭС
0,87	1990	< 0,97	Низкий
0,89	1994	Показатели ИЭС ниже границ	
0,91	1992	доверительного интервала	
0,92	1991		
0,94	2007		
0,96	1995		
0,96	1993		
0,97	1998	0,97 – 1,03	Нормативный
0,98	2006	Показатели ИЭС в пределах	
0,98	1997	границ доверительного интервала	
1,00	1996		
1,03	2004		
1,04	2008	> 1,03	Высокий
1,04	2005	Показатели ИЭС выше границ	
1,06	2002	доверительного интервала	
1,06	2003		
1,09	2001		
1,11	1999		
1,12	2000		

денного подхода с ранжированием ИЭС анализ эпидемиологической ситуации может быть представлен и в виде хронологического динамического ряда (рис. 4).

Рассмотренный нами пример оценки эпидемиологической ситуации по туберкулезу в Республике Саха (Якутия) показал, что, несмотря на разнонаправленную динамику взятых в анализ отдельных статистических показателей, получена обобщенная количественная оценка эпидемиологической ситуации, свидетельствующая о неблагоприятном развитии туберкулезной инфекции в этом крупнейшем регионе страны в последнее десятилетие. Предложенный подход позволил выделить внутри исследуемого периода отдельные годы с относительно низкими или высокими показателями. Используя вышеописанную методику, можно также сделать вывод о том, что в Республике Саха (Якутия) с 1990 до 2000 г. наблюдалось последовательное ухудшение эпидемиологической ситуации по туберкулезу, тогда как период 2000–2008 гг. характеризовался ее улучшением. Причем динамика ИЭС по годам в оба этих периода имела высокодостоверный характер.

Рис. 4. Многолетняя динамика индекса эпидемиологической ситуации по туберкулезу в Республике Саха (Якутия).

Таким образом, использование индекса эпидемиологической ситуации позволяет представить интегральную количественную оценку развития инфекционной заболеваемости на основании отдельных эпидемиологических показателей, причем их количество может быть неограниченным. Данный подход может быть применен для целей эпидемиологического надзора при любой инфекционной патологии и позволяет производить оценку полученных данных в зависимости от принятых условий по отдельным временным периодам (месяцы, годы и т.д.). Все это, в свою очередь, может служить методологической основой в комплексной оценке здоровья населения на любой заданной территории за конкретный промежуток времени.

Глава 4

АНАЛИЗ ВНУТРИГОДОВОЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ

Анализ динамики заболеваемости, т.е. изменений ее интенсивности во времени, относится к числу основных направлений эпидемиологической диагностики. Он включает следующие этапы исследования:

– установление контрольного уровня заболеваемости, сроков начала и окончания сезонных подъемов, сезонных эпидемических надбавок;

– изучение причин вспышечной заболеваемости и др.

Цель анализа внутригодовой динамики заболеваемости – установить причинно-следственные связи в эпидемическом процессе и оценить значимости возможных причин. В зависимости от назначения анализа и исследуемой инфекции при изучении динамики заболеваемости могут быть приняты различные временные интервалы: годы, месяцы, декады, недели, дни и даже часы. Колебания эпидемического процесса обусловлены изменением влияния различного рода факторов, связанных с заболеванием, или условий, при которых действуют эти факторы. Поиск причинных факторов (условий) следует проводить в период, предшествовавший росту или снижению заболеваемости. Следовательно, определение колебаний инфекций во временном аспекте в течение года является первым шагом анализа причин заболеваемости. Собственно в этом и состоит диагностическая ценность временной характеристики эпидемического процесса.

Уровень и динамика заболеваемости связаны со всем многообразием конкретных для каждой территории биологических, социальных и климатических факторов или их сочетанием. Многие из указанных факторов развиваются и действуют на эпидемический процесс в определенные периоды года, поэтому помесячный анализ динамики заболеваемости является непременным требованием изучения внутригодового движения инфекции.

На начальном этапе устанавливаются закономерности помесячного движения заболеваемости всего населения в целом на определенной территории, а в дальнейшем анализируются особенности среди отдельных групп населения, объединенных различными факторами риска (на отдельных микроучастках, среди конкретных возрастных, социальных, профессиональных и других групп), и сопоставляются с конкретными факторами, действующими в определенных условиях жизни людей.

Выявление закономерностей и особенностей помесячной динамики заболеваемости – не цель, а средство для установления причин и условий, формирующих помесячные колебания.

Таким образом, для установления причин, обусловливающих внутригодовую динамику, необходимо получить всесторонние (демографическую, гигиеническую, производственную и т.д.) характеристики населения, среди которого изучается динамика заболеваемости. Кроме того, необходимо исследование формы проявления эпидемического процесса в течение года. Все исходные материалы о заболеваниях и санитарно-эпидемиологическая характеристика должны быть полными и достоверными.

Колебания уровня внутригодовой заболеваемости, как и в многолетней динамике, обусловлены влиянием следующих групп причин: постоянно действующими, периодически активизирующимися, нерегулярными. Указанные причины определяют соответственно три основные формы проявления эпидемического процесса во внутригодовой динамике: 1) круглогодичную заболеваемость; 2) сезонный подъем или просто сезонность; 3) эпизодические вспышки.

4.1. СТАТИСТИЧЕСКАЯ ОЦЕНКА СЕЗОННЫХ ПОКАЗАТЕЛЕЙ ЗАБОЛЕВАЕМОСТИ

Анализ внутригодовой динамики заболеваемости рекомендуется начинать с построения кривой, которую часто называют типовой кривой сезонного распределения заболеваемости. Название не совсем удачно, так как кривая отражает распределение не только в сезонный период заболеваемости, но и в другие эпидемические периоды. Построение типовой кривой позволяет исключить случайные колебания и выявить закономерности помесячной (или более дробной) динамики

заболеваемости на конкретной территории, причем достаточно надежные выводы могут быть получены при использовании данных за 4–6 лет и более. Сравнение динамики, полученной для отдельного года, с типовой кривой для данной местности за ряд лет выявит особенности распределения заболеваемости за анализируемый период. Наиболее целесообразно построение типовой кривой при отсутствии в многолетней динамике выраженных циклов и тенденции к росту или снижению заболеваемости. При их наличии необходимо проводить анализ сезонности для двух типовых кривых, из которых одна представляет заболеваемость в годы подъема, а другая – в годы спада. Это позволяет выявить особенности динамики при резко отличающихся уровнях заболеваемости.

Перед расчетом типовой кривой нужно составить таблицу интенсивных показателей помесячной (или других временных интервалов) заболеваемости. В нее не следует вносить значения, связанные со вспышечной заболеваемостью, так как они могут исказить средние результаты и типичную картину годовой динамики. Однако в эпидемиологической литературе нет однозначного подхода к исключению вспышечной заболеваемости. Одни авторы предлагают отбрасывать показатели вспышечной заболеваемости лишь с достоверно установленными причинами, другие – отрицают возможность вычленения показателей такого рода. Например, Ю.П. Рыкушин и М.С. Исаева [25] обосновывают невозможность исключения вспышечной заболеваемости из внутригодовой динамики по следующим причинам. Они считают, что вспышечная заболеваемость является одной из форм проявления эпидемического процесса и исключение вспышек будет искажать его закономерности. Частота возникновения вспышек имеет прямую связь с интенсивностью эпидемического процесса, поэтому говорить о случайности вспышечной заболеваемости едва ли правомерно. Очевидно, следует исключать только очень крупные вспышки, составляющие 50% и более годовой заболеваемости, которые могут сильно изменить средние показатели. Выделение вспышечной заболеваемости статистическими методами может быть простой «подгонкой» эпидемического процесса под желаемый уровень. Кроме того, по их мнению, исключение всех вспышек невозможно, поскольку большая или меньшая часть из них остается нераспознанной, поэтому показатель «спорадической» заболеваемости становится зависи-

мым от полноты выявления и регистрации вспышек. Наконец, само понятие вспышки субъективно, поскольку нет ее количественной характеристики.

С приведенным положением нельзя не согласиться. Вместе с тем авторы сами себе противоречат, предлагая все же исключать вспышки, составляющие более половины годовой заболеваемости. Отсутствуют четкие рекомендации, как следует поступать, если вспышка составляет, например, 49 % от годовой заболеваемости. Для решения этого вопроса нужен объективный критерий, в качестве которого может выступить статистическая проверка данных на принадлежность к генеральной совокупности. В медицинских исследованиях (в том числе и в эпидемиологии) заключение о принадлежности тех или иных вариант к рассматриваемой совокупности чаще всего делается с 95 %-м уровнем достоверности. В ряде случаев, когда требуется особая надежность выводов, принимается 99 %-й уровень достоверности. Резюмируя вышеизложенное, считаем, что можно рекомендовать исключение из дальнейшего анализа таких показателей вспышечной заболеваемости, которые выходят за грани 99%-го уровня достоверности.

Расчет типовой кривой проводиться следующим образом. Вначале суммируют данные (табл. 27) для каждого месяца отдельно за все годы рассматриваемого периода (например, для января 5,8+12,3+5,1+…6,68 = = 49,4; для февраля 5,3+ 7,3+3,8+…4,4 = 34,2 и т.д.). Затем вычисляют среднемесячные показатели (январь – 49,4:8 = 6,2; февраль – 34,2:8 = 4,3 и т.д.). Они и дают представление о типичном внутригодовом распределении заболеваемости для данной инфекции за определенный период времени на конкретной территории. Целесообразно (особенно при массовых инфекциях) строить типовые кривые не только по данным заболеваемости всего населения в целом, но и отдельно для возрастных и профессиональных групп.

Следующим этапом анализа годовой динамики заболеваемости является выяснение срока начала и окончания сезонного подъема. Для этого суммируют все месячные средние (6,2+4,3+3,7…+6,1=121,4) и делят полученную сумму на 12 (121,4:12=10,1). Полученное значение (10,1) представляет среднемесячный показатель относительно всех лет наблюдения, оно принимается за 100 % и считается границей сезонной волны. Далее методом пропорции рассчитывают, ка-

Внутригодовая динамика заболеваемости дизентерией Зонне в городе

Год	Месяц					
	I	II	III	IV	V	VI
1994	5,8	5,5	4,9	5,2	3,7	3,8
1995	12,3	7,3	7,2	4,2	4,6	4,5
1996	5,0	3,8	4,0	2,7	2,8	4,5
1997	3,4	3,1	3,5	2,1	1,5	4,0
1998	3,9	2,7	3,1	0,8	2,7	2,8
1999	4,6	2,4	1,3	3,4	3,2	4,1
2000	7,6	5,0	2,0	5,4	3,0	3,8
2001	6,9	4,4	3,2	3,6	3,4	5,0
Итого …	49,5	34,2	29,2	27,4	24,9	32,5
Уд. вес, %	5,1	3,5	3,0	2,8	2,6	3,3
Типовая кривая	6,2	4,3	3,7	3,4	3,1	4,1
Индекс сезонности	61,1	42,2	36,0	33,8	30,7	40,1

кой процент составляет каждая из 12-месячных средних

январь: 10,1 – 100 %, а 6,2 – x,

тогда $x = \dfrac{6{,}2 \cdot 100\,\%}{10{,}1} = 61{,}2\,\%$;

февраль: 10,1 – 100 %, а 4,3 – x,

тогда $x = \dfrac{4{,}3 \cdot 100\,\%}{10{,}1} = 42{,}6\,\%$; и т.д.

В результате этих вычислений находим количественный индекс сезонности [11]. Выполненные расчеты можно выразить следующей формулой:

$$\text{индекс сезонности} = \frac{\text{средняя многолетняя за месяц}}{\text{средняя многолетняя}} \cdot 100\,\%.$$

В результате находим количественный индекс сезонности [11].

Если индекс сезонности превышает 100 % (см. табл. 27), это свидетельствует о периоде сезонного подъема заболеваемости (сезонность выражена).

Определить месяцы сезонного подъема заболеваемости можно и другим способом. Для этого сумму показателей за все месяцы при-

Таблица 27

Н. за 1994 – 2001 гг. (на 100 тыс. совокупного населения)

Месяц						Итого за год	Среднемесячная
VII	VIII	IX	X	XI	XII		
8,8	27,6	36,4	21,4	11,4	8,8	143,3	11,9
8,0	38,2	46,2	25,4	11,6	6,6	176,1	14,7
13,4	28,0	20,8	12,0	9,0	3,6	109,6	9,1
12,2	14,4	23,0	13,3	7,9	3,4	91,8	7,7
11,0	22,7	25,2	15,6	11,9	3,8	106,2	8,9
3,7	11,0	17,9	11,5	6,1	5,8	75,0	6,3
7,7	11,6	39,4	26,4	14,4	8,2	134,5	11,2
15,0	24,8	26,4	18,4	16,4	8,2	135,7	11,3
79,8	178,3	235,3	144,0	88,7	48,4	972,2	81,1
8,2	18,3	24,2	14,8	9,1	5,0	100	8,3
10,0	22,3	29,4	18,0	11,1	6,1	121,5	10,1
98,5	220,1	290,4	177,7	109,5	59,7	1200,0	100,0

нимают за 100 %, затем рассчитывают сумму показателей за каждый месяц отдельно (в процентах) (см. табл. 27). Поскольку на каждый из 12 мес приходится по 8,33 % (100 %: 12 = 8,33 %), следовательно, в сезонный подъем включаются лишь те месяцы закономерного роста заболеваемости, в которых среднемесячные показатели превышают 8,33 %. Как видно, этот метод выделения сезонного подъема мало отличается от метода выделения по индексам сезонности. Однако для сравнения данных по сезонности на разных территориях и в разных группах населения все же удобнее принимать общую среднюю за 100 %, чем равную 8,33 %.

Если данные по заболеваемости вызывают сомнение, проверим их на принадлежность к данной выборке. Для этого используем формулу

$$\frac{x_n - x_{n-1}}{x_n - x_1} = \frac{12,3 - 7,6}{12,3 - 3,4} = 0,528.$$

Полученное значение сравним с табличным (см. приложение, табл. 1). Для $n = 8$ и $P = 0,05$ в таблице приведено 0,468, для $P = 0,01$ – 0,590. Оказалось, что рассматриваемый показатель при 99 %-м уровне значимости не может быть исключен из дальнейших расчетов. Таким образом, типовая кривая рассчитывается с учетом всех значений, приведенных в табл. 27.

Согласно индексам сезонности начало сезонного подъема дизентерии Зонне за многолетний период в городе Н. приходится на август, а окончание – на ноябрь. Длительность сезонного периода составляет 4 мес, с максимальной заболеваемостью в сентябре. Аналогичная картина получается при анализе месячных экстенсивных показателей.

Приведенные выше данные позволяют дать усредненную оценку начала сезонного подъема, его окончания и длительности за многолетний период. Для инфекций, имеющих массовый характер (дизентерия, гепатит А, грипп и др.), аналогичный подход можно применять и за каждый отдельно взятый год. В этом случае показатели заболеваемости каждого месяца делятся на среднемесячное значение для этого же года.

В качестве примера подсчитаем индекс сезонности для января 1994 г. (исходные данные см. в табл. 27), который оказался равным 48,6 %. Индекс сезонности для февраля и марта 1994 г. будет равен соответственно 46,5 и 41,1 % и т. д. Аналогичным образом поступаем и при вычислении месячных экстенсивных показателей (определяем частные от деления на годовой показатель заболеваемости).

Рассмотренные выше показатели сезонности достаточно широко применяются в эпидемиологическом анализе. Например, при расследовании причин интенсивного подъема заболеваемости вирусным гепатитом В на одной из территорий нами доказана гипердиагностика этого заболевания. Основанием для этого вывода послужило сравнение внутригодовых распределений заболеваемости гепатитами в 1981, 1982 и 1983 – 1985 гг. Указанные ряды также сравнивались с помесячным распределением количества заболевших гепатитом А (табл. 28). Из приведенных данных ясно, что в 1981, 1982 гг. для гепатита В отсутствует какая-либо закономерность в сезонном распределении заболеваемости, что и является характерной особенностью данной инфекции. Однако в 1983 – 1985 гг. годовое распределение заболеваемости гепатитом В практически полностью повторило динамику заболеваемости гепатитом А, для которого сезонность является характерной чертой.

Определив начало, окончание и длительность сезонного подъема, можно рассчитать количество заболевших (в процентах) за указанный период. В случае заболеваемости дизентерией Зонне (см. табл. 27) с августа по ноябрь оно составило 66,4 %.

Таблица 28

Помесячное распределение заболеваемости вирусными гепатитами, %

Нозологическая форма	Годы	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Гепатит В	1981 – 1982	9,4	9,3	8,5	17,2	5,9	10,1	7,8	6,5	4,5	8,5	5,6	8,8
	1983 – 1985	6,1	5,9	7,7	7,7	8,3	7,9	9,1	7,8	10,9	11,2	8,4	9,6
Гепатит А	1981 – 1985	8,6	7,9	5,6	7,3	6,7	6,4	6,6	8,0	10,6	12,5	11,2	7,1

Можно также определить удельный вес заболеваемости, вызванной влиянием сезонных факторов, по формуле [37]

$$C = \frac{B - \dfrac{(A - B) \cdot M}{12 - M}}{A} \cdot 100\ \%,$$

где A – число заболевших за год; B – то же за период сезонного подъема; M – число месяцев сезонного подъема.

Рассчитываем указанный показатель (C), используя данные табл. 28:

$$A = 121\ {}^{0}/_{0000};\ B = 22{,}3 + 29{,}3 + 18{,}0 + 11{,}1 = 80{,}7\ {}^{0}/_{0000};\ M = 4.$$

Таким образом, в городе Н. по среднемноголетним данным на сезонный период заболеваемости дизентерией Зонне приходится 66,4 % заболевших. Удельный вес заболевших за счет влияния сезонных факторов составляет 49,7 %. Подобные расчеты можно проводить и для каждого года в отдельности.

Рассмотренные выше методы выделения границ сезонного подъема заболеваемости имеют существенный недостаток – они не обеспечивают точных количественных критериев оценки сезонного периода.

Более точную оценку начала и конца сезонного подъема получают с помощью линейно-круговых диаграмм, причем, вычисляют сроки сезонного подъема как за среднемноголетний период, так и за каждый год в отдельности. В первом случае для построения диаграмм используют среднемноголетние данные, а во втором – данные за год. Построение этих диаграмм проводится с учетом среднемесячного или круглогодичного уровней заболеваемости. В настоящем разделе рассмотрен первый из указанных подходов (анализ годовой

Таблица 29

Заболеваемость дизентерией Зонне детского населения в городе Н. по данным за год (в условных показателях)

Месяц													Всего за год	Среднемесячная (радиус окружности *n*/12)
I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII		
84,0	72,2	78,1	28,7	55,4	85,0	99,8	341,0	624,0	439,0	210,0	127,0	2172,0	181,2
6,8	14,4	15,6	5,7	11,1	17,0	20,0	68,2	125,0	88,0	42,0	25,5	439,3	36,6

динамики заболеваемости с учетом круглогодичного уровня описан в следующем разделе).

Определим начало, окончание и длительность сезонного подъема заболеваемости дизентерией Зонне детского населения в городе Н. по данным за один год (табл. 29).

Линейно-круговая диаграмма строится на круге, радиус которого равен среднемесячной заболеваемости ($R = n/12$), где n – заболеваемость, зарегистрированная в течение года. Условно принимается, что один случай заболевания равен 1 мм. Так, в приведенной табл. 23 среднемесячный показатель 176,6 $^{0}/_{0000}$, следовательно, радиус линейно-круговой диаграммы должен быть равным 176,6 мм. Уменьшаем масштаб диаграммы до оптимальных размеров. При этом уменьшается не только радиус окружности (среднемесячная заболеваемость), но соответственно показатели заболеваемости всех месяцев года (см. табл. 29). В нашем случае все показатели уменьшены в 5 раз. Таким образом, 1 мм соответствует пяти случаям дизентерии. Если заболеваемость слишком низкая, то для удобства работы масштаб можно увеличить.

Итак, вычерчивается окружность радиусом, соответствующим среднемесячной заболеваемости. От центра круга через 30° наносятся 12 радиусов – месяцев, которым последовательно по часовой стрелке присваивается номер от 1 (январь) до 12 (декабрь). Далее, на радиусе (месяце) от центра круга откладывается соответствующая ему заболеваемость. В месяц начавшегося сезонного подъема нанесенные отметки будут выходить за пределы круга, так как они превышают среднемесячную заболеваемость. В этом случае длину радиуса следует продолжить. Крайние отметки нанесенных уровней заболеваемости последовательно соединяют (рис. 5).

