НАУКА ВЕЛИЧАЙШИЕ ТЕОРИИ

ФИШЕР

47

Статистический вывод

Возможно да, возможно нет

DeAGOSTINI

ФИШЕР статистический вывод
47

DeA

ФИШЕР
Статистический вывод

ФИШЕР

Статистический вывод

Возможно да, возможно нет

НАУКА. ВЕЛИЧАЙШИЕ ТЕОРИИ

Наука. Величайшие теории: выпуск 47: Возможно да, возможно нет. Фишер. Статистический вывод. / Пер. с итал. — М.: Де Агостини, 2015. — 176 с.

Рональд Фишер — ученый, снабдивший статистику инструментами, благодаря которым она обрела то огромное значение, которое имеет сегодня. Его основной вклад — статистический вывод, инновационный подход, связанный с понятием вероятности, который дал статистике, состоявшей прежде на службе других дисциплин, необходимый импульс для того, чтобы она стала полноправной наукой. Этому британскому математику и биологу мы обязаны статистическим методом, который применяется в планировании научных экспериментов. Он был ярым сторонником евгеники, зародившейся в первой половине XX века, и в этом контексте его исследования касались также генетики и современной эволюционной теории.

ISSN 2409-0069

Иллюстрации предоставлены:
Galton.org: 37, 61 (вверху); Joan Pejoan (инфографика); Rothamsted Research: 79 (внизу); Американский музей естествознания: 121; Американское философское общество: 119; архив RBA: 59, 111, 138; библиотека Лондонской школы экономики: 141; Джон Сноу: 27; кафедра статистики, университет Беркли, Калифорния: 150; колледж Гонвилл и Киз, Кембридж: 79 (вверху); Конрад Якобс: 130; Международный институт статистики: 157 (внизу); Национальная портретная галерея, Лондон: 61 (внизу слева), 135; университет Аделаиды, Австралия: 33, 157 (вверху слева; вверху справа); Университетский колледж, Лондон / лаборатория Колд Спринг Харбор: 61 (внизу справа); Федеральные немецкие архивы: 80; Эллиот и Фрай / Национальная портретная галерея: 45.

Содержание

Введение

В учебниках по статистике обычно дается следующее определение: это наука, изучающая сбор, организацию и интерпретацию данных. Впрочем, в нем отсутствует важное уточнение: статистика говорит на языке вероятности. Она занимается изучением неопределенности самого разного происхождения. Так, например, статистический вывод оценивает расхождения между наблюдаемой реальностью и предсказаниями теоретических моделей, при этом неизбежно опираясь на вычисление вероятностей. Но кто сделал эту концептуальную и вероятностную инъекцию, которую статистика пережила в конце XIX — начале XX века?

В истории статистики немало выдающихся персонажей: Карл Пирсон, Ежи Нейман и Абрахам Вальд — это далеко не полный список. Но есть среди них один гений — Рональд Эйлмер Фишер. Многие ставшие уже привычными статистические методы берут свое начало в работах сэра Рональда, хотя в большинстве учебников этот факт обойден вниманием. Статьи и книги Фишера, в которых среди математических выкладок всегда находят место логические или философские рассуждения, представляют собой захватывающее и увлекательное чтение, хотя нередко вызывают у читателя раздражение из-за колкостей, которые британский статистик высказывал в отношении ряда своих коллег, не скупясь на оскорбления. И все же попасть в круг Фишера означало принять участие в создании современной математической статистики.

Для наиболее важных достижений нашего героя характерен запутанный исторический контекст, мозаика научных концепций и философских идей. Фишер пришел к статистике через три совершенно отличные от нее науки: из астрономии он знал работы Гаусса и Лапласа; благодаря физике газов ему стало известно о приложениях, разработанных Кетле и Максвеллом; а эволюционнная биология открыла ему путь к самым современным статистическим работам конца XIX века, написанным Фрэнсисом Гальтоном и Карлом Пирсоном.

Говоря о Фишере, нельзя не упомянуть о таком титане, как Карл Пирсон. В работах по математической теории эволюции Пирсон предвосхитил некоторые статистические методы, которые сегодня уже стали классикой. И все-таки он не спешил признать талант Фишера, упрямо отказываясь от поправок, которые перспективный исследователь предлагал внести в его работы. Пирсон дорого заплатил за эту ошибку, так как статьи молодого Фишера открывали новые горизонты, расширяли мир статистики и готовили почву для статистического вывода.

В возрасте 19 лет Фишер поступил в Кембридж, а в 21 год получил должность статистика на опытной сельскохозяйственной станции в Ротамстеде. Там, среди картофеля, удобрений и мышей, родилось то, что принесло ему успех и славу как исследователю. В 1920-е годы Фишер принял эстафету у группы ученых-статистиков, объединившихся вокруг Карла Пирсона, и закрепил научный статус статистики, сведя воедино ее математические основы. Английский статистик подарил науке целый ряд характерных понятий и методов. Он переосмыслил и в некоторых случаях дополнил терминологию: совокупность, выборка, параметр, статистика, дисперсия, правдоподобие, проверка значимости, рандомизация... — это лишь вершина айсберга.

Как архитектор, Фишер тогда же заложил основы теории оценок и теории статистических тестов. Первая занимается определением подходящей оценки для какого-либо неизвестного параметра, например для сравнения кандидатов. Вторая имеет дело с эмпирической проверкой гипотез, касающихся конкретных значений параметров. Когда астроном хочет узнать истинное положение звезды с помощью повторных измерений ее положения, он прибегает к теории оценок. Когда два астронома имеют разные данные о положении звезды и решают провести совместное наблюдение, чтобы устранить все сомнения, они используют теорию статистических тестов. Но это не все. Фишер — создатель так называемого «экспериментального дизайна», или «планирования эксперимента», то есть использования статистики при планировании любых экспериментов.

Весь этот потрясающий багаж был собран в книге *«Статистические методы для исследователей»* (*Statistical Methods for Research Workers*), вышедшей в 1925 году и оказавшей значительное влияние на развитие дальнейших исследований статистиков и математиков, а также сельскохозяйственных инженеров, биологов, химиков и ученых в целом. Границы статистики расширились.

Этот обзор не будет полным, если мы не упомянем генетику, которая, как и статистика, владела помыслами Фишера на протяжении всей его жизни. Наш герой был одним из основателей популяционной генетики — науки, позволившей примирить Дарвина и Менделя (то есть естественный отбор и законы наследования) и создавшей таким образом синтетическую теорию эволюции, или неодарвинизм. Тем не менее интерес нашего героя к этой теме был вызван евгеникой, сомнительной доктриной, которая была популярна в первой половине прошлого века, в том числе в нацистской Германии, — Фишер же видел в ней точку соприкосновения статистики и эволюционизма.

В этой книге мы рассмотрим многочисленные философские и научные споры, в которые Фишер был вовлечен, — часть из них продолжаются по сей день, лишь подтверждая жизнеспособность статистики. Классическая теория статистики, какой мы ее знаем сегодня (включая теорию оценок, проверку гипотез, планирование экспериментов и выборочный метод), — результат деятельности двоих людей: Рональда Эйлмера Фишера и Ежи Неймана. Во многом их работы перекликались, дополняли друг друга, но в то же время между ними наблюдались и противоречия. Оба статистика не желали видеть свое имя рядом с именем оппонента, хотя поначалу они состояли в дружеских отношениях. Это противостояние завершилось лишь со смертью Фишера, считавшего, что Нейман только и мог, что критиковать его.

Британский статистик много размышлял о роли статистического вывода в научном методе, включившись таким образом в полемику с большинством своих коллег. Один из любимых всеми философами, от Аристотеля до Юма, вопросов стал навязчивой идеей Фишера-мыслителя. Речь об извечной проблеме индукции, которую он связал с вероятностью и статистикой, и индуктивные умозаключения приводили к вероятностным выводам.

Предположим, перед нами пациент с подозрением на туберкулез. Чтобы подтвердить эту гипотезу, мы проводим осмотр и делаем рентгеновский снимок, но получаем отрицательный результат. Разумеется, он не является окончательным, так как любое медицинское исследование может оказаться ошибочным, дав так называемый «ложноотрицательный» результат (точно так же иногда имеют место «ложноположительные»). Здесь мы сталкиваемся с настоящим статистическим тестом. В этой ситуации можно сформулировать три вопроса.

1. Чему и в какой степени мы должны поверить, исходя из имеющихся данных? Какова вероятность того, что у пациента туберкулез, при отрицательном тесте?

2. Какие из данных свидетельствуют о правдоподобии гипотезы? Можно ли сделать вывод, что пациент здоров?

3. Что мы должны сделать, основываясь на этих данных? Принять или отвергнуть гипотезу, что у пациента туберкулез?

Первый вопрос касается убежденности, второй — доказательства, и третий связан с принятием решения. Как будет видно далее, в основном Фишер пытался дать ответ на второй вопрос. Байесовские статистики, в свою очередь, отвечают на первый вопрос, а последователи Неймана — на третий. Статистики-байесианцы и сторонники частотной интерпретации (в том числе последователи Фишера и Неймана) образуют два противоположных полюса.

Без сомнения, вклад Фишера изменил научную парадигму того времени, но не стоит думать, что только благодаря ему статистика стала самостоятельной наукой, автономной дисциплиной, и прекратила считаться придатком других дисциплин, таких как астрономия, социология и биология. Природа статистики, которая включает множество различных приложений, сложна, и не так легко определить, что кроме набора математических инструментов обеспечивает ее единство.

Конвергенция различных естественнонаучных и социальных дисциплин определила сегодняшний облик статистики и, как ни парадоксально, освобождение статистики от них. От азартных игр законы статистики, которые проявляются на общем, а не индивидуальном уровне, распространились на астрономию, геодезию, социологию, биологию, сельское хозяйство, промышленность и так далее. Монеты, игральные кости, карты и избирательные урны — это модели, которые мы используем в статистических рассуждениях, касающихся звезд, людей, генов, культур и производства автомобилей. В Античности вероятность и статистика проявлялись в наблюдениях за природой. Благодаря Фишеру они в первую очередь ассоциируются с выборочным методом, то есть случайной выборкой из популяции, хотя последняя является не более чем продуктом воображения ученого-статистика.

Рональд Эйлмер Фишер сделал статистику наукой на полпути между математикой и эмпирикой, которую конкретные практические проблемы стимулируют едва ли не больше, чем проблемы теоретические. Демографические, экономические и медицинские данные придают этой науке актуальность и значимость. Без покровительства этого ученого она превратилась бы в маргинальную теоретическую дисциплину. Статистика переплетается со множеством экспериментальных наук, являясь во многом геометрией для умозаключений. И только это объясняет, как она завоевала почти все области знания в XX веке в условиях технологической революции. Это образец объективности и стандартного подхода, который применяется к официальным измерениям, процессам производства и фармацевтическим исследованиям. Вспомним, что понятие населения как конкретной цифры было слабо определено до тех пор, пока не появились статистические учреждения, отвечающие за точную оценку количества жителей, работников и избирателей в государстве. Благодаря статистике цифры распространились по миру вплоть до самых дальних его уголков.

Искра того огня, который согревает нас сегодня, была высечена нашим героем. Ученый, исключительный по своему интеллекту и своей дерзости, как никто другой укрепил фундамент статистики. Его работы образуют остов науки, которую мы знаем. Глубоко вдохните и приготовьтесь погрузиться в океан статистической науки.

1890 17 февраля в пригороде Лондона на свет появился Рональд Эйлмер Фишер.

1909 Поступает в Кембридж, где изучает математику, астрономию, статистическую механику, квантовую теорию и биологию.

1915 Получает первое важное очко в свою пользу, вычислив распределение коэффициента корреляции в выборке. Доказательство опубликовано в журнале Карла Пирсона *Biometrika*.

1917 Между Фишером и Пирсоном нарушается согласие из-за суровой критики, которой они подвергают друг друга.

1919 Фишер начинает работу на опытной сельскохозяйственной станции в Ротамстеде.

1922 Описывает центральные концепции статистического вывода в статье *«О математических основах теоретической статистики»* (*On the Mathematical Foundations of Theoretical Statistics*).

1925 Выходят в свет *«Статистические методы для исследователей»* (*Statistical Methods for Research Workers*) — одна из книг, повлиявших на распространение статистических знаний среди ученых и инженеров.

1930 Печатается монография *«Генетическая теория естественного отбора»* (*The Genetical Theory of Natural Selection*), примиряющая менделевское наследование с теорией Дарвина.

1933 После отставки Карла Пирсона Фишер возглавляет половину отделения в *University College* в Лондоне — кафедру евгеники. Кафедра статистики переходит в руки Эгона Пирсона, сына Карла.

1935 Выходит книга *«Планирование экспериментов»* (*The Design of Experiments*) — основополагающее руководство для ученых, желающих извлечь максимальную пользу из собственных экспериментов при помощи статистических инструментов. Между Ежи Нейманом и Эгоном Пирсоном начинается полемика по поводу доказательства значимости и проверки гипотез.

1943 Возвращается в Кембридж на кафедру генетики.

