# Оглавление

# Предисловие

## Наука, общество и статистика

Современный мир до краев заполнен статистикой. Статистика знает, что мы едим, как тренируемся, с кем дружим, как учим своих детей и какие принимаем лекарства. Очевидно, что статистика вездесуща, как и – к глубокому сожалению – окружающие ее домыслы. В главе 1 мы расскажем о том, как судьи выносят неправильные приговоры, – отправлять человека в тюрьму или нет, – поскольку не понимают даже основ статистики. Мы покажем, что пациенты совершают самоубийство, потому что врачи не умеют интерпретировать результаты анализов. Ученые зачастую ничуть не лучше. Мы знаем коллег, которые слепо доверяли результатам статистических программ, даже когда они не имели ни малейшего смысла. Встречались нам и опубликованные научные статьи, в которых результаты не согласуются с теоретическими выводами авторов.

Есть старое изречение (иногда его приписывают Марку Твену, но, по-видимому, оно все же старше): «Существует три вида лжи: ложь, наглая ложь и статистика». Мы вынуждены признать, что в нем имеется зерно истины. Люди часто неправильно пользуются статистическим анализом. Быть может, до прямой лжи (высказывания заведомо ложного утверждения) и не доходит, но статистика зачастую только запутывает, а не проясняет дело. Название этой книги отражает наши усилия повысить качество использования статистики так, чтобы те, кто занимается анализом, научились лучше интерпретировать свои данные, а те, кто читает статистические результаты, лучше их понимали. Понимание ключевых идей статистики поможет осознать, что многие научные результаты по сути своей бессмысленны, и объяснит, почему многие эмпирические науки в настоящее время сталкиваются с кризисом воспроизводимости. *Вычисление* статистики может оказаться очень трудным делом (отсюда и сложные программы, и толстые книги с глубокими теоремами), но ясное понимание основных принципов статистики вполне доступно каждому.

В 2013 году, вдохновленные этими идеями, мы начали читать курс в Федеральной политехнической школе Лозанны, Швейцария, посвященный концептуальным основам статистики и планирования эксперимента. С годами курс стал довольно популярным, его стали посещать студенты, обучающиеся биологии, нейронаукам, медицине, генетике, психологии и биоинженерии. Обычно такие студенты уже прослушали один или несколько курсов по статистике, на которых им преподавали *детали* статистического анализа. На нашем же курсе и в этой книге акцент сделан на *базовых* принципах этого анализа; мы хотим кратко объяснить, что это такое, и привить читателю понимание возможностей и ограничений анализа.

## Об этой книге

*Предварительные знания и цель.* Как уже было сказано, непонимание статистики стало важной проблемой в нашем обществе. И одно из ее проявлений в том, что вычислить статистические показатели теперь так просто, что наличие хорошего образования кажется ненужным. Однако же все в точности наоборот. Простота использования статистических программ позволят выполнять анализ данных, не понимая, что программа делает и как интерпретировать ее результаты. Итог – несостоятельные выводы. Читатель, вероятно, удивится тому, насколько проблема серьезна и сколь велико количество бесполезных исследований. И еще, наверное, с удивлением узнает, что даже такие базовые термины, как $p$-значение, означают совсем не то, что многим кажется.

Главная цель этой книги – кратко и по существу изложить основы математической статистики. Понимание этих основ подготовит читателя к правильному восприятию и критической оценке научных публикаций во многих отраслях науки. Мы не собираемся учить читателя *вычислять* статистические характеристики. Это вполне можно оставить компьютеру.

*Читательская аудитория.* Эта книга адресована всем гражданам и научным работникам, имеющим желание понять принципы статистики и научиться интерпретировать ее результаты, не вдаваясь в математические детали вычислений. Как ни странно, этой цели можно достичь в очень короткой книжке, содержащей совсем немного уравнений. Мы думаем, что любой человек (а не только студент или ученый), имеющий или не имеющий предварительную подготовку в области статистики, сможет извлечь из этой книги пользу.

Мы свели уровень необходимой математической подготовки к минимуму и всюду, где возможно, обращались к интуиции. Уравнения мы включали только тогда, когда они делали изложение понятнее. Для понимания основных идей достаточно самой элементарной математики и лишь немногих базовых понятий из теории вероятностей. Да и те интуитивно понятны из контекста.