При типичном ходе эпидемического процесса в течение года будет

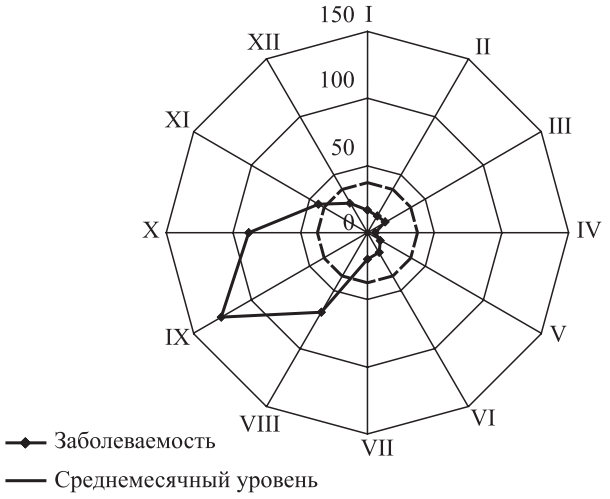

Рис. 5. Графическое определение сроков начала и окончания сезонного подъема заболеваемости дизентерией Зонне детского населения в г. Н. (масштаб 1:5).

два пересечения окружности ломаной линией. Первая точка пересечения соответствует началу, а вторая – окончанию сезонного подъема. Даты определяют по размеру угла от нулевой отметки на радиус, за который принимается январь, до точки пересечения окружности линией (см. рис. 5). Угол измеряется с помощью обычного транспортира. Центр его совмещается с центром окружности по нулевому градусу. В нашем примере угол от нулевого радиуса с точкой, указывающей на начало сезонного подъема, составил 191°, до радиуса с точкой окончания сезонного подъема – 308°.

К последующему этапу работы относится перевод градусов в дни, принимая 1° равным 1 дню. Каждое из полученных чисел (в нашем примере 191 и 308°), необходимо разделить на 30 (по числу градусов, приходящихся на 1 мес). Целое число плюс единица будет соответствовать номеру месяца, а количество градусов в остатке указывает на конкретный день начала или окончания сезонного подъема дизентерии. Из приведенного примера начало сезонного подъема приходится на 11 июля, а окончание – на 8 ноября. Длительность сезонного подъема составила 120 дней, или 3 мес 27 дней. Основные показатели сезонных подъемов заболеваемости оформляются в виде таблицы, примерная форма которой приведена ниже (табл. 30).

Описанный выше метод определения сроков начала и окончания се-

Таблица 30

Основные показатели сезонных подъемов заболеваемости (приводится нозологическая форма) населения города

Год	Дата		Длительность	Месяц максимальной заболеваемости	Удельный вес заболеваемости, вызванной влиянием сезонных факторов
	начала	конца			
1 2 3 ...					
Многолетний показатель					

зонных подъемов не является абсолютно точным. Однако в эпидемиологической диагностике, где широко используется этот метод, основное значение имеет не конкретная дата события, а ее сравнение по годам наблюдений, различным территориям, различным группам населения и т.д.

Линейно-круговые диаграммы применены нами и для выявления роли овощей в сезонных подъемах заболеваемости дизентерией [4]. Мы сравнили сроки сезонного подъема заболеваемости с датами массового поступления овощей в продажу (табл. 31).

Даты начала массовой реализации овощей определились также, как и сроки начала заболеваемости дизентерией, только вместо заболеваемости анализировали сведения о количестве овощей, поступивших в продажу. Во все годы наблюдений, даты начала массовой реализации овощей на 1 – 2 нед опережали сроки начала сезонного подъема заболеваемости дизентерией (см. табл. 31). Причем уровень достоверности влияния реализации овощей на сезонный подъем дизентерии составляет 99 % по критерию знаков при $n = 8$ (по 4 года для каждого города).

Нашими исследованиями также показано, что использование линей-

Таблица 31

Сроки начала сезонных подъемов заболеваемости дизентерией и массовой продажи овощей в городах Иркутской области

Год	Иркутск		Железногорск	
	Начало			
	массовой продажи овощей	сезонного подъема заболеваемости	массовой продажи овощей	сезонного подъема заболеваемости
1982	6.07	20.07	10.08	20.08
1983	17.07	29.07	08.07	15.07
1984	19.07	28.07	18.08	24.08
1985	05.07	19.07	21.07	30.07

но-круговых диаграмм с учетом среднемесячного уровня заболеваемости дизентерией дает возможность выявить последовательность вовлечения в сезонный подъем различных возрастных и профессиональных групп населения [30]. Кроме того, оно позволяет получить важную информацию о типе эпидемического процесса дизентерии на той или иной территории. Например, если за ряд лет в заболеваемость последовательно вовлекается сначала взрослое население, затем дети школьного возраста и последними – дети младших возрастных групп, то можно предполагать, что на данной территории водный путь передачи занимает ведущее место. Если в эпидемический процесс первыми вступают пищевики, а за ними вовлекается все остальное население, это свидетельствует о ведущей роли пищевого пути передачи. Однако эти предположения (о ведущем пути передачи) требуют дополнительного эпидемиологического подтверждения (опрос, лабораторные методы и др.).

Необходимо отметить, что последовательное вовлечение в сезонный подъем различных групп населения удается получить лишь при оценке сезонности по среднемесячному уровню заболеваемости. При круглогодичном уровне эта закономерность четко не проявляется, что, по-видимому, связано с малой интенсивностью факторов, определяющих эту границу.

Определять сроки начала и окончания сезонных подъемов можно с помощью линейной диаграммы. Например, на рис. 6 приведена типовая кривая годовой динамики заболеваемости населения города Н. дизентерией Зонне в 1994 – 2001 гг. Из точек пересечения среднемесячного уровня заболеваемости ($10{,}1^0/_{0000}$) с типовой кривой опущены перпендикуляры на горизонтальную часть графика. С учетом выбранного масштаба 1 мм горизонтальной оси графика равен 3 дням, типовое начало и окончание сезонного подъема дизентерии Зонне в городе Н. приходится соответственно на 01.07 и 06.11 (рис. 6). Для более точной работы рекомендуется строить линейные диаграммы на бумаге с миллиметровыми делениями. Аналогичные расчеты можно проводить и для каждого года в отдельности, используя в этом случае среднемесячный уровень заболеваемости этого же года.

Для большей наглядности графиков на линейной диаграмме размещать материалы следует таким образом, чтобы размер вертикальной части графика относился к горизонтальной примерно как 1: 1,5. Следует также отметить, что выполнение указанных расчетов вручную не

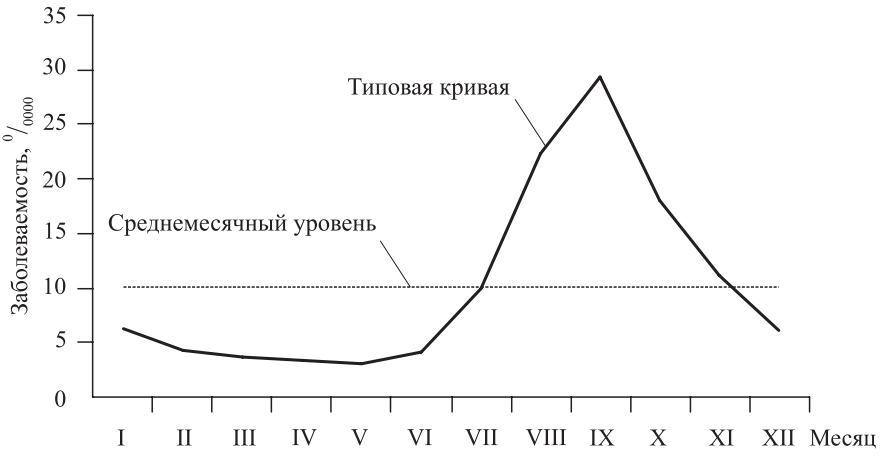

Рис. 6. Типовая кривая годовой динамики заболеваемости дизентерией Зонне в городе Н. в период 1994–2001 гг. (по данным табл. 27).

исключает допущения ошибок, а даты (дни) сезонных подъемов, определенные с помощью линейных или линейно-круговых диаграмм, не соответствуют друг другу.

Все это ограничивает применение указанных методов для углубленного эпидемиологического анализа. В качестве альтернативы может быть предложен расчетный прием определения основных показателей сезонности, который отвечает следующим требованиям:

– исключение графических приемов определения;

– точность и воспроизводимость результатов;

– простота расчетов и легкость их алгоритмизации, что важно при обработке эпидемиологического материала с помощью электронных таблиц.

Нами разработан расчетный метод определения дат начала и окончания сезонных подъемов, основанный на известном положении о пропорциональности сторон подобных треугольников [2]. Для примера воспользуемся данными заболеваемости дизентерией Зонне в городе Н. за 2001 г., а также среднемноголетними показателями за 1994–2001 гг. (см. табл. 27).

Для определения даты начала сезонного подъема рассмотрим динамику заболеваемости дизентерией Зонне в 2001 г. и ее среднемесячный уровень (рис. 7). Для этого впишем в годовую кривую заболеваемости прямоугольный треугольник *ADE*. Катет *AE* соответствует числу дней в месяце начала сезонного подъема, а катет *DE* – разности между показа-

телями заболеваемости месяца, следующего за датой начала сезонного подъема, и месяца, ему предшествующего. Внутри рассматриваемого треугольника образуем подобный ему треугольник *ABC*. Подобность треугольников вытекает из того, что угол *A* для рассматриваемых треугольников является общим, а углы *C* и *E* равны 90°. В треугольнике *ABC* катет *BC* равен разности между среднемесячным уровнем заболеваемости и ее показателем в месяц начала сезонного подъема. Катет *AC* равен числу дней, прошедших от начала месяца до дня начала сезонного подъема, который нам необходимо определить. Исходя из подобия треугольников *ADE* и *ABC*, получаем

$$\frac{AC}{AE} = \frac{BC}{DE},$$

$$AC = \frac{AE \cdot BC}{DE} = \frac{30\,(11{,}3 - 5)}{15 - 5} = 18{,}9.$$

Округляя полученное значение до целого числа, определяем, что дата начала сезонного подъема заболеваемости дизентерией Зонне в городе Н. в 2001 г. приходится на 19.06.

Для нахождения даты окончания сезонного подъема вписываем в годовую кривую заболеваемости треугольник $A'D'E'$ и подобный ему треугольник $B'D'K'$, в котором катет *KB* соответствует числу дней,

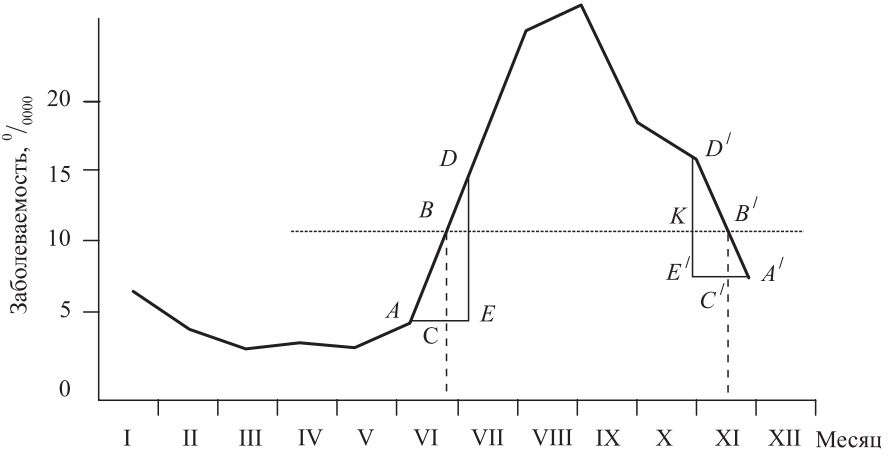

Рис. 7. Кривая годовой заболеваемости дизентерией Зонне и подобные треугольники для определения дат начала и окончания сезонного подъема.

прошедших от начала месяца до дня окончания сезонного подъема. Дата окончания определяется из пропорции

$$\frac{KB'}{E'A'} = \frac{KD'}{E'D'},$$

где KB' – дата окончания сезонного подъема; $E'A'$ – число дней в месяце окончания сезонного подъема; KD' – разность между уровнем заболеваемости в месяц, предшествующий окончанию сезонного подъема и ее среднемесячным уровнем; $E'D'$ – разность между показателями заболеваемости месяца, предшествующего окончанию сезонного подъема, и месяца после его окончания.

$$KB' = \frac{E'A' \cdot KD'}{E'D'} = \frac{30\,(16{,}4 - 11{,}3)}{16{,}4 - 8{,}2} = 18{,}7 \approx 19.$$

Как видно, дата окончания сезонного подъема приходится на 19.11.

При определении рассматриваемых показателей сезонности по проекции перпендикуляра, опущенного из точек пересечения среднемесячного уровня заболеваемости с ее годовой кривой на горизонтальную ось графика, получены те же данные (см. рис. 7). Однако при таком подходе необходима кропотливая работа, которая не избавляет от ошибок, особенно при малом масштабе по оси абсцисс. Расчет этих показателей с использованием предложенного нами способа не требует построения диаграмм, а использование рисунка было необходимо лишь для наглядности. Для более удобного использования предлагаемого метода можно представить расчет дат показателей сезонности в буквенном выражении. В этом случае формула для определения начала сезонного подъема будет выглядеть следующим образом:

$$Д_{н} = \frac{K\,(C - A)}{B - A},$$

где $Д_{н}$ – дата начала сезонного подъема; K – число дней в месяце начала сезонного подъема (30 или 31); C – среднемесячный уровень заболеваемости; A – показатель заболеваемости в месяц начала сезонного подъема; B – показатель заболеваемости в месяц, следующий за датой начала сезонного подъема.

Формула для определения даты окончания этого показателя примет следующий вид:

$$Д_о = \frac{K(A-C)}{A-B},$$

где $Д_о$ – дата окончания сезонного подъема; K – число дней в месяце окончания сезонного подъема (30 или 31); C – среднемесячный уровень заболеваемости; B – показатель заболеваемости в месяц, следующий за датой окончания сезонного подъема.

Для примера рассчитаем даты начала и окончания сезонного подъема заболеваемости дизентерией Зонне за многолетний период, используя формулу

$$Д_н = \frac{K(C-A)}{B-A} = \frac{30\,(10{,}1-10{,}0)}{22{,}3-10{,}0} = 0{,}25\,.$$

Следовательно, начало сезонного подъема приходится на 01.07. Дата окончания сезонного подъема приходится на 06.11:

$$Д_о = \frac{K(A-C)}{A-B} = \frac{30(11{,}1-10{,}1)}{11{,}1-6{,}1} = 6{,}0\,.$$

Используя предложенный принцип расчета, можно определить основные параметры сезонности инфекционной заболеваемости, принимая за границу отсчета любые другие уровни, например круглогодичный. В этом случае в вышеприведенных формулах показатель C будет отражать круглогодичную заболеваемость.

Естественно, что даты (дни) начала и окончания сезонных подъемов на линейно-круговых и линейных диаграммах, а также при расчетном методе не будут точно соответствовать друг другу. Однако повторим еще раз, что в эпидемиологическом анализе (эпидемиологической диагностике) не так важны точные даты, больший интерес представляет сравнение этих сроков в разные годы, на разных территориях и среди различных групп населения. Для сопоставления результатов следует придерживаться лишь одного выбранного метода определения сроков начала и окончания сезонного подъема.

В заключение необходимо остановиться на выборе вида средней величины для построения типовой кривой. Для этой цели в эпидемиологи-

ческой практике широко используются средняя арифметическая и (или) медиана (о вычислении и оценке средних величин см. гл. 1).

Каждая из указанных средних имеет свои преимущества и недостатки. В связи с этим выбор средней в каждом конкретном случае проводится исследователем в зависимости от целей и задач эпидемиологического анализа, а также от объема исходных данных.

Например медиана, как и любой другой непараметрический показатель, значительно меньше зависит от вида распределения по сравнению со средней арифметической. Этот показатель пригоден также для оценки такой совокупности, в которой крайние варианты представлены полуколичественными данными (т.е. в том случае, когда заболеваемость имеет большее или меньшее значение по сравнению с предшествующим показателем). Расчет и особенно нахождение доверительных интервалов медианы проводится значительно проще, чем для средней арифметической. Одним из существенных отличий и преимуществ медианы от средней арифметической является то, что на нее не оказывают влияние крайние значения вариант вариационного ряда. Следовательно, при подготовке исходных данных по заболеваемости при условии, что динамический ряд достаточно велик, можно не проводить работу по выявлению выскакивающих вариант. Однако необходимо иметь в виду, что этот показатель приводит к завышению границ доверительного интервала по сравнению со средней арифметической. Так, при числе вариант 8 и 95%-м уровне достоверности пограничные значения доверительного интервала медианы (см. приложение, табл. 2) охватывают всю совокупность наблюдений. При уровне достоверности 99 % такой максимальный размах имеет место в числе наблюдений, достигающих 11 вариант. Минимальное число вариант для нахождения границ доверительного интервала по медиане – 6. Следовательно, при использовании данных по заболеваемости за 4 – 5 лет нет возможности дать оценку доверительному интервалу. Из ограничений в использовании медианы следует также указать на тот факт, что при анализе годовой динамики заболеваемости по отдельным возрастным или профессиональным группам населения, а также при инфекциях, не имеющих массового распространения, могут иметь место случаи отсутствия заболеваний в различные месяцы изучаемых лет. В этом случае значения медианы могут оказаться равными нулю. Естественно, что такого недостатка лишена средняя арифметическая.

4.2. АНАЛИЗ ГОДОВОЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ С УЧЕТОМ ЕЕ КРУГЛОГОДИЧНОГО УРОВНЯ

Наиболее обоснованным подходом к определению сезонности инфекционных заболеваний являются методы, основанные на вычислении доверительного интервала в отношении уровня круглогодичной заболеваемости [7, 17]. Среди них наиболее статистически оправдан подход А.А. Дегтярева, А.П. Ходырева [18]. Однако его применение ограничено условием многолетней стабильности эпидпроцесса, и поэтому он не может быть использован при значительных колебаниях годовой заболеваемости, а также для изучения сезонности отдельных лет. Кроме того, объем вычислений по указанному методу довольно сложен, что также затрудняет его применение в низовых звеньях эпидемиологической службы.

По нашему мнению, достаточно простой и в то же время логически обоснованной является модификация метода В.Д. Белякова [7], предложенная Ю.П. Рыкушиным и др. [25]. Принцип этого метода состоит в следующем: определяются месяцы круглогодичной заболеваемости, в которых отсутствуют статистически значимые различия месячных показателей по сравнению со средней из 2 мес с наиболее низкой заболеваемостью. Авторы считают, что средняя из 2 мес уменьшит влияние возможной случайности минимального показателя 1 мес.

Для выявления закономерностей в годовой динамике эпидемического процесса необходимо опираться на длительность анализируемого периода, раздельный анализ лет с низкой и высокой заболеваемостью, выбор вида средних и т.д.

По предложению авторов [25], анализ сезонности должен проводиться не по календарному, а по эпидемиологическому году, представляющему собой законченный годовой цикл изменений помесячной динамики заболеваемости. Начало и окончание эпидемиологического года следует определять по многолетним среднемесячным показателям. Таким образом, начинается эпидемиологический год с месяца, следующего за месяцем с минимальным уровнем заболеваемости одного года, и завершается месяцем с минимальным уровнем заболеваемости следующего года.

Проведение анализа заболеваемости с учетом эпидемиологического года предложено и другими исследователями [37]. Справедливо отмечено, что анализ помесячной динамики инфекционных заболеваний по

календарному году нарушает естественное внутригодовое движение заболеваемости [25]. Однако анализ заболеваемости по эпидемиологическому году нельзя рекомендовать в качестве единственно возможной формы. Например, при инфекциях, для которых сезонность в основном укладывается в календарный год, расхождения относительно анализа по эпидемиологическому году не будут носить существенных различий. Кроме этого, необходимо иметь в виду, что в практической работе эпидемиологической службы анализ и отчетность более удобно (а иногда и единственно возможно) проводить по календарному году. Рассмотрим данную методику [25] применительно к динамике заболеваемости дизентерией в городе Х. за 1997 – 2001 гг. (табл. 32).

По среднемноголетним данным видно, что на рассматриваемой территории эпидемиологический год начинается с апреля и заканчивается в марте, на который приходится минимальная заболеваемость. В связи с этим при построении помесячных динамических рядов в строку 1997 – 1998 гг. записываем показатели заболеваемости 1997 г. с апреля по декабрь, а заболеваемость 1998 г. – с января по март. Аналогично поступаем и для следующих эпидемиологических лет.

Проверка исходных данных на наличие выскакивающих значений свидетельствует, что ни один показатель не выходит за принятый нами 99%-й уровень достоверности для вспышечной заболеваемости.