1955 Старые споры с Нейманом и Пирсоном разгораются с новой силой, после того как Фишер представляет в *Королевском статистическом обществе* статью об индуктивном выводе.

1958 Фишер рассуждает о курении и раке легких, утверждая, что между ними отсутствует связь.

1962 29 июля умирает в Аделаиде (Австралия), где последние годы своей жизни провел как уважаемый исследователь.

ГЛАВА 1

Статистика до Фишера

В конце XIX века статистические методы были рассредоточены по довольно далеким друг от друга областям. Астрономия хранила работы Гаусса и Лапласа, связанные с методом наименьших квадратов, законом ошибок и вычислением вероятности. Кривая нормального распределения широко использовалась в социологии и физике газов (если провести параллель между молекулами газа и жителями страны). Но принципиально новые статистические достижения века приготовила эволюционная биология.

Рональд Эйлмер Фишер родился 17 февраля 1890 года в Ист-Финчли (Лондон). После рождения старших Джеффри и Эвелин третьего ребенка родители решили назвать Аланом, однако его ранняя смерть вызвала у них предрассудок, связанный с именами: у всех их сыновей без исключения в имени присутствовала буква *у*, в том числе и у младшего из семерых, Рональда Эйлмера (Ronald Aylmer). Талант к математике проявился у Рональда еще в детстве. Когда мальчику было шесть лет, мать стала читать ему популярную книгу по астрономии, пробудившую в нем интерес, который не оставил его и в отрочестве. Впрочем, в школьном возрасте зрение Рональда серьезно ухудшилось: он страдал ярко выраженной близорукостью, и врачи запретили ему читать при электрическом свете. По вечерам преподаватели занимались с ним без бумаги и ручки, что позволило ему развить исключительную способность решать математические проблемы в уме, основываясь на геометрических догадках, но опуская подробности (эта привычка сохранилась на всю жизнь).

Когда Рональду исполнилось 14, от острого перитонита умерла его мать, а вскоре потерял все свое состояние отец. К счастью, Фишер получил стипендию, позволившую ему оплатить обучение в университете. В Кембридже, куда он поступил в 1909 году, юноша изучал математику и астрономию, но также интересовался биологией. По окончании он завершил работу в области «теории ошибок» — математической теории, широко используемой в астрономии, которая наравне с теорией газов была его первым контактом со статистикой. Может показаться странным, что создатель современной математической статистики встретился с дисциплиной, в которую впоследствии внес революционный вклад благодаря астрономии. Это как если бы звезды хранили секреты опросов и выборов. Чтобы объяснить этот факт, а вместе с ним и величие работ Фишера, следует обратиться в прошлое, в XIX век, и рассмотреть возникновение статистических методов в различных дисциплинах и областях.

Принято считать, что статистика состоит из двух взаимосвязанных, но самостоятельных разделов: описательной статистики, которая занимается эксплоративным анализом данных, и статистического вывода, отвечающего за предсказания в условиях неопределенности. Начало статистического вывода лежит в азартных играх и астрономии, хотя для развития этих концепций и создания социальной среды, в которой зародилась описательная статистика, потребовалось разное время. Эта первая фаза продолжалась приблизительно с 1650 по 1850 год. Наконец, во второй фазе, пришедшейся на вторую половину XIX века, статистические инструменты получили новое распространение: из астрономии и социологии в биологию. Но обо всем по порядку.

ОТ ЛАПЛАСА К СОЦИАЛИЗАЦИИ СТАТИСТИКИ

Представим себе, что статистическая наука — это река, образующаяся из двух независимых притоков. С одной стороны — вычисление вероятности, основа статистического вывода. С другой стороны — «наука о государстве» (от ит. *scienza dello Stato*, от которого и происходит слово «статистика»), более всего связанная с описательной статистикой.

Исчисление вероятности возникло в сфере азартных игр, уже развитой в XVII веке — и это помимо вклада Кардано, Галилея и ряда схоластов. Карты, кости и монеты выступали тогда объектами «геометрии случая», как это свидетельствует из переписки, которую с 1654 года поддерживали строгий янсенист Блез Паскаль и заступник математики Пьер де Ферма, вызвавшиеся решить задачу заядлого игрока Антуана Гомбо, шевалье де Мере. Понятие вероятности (кстати, термин встречается уже у Цицерона) было недоступно грекам из-за отсутствия адекватных арифметических символов вроде симметричных костей (возможные результаты бросания не были равновероятными). Это помешало им вывести правило Лапласа — оно встречается уже у Якоба Бернулли (1654–1705) и Абрахама де Муавра (1667–1754) — как аксиому, гласящую следующее: «Вероятность события равна отношению числа исходов, благоприятствующих событию, к общему числу исходов». Следует уточнить, что понятие вероятности не встречается также в переписке Паскаля и Ферма, и впервые оно появляется в книге *«Искусство предположений»* (*Ars Conjectandi*) Бернулли, изданной после смерти автора в 1713 году.

В этом труде в качестве отправной точки Бернулли использовал проблемы, описанные Кристианом Гюйгенсом в книге *«О расчетах в азартной игре»* (*De ratiociniis in ludo aleae*, 1657), и применил к их решению комбинаторное исчисление. Кроме того, он представил «золотую теорему» (вариант закона стабильности частот) и впервые рассмотрел проблему обратной вероятности: сколько наблюдений требуется для определения вероятности, исходя из частоты? Швейцарский математик впервые рассмотрел возможность вычисления вероятности события *a posteriori* (на основе опыта), когда нельзя вычислить ее *a priori* (до проведения опыта, посредством логических или психологических рассуждений).

На рубеже XVIII и XIX веков Пьер-Симон де Лаплас (1749–1827) завершил эти рассуждения, создав алгебраическое исчисление вероятности с помощью математического анализа в своем труде *«Аналитическая теория вероятностей»* (*Théorie analytique des Probabilités*, 1812). Если до него, за редким исключением, при вычислении вероятностей прибегали к алгебре, то после него для этих целей использовался в основном

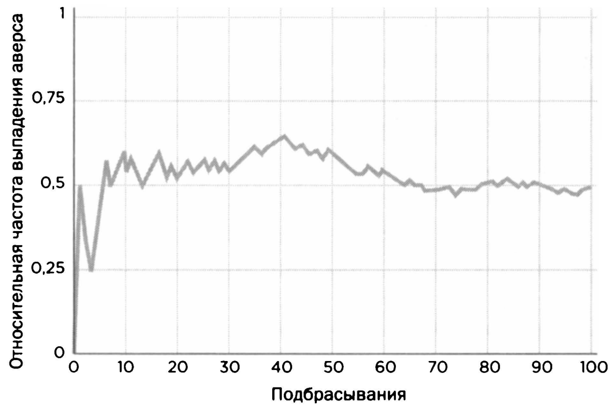
ЗОЛОТАЯ ТЕОРЕМА БЕРНУЛЛИ

Эта теорема, известная сегодня как просто теорема Бернулли, гласит, что относительная частота события стремится к фиксированному числу (вероятности события) по мере того, как увеличивается количество повторений эксперимента. Более формально: для данного события A, его вероятности p и числа независимых попыток эксперимента n, определяющего, произошло A или не произошло, если f — число экспериментов, в которых происходит A из n попыток, и ε — любое положительное число, то вероятность, с которой относительная частота f/n отличается от p более чем на ε (в абсолютном значении), стремится к нулю при n, стремящемся к бесконечности. То есть

$$\lim_{n \to \infty} P\left(\left|\frac{f}{n} - p\right| > \varepsilon\right) = 0.$$

И наоборот, вероятность, с которой относительная частота при большом числе повторений стабилизируется, стремится к 1 (что не означает, будто невозможны редкие отклонения, то есть противоположные серии, «черные лебеди»). Так, относительная частота выпадения аверса при подбрасывании монеты приближается к 0,5 (своей вероятности), когда мы подбрасываем монету достаточное количество раз. Когда-то граф Бюффон подбросил монету 4040 раз, и из них 2048 раз (то есть в 50,69% случаев) выпал аверс. Эта теорема формализует закон случая, или закон стабильности частоты: существует, если употребить термин Бернулли, «моральная уверенность» (вероятность 0,999), что в пределе относительная частота события отличается от его вероятности незначительно (см. рисунок). Это «закон больших чисел» — название введено Симеоном Дени Пуассоном (1781–1840) — в своей более простой форме. Тогда как теорема Бернулли гласит, что относительная частота выпадения аверса при многократном подбрасывании одной и той же монеты стремится к стабилизации, закон больших чисел утверждает, что и относительная частота выпадения аверса при подбрасывании разных монет стабилизируется, даже если у каждой

математический анализ через производящие функции. Лаплас дал строгое определение понятию вероятности и всесторонне рассмотрел проблему обратной вероятности, повторно открыв теорему Байеса (в честь Байеса много лет спустя ее назовет Август де Морган, восстановив таким образом историческую справедливость). Кроме того, он заложил основы байесовского статистического вывода, которым пользовался для предсказа-



Относительная частота выпадения аверса после 100 подбрасываний монеты.

монеты своя вероятность выпадения аверса. Пафнутий Чебышев и русская математическая школа продолжили исследование законов больших чисел, обобщивших золотую теорему. Для Бернулли теорема означала возможность эмпирически вычислить неизвестные вероятности, а также определить вероятность объективно, применив теорему «наоборот». В самом деле, если частота стремится к вероятности пропорционально увеличению числа наблюдений, почему бы не определить вероятность на основании частоты? С помощью индукции казалось возможным определить вероятность как предел частоты, и не только в чисто логическом или субъективном смысле (как степень убежденности). Тем не менее французский математик Абрахам де Муавр, живший в Англии (он был гугенотом), известный своим трактатом *«Доктрина шансов»* (*Doctrine des chances*, 1718), считал, что статистическая закономерность, постулируемая золотой теоремой, для функционирования требует участия Бога. Фишер, как мы объясним далее, перенял этот подход к интерпретации вероятности.

ния процента заключенных браков и выявления соотношения полов среди новорожденных. Также Лаплас обратился к теории вероятностей при решении различных проблем небесной механики, например изучения распределения орбит комет, как если бы это была серия обычных бросков костей. Тем не менее более важные приложения пришли вместе с «теорией ошибок», которую в свое время изучал Фишер.

В период с 1770 по 1820 год в астрономии развивались статистические методы, так как эта наука требовала тщательного изучения ошибок. Речь идет о сведении к минимуму оценки позиции планеты или звезды на основании серии наблюдений. Астрономы хотели определить реальное положение светила после проведения нескольких измерений. Лаплас догадался, что реальное положение звезды было причиной наблюдаемых положений, в то время как ошибки являлись случайными. Умело используя теорему Байеса, он пришел к выводу, что существует кривая, описывающая распределение ошибок во-

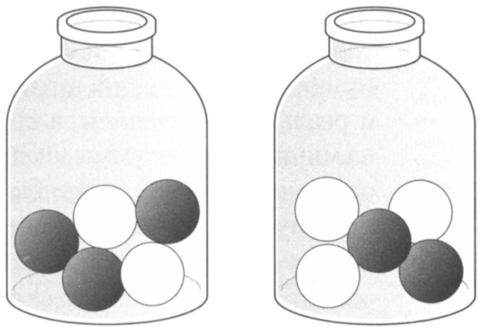
ТЕОРЕМА БАЙЕСА

В книге *«Воспоминания о вероятности причин событий»* (*Mémoire sur la probabilité des causes par les évènemens*, 1773) Лаплас утверждает, что ситуации, связанные со случайностью, бывают двух типов. В ситуациях первого типа случай проявляется в результатах. Известно, что в урне находятся белые и черные шары; вопрос: какой шар мы вытащим? На основе причин (количество белых и черных шаров в урне) вычислим вероятность результата: вытащим ли мы белый или черный шар? В ситуациях второго типа случай проявляется не в результатах, а в причинах. Мы знаем результат опыта (например, мы вытащили черный шар) и хотим вычислить состав содержимого урны, который нам неизвестен. Основываясь на результатах (извлечен черный шар), определим вероятность причин, то есть каждого возможного состава содержимого урны. Таким образом мы перейдем от результатов к причинам. Лаплас изложил и доказал теорему, открытую достопочтенным Томасом Байесом (1702–1761) в посмертно изданных мемуарах (1763), с которой он совершенно точно не был знаком (французские математики обычно не читали труды британцев). Эта теорема гласит, что если $\{A_1, A_2, \ldots, A_n\}$ образуют множество взаимоисключающих и исчерпывающих событий, $P(A_i)$ — вероятности *a priori* и $P(B|A_i)$ — условные вероятности (вероятности наблюдать эффект B при условии A_i), тогда вероятность *a posteriori* каждого события равна

$$P(A_i|B) = \frac{P(A_i)\cdot P(B|A_i)}{\sum_{k=1}^{n} P(A_k)\cdot P(B|A_k)}.$$

круг истинного значения (см. рисунок 1). Кривая симметрична и убывает по мере удаления от центрального значения, то есть чем дальше от него, тем менее вероятны такие ошибки измерения. Вследствие этого вероятнее всего выбранное в качестве реального значение (арифметическое среднее результатов) находится вблизи центрального значения, где кривая достигает максимума. Решив дифференциальное уравнение, Лаплас получил кривую распределения ошибок, которая оказалась экспоненциальной.