*Чего нет в этой книге.* Эта книга не курс математической статистики (например, в ней ничего не говорится о борелевской алгебре). Это и не традиционный учебник статистики, в который принято включать многочисленные критерии и методы. Это не руководство по программам статистического анализа типа SPSS или R. Книга не является полным справочником по статистическим критериям. Мы стремились включить достаточно информации для понимания фундаментальных основ статистики, но не больше.

*О чем эта книга.* В части I мы излагаем философию статистики с минимальным привлечением математики, чтобы были понятны ключевые концепции. Мы познакомимся с самым простым $t$-критерием и покажем, как избежать недоразумений, связанных с вероятностями.

Усвоив материал части I, читатель сможет избежать большинства подводных камней и понять, что на самом деле вычисляют наиболее известные статистические критерии. Мы опишем проверку нулевой гипотезы, не прибегая к сложному математическому аппарату, а применив более простой подход на основе теории обнаружения сигналов (ТОС). Материал части II более традиционный, здесь рассматриваются классические критерии: дисперсионный анализ (ANOVA) и корреляция. В частях I и II описаны стандартные статистики – те, что используются наиболее часто. В части III показано, что кризис науки возник из-за неправильного понимания простейших основ статистики, в частности понятия воспроизводимости. Например, читателя, возможно, удивит, что слишком большое число успешных повторений эксперимента может считаться подозрительным явлением, а не свидетельством бесспорности научного факта. Для понимания части III, включающей идеи, которых не найдешь в других вводных учебниках, достаточно лишь базовых понятий и концепций из главы 3 части I. И хотя книга в основном посвящена статистике, мы продемонстрируем тесную связь статистики с планированием эксперимента. Многих статистических проблем можно было бы избежать при выборе правильного – что чаще всего означает «более простого» – плана.

Мы полагаем, что уникальное сочетание базовых концепций статистики (часть I), краткого описания наиболее распространенных статистических критериев (часть II) и нового метастатистического подхода (часть III) позволит не только достичь основательного понимания статистики, но и по-новому – подчас с неожиданной точки зрения – взглянуть на то, что определяет нашу повседневную жизнь.

*Материалы*. Дополняющие книгу презентации в формате Power Point для преподавателей доступны по запросу на адрес электронной почты michael.herzog@epfl.ch.

**Благодарности.** Мы благодарны Конраду Нойману и Марку Репнову за корректуру рукописи, а Эдди Кристоферу, Алине Критенуд, Марку, Гертруде и Хайке Херцог, Майе Анне Ястржебовска, Слиму Каммоуну, Иларии Риччи, Эвелине Танелл, Ричарду Уолкеру, Хе Сю и Пьеру Дево за полезные замечания. С грустью сообщаем, что во время подготовки книги к печати от нас ушел Аарон Кларк.

| | |
|---|---|
| Лозанна, Швейцария | Майкл Х. Херцог |
| Вест-Лафайетт, Индиана, США | Грегори Фрэнсис |
| Анкара, Турция | Аарон Кларк |

# Часть I

# Принципы статистики

# Глава 1

# Основы теории вероятностей

**Что вы узнаете из этой главы**

Прежде чем ступать на территорию статистики, необходимо вспомнить базовые понятия теории вероятностей. Иначе трудно интерпретировать научные данные, как, впрочем, и информацию в повседневной жизни. В частности, многое из того, что сообщают СМИ, по существу, бесполезно, потому что основано на частичной информации. В этой главе мы объясним, какого рода полная информация нужна для правильных выводов, и познакомимся с теоремой Байеса. Для изложения идей нам понадобятся простые уравнения, а для тех читателей, кто не в ладах с математикой, мы приведем интуитивно понятные соображения на простых примерах и рисунках.