Среднемноголетние показатели каждого месяца, представляющие типовую кривую заболеваемости, определены как обычные среднеарифметические. Например, для апреля (первый месяц эпидемиологического года) среднемесячный показатель равен

Таблица 32

Помесячная заболеваемость в городе Х.
в показателях на 100 тыс. населения (данные условные)

Год	IV	V	VI	VII	VIII	IX	X	XI	XII	I	II	III	Всего
1997 – 1998	8,5	7,9	7,3	26,6	59,0	65,5	40,7	29,8	9,9	11,4	5,9	3,2	275,7
1998 – 1999	13,4	7,4	10,3	9,3	27,5	44,8	28,7	15,2	14,6	18,9	12,5	5,1	207,7
1999 – 2000	10,7	9,8	12,4	37,5	61,9	66,0	46,0	41,1	20,7	20,7	98,0	10,8	347,3
2000 – 2001	10,4	8,4	9,8	23,2	44,3	68,7	45,0	30,5	17,1	17,1	9,8	6,8	290,7
$\pm m$	1,3		1,3							1,7	1,3		
t	1,64		1,29							4,35	1,9		

$$M = \frac{8{,}5 + 13 + 9{,}1 + 1{,}7}{4} = 10{,}4 \text{ и т.д.}$$

Из среднемноголетних показателей выделяем 2 мес с наименьшей заболеваемостью. В нашем случае это март и май с показателями заболеваемости соответственно 6,8 и 8,4 на 100 тыс. населения. Определяем средний показатель этих месяцев и его ошибку из среднемноголетних показателей, выделяем 2 мес с наименьшей заболеваемостью. В нашем случае это март и май с заболеваемостью соответственно 6,8 и 8,4 на 100 тыс. населения. Определяем среднюю арифметическую этих 2 мес и его ошибку ($M \pm m$):

$$M = \frac{6{,}8 + 8{,}4}{2} = 7{,}6,$$

ошибка показателя вычисляется по формуле $m = \sqrt{\frac{pq}{n}}$.

Необходимо помнить, что население города (n) берется в среднем за анализируемые годы. В нашем случае ошибка среднего показателя равна

$$m = \sqrt{\frac{7{,}6\ (100\,000 - 7{,}6)}{586\,000}} = 1{,}1.$$

Итак, средний показатель 2 мес с минимальной заболеваемостью (март, май) и его ошибка составили 7,6 ± 1,1.

Далее определяем ошибку показателя для 2 мес с минимальной заболеваемостью, значимость различий этих показателей со средней 2 мес по критерию Стьюдента. Следующим после марта и мая по абсолютному значению заболеваемости являются февраль и июнь (9,8 на 100 тыс. населения). Ошибка этого показателя равна ± 1,3, а критерий Стьюдента 1,29. Так как эта величина (t = 1,29) меньше 2, следовательно, показатели февраля и июня достоверно не отличаются от средних 2 мес с минимальной заболеваемостью. Аналогичные выводы получены для заболеваемости в апреле (M = 10,4; m = ± 1,3, c = 1,64). Но уже для декабря, показатели заболеваемости которого составляют 16,4 на 100 тыс. населения, отмечается достоверное отличие от средней заболеваемости

марта и мая ($t = 7{,}6$). Таким образом, заболеваемость декабря и в остальных месяцах с высоким уровнем достоверности не относится к периоду низкой заболеваемости.

Такие же выводы получаем путем сравнения средних с их удвоенными ошибками. Для этого необходимо соблюдать следующее условие: $M + 2m \geq P - 2m$, т.е. среднемноголетний показатель 2 мес наименьшей заболеваемости (M) плюс две его ошибки должно быть больше или равно значению сравниваемого показателя заболеваемости (P) минус две его ошибки.

В нашем случае этому условию удовлетворяют показатели июня, февраля и апреля:

июнь и февраль – $9{,}8\ {}^{0}/_{0000}$ –

$7{,}6 + 2 \cdot 1{,}1 > 9{,}8 - 2 \cdot 1{,}3$, т.е $9{,}8 > 7{,}2$;

апрель – $10{,}4\ {}^{0}/_{0000}$ –

$7{,}6 + 2 \cdot 1{,}1 > 10{,}8 - 2 \cdot 1{,}3$, т.е. $9{,}8 > 7{,}8$;

для декабря это условие нарушается –

$7{,}6 + 2 \cdot 1{,}1 < 16{,}4 - 2 \cdot 1{,}7$, т.е. $9{,}8 < 13{,}0$,

и чем выше будут другие среднемноголетние показатели, тем в большей степени отклонения от $M + 2m \geq P - 2m$.

Таким образом, на основании проведенного анализа февраль, март, апрель, май и июнь следует отнести к периоду (месяцам) межсезонной заболеваемости.

Далее из показателей межсезонной заболеваемости рассчитываем круглогодичную заболеваемость. Для этого по фактическим данным месяцев межсезонного периода вычисляем их средний показатель за 1 мес, который надо умножить на 12 (число месяцев в году). В нашем примере средняя межсезонная заболеваемость за ряд лет составила

$$\frac{10{,}4 + 8{,}4 + 9{,}8 + 9{,}8 + 6{,}8}{5} = 9{,}0 \pm 1{,}2.$$

Годовой показатель круглогодичной заболеваемости равен $9{,}012 = 108{,}0$ на 100 тыс. населения, что составляет 37,2 % от среднегодового уровня ($290{,}7 = 100$ %). Межсезонную заболеваемость не следует определять по верхней доверительной границе среднего показателя ($\pm 2m$), поскольку в этом случае ее доля в общегодовой может увеличиваться на 10 – 20 % и более. Сезонная заболеваемость (сезонная «надбавка») определяется как разница годовой и круглогодичной заболева-

емости. В рассматриваемом примере сезонная «надбавка» составляет 290,7 – 108,0 = 182,7 на 100 тыс. населения, или 62,8 % от годового уровня заболеваемости. Такой подход в определении сезонной и предсезонной заболеваемости может быть применен как по средним месячным за ряд лет, так и по месячным показателям за отдельный год.

В качестве примера рассмотрим определение сезонной и круглогодичной заболеваемости за 1997/98 эпидемиологический год. Средняя двух минимальных месяцев (февраль, март) этого периода равна 4,55 ± 0,89. После соответствующих расчетов устанавливаем, что к месяцам межсезонной заболеваемости относятся также апрель, май и июнь. Определяем годовой показатель круглогодичной заболеваемости (см. табл. 26): 6,56 · 12 = 78,72 $^{0}/_{0000}$. Этот показатель составляет 28,6 от годового итога заболеваемости (275,7 = 100 %). Соответственно сезонная заболеваемость составила 197 $^{0}/_{0000}$, или 71,4 % от годового итога.

Следующим этапом анализа круглогодичной заболеваемости является определение сроков начала и окончания сезонного подъема, т.е. выявления моментов времени, когда заболеваемость становиться выше или ниже максимально возможного уровня круглогодичной заболеваемости. Для определения этого уровня предлагается использовать верхний предел (+ 2m) показателя круглогодичной заболеваемости, который в рассматриваемом примере для среднемноголетних данных равен 9,0 + 2 · 1,2 = 11,4 $^{0}/_{0000}$.

Построим график (линейную диаграмму) многолетней годовой динамики заболеваемости и нанесем на него полученное значение верхнего предела круглогодичной заболеваемости (рис. 8). По первой и последним точкам пересечения типовой кривой с прямой, равной верхнему пределу круглогодичной заболеваемости, определяем средние сроки начала и окончания сезонного подъема. Из рис. 7 следует, что типовое начало сезонного подъема заболеваемости дизентерией в городе Х. приходится на 4 июня и окончание – на 24 января. Средняя продолжительность сезонного подъема составила 234 дня, или 7 мес и 20 дней.

При определении рассматриваемых показателей сезонности с помощью предложенного нами расчетного метода [2] получены точно такие же даты начала и окончания сезонного подъема (методику расчета см. разд. 4.1).

Аналогичным образом следует поступать и при определении сроков

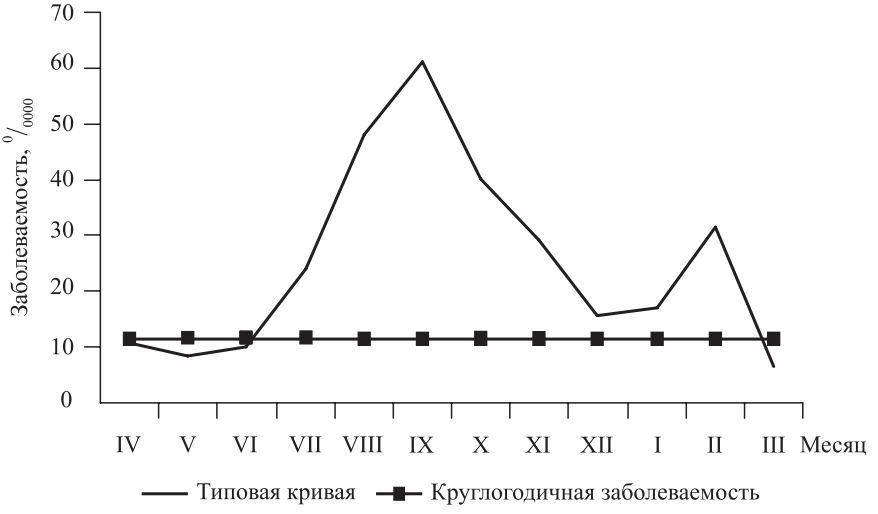

Рис. 8. Типовая кривая годовой динамики заболеваемости дизентерией в городе Х. за 1997 – 2001 гг. (по данным табл. 33).

начала и окончания сезонного подъема для каждого года. В последнем случае верхний предел круглогодичной заболеваемости рассчитывается для каждого отдельного года.

Сравнение круглогодичной заболеваемости по многолетним данным с внутригодовой динамикой (оценкой типичных показателей сезонности) на разных территориях и в различных группах населения за один и тот же период времени представляет как научный, так и практический интерес. Применение круглогодичного показателя средней многолетней в качестве границ сезонности отдельного года методически не совсем верно, особенно в тех случаях, когда наблюдаются значительные колебания заболеваемости в изучаемом ряде лет. При таком подходе средний многолетний показатель круглогодичной заболеваемости может превысить сезонный подъем в годы низкой заболеваемости.

Таким образом, верхняя граница показателя круглогодичной заболеваемости ($+2m$), полученного из средних многолетних данных, в определенной степени может служить точкой отсчета начала и окончания сезонного подъема, но только для сравнительной характеристики сезонности на различных территориях, а не для каждого года изучаемого периода.

По результатам расчетов помесячных показателей заболеваемости по формам проявления эпидемического процесса составляется

таблица (табл. 33). В случае, когда суммарный месячный показатель заболеваемости меньше верхнего предела ее круглогодичного уровня, строка «круглогодичная заболеваемость» представляет фактические значения показателя. При превышении этого уровня в строку «круглогодичная заболеваемость» вносятся значения верхнего предела круглогодичной заболеваемости, а остаток от суммарного месячного показателя считается сезонной надбавкой.

Как видно из приведенной таблицы, в апреле 1998 – 1999 эпидемиологического года показатель общей заболеваемости был выше верхнего пре-

Таблица 33

Уровень заболеваемости дизентерией по формам эпидемического процесса в городе Х. на 100 тыс. населения

Годы	Формы проявления эпидемического процесса	Месяц											
		IV	V	VI	VII	VIII	IX	X	XI	XII	I	II	III
1997 – 1998	Общая заболеваемость	8,5	7,9	7,3	26,6	59,0	65,6	40,7	29,8	9,9	11,4	5,9	3,2
	В том числе:												
	круглогодичная	8,5	7,9	7,3	8,7	8,7	8,7	8,7	8,7	8,7	8,7	5,9	3,2
	сезонная	0,0	0,0	0,0	17,9	50,3	56,9	32,0	21,1	1,2	2,7	0,0	0,0
1998 – 1999	Общая заболеваемость	13,4	7,4	10,3	9,3	27,5	44,8	28,7	15,2	14,6	18,9	12,5	5,1
	В том числе:												
	круглогодичная	10,4	7,4	10,3	9,3	10,4	10,4	10,4	10,4	10,4	10,4	10,4	5,1
	сезонная	0,0	0,0	0,0	0,0	17,1	34,4	18,3	4,8	4,2	8,5	2,1	0,0
	вспышечная	3											
1999 – 2000	Общая заболеваемость	9,1	8,4	9,3	19,3	28,9	98,5	65,9	35,9	20,6	17,2	11,0	8,1
	В том числе:												
	круглогодичная	9,1	8,4	9,3	11,7	11,7	11,7	11,7	11,7	11,7	11,7	11,0	8,1
	сезонная	0,0	0,0	0,0	7,6	17,2	76,8	54,2	24,2	8,9	5,5	0,0	0,0
2001 – 2002	Общая заболеваемость	10,7	9,8	12,4	37,5	61,9	66,0	46,0	41,1	20,6	20,7	9,8	10,8
	В том числе:												
	круглогодичная	10,7	9,8	12,4	13,4	13,4	13,4	13,4	13,4	13,4	13,4	9,8	10,8
	сезонная	0,0	0,0	0,0	24,1	48,5	52,6	32,6	27,7	7,2	7,3	0,0	0,0
Многолетние показатели	Общая заболеваемость	10,4	8,4	9,8	23,2	44,8	68,7	45,3	30,5	16,4	17,1	9,8	6,8
	В том числе:												
	круглогодичная	10,4	8,4	9,8	11,4	11,4	11,4	11,4	11,4	11,4	11,4	9,8	6,8
	сезонная	0,0	0,0	0,0	11,8	32,9	57,3	33,9	19,1	5,0	5,7	0,0	0,0

Таблица 34

Внутригодовая и многолетняя характеристики заболеваемости в городе Х. по формам проявления эпидемического процесса в 1997 – 2001 гг.

Годы	Дата сезонного подъема		Продолжительность подъема, дней		Круглогодичная заболеваемость		Сезонная надбавка		Вспышечная заболеваемость	
	начало	окончание	сезонного	межсезонного	Показатель, $^{0}/_{0000}$	Доля, %	Показатель, $^{0}/_{0000}$	Доля, %	Показатель, $^{0}/_{0000}$	Доля, %
1997 – 1998	03.06	15.01	226	139	78,7	28,6	197,0	71,4	0,0	0,0
1998 – 1999	03.07	09.02	221	144	96,2	46,3	108,5	52,2	3,0	1,4
1999 – 2000	06.06	27.01	235	130	110,2	33,2	220,0	66,8	0,0	0,0
2000 – 2001	02.06	21.01	233	132	128,4	37,0	218,9	63,0	0,0	0,0

дела ее круглогодичного уровня. В этом случае разность между рассматриваемыми величинами характеризует вспышечную заболеваемость. Заметим, что в предполагаемом методе, выделение вспышечной заболеваемости путем статистического анализа возможно лишь в месяцы межсезонного подъема, а в сезонный период вспышечная заболеваемость будет затушевываться. Однако в случае необходимости можно выделять вспышки по данным регистрации, соответственно уменьшая при этом фактические показатели общей месячной заболеваемости.

Годовые итоги интенсивных и экстенсивных показателей, а также другие характеристики внутригодовой заболеваемости оформляются в виде отдельной таблицы (табл. 34). Число показателей в таблице может быть другим, в зависимости от поставленных задач, можно учитывать и вспышечную заболеваемость по данным регистрации и т.д.

4.3. ОПРЕДЕЛЕНИЕ «КОНТРОЛЬНЫХ» УРОВНЕЙ ЗАБОЛЕВАЕМОСТИ

Проведение целенаправленных профилактических мероприятий и изучение закономерностей течения эпидемического процесса невозможно без применения статистических методов исследования. При правильном использовании эти методы позволяют своевременно установить начало подъ-

ема заболеваемости (эпидосложнения), с достаточной надежностью подтвердить или опровергнуть предположения о ведущих путях передачи, а также выявить основные закономерности, присущие эпидемическому процессу. Вышеизложенное определяет важность такой задачи, как оперативное слежение за заболеваемостью и оценка интенсивности эпидемического процесса за месяц (неделю, день). Выбор конкретных интервалов времени зависит от цели и задач эпидемиологического надзора, а также от распространенности той или иной инфекции на рассматриваемой территории. Основой такого анализа является сравнение фактических показателей отдельных инфекций за текущий год с так называемым «контрольным» уровнем заболеваемости. Иногда в литературе этот показатель называют «нормативным» уровнем, хотя едва ли правомерно вводить для инфекционной заболеваемости такое понятие, как «норма».

По нашему мнению, довольно простым и достаточно надежным методом определения «контрольных» уровней заболеваемости является оценка границ доверительных интервалов средних значений, приведенная в монографии И.П. Ашмарина с соавт. [5]. При расчетах необходимо использовать данные заболеваемости за 4 – 6 лет и более.

Рассмотрим определение «контрольных» уровней заболеваемости дизентерией Зонне в городе Н. (см. табл. 27). С этой целью воспользуемся показателями за 7 лет (1994 – 2001 гг.), а данные 2001 г. условно примем за заболеваемость текущего года. В приведенном примере представлена помесячная заболеваемость. Следует отметить, что для оперативной работы эпидемиологов требуются, как правило, более дробные показатели, например недельные, однако методика расчета будет неизменной.

Первый этап анализа – составление таблицы, в которой должны быть приведены интенсивные показатели заболеваемости (1994 – 2001 гг.). Прежде чем перейти к вычислению средних показателей, необходимо проверить эти данные на наличие так называемых выскакивающих значений (о методике расчета выскакивающих вариант см. гл. 1), которые, как правило, представляют собой следствие вспышечной заболеваемости. Выскакивающие варианты с наименьшими значениями исключать из разработки не следует в связи с тем, что минимальные показатели обусловлены профилактическими мероприятиями и определяют круглогодичную заболеваемость.

В представленной таблице обращает на себя внимание заболеваемость в январе 1994 г., значение которой, возможно, является выскакивающим. Рассчитаем отношение

$$\frac{x_n - x_{n-1}}{x_n - x_1} = \frac{12,3 - 7,6}{12,3 - 3,4} = 0,53.$$

Полученное значение оцениваем с помощью табл. 1 приложения. Для 7 вариант критические значения составляют 0,507 для уровня достоверности 95 % и 0,637 – для 99 %. Таким образом, полученная цифра (0,53) оказалась между пограничными значениями для уровня достоверности 99 и 95 %, что не позволяет нам исключить ее из дальнейших расчетов (см. рассуждения о правомочности исключения вспышечной заболеваемости в разд. 4.1).

Следующий этап работы сводится к определению типовой кривой заболеваемости (см. разд. 4.1). Для этого вычисляем средние арифметические показатели (M) за 7-летний период:

$$\text{январь: } \frac{5,8 + 12,3 + 5,0 + 3,4 + 3,8 + 4,6 + 7,6}{7} = 6,1,$$

$$\text{февраль: } \frac{5,5 + 7,3 + 3,8 + 3,1 + 2,7 + 2,4 + 5,0}{7} = 4,3 \text{ и т.д.}$$

Следует помнить, что в случае исключения выскакивающих величин вычисление средних проводится без их участия.

Затем вычисляются доверительные интервалы полученных средних значений. Классические операции расчета доверительного интервала довольно сложны. Однако экспрессивный метод его расчета достаточно прост и основан на оценке различий между значениями крайних вариант в рассматриваемом ряду [5]. Важным условием этого метода является предварительное выполнение описанных выше операций по исключению выскакивающих вариант. Расчет верхней и нижней границы доверительного интервала средней арифметической осуществляется с помощью простых формул:

верхний доверительный интервал – M + КА,

нижний – M – КА,

где M – средняя арифметическая какого-либо месяца; А – амплитуда

(размах) варьирования, т.е. разность между крайними вариантами в упорядоченном ряду; К – коэффициент, который находят по табл. 15 приложения для уровня достоверности 95 или 99 %.

Для нашего примера (n = 7) границы доверительных интервалов средних арифметических для уровня:

январь:

верхний – 6,1+0,34 (12,3 – 3,4) = 9,2 $^{0}/_{0000}$,

нижний – 6,1-0,34 (12,3 – 3,4) = 3,0 $^{0}/_{0000}$;

февраль:

верхний – 4,3+0,34 (7,3 – 2,4) = 6,0 $^{0}/_{0000}$,

нижний – 6,1-0,34 (12,3 – 3,4) = 2,6 $^{0}/_{0000}$ и т.д.

Верхняя граница доверительного интервала и будет являться «контрольным» уровнем заболеваемости дизентерией Зонне для населения города Н. По нижней границе можно давать оценку о значимости проводимых профилактических (прежде всего, специфических) мероприятий, если фактическая заболеваемость окажется ниже этой линии.