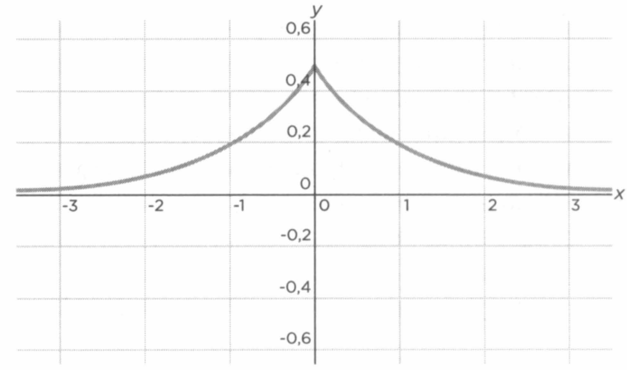
Здесь нас интересует объяснение идеи, скрытой в формуле Байеса и обнаруженной Лапласом, так как она была одним из коньков Фишера. Представим себе урну, в которой могут быть два разных набора шаров: 2 белых и 3 черных или 3 белых и 2 черных, как на рисунке. Вытаскиваем случайный шар, он оказывается черным; какой набор шаров в урне более вероятен? Интуиция подсказывает, что поскольку извлеченный шар оказался черным, первый состав более вероятен, чем второй (так как во втором меньше черных шаров). Теорема Байеса обеспечивает численную оценку этой интуитивной догадке. Две возможные причины события «вытащить черный шар» — два возможных состава урны. Если предположить *a priori*, что два состава равновероятны (вероятность каждого равна 0,5), после применения формулы Байеса вероятность первого состава увеличивается до 0,6 из-за извлечения черного шара, тогда как вероятность второго состава уменьшается до 0,4. Вероятности *a priori* (0,5 и 0,5) корректируются *a posteriori* (0,6 и 0,4). Этот факт кажется неоспоримым, так как в первом составе черных шаров больше, чем во втором. Следовательно, в этих условиях ожидается большая вероятность извлечения такого шара. Для Лапласа, как и для Байеса, эта важная теорема означала возможность обучения через опыт и до некоторого предела оправдывала индукцию.



Если мы вытащили черный шар, то согласно теореме Байеса вероятность *a posteriori* состава на рисунке слева выше, чем вероятность состава на рисунке справа.

Закон ошибок по Лапласу: $\phi(x) = \dfrac{e^{-|x|}}{2}$.

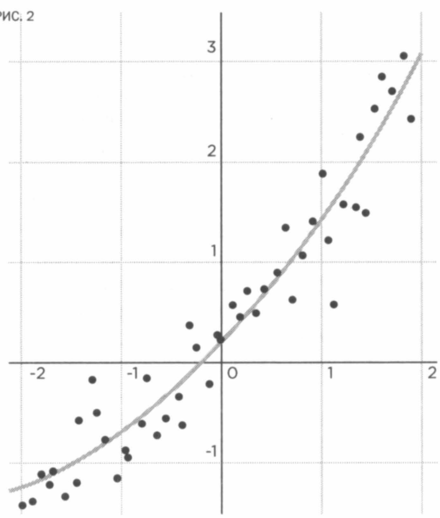
РИС. 1



Метод наименьших квадратов помогает подогнать к наблюдениям кривую, которая соответствует минимальной квадратичной ошибке.

Размещая результаты последовательных наблюдений светила вдоль траектории, Лаплас пытался минимизировать сумму абсолютных ошибок, то есть разниц между наблюдаемым и реальным значением, а другие астрономы сосредоточились на минимизации суммы квадратов ошибок (квадрат имеет одинаковое значение для ошибки как в большую, так и в меньшую сторону). Этот метод оценки удалось распространить на более значительное число переменных, и он оказался проще для расчетов, чем метод Лапласа. Речь идет о методе наименьших квадратов (см. рисунок 2), описанном Адриеном Мари Лежандром в книге *«Новые методы определения орбит комет»* (*Nouvelles méthodes pour la détermination des orbites des comètes,* 1805). Однако молодой немецкий математик Карл Фридрих Гаусс заявил, что был первым, кто использовал этот метод для предсказания орбиты астероида Цереры, открытого в первый день XIX века, 1 января 1801 года.

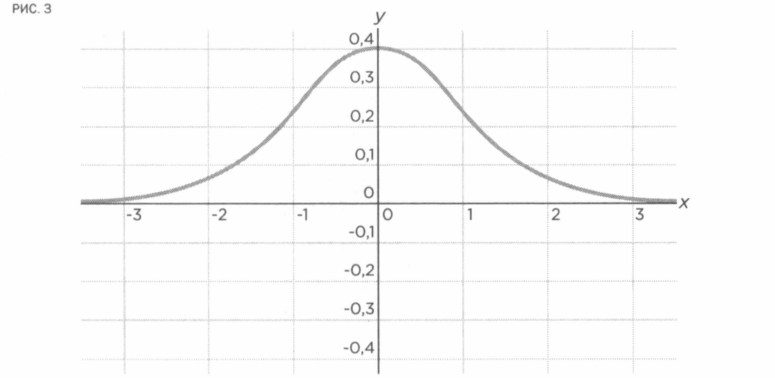
РИС. 2



В своем труде *«Теория движения небесных тел»* (*Theoria motus corporum coelestium*) 1809 года в контексте теории ошибок Гаусс изложил изобретенный им метод подгонки кривой к облаку точек. Он показал, что распределение ошибок связано с методом наименьших квадратов. Определив кривую, соответствующую минимуму квадратичной ошибки, Гаусс заметил, что ошибки подгонки случайно распределяются вокруг среднего значения. Это симметричное колоколообразное распределение было так называемым *нормальным распределением*, или *колоколом Гаусса* (см. рисунок 3), хотя в то время оно было известно как *закон ошибок*. И наоборот, Гаусс показал, что если предположить, что распределение ошибок соответствует этому общему закону, функция минимальных квадратов будет минимизировать вероятность ошибок или, таким же образом, делать наблюдения более правдоподобными (хотя поначалу, используя теорему Байеса, вдохновленный Лапласом, он думал иначе).

Вскоре Лаплас перенес ценное открытие немецкого математика в сферу теории вероятностей, добавив собственное новшество — центральную теорему пределов, которая гласит, что если величина является суммой большого количества факторов, подверженных ошибкам, то она распределена нормально, независимо от распределений каждого отдельного фактора. Эта теорема показала, что стремление биномиального распре-

РИС. 3



Закон ошибок по Гауссу:

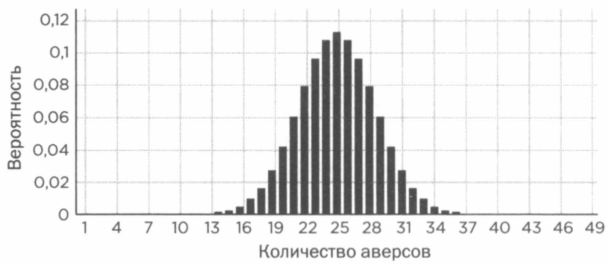
$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

деления к нормальному, использованное де Муавром как вычислительный инструмент без какого-либо вероятностного значения, является всего лишь частным случаем более общего явления. Любая сумма или среднее, а не только количество успехов в n экспериментах (что показал де Муавр), распределены примерно нормально, если n достаточно велико (см. рисунок 4). Другими словами, эта теорема показывала, что в определенных достаточно общих условиях имеет смысл рассматривать переменную как имеющую нормальное распределение. Эту группу методов и теорем историки науки называют синтезом Гаусса — Лапласа.

Если первым предшественником статистики является французская *«математическая теория вероятностей»*, то вторым — «наука о государстве», то есть анализ социоэкономических данных, связанных с расцветом торговли и территориальным устройством Германского государства, точнее, с объединением двух традиций, уходящих корнями в середину XVII века: английской *Political Arithmetic* (*политической арифметики*) и немецкой *Statistik* (*статистики*). Термин «политическая арифметика» введен Уильямом Петти, который утверждал, что с политическим организмом следует работать, имитируя новую натуральную философию, дабы оптимизировать процесс принятия решений. Сюда же относятся наблюдения о таблицах смертности Джона Гонта 1662 года, его исследования демографических данных имели большое значение для подсчета пожизненных рент и страховых премий. Интересно

Вероятность получить определенное число аверсов при 50 подбрасываниях монеты имеет распределение, приближающееся к нормальной кривой.

РИС. 4



ВОЗМОЖНОСТИ СТАТИСТИЧЕСКОГО ГРАФИКА

Знаменитый английский врач Джон Сноу (1813–1858) первым показал графически, что причиной эпидемии холеры, поразившей Лондон в 1854 году, стал зараженный колодец, вокруг которого сконцентрировались инфицированные (представлены точками). То есть инфекция передавалась не воздушным путем, как считали ранее. Более 700 жителей квартала Сохо погибли меньше чем за неделю, употребляя воду из одного зараженного фекалиями источника (изображен крестом на Брод-стрит, в центре рисунка). Иллюстрация соответствует оригинальной карте Джона Сноу. Точки обозначают людей, пострадавших от болезни, а кресты — колодцы, из которых они брали воду.

отметить, что при изучении этих таблиц братья Гюйгенсы догадались, что азартные игры могут дать полезную модель для других областей, и ввели термин «ожидаемая продолжительность жизни» — по аналогии с ожидаемым или наиболее вероятным выигрышем. Немецкое слово *Statistik* возникло в ответ на желание охарактеризовать новые государства, в частности Пруссию: ее статистику, цифры и индексы, так как таможенные сборы внутри германских государств были установлены пропорционально их населению.

Английская и немецкая традиции объединились в конце XVIII века на Британских островах, а французская математика дополнила их лишь на исходе XIX века. С этого момента количественное изучение политики и общества предполагало использование математических инструментов, уже широко применявшихся в учении о случайности и в астрономии. Включение французской теории вероятностей в социологию связано с именем бельгийского астронома Адольфа Кетле (связь с астрономией неслучайна), но принятие английской и прусской науки о государстве потребовало времени, и работы Лапласа и Кетле были замечены в Великобритании астрономом Джоном Гершелем и логиком Августом де Морганом не сразу.

«СРЕДНИЙ ЧЕЛОВЕК» ПО КЕТЛЕ

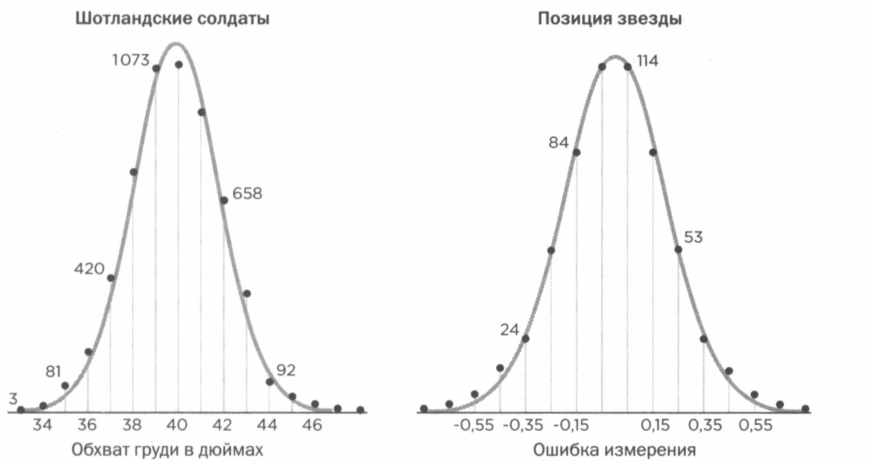
К концу наполеоновской эпохи лавина статистических данных в прессе касалась уже не рождений, смертей и браков, а количества суицидов, убийств и неграмотных. Эти цифры, связанные с преступностью и нехваткой образования, стали культурным ландшафтом, на котором вызрела идея «среднего человека» (*homme moyen*), благоприятствовавшая эрозии детерминизма.

Адольф Кетле (1796–1874) получил образование в Париже, где благодаря своему учителю Жозефу Фурье познакомился с синтезом Гаусса — Лапласа. Кетле приводила в замешательство стабильность статистических данных: усилилась бюрократия, и он увидел, с какой потрясающей точностью воспроизводились преступления, — криминальная статистика во Франции ежегодно давала почти одни и те же цифры. С 1825 по 1830 год число обвиняемых насчитывало около 7100, а число осужденных — 4400. Вернувшись в Брюссель, он заинтересовался технологией переписей и опросов.