## 1.1. Путаница вокруг простых понятий теории вероятностей: условные вероятности

### 1.1.1. Базовый сценарий

**Основные понятия теории вероятностей**

1. **Вероятность.** Событию А назначается вероятность – число от 0 до 1. Например, при бросании кости вероятность выпадения 4 равна $P(4) = 1/6$.
2. **Распределение вероятностей.** В примере выше имеется 6 возможных исходов, каждому из которых назначена вероятность 1/6. Назначение вероятности каждому возможному исходу дает распределение вероятностей.
3. **Условная вероятность.** Условная вероятность $P(\text{A}|\text{B})$ учитывает имеющуюся информацию о событии В. Вертикаль-

ная черта читается «при условии» и означает, что речь идет именно об условной вероятности. Например, мы последовательно вытягиваем две карты из стандартной колоды (52 карты). Вероятность, что первой будет вытянута пиковая масть *P*(первой вытянута пика) = 13/52 = 1/4. Теперь осталась только 51 карта. Вероятность, что во второй раз тоже будет вытянута пика, притом, что она уже была вытянута в первый раз, P(второй вытянута пика|первой вытянута пика) = 12/51. С другой стороны, P(второй вытянута пика|первой вытянута черва) = 13/51. Здесь вероятность вытягивания второй карты зависит от того, какого типа карта была вытянута первой.

4. **Независимые события.** События называются независимыми, если условная вероятность равна безусловной: *P*(A|B) = *P*(A). В этом случае вероятность A не зависит от B. Например, если вытянутая карта возвращается в колоду, то вероятность вытянуть пику во второй раз равна P(второй вытянута пика) = 13/52 вне зависимости от того, какая карта была вытянута первой.

**Определения**

Рассмотрим ситуацию, когда есть подозрение, что пациент инфицирован, и он сдает соответствующий анализ. Возможны четыре исхода.

1. **Чувствительность:** вероятность положительного анализа при условии, что пациент инфицирован.
2. **Специфичность:** вероятность отрицательного анализа при условии, что пациент не инфицирован.
3. **Частота ложноположительных результатов:** вероятность положительного анализа при условии, что пациент не инфицирован.
4. **Частота ложноотрицательных результатов:** вероятность отрицательного анализа при условии, что пациент инфицирован.

Начнем с примера. В 1980-х годах общество охватила паника: обнаружилась новая болезнь, получившая название «синдром приобретенного иммунодефицита» (СПИД), ее вызывал вирус ВИЧ (*англ.* HIV). Ученые разработали высокочувствительный тест для определения наличия вируса в крови. Предположим, что и чувствительность, и специфичность теста на ВИЧ равны 0.9999. Это значит, что тест очень хороший, потому что в большинстве случаев он будет давать положительный результат, если пациент инфицирован, и отрицательный, если не инфицирован. Предположим далее, что коэффициент заболеваемости СПИД составляет 0.0001 в нормальной генеральной совокупности, т. е. 1 из 10 000 человек инфицирован вирусом ВИЧ. Анализ, взятый у случайно выбранного человека, оказывается положительным. Допустим, что вы врач. Что вы скажете пациенту о вероятности заболевания? Математически – какова условная вероятность инфицирования ВИЧ при условии, что тест положителен ($T^+$): $P(HIV|T^+)$?

Поскольку тест очень хорош и почти не дает ошибок, многие полагают, что вероятность $P(HIV|T^+)$ должна быть очень высока, скажем $P(HIV|T^+) = 0.9999$. Однако в действительности $P(HIV|T^+) = 0.5$ – не выше, чем при подбрасывании монеты. Как такое возможно? Мы можем вычислить $P(HIV|T^+)$, воспользовавшись теоремой Байеса, которая ниже сформулирована в общем виде.

Для двух событий A и B

$$P(A \mid B) = \frac{P(B \mid A) \times P(A)}{P(B)}.$$

Теперь подставим в эту формулу значения ($\neg HIV$ означает отсутствие ВИЧ):

$$\begin{aligned} P(HIV \mid T^+) &= \frac{P(T^+ \mid HIV) \times P(HIV)}{P(T^+)} = \\ &= \frac{P(T^+ \mid HIV) \times P(HIV)}{P(T^+ \mid HIV) \times P(HIV) + P(T^+ \mid \neg HIV) \times P(\neg HIV)} = \\ &= \frac{0.9999 \times 0.0001}{0.9999 \times 0.0001 + (1 - 0.9999) \times 0.9999} = 0.5. \end{aligned}$$

Математика дает ответ, но разобраться в ситуации можно и на интуитивном уровне (рис. 1.1). Предположим, что протестировано 10 000 человек. Поскольку коэффициент заболеваемости равен 0.0001, с высокой вероятностью инфицирован только один из них. Поскольку чувствительность теста очень высока (0.9999), вирус, скорее всего, будет обнаружен. Имеется 9999 неинфицированных людей. Хотя специфичность теста тоже очень высока (0.9999), один ложноположительный результат, вероятно, все же будет – потому что протестировано много людей. Таким образом, из 10 000 человек всего у двух будут положительные результаты теста (а у 9998 отрицательные). Поскольку из двух человек действительно инфицирован только один, вероятность инфицирования равна ½, т. е. $P(HIV|T^{+}) = 0.5$.