Как видно из рис. 9 и табл. 35, 2001 г. был неблагополучным по заболеваемости дизентерией Зонне. Заболеваемость выше «контрольных» уровней отмечалась в июне, июле, ноябре и декабре. Вывод о неблагополучии этого года можно сделать и по годовому итогу. Однако для принятия оперативных мер служат «контрольные» уровни заболеваемости,

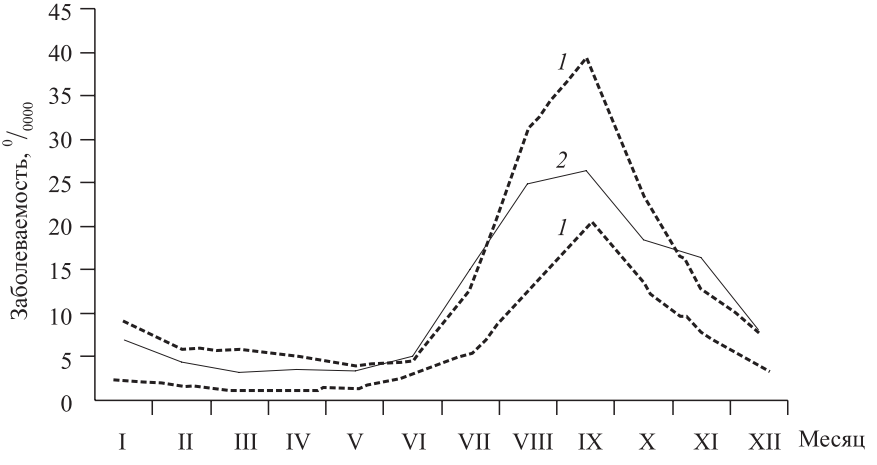

Рис. 9. «Контрольные» уровни (*1*) и фактические показатели (*2*) заболеваемости дизентерией Зонне в городе Н. в 2001 г.

Таблица 35

Помесячные «контрольные» уровни заболеваемости дизентерией Зонне в городе Н. и фактические показатели заболеваемости за 2001 г. (по средним арифметическим)

Показатель	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Типовая кривая	6,1	4,3	3,8	3,4	3,0	3,9	9,3	21,9	29,7	18,0	10,3	5,8
Доверительный интервал:												
верхний	9,2	6,0	5,8	5,0	4,0	4,5	12,6	31,2	39,2	23,1	13,1	7,6
нижний	3,0	2,6	1,8	1,8	2,0	3,3	6,0	12,6	20,1	12,9	7,5	4,0
Заболеваемость в 2001 г.	6,9	4,4	3,2	3,6	3,4	5,0	15,0	24,8	26,4	18,4	16,4	8,2

которые для этих целей должны быть более дробными. Целесообразно также рассчитывать «контрольные» уровни заболеваемости и для различных профессиональных возрастных групп населения (групп риска).

При условии стабилизации эпидемического процесса можно не пересчитывать ежегодно доверительные интервалы. Изменение же уровней заболеваемости влечет за собой и разработку нового «контрольного» уровня в следующем году.

Как уже отмечалось (см. гл. 3), для ряда инфекций характерно наличие упорядоченных, выраженных многолетних циклических колебаний. Это относится прежде всего к вирусному гепатиту А и ряду инфекций с аэрозольным механизмом передачи. При таком течении эпидемического процесса объединение в один ряд лет показателей заболеваемости, находящихся на максимуме или минимуме многолетних циклов, приведет к тому, что разработанные «контрольные» уровни будут недостаточно надежны для лет с высокой заболеваемостью и окажутся практически бесполезны для оценки показателей, находящихся внизу многолетнего цикла.

Таким образом, если ретроспективный эпидемиологический анализ выявил наличие выраженной упорядоченной цикличности, необходимо рассчитать «контрольные» уровни заболеваемости для двух типов лет (об оценке цикличности см. разд. 3.4). В таком случае выбор периодов и лет с минимальными и максимальными показателями заболеваемости должен опираться на данные не менее чем за 10 – 15 лет. Затем нужно выбрать годы с высокими и отдельно с низкими показателями заболеваемости, на основании которых и будут строиться два типа «контрольных» уровней. Выбор того или иного доверительного интервала будет зависеть от того, на каком уровне находилась заболеваемость предшествующего года, по которому можно дать качественный прогноз к дальнейшему росту или снижению заболеваемости.

Для оперативного слежения за заболеваемостью могут быть также использованы медиана и ее доверительные интервалы (об использовании медианы см. разделы 1.2 и 4.1).

Рассмотрим определение «контрольного» уровня для вышеприведенного примера (см. табл. 27 и 35), причем в качестве среднего значения воспользуемся медианой. Упорядочим данные за все январи рассматриваемого периода по возрастанию: 3,4; 3,9; 4,6; 5,0; 5,8; 7,6; 12,3. Медиана в этом случае равна 5,0 $^{0}/_{0000}$. Находим (см. приложение, табл. 2), что при числе вариант 7 доверительный интервал охватывает всю совокупность наблюдений. В связи с этим нижний доверительный интервал равен 3,36, а верхний – 12,32 $^{0}/_{0000}$. Таким образом, Me = 5,0 (3,4 – 12,3) $^{0}/_{0000}$.

Окончательные результаты определения «контрольных» уровней заболеваемости с использованием медианы приведены в табл. 35. Сравнение данных свидетельствует об идентичности выводов об эпидемиологическом неблагополучии 2001 г. Уровень заболеваемости превышал границу верхнего доверительного интервала в июне, июле и ноябре. Лишь для декабря не было такого совпадения. Это объясняется тем, что доверительный интервал медианы шире такового средней арифметической. В нашем примере превышение уровней (верхнего и нижнего) доверительных интервалов имеет место для большинства месяцев.

Анализ годовой динамики заболеваемости должен завершаться обобщением, позволяющим представить более конкретные гипотезы о причинах заболеваемости, чем при анализе многолетней динамики эпидемического процесса. Кроме того, данные внутригодовой динамики заболеваемости позволяют уточнить природу многолетней тенденции и цикличности эпидемического процесса (см. гл. 3). Окончательные выводы о причинах заболеваний могут быть получены при анализе заболеваемости по факторам риска.

Глава 5

АНАЛИЗ ЗАБОЛЕВАЕМОСТИ ПО ФАКТОРАМ РИСКА

Целью анализа заболеваемости по факторам риска являются проверка гипотез о причинах, ее обусловливающих, выявление факторов, определяющих риск заражения и заболевания людей, установление ведущих типов эпидемий [9].

Выбор статистических методов для проверки предположения о факторах риска невелик и определяется методическим подходом к построению гипотезы и особенностями анализируемого материала. Для количественной оценки факторов риска используются корреляционный и регрессионный анализы, показывающие характер и тесноту связи между заболеваемостью и возможной ее причиной, а также методы, позволяющие выявить достоверный уровень различий в показателях заболеваемости при различных значениях фактора, рассматриваемого в качестве причины заболеваемости. С вышеуказанными подходами тесно связан дисперсионный анализ.

Учитывая, что все перечисленные способы достаточно широко освещены в различных руководствах по статистике, мы ограничили свою задачу лишь их применением для эпидемиологического анализа, не вдаваясь при этом в излишнюю детализацию объяснений и расчетов. Статистическая оценка различий в показателях заболеваемости при разных значениях фактора риска подробно изложена в гл. 2.

Для практической работы эпидемиологов, связанной с выявлением факторов риска, наибольшее значение имеет корреляционный анализ. По выражению В.Д. Белякова [8], суть эпидемиологического метода исследования состоит в том, что сначала устанавливаются статистические (корреляционные) связи между факторами риска и показателями, характеризующими уровень, структуру и динамику заболеваемости, с последующим выявлением причинно-следственных и функциональных связей. Необходимо подчеркнуть, что измерение связи допустимо лишь в случае, если ее наличие предполагается хотя бы теоретически. Парал-

лельное изменение сопоставляемых показателей еще не свидетельствует об их взаимозависимости, так как оно может быть обусловлено случайным совпадением, как, например, наличие положительной связи в теплое время года между заболеваемостью острыми кишечными инфекциями и количеством женщин, одетых в яркие платья.

Одним из наиболее совершенных способов измерения связи является вычисление линейного коэффициента корреляции, который часто называют также парным коэффициентом корреляции. Этот показатель используется в том случае, когда связь между признаками *x* и *y* выражается прямой линией (формула Пирсона):

$$r_{xy} = \frac{\sum dx \cdot dy}{\sqrt{\sum d^2x \cdot \sum d^2y}},$$

где r_{xy} – коэффициент корреляции; *x* и *y* – коррелируемые ряды, *dx* и *dy* – отклонение каждого из чисел этих рядов от их средних.

При положительной прямой связи (например, рост заболеваемости дизентерией при увеличении в воде водопровода доли нестандартных проб воды) коэффициент корреляции может принимать значения от 0 до +1. В случае отрицательной (обратной) связи (например, снижение заболеваемости гепатитом В по мере увеличения охвата населения вакцинацией против этой инфекции) коэффициент корреляции выражается отрицательным числом и принимает значения от 0 до –1.

Чем ближе значения коэффициента корреляции к 1, тем теснее измеряемая им связь. Коэффициент корреляции, равный 0, свидетельствует о полном отсутствии связи, а коэффициент +1 или –1 – о наличии между явлениями функциональной (абсолютной) связи. В зависимости от коэффициента корреляции ориентировочно можно сделать следующее заключение:

$0 < r < 0{,}3$ – слабая (малая) связь,

$0{,}3 < r < 0{,}7$ – средняя (умеренная) связь,

$0{,}7 < r < 1$ – сильная (тесная) связь.

Приведем пример вычислений парного коэффициента корреляции по указанной выше формуле (табл. 36). В одном населенном пункте зарегистрировано наличие хронической эпидемии дизентерии Флекснера. Предварительный анализ и лабораторные исследования показали, что в питьевой воде водопроводной сети наблюдаются частые «проскоки» нестандартных проб по бактериологическим показателям (фактор рис-

Таблица 36

Корреляция между заболеваемостью дизентерией Флекснера и количеством нестандартных проб воды в водопроводной сети

Месяц	Число больных дизентерией (x)	Доля нестандартных проб воды (y)	dx $(x - \bar{x})$	dy $(y - \bar{y})$	d^2x	d^2y	$dx \cdot dy$
I	10	0	– 6,7	3,2	44,9	10,2	21,4
II	9	0,5	– 7,7	– 2,7	59,3	7,3	20,8
III	2	1,1	14,7	2,1	216,1	4,4	30,9
IV	7	2,0	– 9,7	–1,2	94,1	1,4	11,6
V	6	1,8	–10,7	1,4	114,5	2,0	15,0
VI	11	2,9	– 5,7	– 0,3	32,5	0,1	1,7
VII	26	6,7	+9,3	+3,5	86,5	12,3	32,6
VIII	32	4,5	+15,3	+1,3	234,1	1,7	19,9
IX	46	8,7	+29,3	+5,5	858,5	30,3	161,2
X	38	7,1	+21,3	+3,9	453,7	15,3	83,1
XI	8	3,2	– 8,7	0	75,7	0,0	0,0
XII	5	0	–11,7	– 3,2	136,9	10,2	37,4
$\sum$	200	38,5	0	0	2406,7	95,2	435,5
$\bar{M}$	16,7	3,2					

ка). Проведем статистическое испытание гипотезы о наличии связи между этими двумя признаками. Естественно, что следует сопоставлять не абсолютные числа нестандартных проб воды, которые зависят от общего количества анализа, а удельные веса нестандартных проб.

Схема этапов выявления наличия связи следующая:

1 – составить рабочую таблицу, в которую заносятся первичные данные (см. табл. 30);

2 – рассчитать средние арифметические показатели для заболеваемости дизентерией Флекснера ($\bar{x}$) и удельного веса нестандартных проб воды по бактериологическим показателям ($\bar{y}$);

3 – рассчитать отклонения значений каждого члена коррелируемых рядов (x и y) от их средней арифметической ($\bar{x}$ и $\bar{y}$) и внести в таблицу (dx и dy);

4 – полученные значения отклонений возвести в квадрат (d^2x и d^2y) и внести в таблицу;

5 – вычислить сумму квадратов отклонений $\sum d^2x$ и $\sum d^2y$;

6 – перемножить попарно значения dx и dy, ($dx \cdot dy$);

7 – суммировать произведения отклонений коррелируемых пар показателей;

8 – вычислить парный коэффициент корреляции, подставляя в формулу итоговые данные табл. 36

$$r_{xy} = \frac{\sum dx \cdot dy}{\sqrt{\sum d^2x \cdot \sum d^2y}} = \frac{453{,}53}{\sqrt{2406{,}68 \cdot 95{,}15}} = +0{,}91.$$

После получения коэффициента корреляции возникает необходимость выяснить, при каком значении он указывает на действительную связь между явлениями, т.е. на достоверность отмеченного факта.

Достоверность коэффициента корреляции определяется сравнением его с вычисляемой средней ошибки. Средняя ошибка ($\pm m_r$) коэффициента корреляции равна

$$m_r = \pm\frac{1-r^2}{\sqrt{n}}.$$

Коэффициент корреляции считается достоверным, если не менее чем в 3 раза превышает свою среднюю ошибку. Иначе необходимо увеличить число наблюдений и вновь вычислить коэффициент и его ошибку, так как из формулы следует, что эта величина обратно пропорциональна $\sqrt{n}$.

Вычислим среднюю ошибку коэффициента корреляции для нашего примера:

$$m_r = \pm\frac{1-r^2}{\sqrt{n}} = \pm\frac{1-0{,}91^2}{\sqrt{12}} = \pm 0{,}05,$$

$$\frac{r_{xy}}{m_r} = \frac{0{,}91}{0{,}05} = 18{,}2.$$

В данном примере коэффициент корреляции превышает свою среднюю ошибку в 18,2 раз, что свидетельствует о его достоверности.

Более просто достоверность коэффициента корреляции определяется по специальным таблицам, которые можно найти в любых руководствах по статистическому анализу, там приведены критические значения *r*, выше которых коэффициент корреляции при данном числе наблюдений признается значимым.

Сравним полученный коэффициент корреляции с его критическим значением (см. приложение, табл. 12). Для числа наблюдений (12) полученное значение коэффициента корреляции превышает табличное для уровня достоверности 99 %. Установив, что полученный коэффициент корреляции достоверен, необходимо перейти к оценке его силы. В нашем примере следует говорить о сильной связи между явлениями, так как коэффициент превышает 0,7.

Таким образом, по результатам корреляционного анализа мы имеем основание говорить, что заболеваемость дизентерией Флекснера в значительной степени зависит от качества питьевой воды, поэтому в данном населенном пункте наши профилактические мероприятия должны быть направлены на обеспечение населения доброкачественной водой.

В этом примере мы разобрали корреляцию между двумя признаками (заболеваемость и фактор риска). Однако часто заболеваемость зависит от многих других варьирующих признаков, поэтому наряду с изучением парных корреляций перед эпидемиологом возникает задача измерения множественных связей между варьирующими факторами риска (так называемая множественная корреляция). Эта задача решается с помощью коэффициента общей (совокупной) корреляции, а также частных (парциальных) коэффициентов корреляции. Совокупный коэффициент корреляции между тремя учитываемыми признаками (x, y и z) вычисляется по формуле

$$r = \sqrt{\frac{r_{xy}^2 + r_{yz}^2 - 2r_{xy} \cdot r_{xz} \cdot r_{yz}}{1 - r_{xy}^2}},$$

где r_{xy},r_{xz} и r_{yz} – парные коэффициенты линейной корреляции между признаками x и y, x и z, y и z.

Совокупный коэффициент корреляции обладает теми же свойствами, которые присущи и парным коэффициентам прямолинейной связи. Но в отличие от последних совокупный коэффициент корреляции всегда имеет только один (положительный) знак. Направление связи (прямая, обратная) определяется исходя из логики события и анализа первичных данных.

Рассмотрим пример анализа при использовании трех признаков. Выше нами показано, что заболеваемость дизентерией Флекснера

Таблица 37

Исходные материалы для расчета совокупного и парциальных коэффициентов корреляции

Месяц	Число больных дизентерией Флекснера	Удельный вес нестандартных проб воды, %	Число аварий на разводящей сети водопровода
I	10	0	18
II	9	0,5	18
III	2	1,1	16
IV	7	2,0	33
V	6	1,8	30
VI	11	2,9	34
VII	26	6,7	46
VIII	32	4,5	23
IX	46	8,7	33
X	38	7,1	23
XI	8	3,2	21
XII	5	0	15

тесно связана с качеством питьевой воды, которая зависит также от таких показателей, как число аварий на разводящей сети водопровода, и (или) длительностью этих аварий. Возьмем для примера число аварий на разводящей сети водопровода по месяцам года в том же населенном пункте. В этом случае имеем три признака: число заболевших дизентерией Флекснера, удельный вес нестандартных проб питьевой воды (по коли-индексу), число аварий на разводящей сети водопровода по месяцам.

Последовательность вычислений будет следующая:

1 – составляем рабочую таблицу исходных данных (табл. 37);

2 – рассчитываем парный коэффициент корреляции между заболеваемостью дизентерией. Флекснера и удельным весом нестандартных проб воды по коли-индексу, т.е. r_{xy} (расчет см. в табл. 38);

3 – аналогично рассчитываем r_{xz} и r_{yz}. В результате вычислений парные коэффициенты корреляции равны: $r_{xy} = 0{,}91$; $r_{xz} = 0{,}43$; $r_{yz} = 0{,}59$. Оцениваем достоверность коэффициентов 0,91 ($P < 0{,}01$) и 0,59 ($P < 0{,}05$) (см. приложение, табл. 12). Коэффициент корреляции, равный 0,43, оказался недостоверен.

4 – найденные значения парных коэффициентов позволяют рассчитать совокупный коэффициент корреляции:

$$r_{xyz} = \sqrt{\frac{0{,}42^2 \cdot 0{,}59^2 - 2 \cdot 0{,}91 \cdot 0{,}43 \cdot 0{,}59}{1 - 0{,}91}} = 0{,}64.$$

В практической работе эпидемиологов большое значение имеет определение не совокупного, а частных коэффициентов корреляции. Это связано с тем, что в большинстве случаев необходимо исследовать связь только двух признаков, при условии постоянного значения третьего учитываемого признака, который находится или может находиться в корреляционной зависимости от первых двух. Частный, или парциальный, коэффициент корреляции между признаками *x* и *y* при исключении влияния на эту связь третьего варьирующего признака *z* равен

$$r_{xy(z)} = \frac{r_{xy} - r_{xz} \cdot r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}.$$

Знак исключенного признака принято заключать в скобки или же обозначать слева (например, ${}_z r_{xy}$).

Соответственно коэффициент парциальной корреляции между признаками *x* и *z* при исключении влияния связанного с ними третьего признака *y* равен

$$r_{xz(y)} = \frac{r_{xz} - r_{xy} \cdot r_{yz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{yz}^2)}}.$$

Наконец, коэффициент парциальной корреляции между *y* и *z* при постоянном значении признака *x* равен

$$r_{yz(x)} = \frac{r_{yz} - r_{xy} \cdot r_{xz}}{\sqrt{(1 - r_{xy}^2)(1 - r_{xz}^2)}}.$$

Из приведенных формул видно, что для вычисления парциальных коэффициентов корреляции прежде всего определяются все три парных коэффициента корреляции. Не приводя конкретных расчетов, отметим лишь, что в разобранном примере корреляционного анализа, при использовании трех признаков, основное значение для эпидемиологов имеет выяснение наличия связи между заболеваемостью и удельным весом нестандартных проб воды, независимо от влияния

числа аварий на водопроводной сети. В нашем случае этот показатель равен +0,90.

При четырех коррелированных признаках (например, x, y, z и v) элиминировать приходится уже по два признака:

$$r_{xy(zv)} = \frac{r_{xy(v)} - r_{xz(v)} \cdot r_{yz(v)}}{\sqrt{(1 - r^2_{xz(v)})(1 - r^2_{yz(v)})}} \text{ и т.д.,}$$

причем эти новые величины $r_{xy(v)}$, $r_{xz(v)}$, $r_{yz(v)}$, в свою очередь, получаются по приведенной выше формуле с использованием трех признаков.

Необходимо отметить, что вычислить парциальные коэффициенты корреляции при 4, 5 и более признаках достаточно сложно и прибегать к нему без использования ПК не рекомендуется. Кроме того, следует иметь в виду, что чем больше признаков принято во внимание, тем строже должна соблюдаться линейность всех корреляций, иначе формулы для перехода от полных коэффициентов корреляции к частным и обратно дают слишком неточные результаты.

В случаях, если заболеваемость можно измерить лишь полуколичественно, или ее значениям можно присвоить порядковые номера (ранг), то гипотезу о корреляционной связи проверяют, вычислив непараметрический коэффициент корреляции рангов Спирмена. Этот метод менее точен по сравнению с разобранным выше «классическим» вариантом нахождения коэффициента корреляции по количественным показателям. Однако из-за своей простоты метод ранговой корреляции находит широкое применение в эпидемиологической практике. Этому во многом способствуют массовость оцениваемых эпидемиологических явлений и значительные колебания исследуемых показателей.

Рассмотрим этапы вычислений коэффициента корреляции рангов Спирмена по материалах разобранного выше примера водной хронической эпидемии дизентерии Флекснера (табл. 38):

1. Составляем рабочую таблицу, в которую заносятся первичные данные (число больных, удельный вес нестандартных проб воды).