Движимый юношеским стремлением стать скульптором, Кетле применил вероятностный подход к пропорциям человеческого тела (именно этому бельгийскому астроному мы обязаны термином «индекс массы тела») и затем с легкостью использовал его в астрономии и геодезии. В 1835 году Кетле заявил, что закон ошибок (или «закон случаев», как он предпочитал его называть) распространяется также на физические и поведенческие характеристики человека, и основным его понятием является *середина*, так как среднее значение распределения изучаемого показателя давало «среднего человека». Некоторые антропометрические показатели — такие как рост французских новобранцев или обхват груди шотландских солдат — имели распределение, сходное с колоколообразной кривой Гаусса. В 1845 году, изучив и графически представив данные по обхвату груди 5738 шотландских солдат, взятые из медицинского журнала того времени, Гаусс заметил, что получен-

Слева: кривая частот, соответствующих обхвату груди 5738 шотландских солдат, по Кетле (1845). Справа: кривая частот ошибок, совершаемых при наблюдении звезды, по Фридриху Бесселю (1818).

РИС. 5



ная кривая частот похожа на кривую, которую дает измерение положения звезд (см. рисунок 5).

И пока астроном много раз определял положение одной и той же звезды, зная реальное значение этого показателя, Кетле изучал данные разных солдат, и его кривая не была подкреплена реальным значением обхвата груди. Кетле утверждал, что измерение обхвата груди множества солдат эквивалентно многократному измерению одного и того же «среднего солдата». И, сделав гигантский онтологический скачок, он предположил, что природа стремится к среднему человеку, а индивиды, находящиеся по бокам колокола, суть случайные отклонения от идеального канона. Его работа положила начало социальной физике и послужила международной оценке значения статистики, предвосхитив, кроме прочих специализированных институтов, основание Статистического общества в Лондоне.

Не следует забывать, что связь вероятности и статистики с обществом, пусть в зачаточной форме, присутствовала уже у Лапласа, с того момента, как французский астроном подхватил эстафету «моральной арифметики», описанной Кондорсе в *«Рассуждении о применении анализа к оценке выборов большинством голосов»* (1785). Чтобы получить представление о сути этой работы, можно обратиться к последней части трактата Бернулли, посвященной применению исчисления вероятностей к вопросам государства, морали и экономики. Так, по его собственным словам, он попытался объединить мудрость философов и благоразумие политиков. В *«Философском эссе о вероятностях»* (*Essai philosophique des probabilités*), первоначально напечатанном как введение ко второму изданию *«Аналитической теории вероятностей»* (*Théorie analytique des Probabilités*, 1814), Лаплас писал, что «фундаментальные проблемы жизни — по сути не что иное, как проблемы вероятности».

Для Лапласа вероятность являлась основой статистического вывода, теории ошибок, философии, причинности и даже квантификации достоверности свидетельств. Если исчисление вероятностей оказалось настолько эффективным в естественных науках, почему оно не могло быть столь же эффективным в науках политических и вопросах морали? В своей брошюре

Лаплас сравнивал решения ассамблеи и судебные приговоры с шарами, извлекаемыми из урны, определяя вероятность ошибки как функцию количества депутатов в ассамблее или количества голосов, которых не хватало для осуждения обвиняемого. Таким образом он улучшил более ранние вычисления Кондорсе. Лапласа забавляло, что наука, начинавшаяся с рассуждений о монетах, костях и игральных картах, со временем стала одним из сосредоточений человеческого знания.

Урна, которую мы исследуем, — сама природа.

Адольф Кетле (1845)

Виднейший ученик Лапласа, Симеон Дени Пуассон, многое сделал, чтобы связать статистику и социологию, продолжив дело Кетле. В 1835 году он исследовал вопросы электоральной математики и юриспруденции и сформулировал «закон больших чисел», который обеспечивал лучшую основу применению исчисления вероятностей к социальным проблемам, объясняя статистическую стабильность в социальных изменениях. Большое количество индивидуумов, действуя в рамках системы, определяют регулярность, которая не зависит от их взаимной координации. Значит, говорить о коллективе можно, даже не имея подробных сведений об индивидуумах. Следовательно, нельзя предсказать поведение индивидуума — только среднее поведение в популяции. Это еще одно проявление статистической регулярности мира. В поведении и качествах миллионов своих сограждан астрономы Пуассон и Кетле видели достоинства и постоянство звезд.

Кетле взял кривую Гаусса, выведенную именно как закон ошибок или распределение в азартных играх, и применил ее к биологическим и социальным явлениям, в которых среднее не является реальной величиной. Среднее не было характеристикой отдельного индивида, но характеристикой популяции, которая упрощала начальные данные. Она помогала представить популяцию как объект исследования таким образом, что различные индивиды считались большими или меньшими отклонениями от среднего человека. Для Кетле наблюдаемый разброс значений был простым изменением, природным отклонением. Не будучи заинтересованным в глубоком изучении вариабельности, бельгийский астроном отождествлял среднее с правильным и корректным. Когда с его работами ознакомились в Великобритании, колоколообразная кривая получила название *нормального закона*. Нормальными считались люди, соответствовавшие центральной тенденции социальных законов, описывающих рост, вес и интеллект. Социология пошла дальше, назвав людей, чьи показатели удалялись от средних, патологическими, «анормальными». Но влияние работ Кетле этим не ограничилось, поскольку они дали дорогу статистической механике Джеймса Клерка Максвелла (1831–1879). Молекулы газа ведут себя, как индивидуумы в популяции, и беспорядок на уровне отдельно взятого человека трансформируется в порядок на уровне популяции. Недаром теория газов стала еще одной дисциплиной (наравне с теорией ошибок), примененной Фишером к изучению методов классической статистики.

СЭР ФРЭНСИС ГАЛЬТОН: «ПОСРЕДСТВЕННЫЙ ЧЕЛОВЕК» И ЕВГЕНИКА

Чтобы понять, каким образом статистические методы из области социальной физики проникли в область физической антропологии и в особенности в эволюционную биологию, обратим внимание на статистическую вариабельность, изучению которой способствовало появление дарвинизма и евгеники. Неполнота генетических теорий Чарльза Дарвина (1809–1882) подтолкнула его кузена Фрэнсиса Гальтона (1822–1911) к попытке разрешить проблемы наследования признаков при помощи математического анализа биологических данных.

Ровесник Грегора Менделя (1822–1884), Гальтон был на 13 лет моложе Дарвина. Получив благодаря щедрому наследству медицинское и математическое образование, он отправился исследовать Африку (среди его изобретений — карты

Портрет молодого Фишера, дипломированного математика. Снимок сделан в 1913 году, вскоре после его перехода в Кембриджский университет, где в результате знакомства со статьями Карла Пирсона проявился его интерес к генетике и эволюции.

антициклонов и спальный мешок). По возвращении в Англию, что совпало с укреплением колониальной антропологии, он заинтересовался эволюцией. Гальтон был очарован первой главой *«Происхождения видов»* (1859), в которой речь шла о колебаниях в случаях одомашнивания животных, и это положило начало регулярной переписке с Дарвином вплоть до смерти последнего. Задумавшись о возможности управления естественным отбором, Гальтон начал разрабатывать программу улучшения человечества посредством селекции человеческих особей. В книге *«Наследственность таланта, ее законы и последствия»* (*Hereditary Genius*) (1869) он утверждает:

> «Точно так же, как методами тщательной и умелой селекции в рамках естественных ограничений удается получить стабильную породу собак или лошадей, обладающих особенными способностями к бегу или к чему-либо еще, представляется вполне возможным произвести высокоталантливую расу людей путем рассчитанных браков в течение нескольких последовательных поколений».

В 1883 году Гальтон ввел термин *евгеника* (наука об улучшении человеческой расы). В конце века это учение получило признание в британском обществе, погруженном в проблемы империи — как внешние (по отношению к другим империям), так и внутренние (усиление люмпен-пролетариата — низших социальных слоев, в которых уровень рождаемости был значительно выше, чем в привилегированных классах). Оно закрепилось в США и нацистской Германии вместе с провозглашением законов о принудительной стерилизации душевнобольных и неимущих. Евгеническое движение развивалось практически без остановки до тех пор, пока не были уничтожены концентрационные лагеря Центральной Европы и разделение человечества по расовому признаку не объявили аморальным мифом довоенной физической антропологии.

Гальтон твердо верил, что английское общество страдает от некоторого регресса, наследственного биологического вырождения, которое проявляется в военных трудностях Британской империи, связанных, вероятно, с врожденной слабостью солдат. Евгеническая наука должна была найти решение проблемы, благоприятствуя развитию полезных качеств и ограничивая размножение в неугодных сословиях.

В отличие от Гальтона, взгляды Дарвина были более осторожными. В книге *«Происхождение человека и половой отбор»* (1871) он рассматривал вопрос рас и, в целом принимая евгенические теории, выражал некоторые сомнения относительно них. То, что Дарвин принимал теории, основанные на видимом наследовании приобретенных качеств, которые разработал Ламарк, может удивлять. Однако с позиции классического дарвинизма механизмом наследования, стоящим за адаптациями, являлись повторяющиеся аномалии. Теория «пангенеза», предложенная Дарвином за неимением лучших альтернатив, согласовывалась с ламаркистской наследственностью (притом что Гальтон распространил противоречащие идее «геммул» результаты серии экспериментов с кроликами, чем вызвал раздражение Дарвина). И только неодарвинизм — результат синтеза классического дарвинизма, менделевой генетики и популяционной генетики — вытеснил ламаркизм с научной арены (причиной наследуемых изменений являются мутации в ДНК).

На склоне лет Гальтон написал утопический роман *Kantsaywhere*, где нарисовал счастливое общество, живущее согласно евгеническим предписаниям жрецов-ученых. Его племянница (у Гальтона не было детей), возмущенная некоторыми непристойными сценами, сожгла отдельные фрагменты рукописи. Влияние идей Гальтона было значительным, они дали стимул социальному дарвинизму и ввели статистику в изучение психологии. К концу столетия антропометрические тесты Гальтона обрели вид популярных интеллектуальных тестов.

ЗАКОН РЕГРЕССИИ К СРЕДНЕМУ И ПОНЯТИЕ КОРРЕЛЯЦИИ

Наиболее долгосрочным влиянием Гальтона стало использование статистики в качестве инструмента для укрощения наследственной изменчивости. Разносторонний британский ученый полагал, что явление можно познать, только если его можно измерить, что в дальнейшем определило путь развития количественной физической антропологии, или антропометрии. Согласно Гальтону, физические характеристики, такие как рост или вес, а также характеристики личности, являются наследуемыми. Гальтон считал, что союз двух умных людей приведет к рождению еще более умных детей, точно так же как у двух высоких людей рождаются еще более высокие дети. Однако эксперименты с наследованием, которые он проводил в течение всей жизни, привели к открытию другой статистической закономерности, отличной от ожидаемой. В своей книге *«Естественное наследование»* (*Natural Inheritance*, 1889) он назвал ее *«возвращением к посредственности»*, а позднее — *«регрессией к среднему»*. Гальтон использовал это понятие, чтобы описать связь между ростом родителей и детей. Он заметил, что у высоких родителей обычно рождаются высокие дети, а у низких родителей дети, как правило, невысокого роста. Но когда родители особенно высокие или особенно низкие, наблюдается возврат к среднему росту в популяции, и дети регрессируют, то есть возвращаются к среднему росту родителей. Гальтон расширил этот результат, сформулировав универсальный закон наследования: каждая черта человека наследуется его потомками, но в целом в незначительной степени (сегодня мы знаем, что это не столько биологическая, сколько исключительно статистическая закономерность: более вероятно, что значения нормальной случайной величины будут ближе к ее среднему, ожидаемому значению).

В 1877 году Гальтон описал это явление, проведя эксперименты с размерами семян в последовательных поколениях гороха. Мендель экспериментировал с качественными свойствами (цвет, шероховатость и так далее) гороха, Гальтон же измерял количественные свойства (размеры, диаметр). Повторив анализ на данных антропометрических регистров (где, кстати, ввел перцентили и дал новую оценку использованию медианы и квартилей), он обнаружил следующее линейное соотношение:

МАШИНА ГАЛЬТОНА, КВИНКОНС

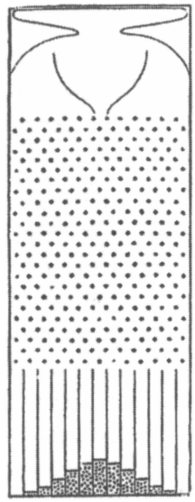


Схема *квинконса* в книге *«Естественное наследование»* (1889).

Разносторонний Гальтон пытался объяснить нормальное распределение некоторых физиологических показателей (таких как рост людей или диаметр горошин). Для доказательства того, что нормальный закон является законом вариации, а не только законом ошибок, в 1873 году он придумал *квинконс* — устройство, где каждый столбик окружен четырьмя другими; его Гальтон использовал для иллюстрации центральной предельной теоремы. Устройство предполагало стол, на который сверху клали бильярдные шары. Они скатывались вниз, отскакивая от столбиков случайным образом и собираясь в отдельные карманы в нижней части стола. С помощью этого устройства Гальтон показал, что шары в нижней части стола образуют колокол нормального распределения, как видно на рисунке. С помощью этого искусного механизма он объяснил преобладание нормального распределения, а кроме того, проиллюстрировал наследование с помощью расположения по фазам. Создавая препятствия для шаров в некоторых зонах, имитируя доминантные воздействия в наследовании, он снова наблюдал образование нормальной кривой, но меньшего размера и менее широкой. Английский ученый был настоящим гением представления абстрактных конструкций с помощью физических моделей. Подобными исследованиями он связал теорию ошибок — согласно которой накопление случайных отклонений порождает нормальное распределение — с наследственностью, наряду со случайными отклонениями обеспечивающей очевидные корреляции, так как каждый организм в какой-то степени напоминает своих предков.