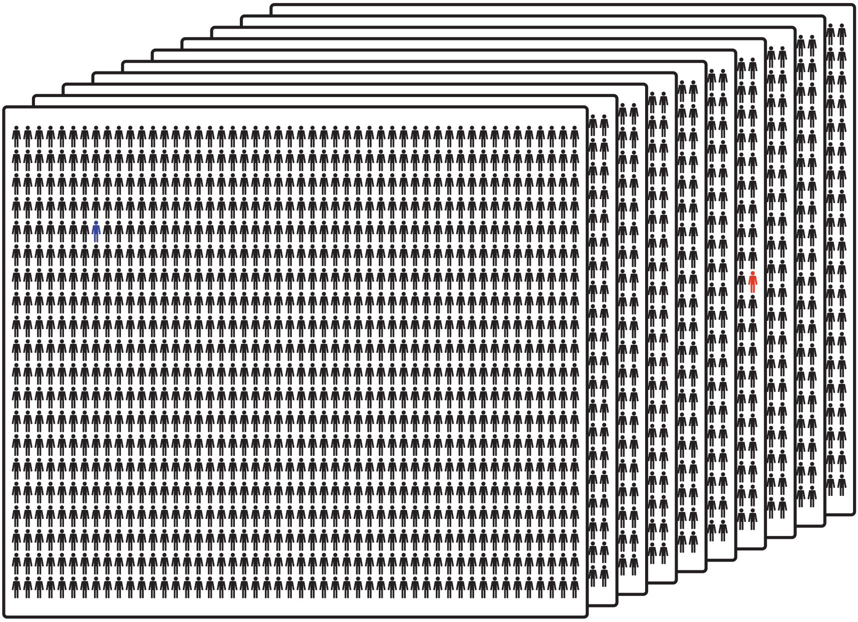

**Рис. 1.1.** В выборке, включающей 10 000 человек, вероятно, имеется только один инфицированный. Поскольку чувствительность теста высока, результат для этого человека с очень большой вероятностью положителен (выделен красным цветом). Если протестировать одного случайно выбранного неинфицированного человека, то результат с очень большой вероятностью будет отрицательным в силу высокой специфичности теста. Однако всего имеется 9999 неинфицированных людей, и, несмотря на высокую специфичность, тест, вероятно, даст один ложноположительный результат (выделен синим цветом). Итак, мы получили два положительных теста, а поскольку инфицирован только один человек, вероятность инфицирования при положительном тесте составляет 1/2: $P(HIV|T^{+}) = 0.5$. Очевидно, что игнорировать коэффициент заболеваемость нельзя. Он так же важен, как чувствительность и специфичность

Предположим, что коэффициент заболеваемости еще ниже, например 1/100 000. Допустим, что протестировано 100 000 человек. Поскольку коэффициент заболеваемости равен 1/100 000, вероятно, один из них инфицирован, и тест этого человека, скорее всего, выявит. Но кроме него на каждые 10 000 человек приходится один ложноотрицательный результат. Следовательно, тест дает 11 положительных результатов[1], и вероятность реального инфицирования при положительном тесте снижается до $P(HIV|T^+) = 1/11 \approx 0.1$. С другой стороны, при коэффициенте заболеваемости 0.5 $P(HIV|T^+) = 0.9999$, т. е. почти 1.0. Следовательно, вероятность $P(HIV|T^+)$ зависит от чувствительности, специфичности *и* коэффициента заболеваемости. Если коэффициент заболеваемости изменяется от 0.0 до 1.0, то $P(HIV|T^+)$ изменяется от 0.0 до 1.0. Для обоснованного заключения необходимо знать все три члена. Если хотя бы один из них неизвестен, все заключения гроша ломаного не стоят. Этот пример иллюстрирует одну из главных тем книги: *помните о частичной информации*!