2. В столбцы «ранги x» и «ранги y» заносим рангоговые номера каждого из сравниваемых рядов в порядке возрастания или убывания (обязательное условие, чтобы оба сравниваемых ряда ранжировались в одну и ту же сторону). Если некоторые варианты повторяются, то

Таблица 38

Корреляция между заболеваемостью дизентерией Флекснера и количеством нестандартных проб воды в водопроводной сети

Месяц	Число больных дизентерией (x)	Удельный вес нестандартных проб воды (y), %	Ранг		$d\,(x-y)$	$d^2\,(x-y)$
			x	y		
I	10	0	7	1,5	5,5	30,25
II	9	0,5	6	3	3	9,00
III	2	1,1	1	4	–3	9,00
IV	7	2,0	4	6	–2	4,00
V	6	1,8	3	5	–2	4,00
VI	11	2,97	8	7	1	1,00
VII	26	6,7	9	10	–1	1,00
VIII	32	4,5	10	9	1	1,00
IX	46	8,7	12	12	0	0
X	38	7,1	11	11	0	0
XI	8	3,2	5	8	–3	9,00
XII	5	0	2	1,5	0,5	0,25
$\sum$						68,5

их рангами будут среднеарифметические показатели из соответствующих порядковых номеров. Например, в нашем случае ноль встречается дважды. Если ранжировать значения в порядке возрастания, то полусумма порядковых чисел 1 и 2 будет равна (1 + 2) / 2 = 1,5

3. Для каждой пары сопоставляемых значений определяется разность рангов и заносится в столбец d.

4. Разность рангов возводим в квадрат и, суммируя, получаем $\sum d^2$, которая в нашем случае равна 68,5.

Устанавливаем коэффициент корреляции рангов Спирмена по формуле

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)},$$

где ρ – коэффициент корреляции рангов; d – разность между порядковыми номерами рядов; n – число пар коррелируемых рядов.

Подставляя в формулу полученную из табл. 38 сумму, производим вычисления:

$$\rho = 1 - \frac{6 \cdot 68{,}5}{12(12^2 - 1)} = 1 - \frac{411{,}0}{1716} = +0{,}76.$$

Определяем среднюю ошибку коэффициента корреляции для оценки его достоверности:

$$m_\rho = \pm\frac{1-0{,}76^2}{\sqrt{12}} = \pm 0{,}12;\ \frac{\rho}{m_\rho} = \frac{0{,}76}{0{,}12} = 6{,}2.$$

В связи с тем что коэффициент корреляции более чем в 3 раза превышает свою ошибку, следует оценить его как значимый. Как и для парного коэффициента корреляции, статистическую значимость рангового коэффициента можно установить по специально разработанным таблицам (см. приложение, табл. 11). Оцениваем полученный коэффициент корреляции. При $n = 12$ для уровня достоверности 95 % критическое значение P равно 0,58, а для 99 % – 0,73. Следовательно, полученный коэффициент корреляции высокодостоверен.

Таким образом, в связи с тем, что также получен достоверный, прямой, сильный коэффициент корреляции, мы должны полностью повторить вывод, сделанный в результате расчетов парного (линейного) коэффициента корреляции. В тех случаях, когда в ряду встречаются два или несколько одинаковых по значению числа, порядковый номер каждого из них обозначается средней от суммы порядковых номеров. Но показано, что при объединении рангов сумма квадратов рангов уменьшается, что влияет на коэффициент корреляции, поэтому с учетом объединенных рангов требуется вычислять коэффициент корреляции по формуле Кендела [16]:

$$r_k = \frac{\frac{n^3-n}{6} - (t_x + t_y) - \sum (x-y)^2}{\sqrt{(\frac{n^3-n}{6} - 2t_x)\cdot(\frac{n^3-n}{6} - 2t_y)}},$$

где t_x, $t_y = (t^3 - t)/12$; t – численность объединенных рядов; n – число наблюдений.

При $t_x = (t_y) = 0$ эта формула превращается в обычную формулу Спирмена.

Рассмотрим пример расчета по формуле Кендела по данным табл. 3. В данном случае ранги первого признака (x) не объединялись $t_x = (t^3 - t)/12 = 0$

Ранги второго признака (y) объединялись один раз. Численность группы объединенных рангов $t_y = 2$.

В таком случае $t_x = (2^3 - 2) / 12 = 0{,}5$.

Теперь вычисляем коэффициент корреляции:

$$r_k = \frac{\frac{12^3 - 12}{6} - (0 + 0{,}5) - 68{,}5}{\sqrt{(\frac{12^3 - 12}{6} - 0) \cdot (\frac{12^3 - 12}{6} - 1)}} = \frac{217{,}0}{285{,}5} = 0{,}76.$$

Коэффициент корреляции рангов по Спирмену при округлении до второго знака не отличается от коэффициента корреляции, полученного по формуле Кендела. Но когда численность объединенных рангов 15 – 20 и более, выявляются существенные различия данных коэффициентов. В этих случаях целесообразно использовать формулу Кендела.

Коэффициент корреляции рангов, как и парный показатель корреляции, сам по себе не указывает причину существующей между явлениями связи. Однако оба эти показателя дают возможность оценить эту связь тогда, когда ее наличие обосновывается гипотезой о причинно-следственной зависимости заболеваемости с определенным фактором риска.

При проведении эпидемиологического анализа может встать вопрос о выделении отдельных внутригодовых периодов (отдельных временных отрезков года), на которые оказывают свое влияние изучаемые факторы риска. Следует иметь в виду, что в ряде случаев отсутствие достоверных коэффициентов корреляции между факторами риска и заболеваемостью может быть следствием того, что в различные эпидемические периоды года взаимосвязь между этими показателями существенно различается. В этих случаях корреляционный анализ необходимо проводить с учетом однородных эпидемических периодов года, объединенных в один ряд за несколько лет, так как внутригодовые эпидемические периоды в отдельные годы слишком коротки для надежных выводов. Указанный подход условно назван нами «вертикальным» анализом.

В качестве примера подобного анализа приведем данные по оценке влияния природно-климатических факторов на интенсивность эпидемического процесса дизентерии в различные периоды года. С этой целью показатели заболеваемости дизентерией в каждом году сгруппируем в однородные эпидемические периоды: предсезонный, сезонный и межсезонный. Эпидемические периоды выделим следующим образом. Предсезонный период объединяет в себе месяцы с заболеваемостью от минимального уровня до

начала сезонного подъема. В сезонный период включим месяцы, в которые заболеваемость превышала среднемесячные показатели или с индексами сезонности более 100 %. Остальные месяцы отнесем к межсезонному периоду. Показатели заболеваемости сравним со среднемесячными показателями природно-климатических факторов тех же месяцев года.

Из основных результатов исследований (табл. 39) следует, что статистически достоверные коэффициенты корреляции между уровнями заболеваемости в отдельные эпидемические периоды и природно-климатическими факторами получены в основном для предсезонного периода. Для дизентерии Зонне такая связь установлена с температурой воды и почвы, а также с абсолютной и относительной влажностью ($P < 0,01$), для дизентерии Флекснера – с температурой воздуха и воды, а также с абсолютной влажностью ($P < 0,05$). Достоверные показатели коэффициентов корреляции для других эпидемических периодов года получены лишь для температуры воды и уровней заболеваемости дизентерией Флекснера в сезонный период и для относительной влажности и заболеваемости дизентерией Зонне в межсезонный период. В остальных случаях значения коэффициентов корреляции в сезонный и межсезонный периоды для обоих видов дизентерии не превышали 0,22 и во всех случаях были меньше соответствующих значений в предсезонный период.

Таблица 39

Коэффициенты корреляции между заболеваемостью дизентерией и природно-климатическими факторами в г. И. (с учетом внутригодовой динамики)

Внутригодовой период	Этиологическая форма	Температура			Влажность	
		воздуха	воды	почвы	абсолютная	относительная
Предсезонный	Дизентерия Зонне	0,32 25	0,55** 25	0,62** 18	0,53** 25	0,63** 25
Сезонный	Дизентерия Флекснера	0,56* 20	0,56* 20	0,45 14	0,48* 20	0,23 20
	Дизентерия Зонне	0,01 31	0,16 31	– 0,08 24	– 0,02 31	– 0,22 31
Межсезонный	Дизентерия Флекснера	0,01 29	0,43* 29	0,07 22	0,02 29	– 0,15 29
	Дизентерия Зонне	– 0,17 40	-	-0,16 30	– 0,18 40	0,50** 40
	Дизентерия Флекснера	– 0,05 47	-	0,16 36	0,01 47	0,06 47

Примечание. В числителе – коэффициенты корреляции, в знаменателе – число месяцев динамического ряда.

* Связь достоверная ($P < 0,05$).
** Связь высокодостоверная ($P < 0,01$).

Таблица 40

Пример вычисления коэффициента корреляции рангов Спирмена между заболеваемостью дизентерией (*x*) и температурой воздуха (*y*) в предсезонный период при «вертикальном» анализе

Год	Месяц	*x*	*y*	Ранг		$x-y$	$(x-y)^2$
				x	*y*		
1977	V	3,68	7,9	14	9,5	4,5	20,25
	VI	3,84	15,4	17	21	– 4	16
	VII	8,80	17,3	28	26	2	4
1978	IV	4,16	2,5	20	7	13	169
	V	4,56	8,8	23	12,5	10,5	110,25
	VI	4,48	16,3	21	23	– 2	4
	VII	8,00	17,5	27	27	0	0
1979	IV	2,72	0,6	4,5	6	– 1,5	2,25
	V	2,76	10,0	6	14	– 8	64
	VI	4,56	18,6	22	28	– 6	36
1980	V	1,52	7,9	2	9,5	– 7,5	56,25
	VI	4,00	16,5	18	24,5	– 6,5	42,25
1981	V	2,72	10,7	4,5	16	– 11,5	132,25
	VI	2,80	14,7	7	18	11	121
1982	III	1,28	– 9,2	1	1	0	0
	IV	3,40	3,4	12	8	4	16
	V	3,16	8,8	9	12,5	– 3,5	12,25
	VI	4,12	15,3	19	19,5	– 0,5	0,25
	VII	3,72	16,5	15	24,5	– 9,5	90,25
1983	III	2,04	– 5,0	3	3	0	0
	IV	5,36	– 1,2	25	5	20	400
	V	2,96	8,6	8	11	– 3	9
	VI	3,76	13,6	17	17	0	0
	VII	7,72	15,3	26	19,5	6,5	42,25
1984	III	3,24	– 9,0	10	2	8	64
	IV	3,64	– 1,5	13	4	9	81
	V	3,36	10,1	11	15	– 4	16
	VI	4,96	15,5	24	22	2	4
Σ							1512,50

Полученные данные свидетельствуют, что условия для запуска сезонных подъемов дизентерии посредством природно-климатических факторов формируются в предсезонном эпидемическом периоде.

Вычисленные коэффициенты корреляции при «вертикальном» анализе приведены (табл. 40) для предсезонного периода. Для других эпидемических периодов года расчеты проводятся аналогичным образом:

$$\rho = 1 - \frac{6 \cdot \sum d^2}{n(n^2 - 1)} = 1 - \frac{6 \cdot 1512{,}5}{28(28^2 - 1)} = 1 - \frac{9075{,}0}{21924} = 1 - 0{,}41 = 0{,}59.$$

Таким образом, коэффициент корреляции позволяет оценить направление и силу связи между признаками. Однако коэффициент корреляции характеризует только линейную связь и непригоден для измерения криволинейной зависимости между переменными x и y. В таких случаях используется показатель, предложенный Пирсоном, который называется корреляционным отношением и обозначается η. Корреляционное отношение характеризует связь между варьирующими признаками двусторонне – y по x и x по y, а потому и выражается не одним, а двумя показателями $\eta_{y/x}$, $\eta_{x/y}$[13].

Получить представление о том, насколько в среднем может измениться один из признаков при изменении другого, связанного с ними признака, можно путем вычисления коэффициента регрессии

$$R_{x/y} = r\frac{\sigma_x}{\sigma_y} \text{ и } R_{y/x} = r\frac{\sigma_y}{\sigma_x},$$

где R – коэффициент регрессии; r – коэффициент корреляции; σ_x и σ_y – среднеквадратические отклонения для ряда x и y соответственно.

Как видно из приведенных формул, регрессионный анализ неотделим от корреляционного. Однако в отличие от последнего показатели регрессии измеряют отношения между коррелированными признаками x и y двусторонне, т.е. учитывают изменения x в зависимости от изменений y и наоборот – изменения y от x. Исключением служат ряды динамики, показывающие изменение признаков во времени. Регрессия таких рядов оказывается односторонней.

Рассмотрим (см. табл. 36) насколько в среднем увеличивается число больных дизентерией Флекснера при увеличении на 1 % удельного веса нестандартных проб воды (по коли-индексу). Из промежуточных этапов вычисления коэффициента корреляции (см. табл. 36) легко получить соответствующие значения средних квадратических отклонений:

$$\sigma_x = \sqrt{\frac{\sum d^2 x}{n-1}} = \sqrt{\frac{2406{,}68}{11}} = 14{,}8,$$

$$\sigma_y = \sqrt{\frac{\sum d^2 y}{n-1}} = \sqrt{\frac{90{,}07}{11}} = 2{,}9.$$

Используя полученные данные, находим исходный коэффициент регрессии:

$$R_{x/y} = 0,91 \cdot \frac{14{,}8}{2{,}9} = 4{,}6 \text{ случая},$$

т.е. с увеличением удельного веса нестандартных проб питьевой воды на 1 % число больных дизентерией Флекснера увеличивается в среднем на 4,6 случая. Соответственно этому увеличение числа больных на 1 случай свидетельствует об увеличении удельного веса нестандартных проб воды на

$$R_{y/x} = 0{,}91 \cdot \frac{2{,}9}{14{,}8} = 0{,}2\,\%.$$

Таким образом, показатели регрессии – величины именованные и характеризуют зависимость между признаками. Показатели же корреляции – величины относительные, измеряющие связь между показателями в долях единицы.

Причинно-следственные отношения между явлениями сложны и многообразны. Известно, что вариации биологических процессов связаны с влиянием самых различных причин. Оценивая с помощью критерия Стьюдента или Фишера статистическую значимость различий двух показателей, полученных для сравниваемых групп, нельзя одновременно оценить степень влияния отдельных факторов на их уровень. Даже если достоверно установлена связь рассматриваемого фактора с уровнем заболеваемости, то еще остается невыясненным, какова доля этого влияния среди прочих факторов. Эти вопросы решаются с помощью дисперсионного анализа, причем наилучшие результаты получаются при сборе первичных данных с учетом большинства факторов, систематически влияющих на здоровье изучаемых контингентов.

Сущность дисперсионного анализа сводится к разложению общей дисперсии некоторого комплекса наблюдений на его составные части, из которых выявляется влияние отдельных факторов на изучаемый признак.

С помощью этого анализа можно определить долю влияния нескольких факторов, вместе взятых, и роль каждого из них – сравнить действие отдельных факторов между собой. Таким образом, дисперсионный анализ позволяет более глубоко изучить влияние различных условий и их комбинированное воздействие, а чем больше учтено факторов, формирующих уровни заболеваемости, тем более точна будет информация о причинах заболевания и более обоснованы профилактические рекомендации.

Следует отметить, что для решения дисперсионных комплексов необходимо выполнение большого объема сложных вычислений, поэтому проведение дисперсионного анализа целесообразно с использованием ПК.

Более подробно методика дисперсионного анализа приведена в различных пособиях по статистике [12, 21, 34].

Глава 6

ПРОГНОЗИРОВАНИЕ ЗАБОЛЕВАЕМОСТИ

Прогнозирование инфекционной заболеваемости является одним из самых сложных вопросов эпидемиологии. Знание и изучение его чрезвычайно важно в связи с тем, что для санитарно-эпидемиологической службы он является основой (первым этапом) рационального планирования профилактических и противоэпидемических мероприятий.

В научной литературе предложены различные сроки прогнозирования эпидситуации: от оперативного (до 1 мес) до долгосрочного (более 20 лет). Для эпидемиологических целей можно воспользоваться следующими прогнозами: краткосрочным (в течение года), среднесрочным (от 1 до 2 лет) и долгосрочным (с заблаговременностью свыше 2 лет). Долгосрочное прогнозирование важно при планировании кардинальных мероприятий, краткосрочное – для проведения оперативных мер [1].

В настоящее время имеется ряд методов научно обоснованного качественного и количественного прогнозирования инфекционной заболеваемости, которые позволяют предсказать с определенной вероятностью время (сроки) эпидемических подъемов и спадов заболеваемости (качественный прогноз), а также уровень заболеваемости в тот или иной момент (количественный прогноз).

Качественный прогноз может осуществляться по линии многолетней тенденции и типу периодических многолетних колебаний заболеваемости (см. гл. 3), а также по характеру типовой кривой годовой ее динамики или по графикам «контрольных» уровней при оперативном слежении за уровнем заболеваемости (см. гл. 4).

Для количественного прогноза отдельных этиологических форм инфекционных заболеваний предложено большое количество методов. Однако большая их часть не используется в эпидемиологической практике из-за сложности математического аппарата и высокой погрешности методов. Одним из немногих исключений являются методы прогнозирования заболеваемости вирусным гепатитом А и дизентерией, использующие в

качестве исходного параметра заболеваемость за определенный отрезок предсезонного периода [1]. Отличительной особенностью этих методов является небольшая трудоемкость вычислений, что позволяет широко использовать их в работе практических органов здравоохранения.

Для краткосрочного прогноза годовых показателей заболеваемости вирусным гепатитом А и дизентерией указанными авторами предложено два методических подхода: с использованием определения «максимальной стабильности» и регрессионного анализа.

Необходимые условия для краткосрочного прогнозирования заболеваемости по предлагаемым методам:

1. Более высокая точность прогнозов достигается на территориях с большим абсолютным числом заболевших, поэтому прогнозирование возможно только для крупных городов и административных территорий, так как на большой территории или при большой численности населения нивелируется элемент случайности, который может привести к ошибочным результатам прогнозирования. При этом минимальная численность населения района не должна быть ниже 60 тыс. чел.

2. Необходимо использовать данные месячной заболеваемости за 15 – 20-летний период. Не следует брать слишком короткие (менее 10 лет) и слишком длинные (более 20 лет) временные ряды: короткий ряд может не включить полного эпидемического цикла на некоторых территориях, слишком длинный нежелателен из-за изменения характера эпидемического процесса на протяжении более чем 20-летнего промежутка времени.

3. Следует анализировать те инфекции, при которых в текущем году не применялись эффективные средства иммунопрофилактики. Так, например, проведение в прогнозируемом году массовой предсезонной вакцинопрофилактики детского населения с целью снижения заболеваемости вирусным гепатитом А может сопровождаться значительным (по сравнению с прогнозируемым) снижением уровня заболеваемости.

Рассмотрим поэтапно оба метода краткосрочного прогнозирования.

Метод прогнозирования с использованием определения максимальной стабильности основан на выявлении в первом полугодии периода (месяца, квартала) с наименьшей колеблемостью уровня заболеваемости вирусным гепатитом или дизентерией. Пример расчета прогноза заболеваемости вирусным гепатитом на 1981 г. заимствован из работ М.Д. Алейник и др. [1] (табл. 41). Порядок вычислений следующий:

1. Составить таблицу помесячной заболеваемости (в интенсивных показателях) за выбранный период времени.

2. Определить долю заболеваемости по месяцам (в % от годового уровня) за первое полугодие (январь – июнь). Результаты занести в таблицу.

3. Вычислить среднемесячную долю заболеваемости (M) каждого месяца по суммарным данным анализируемых лет по формуле

$$M = \frac{\text{сумма за месяц}}{\text{число анализируемых лет}},$$

пример за январь:

$$M = \frac{156{,}7}{16} = 9{,}8.$$

4. Определить колеблемость долевых показателей каждого ме-

Таблица 41

Пример расчета краткосрочного прогноза заболеваемости вирусным гепатитом с использованием метода «максимальной стабильности», % от годовой заболеваемости

Год	Месяц					
	I	II	III	IV	V	VI
1965	7,6	7,4	6,6	4,8	5,6	5,6
1966	12,3	11.0	9,7	5,7	4,7	4,4
1967	15,5	9,8	10,3	7,5	5,2	6,3
1968	6,9	6,4	6,4	5,8	5,2	6,4
1969	10,5	9,2	7,9	6.1	4,8	4,4
1970	10,6	8,8	8,3	6,0	7,4	6,0
1971	6,0	5,8	4,9	3,4	3,4	3,0
1972	14,6	8,4	8,5	6,1	6,1	5,7
1973	11,9	10,2	9,8	6,8	7,0	6,3
1974	13,2	8,0	8,2	7,7	5,8	5,2
1975	7,9	7,2	8,5	6,0	5,5	5,7
1976	6,5	8,3	6,5	5,6	6,5	6,5
1977	7,4	7,5	6,1	5,3	5,4	5,5
1978	9,6	7,6	9,1	8,2	6,9	7,5
1979	6,6	3,7	4,5	6,7	10,7	4,6
1980	9,6	6,7	8,9	5,6	4,5	3,3
M	9,8	7,9	7,8	5,7	5,9	5,4
Колеблемость доли месяца	6 – 15,5	3,7 – 11	4,5 – 10,3	3,4 – 8,2	3,4 – 10,7	3 – 7,5
A	9,5	7,3	5,8	4,8	7,3	4,5
S	9,2	6,7	4,3	4,0	9,0	3,8

сяца путем нахождения минимального и максимального значения в анализируемом ряду. Результаты записать в соответствующую строку таблицы.