рост ребенка, см = 85 см + 0,5 × рост родителя, см.

Это и была одна из прямых регрессии. Кроме того, он предположил, что связь между двумя переменными (рост родителя и рост ребенка) может быть охарактеризована количественно. Речь идет о большом достижении статистики того века — о корреляции.

Работа Гальтона о регрессии происходила из его исследований наследственности, теория корреляции родилась как ответ на проблему идентификации преступников (он впервые предложил снимать отпечатки пальцев). Гальтон быстро понял, что система идентификации, предложенная французским полицейским Альфонсом Бертильоном (1853–1914), избыточна. Бертильон предлагал регистрировать рост, размер стоп, предплечий и пальцев каждого человека. Эти четыре параметра не являются независимыми, так как у высоких людей обычно большие ступни и длинные предплечья и пальцы. Гальтон предположил, что по сути это тот же вопрос, который возник у него при исследовании регрессии: корреляция между переменными. В статье 1888 года он представил первый математический инструмент для измерения корреляции, то есть зависимости между переменными. Хотя коэффициентом корреляции в 1892 году его назвал экономист Фрэнсис Эджуорт и в 1896 году окончательно оформил в своих работах математик Карл Пирсон (о нем речь пойдет в следующей главе), разделив лавры открытия с французским астрономом Огюстом Брависом, который еще в 1846 году предложил аналогичную математическую формулу, изучая ошибки корреляции между координатами объектов. Сегодня этот показатель известен как *коэффициент линейной корреляции Пирсона*, он позволяет охарактеризовать положительную и отрицательную корреляцию (по-видимому, Гальтон не рассматривал случай, когда увеличение одной переменной связано с уменьшением другой).

Гальтон всегда говорил, что евгеника, стремление улучшить физические или умственные качества человеческой расы, стала импульсом, подтолкнувшим его к изучению побочной проблемы статистической вариации. До тех пор статистические методы занимались лишь обобщенными материями, не интересуясь индивидуальными отличиями. Для Кетле средний человек был центром тяжести социального тела, вокруг которого вращались социальные атомы, отдельные люди. Этот средний человек был воплощенным совершенством, так как у него не было никаких отклонений — ни в большую, ни в меньшую сторону. Гальтон признавал заслуги Кетле, называя его величайшим авторитетом в социальной статистике, ведь с него началось использование нормальной кривой не как закона ошибок, а как описания распределения. Но у этих двух ученых возникло фундаментальное расхождение в понимании статистических законов, во многом связанных с интересом Гальтона к исключительному, в противовес интересу Кетле к средним значениям.

Кетле занимался центральной тенденцией, то есть средним, а Гальтон, исследуя крайности, сосредоточивался на хвостах распределения и дисперсии. Гальтон обращал внимание на значительно отличающихся от среднего индивидуумов: средний человек Кетле был не прообразом совершенства, а посредственностью, которая должна была развиваться. Прекрасное находилось на одном из краев нормальной кривой таланта. Этот подход был признан, только когда нормальность стала считаться посредственностью, и в результате естественный отбор Дарвина и евгеническая реформа вновь пробудили интерес к вариабельности: исключительность была не ошибкой природы, отклонением от идеального среднего человека, но важным для улучшения расы колебанием. Так, из инструмента, созданного для минимизации ошибок, статистика превратилась в модель для представления вариабельности, связанной со случайностью. Переосмысление нормальной кривой как закона о вариации, а не ошибки, стало центральным результатом статистической мысли XIX века.

> Если бы греки знали о нормальном законе, они боготворили бы его.
>
> **Фрэнсис Гальтон, «*Естественное наследование*» (1889)**

Наш герой, Фишер, познакомился с секретами статистики в результате любопытного слияния астрономического, физического и биологического знания. Через астрономическую теорию ошибок он узнал о синтезе Гаусса — Лапласа, об объединении исчисления вероятностей, метода наименьших квадратов и закона ошибок. Благодаря кинетической теории газов он научился моделировать обобщенные свойства с помощью нормального распределения. И наконец, прогресс биологии и антропологии при участии Гальтона помог ему узнать о главном достижении статистики XIX века — корреляции.

ГЛАВА 2

Карл Пирсон и биометрическая школа

Труд Фишера невозможно оценить без сопоставления с наследием его предшественника, Карла Пирсона. Стремясь вывести математическую теорию эволюции, Пирсон разработал несколько ставших классическими понятий (гистограмма и стандартное отклонение) и методов статистики (регрессионный анализ и критерий «хи-квадрат»). Поправки, которые молодой Фишер внес в ряд работ Пирсона, привели к их ожесточенному соперничеству, продолжавшемуся в течение всей жизни.

Еще в Кембридже Фишер прочитал статьи математика Карла Пирсона, собранные под заголовком *«Математический вклад в теорию эволюции»* (*Mathematical Contributions to the Theory of Evolution*). Воодушевленный этим сборником, объединяющим два его основных увлечения (статистику и биологию), Фишер выполнил свое первое научное исследование. Шел 1912 год, и студенту Фишеру было всего 20 лет.

К моменту окончания университета финансовое положение его семьи пошатнулось, и Фишер, не теряя времени, устроился статистиком в торговое общество и какое-то время работал на фабрике в Канаде. В 1914 году, по возвращении в Англию, что совпало с началом Первой мировой войны, он пытался записаться добровольцем, но был признан негодным к строевой службе ввиду слабого зрения. В 1917 году он тайно женился на 17-летней Рут Эйлин, которая впоследствии родила ему восьмерых детей — двоих сыновей и шестерых дочерей (старшая, Джоан, вышла замуж за статистика Джорджа Бокса). В 1919 году, когда Фишер уже несколько лет работал преподавателем физики и математики в разных школах, ему представились сразу две возможности сделать карьеру. Пирсон предложил ему пост статистика в лаборатории Гальтона, и одновременно освободилось место на экспериментальной сельскохозяйственной станции в Ротамстеде — исследовательском центре с древнейшими в Соединенном Королевстве традициями.

Фишер разрешил дилемму, склонившись ко второму варианту — Ротамстеду. Основной причиной для этого послужило то, что работа в лаборатории Гальтона означала бы, что все его публикации станет утверждать Пирсон, и это условие было для Фишера абсолютно неприемлемым. Особенно если учесть, что за несколько лет разногласия между Фишером и Пирсоном лишь нарастали и продолжили нарастать дальше: *верное* распределение коэффициента корреляции, *точное* число степеней свободы в критерии χ^2 («хи-квадрат»), *эффективность* метода оценки моментов... То, что началось с дружеских отношений, испортилось из-за недопонимания. Несмотря на молодой возраст, Фишер вносил разнообразные исправления в работы Пирсона и своих ближайших коллег, что сам Пирсон находил возмутительным, а высокомерный Фишер не пытался улучшить давшие трещину отношения. Чтобы уяснить, каким образом теоретические ошибки Карла Пирсона способствовали подъему исследований Фишера, следует обратиться к основному действующему лицу статистики викторианской эпохи и его главному труду.

МЕЖДУ ЭЛАСТИЧНОСТЬЮ И БИОМЕТРИЕЙ

С 1884 года Пирсон был профессором прикладной математики и механики в Университетском колледже Лондона. Получив кафедру, он специализировался на теории эластичности, так как во второй половине XIX века эластичность являлась проблемой для развития космологии, поскольку электромагнитная передача требовала наличия эластичного эфира. Но призвание Пирсона не было строго научным. В значительной мере его притягивали гуманитарные области: он любил литературу, историю и философию, даже когда полностью сосредоточился на развитии статистических методов в эволюционной биоло-

НЕВЕРОЯТНЫЙ ЧЕЛОВЕК

Карл Пирсон (1857–1936) родился в Лондоне. Его семья принадлежала к среднему классу, что позволило юноше изучать математику в Кембридже, а после окончания в 1879 году продолжить обучение в университетах Гейдельберга и Берлина. Там он без устали читал и писал. Его интересовали поэзия, театр, этика, социализм, права женщин и так далее. В 1880 году под псевдонимом *Loki* он даже издал драму *«Новый Вертер»* (*The New Werther*). В 1892 году Пирсон опубликовал *«Грамматику науки»* (*The Grammar of Science*), в которой представил свою философию науки, объединив идеализм неокантианского философа Куно Фишера и позитивизм Эрнста Маха, который он изучил в Германии (неслучайно Пирсон написал свое имя на немецкий манер — Karl). В дальнейшем эта книга выдержала множество переизданий, снискав большой успех. Так, в 1902 году Альберт Эйнштейн создал в Берне небольшую группу по изучению этого текста, повлиял этот труд и на формулировку специальной теории относительности. Одна из центральных идей книги: функция науки должна быть ограничена описанием наблюдаемых фактов, избегая любых отклонений в метафизику. Научные законы являются не объяснением причины, а упорядоченными утверждениями явлений. Другими словами, они не объясняют, почему явления происходят, а только описывают, каким образом. Пирсон стремился продвигать через науку идею национального благосостояния и считал, что наука должна формировать базовые культурные ценности общества. Как и Гальтон, он отстаивал евгенику и не раз выражал желание подвергнуть членов общества с тяжелыми физическими или умственными отклонениями более жесткому отбору. Английский ученый был озабочен упадком британской нации, который, по его мнению, был связан со снижением рождаемости в либеральных классах. Однако его вера в научную евгенику сочеталась с пламенной поддержкой социализма. В дарвинистской борьбе за существование между нациями социализм казался историческим уроком.

гии. В книге *«Новый Вертер»* (*The New Werther*, 1880) Пирсон восклицает:

> «Гиганты литературы, тайны многомерного пространства, попытки Больцмана и Крукса исследовать лабораторию природы, теория вселенной Канта и последние открытия в эмбриологии, удивительные данные о развитии жизни... какие ошеломляющие глубины за пределами нашего понимания!»

Превращение этого математика из эксперта по теории эластичности в первого статистика в современном понимании можно объяснить только тем, что он был необычайным — пусть это и анахронизм — ученым Возрождения, одержимым числовой и духовной истиной. Неслучайно одной из целей, к которым стремился Пирсон, было объединение статистиками будущего двух пластов (научного и гуманистического) в процессе либо решения проблем, либо истории формирования дисциплины: например, написанная им в молодости хронология теории эластичности, в зрелые годы — амбициозная трехтомная биография Фрэнсиса Гальтона, которым он восхищался, а также сборник лекций о происхождении статистики в свете религиозной мысли.

К 1892 году научные интересы Пирсона резко изменились. Благодаря дружбе с Уолтером Фрэнком Рафаэлем Уэлдоном (1860–1906) — профессором зоологии Университетского колледжа, с которым он познакомился годом ранее на съезде, связанном с реформой университета, — он заинтересовался разработкой статистических методов. Это должно было придать импульс исследованиям наследственности и эволюции, так как после смерти Дарвина в этой сфере наблюдался застой (за исключением деятельности Гальтона). Интересно отметить, что после обучения в Германии он вернулся не только пламенным социалистом, но прежде всего убежденным дарвинистом, поскольку в Берлине он ассистировал на лекциях Эмиля Дюбуа-Реймона, брата математика Поля Дюбуа-Реймона.

Рафаэлю Уэлдону требовалась помощь с анализом зоометрических данных, собранных с целью прояснить протекание естественного отбора. В 1890 году, основываясь на измерениях десятиногих ракообразных, он продемонстрировал, что распределение вариабельности у этих животных примерно соответствует распределению, которое наблюдали Кетле и Гальтон у людей, — нормальному закону. Это стало первым в биологии применением статистических техник, разработанных Гальтоном в сфере антропологии. Впервые был вычислен коэффициент корреляции между размерами двух органов. Гальтон, выступивший рецензентом доклада, незамедлительно написал Уэлдону, который благодаря исследованиям с крабами был убежден, что эволюция — вопрос исключительно статистический. Два механизма в теории эволюции — изменчивость и естественный отбор в ходе борьбы за существование — имеют с этой точки зрения несомненную связь. Возникновение изменчивости зависит от случая, от исчисления вероятностей, естественный отбор — от изучения популяций, так как отдельные особи эволюционируют в совокупности. И вот тут Уэлдону срочно потребовалась помощь коллеги-математика.