> **Замечание 1.** Эта демонстрация показывает, насколько важно понимать основы статистических рассуждений. Для пациента умение интерпретировать положительный результат теста имеет огромное значение. Например, в 1987 году 22 реципиента переливания крови получили положительный тест на ВИЧ, и семеро из них покончили жизнь самоубийством [1]. А в одном исследовании отмечается, что в Германии 16 из 20 врачей говорят пациентам, что тест на ВИЧ практически не дает ложноположительных результатов [2].

> **Замечание 2.** Важно, что если вы врач, то ситуация отличается от описанной в примере выше, потому что люди, у которых есть причины подозревать у себя инфекцию, с большей вероятностью сдают тест, чем те, кто более-менее уверен, что не инфицирован. Поэтому коэффициент заболеваемости в больнице, вероятно, выше, чем в примере. Это означает, что $P(HIV|T^+)$ может быть больше 0.5: этот озадачивающий вывод показывает, почему интуитивные представления людей о статистике зачастую не имеют под собой основания.

### 1.1.2. Второй тест

Правильное понимание природы вероятности также помогает собирать более качественную информацию. Что будет, если мы второй

---

[1] Мы можем также протестировать 10 000 человек. Поскольку коэффициент заболеваемости равен 1/100 000, с вероятностью 0.1 в этой выборке будет один инфицированный. А при коэффициенте заболеваемости 1/10 000 будет один ложноположительный результат. Поэтому мы должны вычислить частное 0.1/1.1 и получим точно такой же результат – приблизительно 0.1.

раз возьмем тест только у двух человек с положительным первым тестом?[1] Чему теперь будет равна вероятность инфицирования при положительном тесте? Интересующая нас величина равна

$$P(HIV \mid T^{+}) = \frac{0.9999^2 \times 0.0001}{0.9999^2 \times 0.0001 + (1-0.9999)^2 \times 0.9999} =$$

$$= \frac{0.9999}{0.9999 + 0.0001} = 0.9999.$$

Теперь положительный результат означает, что человек почти наверняка инфицирован.

Это равенство можно объяснить на интуитивном уровне. Первый тест дал два положительных результата. Во второй раз только эти два человека и тестировались. Поскольку тест очень хорош, он почти наверняка определит инфицированного человека и почти наверняка даст отрицательный результат для неинфицированного. Поэтому для человека, сдавшего два положительных теста, вероятность инфицирования близка к 1.0.

**Замечание 1.** В действительности, назначая повторный тест, врач обнаруживает, что $P(HIV|T^{2+})$ меньше 0.9999. Причина в том, что для некоторых людей тест упорно дает положительный результат, даже если они не инфицированы. Видимо, в их организме имеются молекулы, похожие на антитела, к которым чувствителен тест на ВИЧ.

**Замечание 2.** Недоразумения вокруг статистики возникают во всех отраслях знания. Корали Колмез и Лейла Шнепс посвятили целую книгу «Math on Trial» анализу судебных дел. В этой книге показано, как непонимание простых положений статистики может привести (и приводило) к неверным выводам. Рассматривается, в частности, дело студентки Аманды Нокс, которая обвинялась в убийстве соседки по квартире. Генетический анализ дал некоторые свидетельства в пользу того, что соседка была убита ножом, на котором присутствовали отпечатки пальцев Аманды. Изучив, какова вероятность точного результата теста, судья решил не проводить повторный тест – несмотря на то, что, как показано выше, повторный анализ мог бы дать совершенно иной результат. Судья был попросту недостаточно знаком с основами статистики [3].

[1] Предполагается, что тесты для данного человека независимы.