5. Определить амплитуду колеблемости (*A*) в пределах ряда каждого месяца по формуле:

A = максимальный показатель – минимальный показатель.

Результаты занести в таблицу.

6. Вычислить показатель стабильности (*S*) для каждого месяца по формуле

$$S = A^2/M.$$

Результаты записать в соответствующую строку таблицы.

Пример за январь: $5 = 9{,}5^2/9{,}8 = 9{,}2$ и т. д,

7. Провести расчет прогнозируемого уровня заболеваемости вирусным гепатитом текущего года (в нашем случае 1981 г.). Для этого определяется месяц с наименьшей стабильностью. В рассматриваемом примере максимальная стабильность приходится на июнь (3,8), а также весьма близкий к нему показатель в апреле (4,0). Прогнозирование по апрелю предпочтительнее как более раннее.

Показатель заболеваемости в апреле 1981 г. составлял 8,8 на 100 тыс., а среднемесячная доля апреля в годовой заболеваемости (*M*) – 5,7 %. Отсюда путем решения пропорции получаем годовой прогнозируемый показатель – 154 на 100 тыс.:

$$8{,}8 – 5{,}7\ \%$$
$$x – 100\ \%.$$

Фактически по итогам года зарегистрирован очень близкий показатель – 157 на 100 тыс. Таким образом, в данном случае апрель может быть принят за «ключевой» период для прогнозирования годовой заболеваемости.

В рассматриваемом выше методе качество прогноза зависело от степени стабильности какого-то начального периода. Исходя из этой предпосылки, можно независимо от рассмотренного выше использовать метод краткосрочного прогнозирования с применением уравнения регрессии. Предлагаемый метод позволяет по заболеваемости месяцев «ключевого» периода определить ожидаемый уровень заболеваемости (годовой показатель) текущего года.

Предварительно рассчитывается коэффициент корреляции между заболеваемостью I или II квартала и годовой в относительных показателях (на большинстве территорий можно использовать показатели I квартала). Обычно избирается период максимальной стабильности (см. табл. 41).

Рассмотрим применение этого метода по этапам, воспользовавшись примером уже указанных авторов [1]. Алгоритм расчета следующий:

1. Составить две матрицы аналитических таблиц для I и II кварталов анализируемых лет (в нашем случае составлена одна таблица для I квартала, на примере которой будут приведены все расчеты (табл. 42).

2. Внести в таблицу анализируемые годы, годовые показатели заболеваемости, квартальные показатели каждого года.

3. Определить среднеарифметическую:

а) годовых показателей заболеваемости ($\bar{x}$);

б) квартальных показателей заболеваемости ($\bar{y}$).

4. Вычислить отклонения от среднеарифметической:

Таблица 42

Пример расчета краткосрочного прогноза заболеваемости вирусным гепатитом с использованием уравнения регрессии

Год	За год ($\bar{x}$)	За I квартал ($\bar{y}$)	Отклонение от средней		dx^2	dy^2	$dx \cdot dy$	Прогноз x_1
			dx	dy				
1967	182	52	14	14	196	196	196	220
1968	216	57	48	19	2304	361	912	239
1969	225	48	57	10	3249	100	570	205
1970	270	56	102	18	10440	324	1836	235
1971	220	55	52	17	2704	289	884	232
1972	215	44	47	6	2209	36	282	190
1973	198	43	30	5	900	25	150	186
1974	147	25	-21	-13	441	169	273	119
1975	109	23	-59	-15	3481	225	885	111
1976	94	24	-74	-14	5476	196	1036	115
1977	95	20	-73	-18	5329	324	1314	100
1978	98	24	-70	-14	4900	196	980	115
1979	141	33	-27	-5	729	25	135	149
1980	135	27	-33	-11	1089	121	363	126
S					43411	2587	9816	
Среднеарифметическая	168	38						

а) годовых показателей (dx) и внести эти данные в таблицу;

б) квартальных, показателей (dy) и внести эти данные в соответствующую графу.

5. Определить квадрат отклонения годовых и квартальных показателей заболеваемости и результаты записать (dx^2 и dy^2).

6. Вычислить сумму квадратов отклонений годовых показателей заболеваемости ($\Sigma\, dx^2$) и показателей за квартал ($\Sigma\, dy^2$) от их среднемноголетних. Результаты записать в графы таблицы.

7. Определить произведение отклонений годовых и квартальных показателей ($dx \cdot dy$) в каждом ряду с последующим вычислением суммы произведения отклонений ($\Sigma\, dx \cdot dy$). Результаты записать в графу таблицы.

8. Определить коэффициент корреляции (r) среднеквартального показателя со среднегодовым уровнем по общепринятой формуле

$$r = \frac{\sum dx \cdot dy}{\sqrt{\sum dx^2 \cdot \sum dy^2}}.$$

9. По наибольшему коэффициенту корреляции принять решение – уровень заболеваемости какого квартала (I или II) можно использовать для вычисления прогнозируемого годового уровня заболеваемости. В нашем примере этот показатель равен + 0,92, что свидетельствует о тесной прямой связи между сравниваемыми величинами, и прогноз заболеваемости следует рассчитывать исходя из показателя I квартала.

10. Вычислить прогнозируемый годовой уровень заболеваемости, используя уравнения регрессии

$$r\frac{\sigma_x}{\sigma_y}(y - \bar{y}) + \bar{x}.$$

Уравнение решается на основании данных табл. 35, откуда определяются среднегодовая ($\bar{x}$) и среднеквартальная ($\bar{y}$) заболеваемость за ряд лет и среднеквадраческие отклонения от них (σ_x и σ_y) по формулам

$$\sigma_x = \sqrt{\frac{\sum dx^2}{n}} = 55{,}7,$$

$$\sigma_y = \sqrt{\frac{\sum dy^2}{n}} = 13{,}6.$$

В результате преобразования уравнения регрессии получаем формулу прогноза: 3,76 y + 24,8, позволяющую рассчитать заболеваемость на текущий год, подставив показатель заболеваемости в I квартале текущего года. Так, показатель заболеваемости вирусным гепатитом на 1980 г. равен 3,76 · 27 + 24,8 = 126,32 ≈ 126 $^{0}/_{0000}$. Расхождение с фактической заболеваемостью составляет 6,7 %.

11. Поскольку полученное значение не может точно совпадать с фактически зарегистрированным годовым показателем заболеваемости, то следует рассчитать ошибку прогноза по формуле

$$m = \pm \sigma_x \sqrt{1 - r^2}.$$

С учетом изменяющихся особенностей эпидемического процесса, формула прогноза на каждой территории составляет 1 раз в 3 года.

Для оценки достоверности прогноза заболеваемости авторами предложены следующие критерии по совпадению показателя прогнозируемой и регистрируемой заболеваемости: отклонения до 15 % – хороший результат, от 16 до 25 % – удовлетворительный, более 25 % – неудовлетворительный.

Апробация рассмотренных методов прогнозирования на различных административных территориях показала, что хорошие и удовлетворительные результаты при вирусном гепатите отмечены в 93 %, а при дизентерии – в 80 %.

Среднесрочный прогноз заболеваемости разработан этими же авторами [1]. Необходимые условия для прогнозирования заболеваемости на следующий год те же, что и для краткосрочного прогноза. Следует лишь отметить, что среднесрочный прогноз применим только для вирусного гепатита. При этом минимальная численность населения административной территории должна быть не менее 200 тыс. Среднесрочный прогноз расчитывается с использованием одноступенчатой парной регрессии. В этом случае наряду с заболеваемостью периода «максимальной стабильности» первого полугодия текущего года (x_1) прогнозируется годовой показатель последующего (x_2) года.

Последовательность расчетов такая же, как при краткосрочном прогнозе. Единственное отличие состоит в том, что при расчете формулы про-

гноза в исходной таблице x последующего года должна соответствовать строке y предыдущего (например x 1967 г. – y 1966 г. и т. д.).

Этапы расчета среднесрочного прогноза приведены в табл. 43 с использованием данных предыдущего примера. Не останавливаясь на деталях расчета, ибо они были подробно изложены ранее, отметим, что коэффициент корреляции в этом случае равен +0,90, σ_z – 57,63 и σ_y – 13,75. Подставляем в уравнение регрессии данные табл. 43:

$$z_1(\textit{прогнозируемая}) = 0{,}90 \cdot \frac{57{,}63}{13{,}75}\,(y - 39) + 166.$$

В результате преобразования уравнения регрессии получаем формулу прогноза: $z_1 = 3{,}77y + 18{,}97$, которая позволяет рассчитать заболеваемость на следующий год по данным заболеваемости I квартала текущего года. Например, прогнозируемый показатель заболеваемости вирусным гепатитом на 1980 г. равен

$$3{,}77 \cdot 33 + 18{,}97 = 143{,}38 \approx 143\ {}^{0}/_{0000}.$$

Расхождение с фактической заболеваемостью составляет 5,9 %.

Таблица 43

Пример расчета среднегодового прогноза заболеваемости вирусным гепатитом А

Год	Заболеваемость за I квартал (y)	Год	Заболеваемость за год (x)	Отклонение от средней		dx^2	dy^2	$dx \cdot dy$	Прогноз x_1
				dx	dy				
1967	52	1968	216	50	13	2500	169	650	215
1968	57	1969	225	59	18	3481	324	1062	234
1969	48	1970	270	104	9	10816	81	936	200
1970	56	1971	220	54	17	2916	289	918	230
1971	55	1972	215	49	16	2401	256	784	226
1972	44	1973	198	32	5	1024	25	160	185
1973	43	1974	147	– 19	4	361	15	– 76	181
1974	25	1975	109	– 57	– 14	3249	196	798	113
1975	23	1976	94	– 72	– 16	5184	256	1152	106
1976	24	1977	95	– 71	– 15	5041	225	1065	109
1977	20	1978	98	– 68	– 19	4624	361	1292	94
1978	24	1979	141	– 25	– 15	625	225	375	109
1979	33	1980	135	– 31	– 6	961	36	186	143
Сумма…						43183	2459	9302	
Среднеарифметическое	39		166						

Учитывая, что с увеличением долгосрочности прогноза возрастает и его ошибка, авторами предложена менее «строгая» шкала для оценки оправдываемости прогноза: отклонение до 20 % – хороший результат, до 30 % – удовлетворительный, более 30 % – неудовлетворительный. Хорошие и удовлетворительные результаты прогнозов получены в 80,2 % при их проверке на территориях России.

Также предложены методы прогнозирования годовых показателей заболеваемости (краткосрочный прогноз) отдельно для дизентерии Зонне [15] и для дизентерии Флекснера [19]. Показано, что наиболее сильное влияние как на сезонный подъем, так и на годовой уровень заболеваемости дизентерией Зонне оказывают среднемесячная температура воздуха июня – июля и число дней в эти месяцы с температурой выше 25 °С. Для дизентерии Флекснера имеется тесная связь с обратным плювеотермическим коэффициентом (ОПТК) июня – июля текущего года, взятого в виде его аномалий, а также с числом случаев дизентерии с января по июль, учтенная с диагнозом, подтвержденным лабораторными методами (выделение возбудителя, обнаружение антител). ОПТК представляет собой отношение среднемесячной температуры воздуха к сумме осадков в этом же месяце. Оба эти метода (для дизентерии Зонне и Флекснера) более трудоемки, чем ранее рассмотренные. Кроме этого, прогноз на текущий год может быть выдан не раньше августа, а методы М.Д. Алейник и др. позволяют выдать прогноз, как правило, не позже апреля.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. **Алейник М.Д., Кривицкая В.С., Стриковская И.Х., Трифонов В.А.** Разработка методов прогнозирования заболеваемости вирусным гепатитом // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1983. – № 2. – С. 107 – 112.

2. **Астафьев В.А., Савилов Е.Д.** Определение сроков сезонных подъемов заболеваемости расчетным методом // Там же. – 1991. – № 2. – С. 36 –38.

3. **Астафьев В.А., Савилов Е.Д., Лебедев В.И. и др.** Многолетняя динамика сезонной и межсезонной заболеваемости дизентерией в г. Иркутске // Бюл. СО РАМН СССР. – 1986. – № 2 – С. 59 – 62.

4. **Астафьев В.А., Савилов Е.Д., Попов Л.В. и др.** Роль овощей и фруктов в сезонной заболеваемости дизентерией // Окружающая среда и здоровье человека. – Иркутск, 1990. – С. 442 – 447.

5. **Ашмарин Н.П., Васильев Н.Н., Амбросов В.А.** Быстрые методы статистической обработки и планирование экспериментов. – Л.: Изд-во Ленингр. ун-та, 1975. – 80 с.

6. **Ашмарин Н.П., Воробьев А.А.** Статистические методы в микробиологических исследованиях. – Л.: Медгиз., 1962. – 180 с.

7. **Беляков В.Д.** Сезонность дизентерии. Выбор способа определения начала сезонных подъемов заболеваемости: Сообщение 2 // Мед. паразитология и паразитарные болезни. – 1970. – Т. 39, № 6. – С. 722 –727.

8. **Беляков В.Д.** Эпидемический процесс: Теория и метод изучения. – М., 1964. – 244 с.

9. **Беляков В.Д., Дегтярев А.А., Иванников Ю.Г.** Качество и эффективность противоэпидемических мероприятий. – Л., 1981. – 340 с.

10. **Бессмертный Б.С.** Математическая статистика в клинической, профилактической и экспериментальной медицине. – М.: Медицина, 1967. – 303 с.

11. **Бессмертный Б.С., Ткачева М.Н.** Статистические методы в эпидемиологии / под ред. А. М. Меркова. – М.: Гос. изд-во мед. лит-ры, 1961. – 204 с.

12. **Венецкий И.Г.** Корреляционный анализ и его применение в статистических исследованиях // Вестн. статистики. – 1981. – № 7. – С. 40 – 47.

13. **Венецкий И.Г.** Корреляционный анализ и его применение в статистических исследованиях // Вестн. статистики. – 1982. – № 3. – С. 37 – 50.

14. **Венецкий И.Г., Венецкая В.И.** Основные математико-статистические понятия и формулы в экономическом анализе: справочник. – 2-е изд. – М.: Статистика, 1979. – 448 с.

15. **Власов В.И., Замотин Б. А.** Краткосрочное прогнозирование годовых показателей заболеваемости дизентерией вида Зонне в отдельных группах населения: метод. рекомендации. – Рязань, 1981. – 14 с.

16. **Гланц С.** Медико-биологическая статистика: Пер. с анг. – М.:, Практика, 1998. – 459 с.

17. **Гублер Е.В., Генкин А.А.** Применение непараметрических критериев статистики и медико-биологических исследованиях. – М.: Медицина, 1973. – 142 с.

18. **Дегтярев А.А., Ходырев А.П.** Методика анализа годовой динамики инфекционной заболеваемости // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1976. – № 2. – С. 97 –102.

19. **Краткосрочное** прогнозирование годовых показателей заболеваемости дизинтерией Флекснера: метод. рекомендации – Новокузнецк, 1984. – 11 с.

20. **Лакин Г.Ф.** Биометрия. – 2-е изд. – М.: Высш. шк., 1973. – 344 с.

21. **Мерков А.М., Поляков Л.Е.** Санитарная статистика: пособие для врачей. – Л.: Медицина, 1974. – 384 с.

22. **Ноткин Е.Л.** Статистика в гигиенических исследованиях. – М.: Медицина, 1965. – 259 с.

23. **Плохинский Н.А**. Биометрия. – М.: Изд-во МГУ, 1970. – 368 с.

24. **Прогнозирование** заболеваемости вирусным гепатитом: метод. рекомендации. – Горький, 1984. – 15 с.

25. **Рыкушин Ю.П., Исаева М.С.** О контрольных уровнях заболеваемости дизентерией для эпиднадзора за этой инфекцией // Острые кишечные инфекции: респ. сб. науч. тр. – Л., 1986. – Вып. 10. – С. 17 – 21.

26. **Савилов Е.Д., Алексеева Г.И., Мальцева М.В. и др.** Методический подход к оценке эпидемиологической ситуации по обобщенному критерию // Эпидемиология и инфекционные болезни. – 2011. – № 2. – С. 17 – 20.

27. **Савилов Е.Д., Астафьев В.А.** Внутригодовая динамика заболеваемости дизентерией на фоне ее многолетнего снижения // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1989. – № 3. – С. 114 –115.

28. **Савилов Е.Д., Астафьев В.А.** Количественная оценка достоверности тенденций в ретроспективном эпидемиологическом анализе // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1989. – № 4. – С. 107 – 108.

29. **Савилов Е.Д., Астафьев В.А.** Применение непараметрических критериев статистики в эпидемиологических исследованиях // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1987. – № 11. – С. 36 – 38.

30. **Савилов Е.Д., Астафьев В.А., Попова О.А., Музыка Ю.С.** О сроках наступления и окончания сезонных подъемов дизентерии в различных возрастных группах населения // Бюл. СО АМН СССР. – 1984. – № 6. – С. 81 – 83.

31. **Савилов Е.Д., Мамонтова Л.М., Астафьев В.А., Жданова С.Н.** Применение статистических методов в эпидемиологическом анализе. – М.: МЕДпресс-информ, 2004. – 112 с.

32. **Сергиенко В.И., Бондарева И.Б.** Математическая статистика в клинических исследованиях. – М.: ГЭОТАР-Мед., 2001. – 256 с.

33. **Статистический словарь.** – 2-е изд. – М.: Финансы и статистика, 1989. – 623 с.

34. **Урбах В.Ю.** Биометрические методы: Статистическая обработка опытных данных в биологии, сельском хозяйстве и медицине. – М.: Наука, 1964. – 416 с.

35. **Урбах В. Ю.** Статистический анализ в биологических и медицинских исследованиях. – М.: Медицина, 1975. – 296 с.

36. **Шаханина И.Л., Кучеровская Т. В., Чернова Т. П.** Применение различных статистических методов эпидемиологического анализа при оценке сезонности и территориального распространения инфекционных болезней // Журн. микробиологии, эпидемиологии и иммунобиологии. – 1990. – № 5. – С. 43 –47.

37. **Шляхов Э.Н.** Практическая эпидемиология. – Кишинев: Штиинца, 1986. – 528 с.

38. **Юзбашев М., Афанасьев В.** Коэффициент корреляции рангов как показатель устойчивости динамики // Вестн. статистики. –1983. – № 11. – С. 39 – 40.

39. **Ягодинский В.Н.** Динамика эпидемического процесса. – М., 1977. – 240 с.