В возрасте 35 лет Пирсон начал изучать статистические методы, описанные во многих европейских руководствах, посвященных демографии. Он перечитал книги Гальтона (с которым познакомился в 1894 году через Уэлдона), так как первое прочтение *«Естественного наследования»* (1889) не дало результата, судя по его заявлению в лондонском клубе *Men and Womens'*, членом которого состоял:

> «Считаю нужным сказать, что есть определенная опасность в применении методов точных наук к вопросам описательной науки, будь то проблемы наследования или политической экономии».

Кроме того, в сохранившемся экземпляре книги Гальтона Пирсон оставил на полях пометку, выражающую его раздражение аргументами автора: с его точки зрения, речь шла лишь об аналогиях, не имеющих какой-либо научной ценности. Несмотря на эти свидетельства, все еще распространено мнение, что статистический порыв Пирсона уходит корнями в работы Гальтона, чьим учеником Пирсона ошибочно считают. Возможно, эта перемена мнения произошла из-за Уэлдона, которому удалось проиллюстрировать на конкретных примерах, что статистические техники, описанные Гальтоном, могут успешно применяться к биологическому материалу.

Спустя годы, переосмыслив свое сближение с работами Гальтона, Пирсон продолжал удивляться открытию этого ученого: существует категория, более широкая, чем причинность — а именно корреляция, — и причинность является лишь ее границей. Благодаря этому новому пониманию стала возможной связь между социологией, психологией, антропологией, биологией и математикой.

Физик полагает, что некоторое значение x обусловливает определенное значение y, статистик считает, что взаимосвязь между x и y более размытая, основанная на вероятности. Гальтон избавил Пирсона от предрассудка, что математика может применяться к природным явлениям только в рамках причинности. Конечно, его вновь возросший интерес к работам Гальтона был отчасти связан с их общим интересом к евгенике.

Желание проводить совместные исследования привело Уэлдона и Пирсона в 1893 году к учреждению Школы биометрики при прямом участии Гальтона. Термин *биометрия*, введенный Пирсоном, означает «наука об измерениях живого».

Школа заложила основы математической статистики между 1895 и 1915 годами, хотя дисциплина достигла своего развития только в период между 1915 и 1935 годами, уже под руководством Фишера. В обоих случаях существовала необходимость разрешения биологических вопросов (связанных сперва с наследственностью и эволюцией, а впоследствии — с генетикой и сельскохозяйственными экспериментами) для ускорения процесса создания новых статистических инструментов.

МАТЕМАТИЧЕСКИЙ ВКЛАД В ТЕОРИЮ ЭВОЛЮЦИИ

Пирсон начал с огромной скоростью вводить новые понятия и методы, и вскоре их уже использовали в других сферах, в которых применялась статистика. Многие свои амбициозные идеи, еще до публикации, Пирсон испытывал на аудитории в рамках цикла вечерних собраний, которые проводились в Грэшем-колледже с 1891 по 1894 год. Первые восемь встреч были посвящены фундаментальным аспектам философии науки, позже объединенным в книге *«Грамматика науки»* (*The Grammar of Science*, 1892). В переиздании 1900 года он писал:

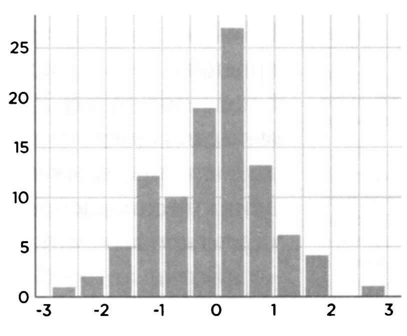
> «Мы почти ничего не можем сказать об отдельном атоме, точно так же мало мы можем судить об отдельном живом организме. Мы можем использовать только статистические данные о его среднем поведении. Но у нас есть несколько законов изменчивости и наследственности, почти столько же определенных и общих, как законы физики».

Остальные 30 собраний отводились «геометрии статистики» и «геометрии случая», если использовать авторские термины. Английский математик выбрал эти слова, потому что многие его слушатели работали в Сити, финансовом центре Лондона, и он справедливо полагал, что они одобрят графическое представление статистики. На одном из таких собраний он продемонстрировал *гистограммы* (см. рисунок 1) — графики, которые могут быть полезны для исторической (как это следует из названия) науки, иллюстрируя изменения численности населения или уровня государственных доходов за определенный промежуток времени. Эти лекции открыли новую эру в теории и практике статистики. И Пирсон заявлял слушателям, что у этой науки большое будущее, ведь она добавила блеска и математике, и даже биологии.

Одним из первых введенных им понятий было «стандартное отклонение» (или «типичное отклонение»), которым в 1893 году он заменил термин «возможная ошибка»,

На гистограммах, в отличие от столбчатых диаграмм (используются для представления негруппированных данных), классы не отдельные, а протяженные.

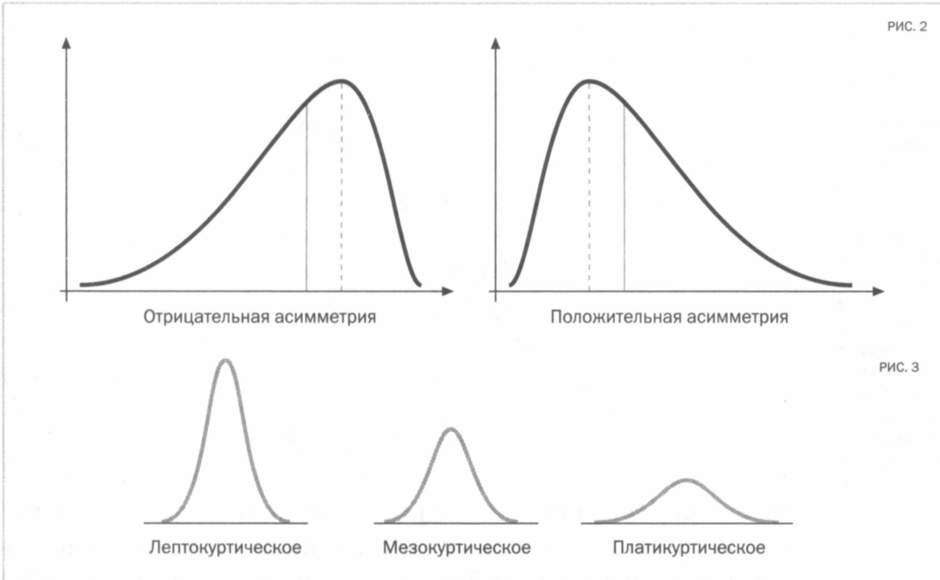
РИС. 1



сформулированный астрономом Фридрихом Бесселем примерно в 1815 году для измерения биологической изменчивости. Пока большая часть математиков и астрономов XIX века занимались изучением концентрации и положения данных, Пирсон обратился к измерению их дисперсии и вариабельности. Если Кетле переосмыслил использование среднего, а Гальтон — медианы (показателя, предложенного Антуаном Огюстеном Курно), квартилей и перцентилей, то Пирсон назвал квадратный корень из среднего квадратов отклонений каждого элемента данных от среднего *стандартным отклонением* (в то время оно называлось «среднеквадратичной ошибкой») и обозначил символом σ, чтобы подчеркнуть, что вариабельность не всегда должна считаться ошибкой.

Свойство возможной ошибки заключается в том, что она разделяет наблюдения за светилом (распределенные в соответствии с кривой Гаусса вокруг истинного значения) на два равновероятных класса: на достаточно большом количестве данных половина наблюдений окажется в окрестности среднего не дальше возможной ошибки, и другая половина будет более удаленной, в большую или меньшую сторону. Возможную ошибку сегодня иногда называют *абсолютным отклонением от медианы*. Стандартное отклонение в серии наблюдений можно вычислить более простым способом, и у него есть важное свойство: стандартное отклонение распределения теоретической ошибки, модели вероятности, есть не что иное, как непрерывная версия только что описанной дискретной формулы. В нормальном распределении возможная ошибка составляет 0,6745 от стандартного отклонения. То есть если в окрестности среднего в пределах возможной ошибки оказывается 50% наблюдений, то в границах одного стандартного отклонения оказывается примерно 68% наблюдений, и на расстоянии не более двух стандартных отклонений — немногим менее 95% наблюдений (если распределение не является нормальным, можно гарантировать, что там будет как минимум 75% наблюдений).

Более того, Пирсон придумал коэффициент вариации, определив его как связь между стандартным отклонением и средним в абсолютном значении, и использовал его для



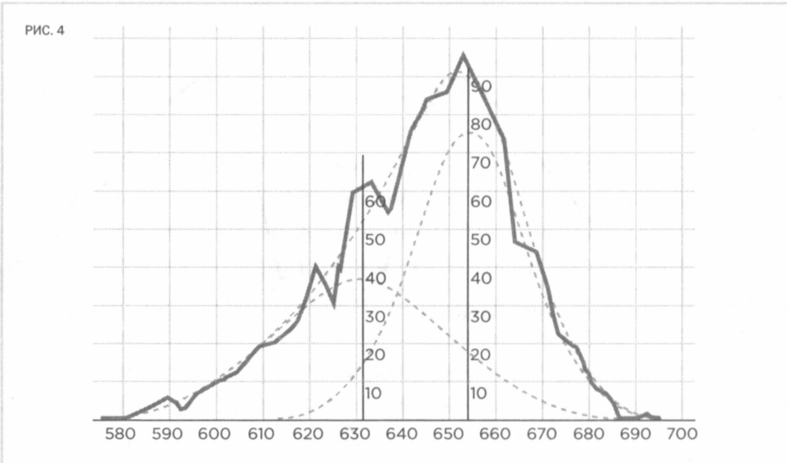
сравнения изменчивости в различных группах данных, характеризуя в каждой из них степень, в которой данные отражают среднее, то есть более или менее кучно распределены вокруг среднего. И наконец, Пирсон разработал еще два описательных показателя, коэффициент асимметрии (см. рисунок 2) и *куртозис* (см. рисунок 3), для измерения формы распределения: симметрично ли оно относительно среднего и более или менее вытянуто по сравнению с нормальным распределением. В конечном счете Пирсон вывел целую серию показателей, столь удобных для описательной статистики и поискового анализа данных.

Но это еще не все. Уэлдон попросил у Пирсона совета при анализе данных по размерам крабов (диаметр панциря, длина клешней и так далее), которые получил, проводя отпуск в бухте Неаполитанского залива. Наблюдения не подчинялись нормальному закону. Их распределение не было симметричным: вместо одного пика, как у нормального распределения, выделялись два горба (см. рисунок 4). Пирсон помог Уэлдону разделить распределение на два нормальных компонента, согласно

РИСУНОК 2: **Слева — отрицательно асимметричное распределение, справа — положительно асимметричное распределение. Непрерывная линия — среднее, пунктирная линия — мода (разница между этими показателями позволяла Пирсону установить тип асимметрии).**

РИСУНОК 3: **Куртозис (от греч. «выгнутый», «дугообразный») показывает степень сплющивания распределения по отношению к нормальному, называемому мезокуртическим.**

График III из статьи **«О некоторых корреляциях у Carnicus maenas» (On certain correlated variations in Carnicus maenas)**, опубликованной Уэлдоном в 1893 году. Здесь представлено асимметричное распределение (разделенное на две нормальные наложенные друг на друга кривые) размеров неаполетанских крабов.



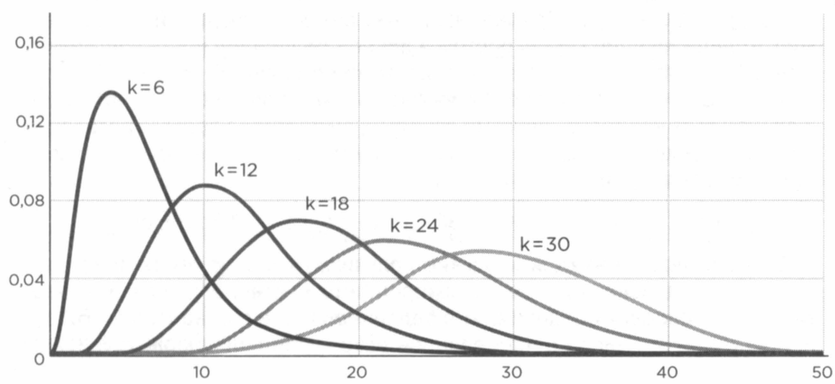
логике Гальтона (то есть все распределения были нормальными или суммой нормальных), и тот пришел к выводу, что перед ним два разных вида крабов, которых он по незнанию относил к одному, или же один вид, но находящийся в процессе разделения на два разных вида. Но английский математик хотел найти способ интерпретировать данные без принудительной подгонки к нормальному распределению, без нарушения формы кривой частот. Он не желал отбрасывать возможность асимметрии в исходных данных.