### 1.1.3. Еще пример: синдром Гийена–Барре

Вакцинация (V) от свиного гриппа (SF) может вызвать в качестве побочного эффекта синдром Гийена–Барре (GB) в одном случае на миллион, т. е. $P(GB|V) = 1/1\,000\,000$. В тяжелых случаях СГБ напоминает синдром «запертого человека», когда пациент утрачивает способность двигаться и даже разговаривать. Учитывая ужасающие последствия СГБ, стоит ли вообще делать вакцинацию? И на этот раз мы не можем ответить на вопрос, потому что обладаем лишь частичной информацией. Необходимо знать, какова вероятность заболеть синдромом Гийена–Барре без вакцинации ($\neg V$). Предположим, что, помимо вакцинации, СГБ возникает только как осложнение после свиного гриппа (и дополнительно предположим, что вакцина против свиного гриппа стопроцентно эффективна). Вероятность получить СГБ после свиного гриппа довольно высока: 1/3000. Это гораздо больше, чем вероятность 1/1 000 000. Похоже, что вакцинироваться все-таки стоит. Однако следует принять во внимание уровень заражения свиным гриппом, потому что не каждый человек заражается. Этот уровень в каждой новой эпидемии изменяется; предположим, что для случайно выбранного невакцинированного человека вероятность заболеть составляет 1/300. Тогда вероятность получить синдром Гийена–Барре для невакцинированного равна

$$P(GB \mid \neg V) = P(GB \mid SF) \times P(SF \mid \neg V) P(\neg V) = \frac{1}{3000} \times \frac{1}{300} \times 1 = \frac{1}{900\,000}.$$

Следовательно, в такой ситуации вероятность получить СГБ для невакцинированного лишь ненамного больше, чем для вакцинированного. А вакцина заодно защищает от свиного гриппа.

Важно здесь то, что нельзя принять хорошее решение, основываясь только на одной вероятности (получить синдром Гийена–Барре в результате вакцинации). Нужно также учитывать вероятность дополнительного события (получить синдром Гийена–Барре без вакцинации).

## 1.2. Недоразумения вокруг вероятностей: отношение шансов

### 1.2.1. Основные сведения об отношении шансов (ОШ)

Многие курильщики умирают от инфаркта. Бросать курить? Это частичная информация! Встречный вопрос: сколько некурящих умирает от инфаркта? Без этой информации попытка ответить на первый вопрос будет ничуть не лучше утверждения «100 % курильщиков однажды умрут – равно как и 100 % некурящих».

Обобщим эту ситуацию, введя в рассмотрение понятие шанса. Гипотетический пример: из 107 курящих семь перенесли инфаркт, т. е. 100 не переносили (табл. 1.1A). Шансом называется отношение 7/100. Для некурящих инфаркт перенес 1 человек из 100, поэтому шанс равен 1/100. Идея отношения шансов (ОШ) – сравнить две дроби, поделив одну на другую. Это отношение двух отношений говорит нам, в какой степени курящие страдают от инфаркта чаще, чем некурящие: 7/100 / 1/100 = 7. Таким образом, у курящего шанс получить инфаркт в семь раз больше, чем у некурящего, – немало. Для сравнения – если бы никакого эффекта не было, т. е. инфаркт случался бы у курящих и некурящих с одинаковой частотой, то ОШ = 1.0.

**Таблица 1.1.** Гипотетический пример

| **A** | **Курящие** | **Некурящие** | | **B** | **Курящие** | **Некурящие** |
|---|---|---|---|---|---|---|
| Инфаркт был | 7 | 1 | | Инфаркт был | 7 | 1 |
| Инфаркта не было | 100 | 100 | | Инфаркта не было | 10 000 | 10 000 |

A) Каковы шансы получить инфаркт, не будучи курильщиком? Предположим, что из 107 курящих семеро перенесли инфаркт, а из 101 некурящего – только один. Насколько шансы курящего получить инфаркт выше, чем у некурящего? Для вычисления отношения шансов мы сначала вычисляем 7/100 и 1/100, а затем делим эти отношения: (7/100) / (1/100) = (7*100) / (1*100) = 7/1= 7. Следовательно, шансы в семь раз выше, что представляется существенным.

B) Теперь предположим, что в группах курящих и некурящих по 10 000 человек, не перенесших инфаркт. Отношение шансов равно (7/10 000)/(1/10 000) = 7/1 = 7, т. е. не изменилось. Таким образом, отношение шансов не зависит от коэффициента заболеваемости. Однако шанс получить инфаркт уменьшился примерно в 100 раз. Получить ли инфаркт в 7 из 107 случаев или в 7 из 10 007 – «две большие разницы». Отношение шансов дает лишь частичную информацию!