ПРИЛОЖЕНИЕ

Таблица 1

Критерии для исключения выскакивающих значений (максимальных и минимальных) [5]

n	Уровень достоверности, %		*n*	Уровень достоверности, %	
	95	99		95	99
3	0,941	0,988	10	0,412	0,527
4	0,765	0,889	11	0,392	0,502
5	0,642	0,780	12	0,376	0,482
6	0,560	0,698	15	0,338	0,438
7	0,507	0,637	20	0,300	0,391
8	0,468	0,590	24	0,281	0,367
9	0,437	0,555	30	0,260	0,341

Таблица 2

Норма опытов, результаты которых могут быть приняты в качестве доверительного интервала медианы при условии, что они расположены и пронумерованы в порядке возрастающих значений [10]

n	Уровень достоверности, %				*n*	Уровень достоверности, %			
	95		99			95		99	
	Граница					Граница			
	нижняя	верхняя	нижняя	верхняя		нижняя	верхняя	нижняя	верхняя
1	2	3	4	5	6	7	8	9	10
6	1	6	-	-	26	8	19	7	20
7	1	7	0	0	27	8	20	7	21
8	1	8	1	8	28	9	20	7	22
9	2	8	1	9	29	9	21	8	22
10	2	9	1	10	30	10	21	8	23
11	2	10	1	11	31	10	22	8	24
12	3	10	2	11	32	10	23	9	24
13	3	11	2	12	33	11	23	9	25
14	3	12	2	13	34	11	24	10	25

Окончание табл. 2

1	2	3	4	5	6	7	8	9	10
15	4	12	3	13	35	12	24	10	26
16	4	13	3	14	36	12	25	10	27
17	5	13	3	15	37	13	25	11	27
18	5	14	4	15	38	13	26	11	28
19	5	15	4	16	39	13	27	12	28
20	6	15	4	17	40	14	27	12	29
21	6	16	5	17	41	14	28	12	30
22	6	17	5	18	42	15	28	13	30
23	7	17	5	19	43	15	29	13	31
24	7	18	6	19	44	16	29	14	31
25	8	18	6	20	45	16	31	14	32

Таблица 3

Коэффициент А для вычисления среднего квадратического отклонения по “размаху”[33]

n	*A*	*n*	*A*	*n*	*A*
6	2,53	18	3,64	40	4,32
7	2,70	19	3,69	50	4,50
8	2,85	20	3,74	60	4,64
9	2,97	22	3,82	70	4,76
10	3,08	24	3,90	80	4,85
11	3,17	26	3,69	90	4,94
12	3,26	28	4,03	100	5,01
13	3,34	30	4,09	120	5,15
14	3,41	32	4,14	140	5,26
15	3,47	34	4,19	160	5,35
16	3,53	36	4,24	180	5,43
17	3,59	38	4,28	200	5,50

Таблица 4

Коэффициент В для вычисления ошибки средней по «размаху»[5]

n	*B*	*n*	*B*	*n*	*B*
6	6,61	18	15,40	40	27,30
7	7,16	19	16,10	50	31,80
8	8,05	20	16,70	60	35,90
9	8,90	22	17,90	70	39,80
10	9,70	24	19,00	80	43,30
11	10,05	26	20,20	90	46,90
12	11,20	28	21,20	100	50,10
13	12,00	30	22,40	120	56,30
14	12,70	32	23,40	140	62,30
15	13,40	34	24,60	160	67,60
16	14,10	36	25,50	180	73,00
17	14,80	38	26,40	200	77,80

Таблица 5

Критические значения критерия Стьюдента (*t*) [16]

Число степеней свободы	Уровень достоверности, %		Число степеней свободы	Уровень достоверности, %	
	95	99		95	99
1	2	3	4	5	6
2	4,30	9,93	8	2,31	3,36
3	3,18	5,84	9	2,26	3,25
4	2,78	4,60	10	2,23	3,17
5	2,57	4,03	11	2,20	3,11
6	2,45	3,71	12	2,18	3,06
7	2,37	3,50	13	2,16	3,01

Окончание табл. 5

1	2	3	4	5	6
14	2,14	2,98	27	2,05	2,77
15	2,13	2,95	28	.2,05	2,76
16	2,12	2,92	29	2,04	2,76
17	2,11	2,90	30	2,04	2,76
18	2,10	2,88	40	2,02	2,70
19	2,09	2,86	50	2,01	2,68
20	2,09	2,84	60	2,00	2,66
21	2,08	2,83	80	1,99	2,64
22	2,07	2,82	100	1,98	2,63
23	2,07	2,81	120	1,97	2,60
24	2,06	2,80	200	1,97	2,60
25	2,06	2,79	500	1,96	2,59
26	2,06	2,78		1,96	2,58

Примечание. Нулевая гипотеза принимается при $t_{\text{факт}} \le t_{0,05}$ и отвергается при $t_{\text{факт}} > t_{0,05}$.

Таблица 6

Критические значения критерия Фишера (*F*) [16]

Число степеней свободы знаменателя ($K_2 = n - 1$)	Число степеней свободы числителя ($K_1 = n - 1$)										
	1	2	3	4	5	6	7	8	9	10	∞
1	2	3	4	5	6	7	8	9	10	11	12
2	18,51 98,49	19,00 99,01	19,16 99,17	19,25 99,25	19,30 99,30	19,33 99,33	19,36 99,34	19,37 99,36	19,38 99,38	19,39 99,40	19,50 99,50
3	10,13 34,12	9,55 30,81	9,28 29,46	9,12 28,71	9,01 28,24	8,94 27,91	8,88 27,67	8,84 27,49	8,81 27,34	8,78 27,23	8,53 26,12
4	7,71 21,20	6,94 18,00	6,59 16,69	6,39 15,98	6,26 15,52	6,16 15,21	6,09 14,98	6,05 14,80	6,00 14,66	5,96 14,54	5,63 13,46
5	6,61 16,26	5,59 13,27	5,41 12,06	5,19 11,39	5,05 10,97	4,95 10,67	4,88 10,45	4,82 10,27	4,78 10,23	4,74 10,05	4,36 9,02
6	5,99 13,74	5,14 10,92	4,76 9,78	4,53 9,15	4,39 8,75	4,28 8,47	4,21 8,26	4,15 8,10	4,10 7,98	4,06 7,87	3,67 6,88
7	5,99 12,25	4,74 9,55	4,35 8,45	4,12 7,85	3,97 7,46	3,87 7,19	3,79 7,00	3,73 6,84	3,68 6,71	3,63 6,62	3,23 5,65
8	5,32 11,26	4,46 8,65	4,07 7,59	3,84 7,01	3,69 6,63	3,58 6,37	3,50 6,19	3,44 6,03	3,39 5,91	4,34 5,82	2,93 4,86
9	5,12 10,56	4,26 8,02	3,86 6,99	3,63 6,42	3,43 6,06	3,37 5,80	3,29 5,62	3,23 5,47	3,18 5,35	3,13 5,26	2,71 4,31

Окончание табл. 6

1	2	3	4	5	6	7	8	9	10	11	12
10	4,96 10,04	4,10 7,56	3,71 6,56	3,48 5,99	3,33 5,64	3,22 5,39	3,14 5,21	3,07 5,06	3,02 4,95	2,97 4,85	2,54 3,91
12	4,75 9,33	3,88 6,93	3,49 5,95	3,26 5,41	3,11 5,06	3,00 4,82	2,92 4,65	2,85 4,50	2,80 4,39	2,76 4,30	2,30 3,36
14	4,60 8,86	3,74 6,51	3,34 5,56	3,11 5,03	2,96 4,69	2,85 4,46	2,77 4,28	2,70 4,14	2,65 4,03	2,60 3,94	2,13 3,00
16	4,49 8,53	3,63 6,23	3,24 5,29	3,01 4,77	2,85 4,44	2,74 4,20	2,66 4,03	2,59 3,89	2,54 3,78	2,49 3,69	2,01 2,75
18	4,41 8,28	3,55 6,01	3,16 5,09	2,93 4,58	2,77 4,25	2,66 4,01	2,58 3,85	2,51 3,71	2,46 3,60	2,41 3,51	1,92 2,57
20	4,35 8,10	3,49 5,85	3,10 4,94	2,87 4,43	2,71 4,10	2,60 3,87	2,62 3,71	2,45 3,56	2,40 3,45	2,35 3,37	1,84 2,42
24	4,26 7,82	3,40 5,61	3,01 3,72	2,78 4,22	2,62 3,90	2,51 3,67	2,43 3,50	2,36 3,36	2,30 3,25	2,26 3,17	1,73 2,21
30	4,17 7,56	3,32 5,39	2,92 4,51	2,69 4,02	2,53 3,70	2,42 3,47	2,34 3,30	2,27 3,17	2,21 3,06	2,16 2,98	1,62 2,01
40	4,08 7,31	3,23 5,18	2,84 4,31	2,61 3,83	2,45 3,51	2,34 3,29	2,25 3,12	2,18 2,99	2,12 2,88	2,07 2,80	1,51 1,81
60	4,00 7,08	3,15 4,98	2,75 4,13	2,52 3,65	2,37 3,34	2,25 3,12	2,17 2,95	2,10 2,82	2,05 2,72	1,99 2,63	1,39 1,60
100	3,94 6,90	3,09 4,82	2,70 3,98	2,46 3,51	2,30 3,20	2,19 2,99	2,10 2,82	2,03 2,69	1,97 2,59	1,92 2,51	1,28 1,43
∞	3,84 6,63	2,99 4,60	2,60 3,78	2,37 3,32	3,21 3,02	2,09 2,80	2,01 2,64	1,94 2,51	1,88 2,41	1,83 2,32	1,00 1,09

Примечание. Верхнее число – $F_{0,05}$, нижнее – $F_{0,01}$. Нулевая гипотеза принимается при $F_{\text{факт}} \le F_{0,05}$ и отвергается при $F_{\text{факт}} > F_{0,05}$.

Таблица 7

Минимальная разность, достижение которой означает, что превышение одного показателя над другим не случайное [21]

Меньший из обоих показателей, %	Число наблюдений, лежащих в основе каждого показателя																														
	20	30	40	50	60	70	80	90	100	200	300	400	500	600	700	800	900	1000	2000	3000	4000	5000	6000	7000	8000	9000	10000	20000	30000	40000	50000	100000
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	19	14	11	9,0	7,9	6,9	6,2	5,7	5,3	3,2	2,4	2,0	1,7	1,5	1,4	1,3	1,0	0,9	0,7	0,6	0,5	0,4	0,4	0,4	0,3	0,3	0,3	0,2	0,2	0,1	0,1	0,1
2	20	15	12	10	8,9	7,9	7,2	6,7	6,3	3,9	3,0	2,5	2,2	1,9	1,8	1,7	1,6	1,5	1,0	0,8	0,7	0,6	0,5	0,5	0,5	0,4	0,4	0,3	0,2	0,2	0,2	0,1
3	21	16	13	11	9,8	8,8	8,0	7,4	6,9	4,5	3,4	2,9	2,6	2,3	2,1	2,0	1,8	1,7	1,2	0,9	0,8	0,7	0,6	0,6	0,6	0,5	0,5	0,4	0,3	0,2	0,2	0,2
4	22	17	14	12	11	9,5	8,7	8,0	7,5	4,9	3,8	3,2	2,9	2,6	2,4	2,2	2,1	1,9	1,3	1,1	0,9	0,8	0,7	0,7	0,7	0,6	0,6	0,4	0,4	0,3	0,2	0,2
5	23	18	14	13	11	10	9,3	8,6	8,1	5,3	4,2	3,5	3,1	2,8	2,6	2,4	2,3	2,1	1,5	1,2	1,0	0,9	0,8	0,7	0,7	0,6	0,6	0,4	0,4	0,3	0,3	0,3
6	24	18	15	13	12	11	9,8	9,1	8,5	5,6	4,5	3,8	3,4	3,1	2,8	2,6	2,5	2,3	1,6	1,3	1,1	1,0	0,9	0,8	0,8	0,7	0,7	0,5	0,4	0,3	0,3	0,2
7	24	19	16	14	12,9	11	10	9,6	9,0	6,0	4,7	4,0	3,6	3,3	3,0	2,8	2,7	2,4	1,7	1,4	1,2	1,1	1,0	0,9	0,8	0,7	0,7	0,5	0,4	0,3	0,3	0,2
8	25	19	16	14	13	12	11	10	9,4	6,3	5,0	4,3	3,8	3,4	3,2	2,9	2,8	2,5	1,8	1,5	1,3	1,1	1,0	0,9	0,9	0,8	0,8	0,6	0,5	0,4	0,3	0,2
9	25	20	17	15	13	12	11	10	9,7	6,5	5,2	4,5	4,0	3,6	3,3	3,1	2,9	2,7	1,9	1,5	1,3	1,2	1,1	1,0	0,9	0,8	0,8	0,6	0,5	0,4	0,4	0,3
10	26	20	17	15	14	12	11	11	10	6,8	5,4	4,6	4,1	3,7	3,4	3,2	3,0	2,8	2,0	1,6	1,4	1,2	1,1	1,0	1,0	0,9	0,9	0,6	0,5	0,4	0,4	0,3
15	28	22	18	17	15	14	13	12	11	7,8	6,3	5,4	4,8	4,3	4,0	3,7	3,5	3,3	2,3	1,9	1,6	1,5	1,3	1,2	1,1	1,0	1,0	0,7	0,6	0,5	0,5	0,3
20	29	24	20	18	16	15	14	13	12	8,6	6,9	5,9	5,3	4,8	4,4	4,1	3,9	3,7	2,6	2,1	1,8	1,6	1,4	1,3	1,2	1,2	1,1	0,8	0,7	0,6	0,5	0,4
25	30	25	21	19	17	16	15	14	13	9,1	7,4	6,3	5,7	5,2	4,8	4,4	4,2	4,0	2,8	2,3	2,0	1,8	1,5	1,4	1,4	1,3	1,2	0,9	0,8	0,7	0,6	0,4
30	31	25	22	19	18	16	15	14	14	9,5	7,7	6,6	5,9	5,6	5,0	4,7	6,2	4,1	2,9	2,4	2,1	1,9	1,7	1,6	1,5	1,4	1,3	0,9	0,8	0,7	0,6	0,4
35	31	25	22	20	18	17	16	15	14	9,8	7,9	6,9	6,1	5,6	5,2	6,8	6,5	4,3	3,0	2,5	2,2	1,9	1,8	1,6	1,5	1,4	1,3	1,0	0,8	0,7	0,6	0,4
40	32	25	22	20	18	17	16	15	14	9,9	8,1	7,0	6,4	5,7	5,3	4,9	4,6	4,4	3,1	2,5	2,2	2,0	1,8	1,7	1,6	1,5	1,4	1,0	0,8	0,7	0,6	0,4
45	30	25	22	20	18	17	16	15	14	10	8,1	7,1	6,3	5,7	5,3	5,0	4,7	4,5	3,2	2,5	2,2	2,0	1,8	1,7	1,6	1,5	1,4	1,0	0,8	0,7	0,6	0,4

Окончание табл. 7

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
50	30	25	22	19	18	17	16	15	14	9,9	8,1	7,0	6,3	5,7	5,3	5,0	4,7	4,5	3,2	2,6	2,2	2,0	1,8	1,7	1,6	1,5	1,4	1,0	0,8	0,7	0,6	0,4
55	29	24	21	19	17	16	15	14	10	9,7	8,0	6,9	6,2	5,6	5,2	4,9	4,6	4,4	3,1	2,5	2,2	2,0	1,8	1,7	1,6	1,5	1,4	1,0	0,8	0,7	0,6	0,4
60	27	23	20	18	17	16	15	14	13	9,5	7,8	6,8	6,2	5,5	5,1	4,8	4,5	4,4	4,0	2,5	2,2	2,0	1,8	1,7	1,6	1,5	1,4	1,0	0,8	0,7	0,6	0,4
65	26	22	19	17	16	15	14	13	13	9,2	7,6	6,5	5,9	5,4	5,0	4,7	4,4	4,2	2,9	2,4	2,1	1,9	1,7	1,6	1,5	1,4	1,3	1,0	0,8	0,7	0,6	0,4
70	24	20	18	16	15	14	13	13	12	8,7	7,1	6,2	5,6	5,2	4,7	4,5	4,2	3,9	2,8	2,3	2,0	1,8	1,7	1,5	1,4	1,4	1,3	0,9	0,8	0,7	0,6	0,4
75	22	19	17	15	14	13	12	12	11	8,1	6,7	5,8	5,3	4,9	4,4	4,2	4,0	3,8	2,7	2,2	1,9	1,7	1,6	1,4	1,3	1,3	1,2	0,9	0,8	0,7	0,6	0,4
80	18	16	15	13	13	12	11	10	10	7,3	6,1	5,3	4,8	4,4	4,1	3,8	3,6	3,4	2,5	2,0	1,8	1,6	1,5	1,3	1,2	1,2	1,1	0,8	0,7	0,6	0,5	0,3
85	-	14	13	12	11	10	9,5	9,0	7,9	6,4	5,3	4,7	4,2	3,8	3,6	3,4	3,2	3,0	2,2	1,8	1,6	1,4	1,3	1,2	1,1	1,0	0,9	0,7	0,6	0,5	0,5	0,3
90	-	-	10	9,0	8,5	8,0	7,6	7,2	6,9	5,2	4,4	3,8	3,5	3,2	3,0	2,8	2,6	2,5	1,8	1,5	1,3	1,2	1,1	1,0	1,0	0,9	0,8	0,6	0,5	0,4	0,4	0,3
91	-	-	9,0	8,4	7,9	7,5	7,1	6,8	6,5	4,9	4,1	3,6	3,3	3,0	2,8	2,6	2,5	2,4	1,7	1,4	1,2	1,1	1,1	1,0	0,9	0,8	0,8	0,6	0,5	0,4	0,4	0,2
92	-	-	8,0	7,8	7,4	7,0	6,7	6,3	6,1	4,6	3,9	3,4	3,1	2,9	2,7	2,5	2,3	2,0	1,6	1,3	1,2	1,1	1,0	0,9	0,9	0,8	0,8	0,5	0,5	0,5	0,3	0,2
93	-	-	-	7,0	6,8	6,4	6,1	5,8	5,6	4,2	3,6	3,2	2,9	2,7	2,5	2,4	2,2	2,1	1,5	1,3	1,1	1,0	1,0	0,8	0,8	0,7	0,7	0,5	0,4	0,3	0,3	0,2
94	-	-	-	-	-	5,8	5,5	5,2	5,1	3,8	3,3	3,0	2,6	2,5	2,3	2,2	2,1	1,9	1,4	1,0	0,9	0,9	0,8	0,8	0,7	0,7	0,7	0,5	0,4	0,3	0,3	0,2
95	-	-	-	-	-	-	4,9	4,7	4,5	3,5	3,0	2,6	2,4	2,2	2,1	1,9	1,8	1,8	1,3	1,1	0,9	0,8	0,8	0,7	0,7	0,6	0,4	0,4	0,3	0,3	0,3	0,2
96	-	-	-	-	-	-	-	4,0	3,9	3,0	2,6	2,3	2,1	2,0	1,9	1,7	1,6	1,6	1,2	0,9	0,8	0,7	0,7	0,6	0,6	0,6	0,6	0,4	0,4	0,3	0,2	0,2
97	-	-	-	-	-	-	-	-	-	2,7	2,2	2,0	1,8	1,7	1,6	1,5	1,5	1,3	1,0	0,8	0,7	0,6	0,6	0,5	0,5	0,5	0,5	0,3	0,3	0,2	0,2	0,2
98	-	-	-	-	-	-	-	-	-	2,0	1,8	1,5	1,4	1,3	1,2	1,2	1,1	1,0	0,8	0,7	0,6	0,5	0,5	0,4	0,4	0,4	0,4	0,3	0,2	0,2	0,2	0,1
99	-	-	-	-	-	-	-	-	-	-	-	1,0	0,9	0,8	0,8	0,7	0,6	0,5	0,5	0,5	0,4	0,4	0,3	0,3	0,3	0,3	0,3	0,2	0,2	0,1	0,1	0,1

Таблица 8

Критические значения χ^2 [16]

Число степеней свободы	Уровень достоверности, %		Число степеней свободы	Уровень достоверности, %	
	95	99		95	99
1	3,84	6,63	26	38,9	45,6
2	5,99	9,21	27	40,1	47,0
3	7,81	11,3	28	41,3	48,3
4	9,49	13,3	29	42,6	49,6
5	11,1	15,1	30	43,8	50,9
6	12,6	16,8	31	45,0	52,2
7	14,1	18,5	32	46,2	53,5
8	15,5	20,1	33	47,4	54,8
9	16,9	21,7	34	48,6	56,1
10	18,3	23,2	35	49,8	57,3
11	19,7	24,7	36	51,0	58,6
12	21,0	26,2	37	52,2	59,9
13	22,4	27,7	38	53,4	61,2
14	23,7	29,1	39	54,6	62,4
15	25,0	30,6	40	55,8	63,7
16	26,3	32,0	41	56,9	65,0
17	27,6	33,4	42	58,1	66,2
18	28,9	34,8	43	59,3	67,5
19	30,1	36,2	44	60,5	68,7
20	31,4	37,6	45	61,7	70,0
21	32,7	38,9	46	62,8	71,2
22	33,9	40,3	47	64,0	72,4
23	35,2	41,6	48	65,2	73,7
24	36,4	43,0	49	66,3	74,9
25	37,7	44,3	50	67,5	76,2

Примечание. Нулевая гипотеза принимается при $\chi^2_{\text{факт}} \leq \chi^2_{0,05}$ и отвергается при факт $\chi^2_{\text{факт}} > \chi^2_{0,05}$.

Таблица 9

Критерий знаков [17]

n	Уровень значимости α, %		n	Уровень значимости α, %	
	5	1		5	1
1	2	3	4	5	6
5	0	-	23	7	5
6	0	-	24	7	5
7	0	0	25	7	6
8	1	0	26	8	6

Окончание табл. 9

1	2	3	4	5	6
9	1	0	27	8	7
10	1	0	28	8	7
11	2	1	29	9	7
12	2	1	30	10	8
13	3	1	31	10	8
14	3	2	32	10	8
15	3	2	33	11	9
16	4	2	34	11	9
17	4	3	35	12	10
18	5	3	36	12	10
19	5	4	37	13	10
20	5	4	38	13	11
21	6	4	39	13	11
22	6	5	40	14	12

Примечание. Максимальное значение знаков (менее часто встречающихся), при которых различия в парных сравнениях можно считать существенными, $P \leq 0{,}05$ и $P \leq 0{,}01$.

Таблица 10

Критические значения парного критерия Вилкоксона (*T*) [17]

n	Уровень значимости α, %		*n*	Уровень значимости α, %	
	5	1		5	1
5	0	-	13	21	12
6	2	0	14	25	16
7	3	0	15	30	19
8	5	1	16	35	23
9	8	3	17	41	28
10	10	5	18	47	33
11	12	7	19	53	38
12	17	10	20	60	42

Примечание. Максимальное значение *T*, при которых различия между двумя группами можно считать существенными, $P \leq 0{,}05$ и $P \leq 0{,}01$.