В 1894 году в своей первой публикации о статистике Пирсон обозначил целую систему частотных кривых, которые можно было использовать в биологических исследованиях. Он хотел предоставить биометристам целый набор моделей для получения всей информации из данных без их деформации. Система частотных кривых описывала сразу серию распределений вероятности, которые могли быть применены к различным случайным явлениям. Среди них были некоторые распределения, которые уже показали свою мощь в статистических методах: например, бета- и гамма-распределения, распределение χ^2 (см. рисунок 5). Это семейство асимметричных распределений представляло собой альтернативу нормальному распределению, доминировавшему со времен Кетле, и могло быть адаптировано к практическим ситуациям. Чтобы решить, какая кривая лучше соответствует данным в каждом случае, Пирсон разработал метод моментов, позволявших оценивать параметры, определяющие каждую кривую — так называемые *моменты* (термин он позаимствовал в механике) — на основе имеющихся данных. Это самый первый из известных способов оценки параметров, и заключается он в приравнивании подходящих моментов из теоретического распределения к соответствующим моментам, вычисленным по наблюдаемым данным, и затем нахождении неизвестных параметров. В частности, оценка начиналась с вычисления четырех моментов, связанных соответственно со средним, стандартным отклонением, асимметрией и куртозисом (хотя до 1905 года этот термин не употреблялся), которые определяли форму кривой частот.

Пирсон пытался отнять у нормального распределения его ведущую роль в биологии, предложив ряд альтернативных кривых для описания асимметричных и даже бимодальных распределений. В течение многих лет любое эмпирическое распределение описывалось кривой Гаусса, так как это был единственный доступный вариант. Гальтон наивно полагал, что любые данные должны соответствовать нормальному распределению. Пирсон, напротив, подчеркивал, что эмпирические распределения частот могут иметь любую форму. Нормальная кривая не является канонической, и эпоха нормального зако-

Распределение χ^2. Вернее, семейство распределений, каждое из которых зависит от параметра, называемого «числом степеней свободы» (по мере увеличения этого параметра кривая теряет асимметрию и приближается к нормальному распределению).

РИС. 5



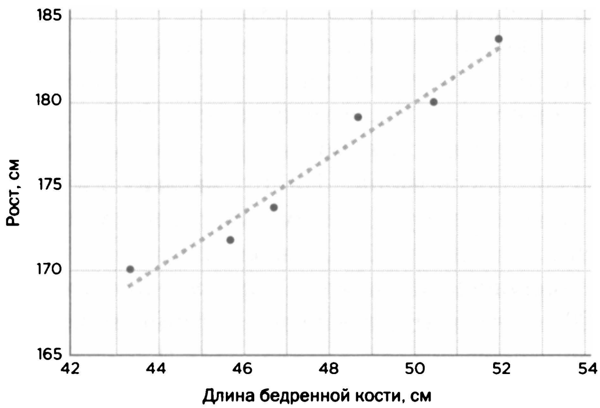
на подошла к своему завершению вместе со столетием, когда Пирсон добился свержения этого монолитного представления. Кроме биномиального распределения Бернулли и возведенного на трон нормального распределения (связанных между собой через центральную предельную теорему), до появления системы кривых Пирсона существовало не так много альтернативных моделей вероятности. Среди них — равномерное распределение, экспоненциальное распределение, дополненное распределение Пуассона, известное также как закон «редких событий», получившее известность из-за представления доли прусских офицеров, пострадавших в 1890-х от копыт своих лошадей.

РОСТ НЕАНДЕРТАЛЬЦА

Карл Пирсон применил коэффициент корреляции и прямые регрессии к данным о росте родителей и детей, собранным Гальтоном. Рост детей связан с ростом родителей: дети высоких родителей обычно высокие. И это не жесткая математическая связь, а тенденция, ее можно измерить с помощью «коэффициента корреляции Пирсона» (который определяется как частное от деления произведений моментов, или ковариации, и стандартных отклонений двух анализируемых переменных). Коэффициент принимает значения в диапазоне от −1 до +1. Если значение коэффициента близко к 1, то когда переменная «рост родителей» увеличивается, переменная «рост детей» увеличивается тоже. В 1898 году Пирсон предположил, что подобная связь существует между ростом человека и длиной его бедренных костей. Проанализировав сотни наблюдений, он обнаружил, что корреляция между ростом и длиной бедренной кости составляет 0,8048. Это свидетельствовало о достаточно сильной связи, и он сделал вывод, что существует взаимосвязь между длиной бедренной кости и длиной тела человека. Другими словами, он построил прямую регрессии роста по отношению к длине бедренной кости, которая для человеческого вида имеет следующие параметры:

$$\text{Рост, см} = 81{,}31\ \text{см} + 1{,}88 \cdot \text{длина бедренной кости, см.}$$

Пирсон также указал, как можно получить данные о росте доисторических людей, основываясь на измерении их костей. В случае неандертальского человека средняя длина бедренной кости составляла 44,52 см, и, подставив это число в приведенное выше уравнение, можно увидеть, В 1896 году Пирсон математически строго вывел коэффициент корреляции и линейной регрессии, которые Гальтон использовал эмпирически. Гальтон не применял сложных вычислений и редко анализировал наборы более чем из 100 наблюдений (для простоты вычисления), однако Пирсон сделал использование абстрактной математики в статистике обязательным и анализировал большие наборы данных (более 1000 объектов). Он предложил форму коэффициента корреляции с использованием произведения моментов (Фишер и его круг называли его *ковариацией*), а также явные уравнения прямых регрессии, хотя и завершил формирование теории нелинейной регрессии только к 1905 году. Его ассистент, инже-



На диаграмме разброса представлены длина бедренной кости и рост (в см) в выборке из шести индивидуумов. Между двумя переменными наблюдается сильная линейная связь (пунктиром обозначена прямая регрессии).

что средний рост — 165,01 см. С другой стороны, средний рост кроманьонского человека близок к 172,15 см, так как средняя длина сохранившихся бедренных костей достигает 48,32 см. И неандертальский, и кроманьонский человек были значительно ниже, чем современные люди. Сегодня эту почти не изменившуюся методику применяют в палеоантропологии для вычисления характеристик исчезнувших видов гоминидов, обнаруживаемых при раскопках.

нер и впоследствии профессор статистики Джордж Удни Юл в 1897 году разработал множественную регрессию (двух и более переменных, когда предполагается, что изучаемая переменная зависит от двух и более других переменных), связав ее с методом наименьших квадратов и синтезом Гаусса — Лапласа. Немногие знают, что Пирсон первым предупреждал об опасности выявления «ложных корреляций» (характерных для второй половины XX века): две переменные могут коррелировать между собой в отсутствие причинно-следственной связи или общей причины (например, в случае количества аистов и количества младенцев, рождающихся в Лондоне за неделю).

Наконец, в 1900 году Пирсон вывел критерий хи-квадрат (χ^2) для определения качества подгонки наблюдаемого и теоретического, или ожидаемого, распределения. Критерий использовался не только для определения соответствия данных распределению, но был также распространен на оценку гомогенности различных выборок и независимости переменных (хотя точное число степеней свободы распределения χ^2 было вычислено в 1920-е годы Фишером). Вследствие этого назначение нормального распределения перестало быть вопросом визуального сходства графиков и получило количественное статистическое обоснование. На тот момент это был один из самых прочных мостов между описательной статистикой и статистическим выводом. И действительно, в конце XX века известный американский научный журнал назвал критерий χ^2 одним из 20 научных открытий века, которые значительно изменили нашу жизнь.

Среди более прозаических достижений Пирсона и его сотрудников — публикация сборника таблиц для биометристов и статистиков, столь удобных для подгонки кривых, созданных с использованием первых вычислительных машин. Не следует забывать, что до изобретения компьютеров эти таблицы значительно упрощали труд статистиков, позволяя им быстро свериться с результатом сложных вероятностных вычислений. Этот исключительный набор данных был опубликован в виде 18 статей, написанных Пирсоном с 1894 по 1912 годы, под общим заголовком *«Математический вклад в теорию эволюции»*.

Сегодня эти тексты ясно свидетельствуют об экстраординарной работоспособности и умении сопоставлять различные дисциплины, которыми обладал Пирсон.

ИНСТИТУЦИОНАЛИЗАЦИЯ СТАТИСТИКИ

Первые статьи Пирсона увидели свет в *«Философских трудах Королевского общества»*. Однако сопротивление биологов пространным математическим выкладкам (натуралисты не соглашались с биологическими умозаключениями, сделанными на основе статистических рассуждений) в 1901 году заставило Уэлдона и Пирсона, при участии Фрэнсиса Гальтона, основать журнал *«Биометрика»* (*Biometrika*). Идея создать собственный журнал принадлежит Уэлдону, но название предложил Пирсон. Для обоих ученых проблема эволюции была проблемой статистики. Дарвин представил свою биологическую теорию, не прибегая к математике, но каждое ее понятие, начиная с изменчивости и отбора и заканчивая наследственностью и регрессией, можно было определить математически и проанализировать статистически.

В первой редакционной статье журнала Уэлдон и Пирсон обозначили план действий и предсказали наступление эпохи, когда появятся математики, являющиеся одновременно компетентными биологами, и биологи, являющиеся компетентными математиками. В *«Биометрике»* десятилетиями публиковали результаты статистического анализа самых разнообразных данных, таких как размах крыльев экзотических птиц, рост албанских новобранцев, длина большеберцовой кости африканца или размер пениса пигмеев.

Карл Пирсон оставался постоянным редактором журнала *«Биометрика»* с самого первого номера, вышедшего в октябре 1901 года, вплоть до своей смерти 35 лет спустя. После скоропостижной кончины Рафаэля Уэлдона, в 1906 году получившего травму, катаясь на лыжах, Пирсон отдалился от эволюционной биологии. Без неоценимого участия своего коллеги он

ТЕСТ χ^2 И ОБСТРЕЛ АНГЛИИ РАКЕТАМИ V2

Во время Второй мировой войны немцы обрушили на Лондон шквал ракет V2. Статистики, сотрудничавшие с войсками ПВО, поделили карту города на 576 клеток, каждая площадью 1/4 км^2, и подсчитали количество бомб, попавших в каждую клетку во время бомбардировки. Было замечено, что в 229 клеток не попала ни одна бомба, в 211 — только одна, и так далее. Результаты выглядели так.

Количество попаданий в клетку	0	1	2	3	4	5
Наблюдаемая частота	229	211	93	35	7	1

Статистики хотели понять, были бомбардировки случайными или же нацеленными на конкретные военные объекты (последнее означало бы, что немцы не могли контролировать полет V2). Для этого они использовали критерий χ^2 Пирсона, проверяя соответствие наблюдаемого распределения теоретически предсказанному, которое в данном случае было распределением Пуассона, или распределением «редких событий», измеряющим вероятность получения определенного числа событий — гипотетически они являются «редкими», маловероятными — на протяжении определенного периода времени. Распределение Пуассона зависит только от одного параметра (λ), который соответствует средней частоте возникновения события. Вычисленное на основании эмпирических данных значение λ:

$$\lambda = \frac{0 \cdot 229 + 1 \cdot 211 + 2 \cdot 93 + \ldots + 5 \cdot 1}{576} = 0{,}929$$

(в среднем ожидается примерно одно попадание в каждую клетку). Значит, если бомбардировки следуют этому распределению, ожидаемые частоты будут такими (формула для расчета этих значений довольно сложна, но ее легко доказать, только мы не станем вдаваться в такие подробности):

	Число попаданий в клетку					
	0	1	2	3	4	5
Ожидаемая частота	227,5	211	98	30	7	1,5
Наблюдаемая частота	229	211	93	35	7	1
Разница	1,5	0	–5	5	0	–0,5

Статистики подсчитали значение «статистики хи-квадрат» — суммарной меры расхождения, которая вычисляется сложением квадратов разницы между наблюдаемой и ожидаемой частотой (таким образом расхождения в большую сторону не будут компенсироваться расхождениями в меньшую сторону), поделенных на ожидаемую частоту:

$$\chi^2 = \sum \frac{(\text{Разница})^2}{\text{Ожидаемая частота}} = \frac{1{,}5^2}{227{,}5} + \ldots + \frac{(-0{,}5)^2}{1{,}5} = 1{,}27.$$

Если распределение Пуассона оказалось подходящим, значение статистики являлось бы значением распределения хи-квадрат с 6 – 2 = 4 степенями свободы (в общем случае число степеней свободы на одну меньше, чем количество исходных классов, но так как мы оценивали значение λ на основании данных, следует вычесть еще одну, как впоследствии доказал Фишер). Сверившись с таблицами, статистики выяснили: вероятность того, что χ^2_4 примет значение больше или равное 1,27, равна 0,87. Другими словами, вероятность получить такую разницу высока, если предположить, что бомбардировки были случайными, без конкретной цели. Лондонцы могли вздохнуть с облегчением.

Ракета V2 на платформе.

не ощущал в себе достаточно сил для продвижения статистики в области эволюции и наследственности в одиночку. Тем не менее он удвоил усилия по организации центра, который преобразовал бы статистику в самостоятельную ветвь прикладной математики с набором независимых методов, что позволило бы статистикам по праву считаться «людьми науки».