**Таблица 1.2.** Члены, вносящие вклад в отношение шансов[a]

| | С фактором риска | Без фактора риска |
|---|---|---|
| Болен | a | b |
| Не болен | c | d |

[a] Небольшое замечание: для вычисления отношения шансов производится деление $a/b$ на $c/d$. Можно было бы также воспользоваться пропорциями $a / (a + b)$, $c / (c + d)$ и взять их отношение.

В общем виде (табл. 1.2) отношение шансов $a/c / b/d = a * d / b * c$.

Отношение шансов – очень компактный способ сравнения экспериментального и контрольного условия. Оно чаще других показателей используется в медицине и биологии. Например, влияние гена на заболевание обычно выражают в терминах ОШ. Однако решения, принимаемые на базе ОШ, основаны на частичной информации. И вот почему. Увеличим количество людей, перенесших инфаркт, в обеих группах в 100 раз. Отношение шансов при этом не изменится (см. табл. 1.1B).

Очевидно, что отношение шансов не зависит от доли незатронутых людей, пусть даже вероятность получить инфаркт существенно изменилась. Поскольку ОШ не зависит от коэффициента заболеваемости, высокое ОШ почти ничего не говорит, если, к примеру, заболевание редкое.

Как интерпретировать отношения шансов? Во-первых, высокое ОШ – причина для тревоги, только если основной эффект, $a/c$, тоже велик. Например, отношение (число курящих, перенесших инфаркт) / (число курящих, не перенесших инфаркт) = 7/10 000 нельзя считать значительным эффектом, пусть даже ОШ, равное 7, велико. В табл. 1.1B инфаркты просто не очень часты. Только 8 человек из 20 008 перенесли инфаркт. Поэтому крайне маловероятно, что у кого-нибудь вообще был инфаркт, – в отличие от случая в табл. 1.1, где инфаркт перенесли 8 из 208 человек. И в таком случае есть повод для беспокойства. Во-вторых, высокий основной эффект $a/c$ – причина беспокоиться, только если ОШ также велико. Вот крайний пример. Если у вас голубые глаза, то вы умрете с очень высокой вероятностью (100 %). Однако для кареглазых вероятность умереть также равна 100 %. Следовательно, ОШ = 1.0, это мало. Можно тревожиться по поводу смерти, но не по поводу цвета глаз.

### 1.2.2. Частичная информация и мир, полный болезней

Ситуация в целом может быть и еще более запутанной. Мы обсудили влияние одного фактора (курения) на исход (инфаркт). Но курение может также влиять на другие заболевания положительно или отрицательно (даже курение не всегда вредно). Поэтому, чтобы формально ответить на вопрос, следует ли бросать курить, нужно принять во внимание все заболевания, в том числе потенциально неизвестные. Кроме того, нужно учесть затраты на лечение – кариес все же не так серьезен, как инфаркт. Таким образом, нужно вычислить своего рода эффект заболеваемости, который принимает во внимание стоимость различных заболеваний и вероятность их возникновения для данного фактора:

$$\text{Заболеваемость(Фактор)} = \sum_s P(\text{заболевание } S|\text{Фактор}) \times \text{Стоимость(заболевание } S\text{)}.$$

Итак, нужно учитывать все болезни, даже еще не открытые. Однако решить, стоит ли бросать курить или менять диету, почти невозможно, если величина эффекта невелика. На практике вся эта информация никогда не бывает доступна. Но это не значит, что статистические рассуждения никогда нельзя использовать для принятия решения. Нужно лишь понимать, что такие решения основаны на неполной информации. И это понимание должно побудить вас собирать как можно больше информации.

**Что следует запомнить**

1. Не забывайте о частичной информации и старайтесь получить полную информацию, чтобы делать правильные выводы. Например, отношение шансов обычно несет слишком мало информации.
2. Коэффициенты заболеваемости различными болезнями обычно малы, за исключением таких как кариес.

### Литература

1. Gigerenzer G, GaissmaierW, Kurz-Milcke E, Schwartz L,Woloshin S. Glaub keiner Statistik, die du nicht verstanden hast. Geist und Gehirn. 2009;10:34–39.
2. Gigerenzer G, Hoffrage U, Ebert A. AIDS counselling for low-risk clients. AIDS Care. 1998;10:197–211.
3. Colmez C, Schneps L. Math on trial: how numbers get used and abused in the courtroom. Basic Books: New York; 2013.