Таблица 11

Критические значения коэффициента ранговой корреляции Спирмена (ρ)[16]

n	Уровень значимости α, %		n	Уровень значимости α, %	
	5	1		5	1
5	0,99	-	29	0,37	0,48
6	0,89	1,00	30	0,36	0,47
7	0,79	0,93	31	0,36	0,46
8	0,73	0,88	32	0,36	0,45
9	0,70	0,83	33	0,33	0,45
10	0,65	0,79	34	0,33	0,44
11	0,62	0,76	35	0,33	0,43
12	0,59	0,73	36	0,33	0,43
13	0,56	0,70	37	0,33	0,42
14	0,54	0,68	38	0,32	0,41
15	0,52	0,66	39	0,32	0,41
16	0,50	0,64	40	0,31	0,40
17	0,48	0,62	41	0,31	0,40
18	0,47	0,60	42	0,31	0,40
19	0,46	0,58	43	0,30	0,39
20	0,45	0,57	44	0,30	0,39
21	0,44	0,56	45	0,29	0,38
22	0,43	0,54	46	0,29	0,38
23	0,42	0,53	47	0,29	0,37
24	0,41	0,52	48	0,29	0,37
25	0,40	0,51	49	0,28	0,37
26	0,39	0,50	50	0,28	0,36
27	0,38	0,49			
28	0,38	0,48			

Если объем выборки больше 50, применяется критерий Стьюдента с числом степеней свободы $v = n - 2$

$$t = \frac{r_s}{\sqrt{\frac{1 - r_s^2}{n - 2}}}$$

Примечание. ρ незначим при $\rho \leq \rho_\alpha$ и значим при $\rho > \rho_\alpha$.

Таблица 12

Критические значения выборочного коэффициента корреляции Пирсона (r) [16]

n	Уровень значимости α, %		n	Уровень значимости α, %	
	5	1		5	1
4	0,950	0,990	26	0,388	0,496
5	0,878	0,959	27	0,381	0,487
6	0,811	0,917	28	0,374	0,478
7	0,754	0,874	29	0,367	0,470
8	0,707	0,834	30	0,361	0,463
9	0,666	0,798	35	0,332	0,435
10	0,632	0,765	40	0,310	0,407
11	0,602	0,35	45	0,292	0,384
12	0,576	0,708	50	0,277	0,364
13	0,553	0,684	60	0,253	0,333
14	0,532	0,661	70	0,234	0,308
15	0,514	0,641	80	0,219	0,288
16	0,497	0,623	90	0,206	0,272
17	0,482	0,606	100	0,196	0,258
18	0,468	0,590	125	0,175	0,230
19	0,456	0,575	150	0,160	0,210
20	0,444	0,561	200	0,138	0,182
21	0,433	0,549	250	0,124	0,163
22	0,423	0,587	300	0,113	0,148
23	0,413	0,526	400	0,098	0,128
24	0,404	0,516	500	0,088	0,115
25	0,396	0,505	1000	0,062	0,081

Примечание. r незначим при $r \leq r_\alpha$ и значим при $r > r_\alpha$.

Таблица 13

Критические значения критерия *U* Вилкоксона – Манна-Уитни [16]
Двусторонний критерий

Численность группы		Уровень значимости					
меньшей	большей	0,05, значения			0,01, значения		
		критические		точное	критические		точное
3	4	6	18	0,057			
	5	6	21	0,036			
	5	7	20	0,071			
	6	7	23	0,048	6	24	0,024
	7	7	26	0,033	6	27	0,017
	7	8	25	0,067			
	8	8	28	0,042	6	30	0,012
4	4	11	25	0,057	10	26	0,026
	5	11	29	0,032	10	30	0,016
	5	12	28	0,063			
	6	12	32	0,038	10	34	0,010
	7	13	35	0,042	10	38	0,012
	8	14	38	0,048	11	41	0,008
	8				12	40	0,016
5	5	17	38	0,032	15	40	0,008
	5	18	37	0,056	16	39	0,016
	6	19	41	0,052	16	44	0,010
	7	20	45	0,048	17	48	0,010
	8	21	49	0,045	18	52	0,011
6	6	26	52	0,041	23	55	0,009
	6				24	54	0,015
	7	28	56	0,051	24	60	0,008
	7				25	59	0,014
	8	29	61	0,043	25	65	0,008
	8	30	60	0,059	26	64	0,013
7	7	37	68	0,053	33	72	0,011
	8	39	73	0,054	34	78	0,009
	8	49	87	0,050	44	92	0,010

Таблица 14

Граничные значения оценок коэффициентов асимметрии *As* и эксцесса *E* [34]

n	*As*		*E*	
	Уровень значимости α			
	0,05	0,01	0,05	0,01
10	1,13	1,49	1,43	-
20	0,92	1,21	1,41	1,95
30	0,79	1,05	1,31	1,78
40	0,71	0,93	1,19	1,62
50	0,63	0,84	1,11	1,50
60	0,59	0,78	1,05	1,42
80	0,52	0,68	0,94	1,25
100	0,47	0,62	0,85	1,14
200	0,37	0,44	0,63	0,83
500	0,21	0,28	0,42	0,55
1000	0,15	0,20	0,30	0,40

Примечание. Если $A_{\text{факт}} \leq A_{\alpha}$ и $E_{\text{факт}} \leq E_{\alpha}$, то нулевая гипотеза принимается и испытуемое распределение можно отнести к нормальному; если $A_{\text{факт}} > A_{\alpha}$ и $E_{\text{факт}} > E_{\alpha}$, то нулевая гипотеза отвергается.

Таблица 15

Значения коэффициента К для расчета границ доверительных интервалов средних значений по размаху варьирования [34]

n	Уровень значимости α, %		*n*	Уровень значимости α, %	
	5	1		5	1
3	1,30	3,00	12	0,19	0,28
4	0,72	1,32	13	0,18	0,26
5	0,51	0,84	14	0,17	0,24
6	0,40	0,63	15	0,16	0,22
7	0,33	0,51	16	0,15	0,21
8	0,29	0,43	17	0,14	0,20
9	0,25	0,37	18	0,14	0,19
10	0,23	0,33	19	0,13	0,18
11	0,21	0,30	20	0,13	0,17

Таблица 16

Критические значения критерия Ньюмана – Кейсла (q) для $\alpha' = 0{,}05$ [16]

v	Интервал сравнения l								
	2	3	4	5	6	7	8	9	10
1	17,97	26,98	32,82	37,08	40,41	43,12	45,40	47,36	49,07
2	6,0,85	8,331	9,798	10,88	11,74	12,44	13,03	13,54	13,99
3	4,501	5,910	6,825	7,502	8,037	8,478	8,853	9,177	9,462
4	3,927	5,040	5,757	6,287	6,707	7,053	7,347	7,602	7,826
5	3,635	4,602	5,218	5,673	6,033	6,330	6,582	6,802	6,995
6	3,461	4,339	4,896	5,305	5,628	5,895	6,122	6,319	6,493
7	3,344	4,165	4,681	5,060	5,359	5,606	5,815	5,998	6,158
8	3,261	4,041	4,529	4,886	5,167	5,399	5,597	5,767	5,918
9	3,199	3,949	4,415	4,756	5,024	5,244	5,432	5,595	5,739
10	3,151	3,877	4,327	4,654	4,912	5,124	5,305	5,461	5,599
11	3,113	3,820	4,256	4,574	4,823	5,028	5,202	5,353	5,487
12	3,082	3,773	4,199	4,508	4,751	4,950	5,119	5,265	5,395
13	3,055	3,735	4,151	4,453	4,690	4,885	5,049	5,192	5,318
14	3,033	3,702	4,111	4,407	4,639	4,829	4,990	5,131	5,254
15	3,014	3,674	4,076	4,367	4,595	4,782	4,940	5,077	5,198
16	2,998	3,649	4,046	4,333	4,557	4,741	4,897	5,031	5,150
17	2,984	3,628	4,020	4,303	4,524	4,705	4,858	4,991	5,108
18	2,971	3,609	3,997	4,277	4,495	4,673	4,824	4,956	5,071
19	2,960	3,593	3,997	4,253	4,469	4,645	4,794	4,924	5,038
20	2,950	3,578	3,958	4,232	4,445	4,620	4,768	4,896	5,008
24	2,919	3,532	3,901	4,166	4,373	4,541	4,684	4,807	4,915
30	2,888	3,486	3,845	4,102	4,302	4,464	4,602	4,720	4,824
40	2,858	3,442	3,791	4,039	4,232	4,389	4,521	4,635	4,735
60	2,829	3,399	3,737	3,977	4,163	4,314	4,441	4,550	4,646
120	2,800	3,356	3,685	3,917	4,096	4,241	4,363	4,468	4,560
∞	2,772	3,314	3,633	3,858	4,030	4,170	4,286	4,387	4,474

Таблица 17

Критические значения критерия Ньюмана – Кейсла (q) для $\alpha' = 0{,}01$ [16]

v	Интервал сравнения l								
	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	90,03	135,0	164,3	185,6	202,2	215,8	227,2	237,0	245,6
2	14,04	19,02	22,29	24,72	26,63	28,20	29,53	30,68	31,69
3	8,261	10,62	12,17	13,33	14,24	15,00	15,64	16,20	16,69
4	6,512	8,120	9,173	9,958	10,58	11,10	11,55	11,93	12,27
5	5,702	6,976	7,804	8,421	8,913	9,321	9,669	9,972	10,24
6	5,243	6,331	7,033	7,556	7,973	8,318	8,613	8,869	9,097

Окончание табл. 17

1	2	3	4	5	6	7	8	9	10
7	4,949	5,919	6,543	7,005	7,373	7,679	7,939	8,166	8,368
8	4,746	5,635	6,204	6,625	6,960	7,237	7,474	7,681	7,863
9	4,596	5,428	5,957	6,348	6,658	6,915	7,134	7,325	7,495
10	4,482	5,270	5,769	6,136	6,428	6,669	6,875	7,055	7,213
11	4,392	5,146	5,621	5,970	6,247	6,476	6,672	6,842	6,992
12	4,320	5,046	5,502	5,836	6,101	6,321	6,507	6,670	6,814
13	4,260	4,964	5,404	5,727	5,981	6,192	6,372	6,528	6,667
14	4,210	4,895	5,322	5,634	5,881	6,085	6,258	6,409	6,543
15	4,168	4,836	5,252	5,556	5,796	5,994	6,162	6,309	6,439
16	4,131	4,786	5,192	5,489	5,722	5,915	6,079	6,222	6,349
17	4,099	4,742	5,140	5,430	5,659	5,847	6,007	6,147	6,270
18	4,071	4,703	5,094	5,379	5,603	5,788	5,944	6,081	6,201
19	4,046	4,670	5,054	5,334	5,554	5,735	5,889	6,022	6,141
20	4,024	4,639	5,018	5,294	5,510	5,688	5,839	5,970	6,087
24	3,956	4,546	4,907	5,168	5,374	5,542	5,685	5,809	5,919
30	3,889	4,455	4,799	5,048	5,242	5,401	5,536	5,653	5,756
40	3,825	4,367	4,696	4,931	5,114	5,265	5,392	5,502	5,559
60	3,762	4,282	4,595	4,818	4,991	5,133	5,253	5,356	5,447
120	3,702	4,200	4,497	4,709	4,872	5,005	5,118	5,214	5,299
∞	3,643	4,120	4,403	4,603	4,757	4,882	4,987	5,078	5,157

Таблица 18

Критические значения критерия Даннета (q') для $\alpha' = 0{,}05$ [16]

v	Интервал сравнения l													
	2	3	4	5	6	7	8	9	10	11	12	13	16	21
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	2,57	3,03	3,29	3,48	3,62	3,73	3,82	3,90	3,97	4,03	4,09	4,14	4,26	4,42
6	2,45	2,86	3,10	3,26	3,39	3,49	3,57	3,64	3,71	3,76	3,81	3,86	3,97	4,11
7	2,36	2,75	2,97	3,12	3,24	3,33	3,41	3,47	3,53	3,58	3,63	3,67	3,78	3,91
8	2,31	2,67	2,88	3,02	3,13	3,22	3,29	3,35	3,41	3,46	3,50	3,54	3,64	3,76
9	2,26	2,61	2,81	2,95	3,05	3,14	3,20	3,26	3,32	3,36	3,40	3,44	3,53	3,65
10	2,23	2,57	2,76	2,89	2,99	3,07	3,14	3,19	3,24	3,29	3,33	3,36	3,45	3,57
11	2,20	2,53	2,72	2,84	2,94	3,02	3,08	3,14	3,19	3,23	3,27	3,30	3,39	3,50
12	2,18	2,50	2,68	2,81	2,90	2,98	3,04	3,09	3,14	3,18	3,22	3,25	3,34	3,45
13	2,16	2,48	2,65	2,78	2,87	2,94	3,00	3,06	3,10	3,14	3,18	3,21	3,29	3,40
14	2,14	2,46	2,63	2,75	2,84	2,91	2,97	3,02	3,07	3,11	3,14	3,18	3,26	3,36
15	2,13	2,44	2,61	2,73	2,82	2,89	2,95	3,00	3,04	3,08	3,12	3,15	3,23	3,33
16	2,12	2,42	2,59	2,71	2,80	2,87	2,92	2,97	3,02	3,06	3,09	3,12	3,20	3,30
17	2,11	2,41	2,58	2,69	2,78	2,85	2,90	2,95	3,00	3,03	3,07	3,10	3,18	3,27
18	2,10	2,40	2,56	2,68	2,76	2,83	2,89	2,94	2,98	3,01	3,05	3,08	3,16	3,25
19	2,09	2,39	2,55	2,66	2,75	2,81	2,87	2,92	2,96	3,00	3,03	3,06	3,14	3,23

Окончание табл. 18

1	2	3	4	5	6	7	8	9	10	11	12	13	16	21
20	2,09	2,38	2,54	2,65	2,73	2,80	2,86	2,90	2,95	2,98	3,02	3,05	3,12	3,22
24	2,06	2,35	2,51	2,61	2,70	2,76	2,81	2,86	2,90	2,94	2,97	3,00	3,07	3,16
30	2,04	2,32	2,47	2,58	2,66	2,72	2,77	2,82	2,86	2,89	2,92	2,95	3,02	3,11
40	2,02	2,29	2,44	2,54	2,62	2,68	2,73	2,77	2,81	2,85	2,87	2,90	2,97	3,06
60	2,00	2,27	2,41	2,51	2,58	2,64	2,69	2,73	2,77	2,80	2,83	2,86	2,92	3,00
120	1,98	2,24	2,38	2,47	2,55	2,60	2,65	2,69	2,73	2,76	2,79	2,81	2,87	2,95
∞	1,96	2,21	2,35	2,44	2,51	2,57	2,61	2,65	2,69	2,72	2,74	2,77	2,83	2,91

Таблица 19

Критические значения критерия Даннета (q') для $\alpha' = 0{,}01$ [16]

v	Интервал сравнения l													
	2	3	4	5	6	7	8	9	10	11	12	13	16	21
5	4,03	4,63	4,98	5,22	5,41	5,56	5,69	5,80	5,89	5,98	6,05	6,12	6,30	6,52
6	3,71	4,21	4,51	4,71	4,87	5,00	5,10	5,20	5,28	5,35	5,41	5,47	5,62	5,81
7	3,50	3,95	4,21	4,39	4,53	4,64	4,74	4,82	4,89	4,95	5,01	5,06	5,19	5,36
8	3,36	3,77	4,00	4,17	4,29	4,40	4,48	4,56	4,62	4,68	4,73	4,78	4,90	5,05
9	3,25	3,63	3,85	4,01	4,12	4,22	4,30	4,37	4,43	4,48	4,53	4,57	4,68	4,82
10	3,17	3,53	3,74	3,88	3,99	4,08	4,16	4,22	4,28	4,33	4,37	4,42	4,52	4,65
11	3,11	3,45	3,65	3,79	3,89	3,98	4,05	4,11	4,16	4,21	4,25	4,29	4,30	4,52
12	3,05	3,39	3,58	3,71	3,81	3,89	3,96	4,02	4,07	4,12	4,16	4,19	4,29	4,41
13	3,01	3,33	3,52	3,65	3,74	3,82	3,89	3,94	3,99	4,04	4,08	4,11	4,20	4,32
14	2,98	3,29	3,47	3,59	3,69	3,76	3,83	3,88	3,93	3,97	4,01	4,05	4,13	4,24
15	2,95	3,25	3,43	3,55	3,64	3,71	3,78	3,83	3,88	3,92	3,95	3,99	4,07	4,18
16	2,92	3,22	3,39	3,51	3,60	3,67	3,73	3,78	3,83	3,87	3,91	3,94	4,02	4,13
17	2,90	3,19	3,36	3,47	3,56	3,63	3,69	3,74	3,79	3,83	3,86	3,90	3,98	4,08
18	2,88	3,17	3,33	3,44	3,53	3,60	3,66	3,71	3,75	3,79	3,83	3,86	3,94	4,04
19	2,86	3,15	3,31	3,42	3,50	3,57	3,63	3,68	3,72	3,76	3,79	3,83	3,90	4,00
20	2,85	3,13	3,29	3,40	3,48	3,55	3,60	3,65	3,69	3,73	3,77	3,80	3,87	3,97
24	2,80	3,07	3,22	3,32	3,40	3,47	3,52	3,57	3,61	3,64	3,68	3,70	3,78	3,87
30	2,75	3,01	3,15	3,25	3,33	3,39	3,44	3,49	3,52	3,56	3,59	3,62	3,69	3,78
40	2,70	2,95	3,09	3,19	3,26	3,32	3,37	3,41	3,44	3,48	3,51	3,53	3,60	3,68
60	2,66	2,90	3,03	3,12	3,19	3,25	3,29	3,33	3,37	3,40	3,42	3,45	3,51	3,59
120	2,62	2,85	2,97	3,06	3,12	3,18	3,22	3,26	3,29	3,32	3,35	3,37	3,43	3,51
∞	2,58	2,79	2,92	3,00	3,06	3,11	3,15	3,19	3,22	3,25	3,27	3,29	3,35	2,42

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ. 3

Глава 1
ОСНОВНЫЕ КРИТЕРИИ ЭПИДЕМИОЛОГИЧЕСКОГО АНАЛИЗА 7
1.1. Эпидемиологические показатели 7
1.2. Статистическая оценка эпидемиологических показателей . . . 10
1.3. Средние величины, их размеры и количественные отношения признаков 15

Глава 2
СРАВНЕНИЕ ДВУХ ВЫБОРОК 28
2.1. Параметрические методы сравнения двух выборок 28
2.2. Непараметрические методы сравнения двух выборок 39

Глава 3
АНАЛИЗ МНОГОЛЕТНЕЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ В ЭПИДЕМИОЛОГИЧЕСКОЙ ДИАГНОСТИКЕ 52
3.1. Обобщающая характеристика многолетней динамики заболеваемости и измерение ее интенсивности 54
3.2. Методы анализа многолетней тенденции 57
3.3. Оценка многолетней периодичности 69
3.4. Анализ нерегулярных колебаний заболеваемости 73
3.5. Оценка эпидемиологической ситуации по обобщенному критерию 76

Глава 4
АНАЛИЗ ВНУТРИГОДОВОЙ ДИНАМИКИ ЗАБОЛЕВАЕМОСТИ 82
4.1. Статистическая оценка сезонных показателей заболеваемости . 83
4.2. Анализ годовой динамики заболеваемости с учетом ее круглогодичного уровня 99
4.3. Определение «контрольных» уровней заболеваемости 106

Глава 5
АНАЛИЗ ЗАБОЛЕВАЕМОСТИ ПО ФАКТОРАМ РИСКА 112

Глава 6
ПРОГНОЗИРОВАНИЕ ЗАБОЛЕВАЕМОСТИ 128

БИБЛИОГРАФИЧЕСКИЙ СПИСОК 137

ПРИЛОЖЕНИЕ 140

Научное издание

Савилов Евгений Дмитриевич
Астафьев Виктор Александрович
Жданова Светлана Николаевна
Заруднев Евгений Алексеевич

ЭПИДЕМИОЛОГИЧЕСКИЙ АНАЛИЗ

МЕТОДЫ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ МАТЕРИАЛА

Редактор *Т.А. Никитина*
Художник *Ю.В. Антипова*
Корректор *О.Ю. Якушева*
Оператор электронной верстки *Ю.В. Антипова*

Сдано в набор 02.07.11. Подписано в печать 28.09.11. Бумага ВХИ.
Формат 60 × 90 1/16. Офсетная печать. Гарнитура Таймс.
Усл. печ. л. 9,75. Уч-изд. л. 7,5. Тираж 300 экз. Заказ №

ООО «Наука-Центр». 6300090. Новосибирск, ул. акад. М.А. Лаврентьева, 6

ISBN 978-59554-0024-2