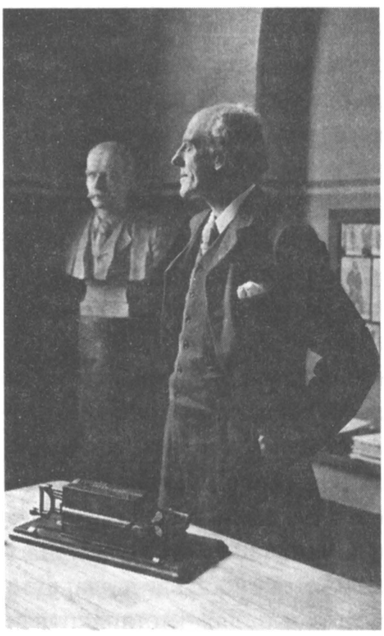
Наука будущего будет называться биометрией, и ее официальным печатным органом станет «Биометрика».

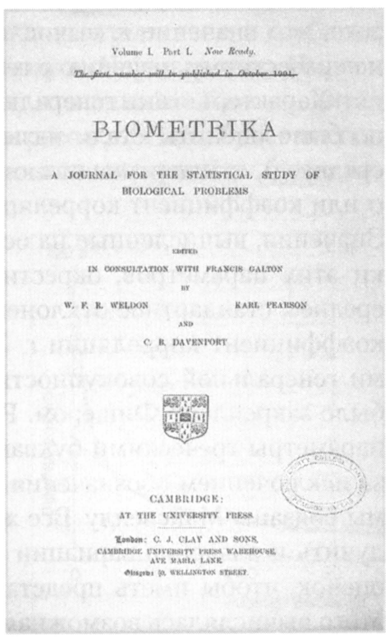
Карл Пирсон

Сэр Фрэнсис Гальтон умер в 1911 году, оставив после себя кафедру евгеники в Университетском колледже Лондона, и перешла она к его протеже Пирсону, который смог осуществить свою мечту, основав отделение прикладной статистики, совмещенное с биометрической лабораторией (ею он руководил с момента ее создания в 1903 году) и национальной евгенической лабораторией Гальтона (была создана в 1907 году как преемница *«Бюро евгенических данных»* (*Eugenics Record Office*), учрежденного Гальтоном в 1904 году). Биометрическая лаборатория занималась разработкой статистических методов в биологическом контексте, а евгеническая лаборатория применяла их при изучении «национальной деградации» (связывая показатели рождаемости с социальным статусом или алкоголизм — с его воздействием на здоровье и способности потомства). В 1925 году Пирсон учредил *«Анналы евгеники»* (*Annals of Eugenics*, сегодня известные как *«Анналы генетики человека»* — *Annals of Human Genetics*), чтобы продолжать публиковать отчеты исследований в области евгеники.

ОЖЕСТОЧЕННЫЕ СПОРЫ

В 1914 году Пирсон получил для публикации в *«Биометрике»* статью, подписанную 24-летним студентом по имени Р.А. Фи-





Volume I. Part I. *Now Ready.*

~~*The first number will be published in October 1901.*~~

BIOMETRIKA

A JOURNAL FOR THE STATISTICAL STUDY OF
BIOLOGICAL PROBLEMS

EDITED
IN CONSULTATION WITH FRANCIS GALTON
BY
W. F. R. WELDON KARL PEARSON
AND
C. B. DAVENPORT

CAMBRIDGE:
AT THE UNIVERSITY PRESS

London: C. J. CLAY AND SONS,
CAMBRIDGE UNIVERSITY PRESS WAREHOUSE,
AVE MARIA LANE.
Glasgow: 50, WELLINGTON STREET.

ВВЕРХУ:
Гальтон в возрасте 87 лет и Карл Пирсон. Фото 1909 года.

СЛЕВА ВНИЗУ:
Карл Пирсон с бюстом Рафаэля Уэлдона. Фото 1910 года.

СПРАВА ВНИЗУ:
Титульный лист *«Биометрики»* — журнала, издаваемого Уэлдоном и Пирсоном при поддержке Гальтона в сотрудничестве с Чарльзом Девенпортом (1866–1944), влиятельным американским биологом, который разделял их биометрический подход и евгенические убеждения.

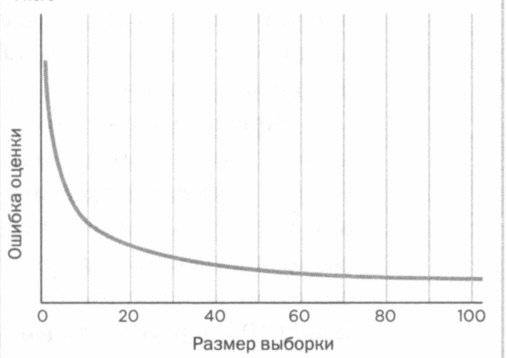
шер. На сплошь исписанных страницах Фишер приходит к выводу, который Пирсон и его команда систематически упускали из виду: истинное распределение выборочного коэффициента корреляции r, необходимое условие для определения возможной ошибки при оценке коэффициента корреляции в генеральной совокупности ρ. Вопрос выборочных распределений превращался в важную тему для дальнейшего развития статистического вывода, так как позволял количественно оценить надежность предположений, сделанных на основании репрезентативной выборки, имеющей своей целью узнать характеристики генеральной совокупности, набора объектов, который считался слишком большим, чтобы исследовать его полностью. Определять коэффициент корреляции ρ во всей генеральной совокупности на основании корреляции r, наблюдаемой в данных выборки, было ошибкой, и без указания точности такого коэффициента это оказывалось бесполезным. Изучение выборочного распределения, то есть распределения, возникающего в последовательности всех возможных случайных выборок генеральной совокупности, позволяло определить вероятность того, что значение r, вычисленное по выборке, было близко к неизвестному значению ρ генеральной совокупности.

Характеристики генеральной совокупности, которые нужно было оценить, стали называть *параметрами*: генеральное среднее μ, стандартное отклонение генеральной совокупности σ или коэффициент корреляции генеральной совокупности ρ. Значения, вычисленные на основе данных выборки для оценки этих параметров, окрестили *статистиками*: выборочное среднее, стандартное отклонение выборки S или выборочный коэффициент корреляции r. Это различие между параметрами генеральной совокупности и выборочными статистиками было закреплено Фишером. Равно как и традиция обозначать параметры греческими буквами, а статистики — латинскими, за исключением обозначения выборочного среднего, которым мы обязаны Максвеллу. Все же вместе с оценкой хотелось получить и значение вариации или дисперсии всех возможных оценок, чтобы иметь представление о точности вывода. Для этого вычислялась возможная ошибка, или даже «стандартная ошибка» оценки, которая является простым стандартным отклонением распределения статистики в выборке (то есть распределением, характеризующим вероятность того, что статистика принимает то или иное значение в зависимости от данных выборки, считающейся извлеченной случайным образом из генеральной совокупности). Это число позволяло судить о корректности вывода: чем меньше ошибка, тем лучше оценка. Кроме того, ошибка имеет тенденцию зависеть от квадратного корня из размера выборки, поэтому при увеличении размера выборки точность оценки тоже увеличивается, так как ошибка уменьшается пропорционально квадратному корню из размера выборки (см. рисунок 6).

За несколько лет до этого, в 1896 году, Пирсон бездоказательно заявил (доказательство выполнил Фишер), что наилучшей оценкой ρ — коэффициента корреляции генеральной совокупности — является r — коэффициент корреляции, вычисленный на основании данных выборки (хотя нотация Пирсона не всегда различала два этих значения, параметр совокупности и выборочную статистику). Пирсон с энтузиазмом ответил Фишеру, поздравил его с доказательством и сообщил, что статья обязательно будет принята. Через неделю Пирсон снова написал Фишеру, сообщая, что внимательно прочитал работу, увидел определенный прогресс и для него будет честью опубликовать этот текст, если финальные страницы будут немного дополнены. Фишер изучал распределение выборочного коэффициента корреляции геометрически (представляя выборку как n-размерный вектор и распределение как дифференцируемое многообразие), кроме того, прибегал к алгебраической трансформации, так что Пирсону было сложно проследить доказательство, в котором рассуждение шло не от r, а от функции, определенной

Ошибка оценки быстро уменьшается при увеличении размера выборки до некоторой точки, за которой увеличение выборки не будет давать ощутимого уменьшения ошибки.

РИС. 6



на r. Фишер принял совет, и новый вариант его статьи был опубликован в *«Биометрике»* в 1915 году.

Вплоть до 1917 года отношения двух математиков были дружескими, но весной того года Пирсон и его коллеги опубликовали совместное исследование, в котором Пирсон выступил против Фишера, посвятив несколько страниц критике предполагаемой ошибки, допущенной молодым ученым в статье 1915 года. Возможно, поводом к этому послужила записка, отправленная Фишером, в которой тот затронул исследование докторантки из Дании, работавшей в лаборатории Пирсона; кроме того, он ставил под сомнение достоинства критерия χ^2 и метода моментов при построении оценок. В упомянутой статье 1915 года Фишер привел полное доказательство того, что Пирсон установил за несколько лет до этого: наиболее вероятным значением коэффициента корреляции в генеральной совокупности ρ является коэффициент корреляции r, вычисляемый по выборке (при увеличении объема выборки, так как обычно r несколько больше, чем ρ). Пирсон утверждал, что Фишер использовал в доказательстве методы обратной вероятности, то есть теорему Байеса, и упрекнул его, указав на произвол в подходе, так как следовало использовать распределение, равномерное *a priori*, исходя из гипотезы полного незнания. Тем не менее Фишер выбрал не такой подход. Как будет разъяснено в главе 5, он не просто разделял эту радикальную оппозицию по отношению к байесовскому выводу, но и использовал другой, новый метод, который мы опишем в следующей главе, — «метод наибольшего правдоподобия», не имеющий с этим почти ничего общего, но наверняка разъясненный довольно туманно.

Фишеру было неприятно читать об этом в статье, и вполне логично, что инцидент оказался решающим, когда в 1919 году Фишер отказался работать под началом Пирсона в лаборатории Гальтона и принял предложение экспериментальной сельскохозяйственной станции в Ротамстеде. Более того, он написал ответ в виде статьи, которую Пирсон получил в 1920 году. В ней Фишер углублял изучение коэффициента корреляции небольших выборок и одновременно указывал, что в статье

1915 года не использовал теорему Байеса. И хотя он утверждал, что не собирался критиковать авторов-статистиков (в том числе, разумеется, Пирсона), статью он закончил высмеиванием приводимых ими примеров, завершив свой ответ примечанием о путанице между правилом Байеса и его новым методом построения оценок. Пирсон категорически отказался публиковать эту статью, вернул ее автору и попросил не настаивать.

Основным результатом этого неудачного спора стала неприязнь, которая продолжалась много лет, в течение которых оба не упускали возможности подвергнуть критике соперника. Так, Фишер в трех статьях, опубликованных с 1922 по 1924 год, совершенствовал критерий хи-квадрат, указав точное количество степеней свободы, но Пирсон так и не счел эту модификацию правильной. И наоборот, когда в 1945 году Фишера попросили написать статью о Пирсоне для биографического словаря, редактору пришлось отвергнуть этот текст из-за клеветнического тона. В любом случае, даже отвлекаясь от академических споров, нужно отметить резкий контраст между взглядами Пирсона и Фишера на статистику. Один использовал большие выборки, а второй, находясь под влиянием Уильяма Сили Госсета (известного как *Стьюдент*), предпочитал работать с малыми выборками, так как упомянутый статистик утверждал: чтобы попробовать суп, достаточно одной ложки.

Карл Пирсон сыграл огромную роль в определении содержания и организации современных статистических исследований благодаря своим научным работам, преподаванию, основанию лаборатории и обширному списку публикаций. К невероятной для математика плодовитости следует добавить поразительную работоспособность, которую сам Пирсон часто шутя объяснял своей привычкой не отвечать на телефонные звонки и никогда не участвовать в оргкомитетах.

На руках уже имелись все элементы головоломки. Все было готово к реорганизации статистических материалов, к чему Фишер и приступил. В результате статистические суждения вышли на новый уровень, где в теоретическом плане к ним были присоединены аксиоматические математические рассуждения, а в экспериментальном — гипотетическо-дедуктивный

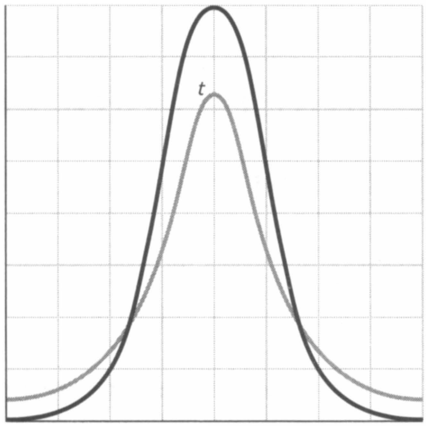
СТЬЮДЕНТ И ПИВОВАРНЯ ГИННЕСС

Уильям Сили Госсет (1876–1937) по образованию был химиком, а познакомился со статистикой после работы в биометрической лаборатории Пирсона. В 1908 году он издал свою знаменитую статью *«Возможная ошибка среднего»* (*The probable error of a mean*) под псевдонимом *Стьюдент*. Причиной такой таинственности послужило то, что дублинская пивоварня Гиннесс, где он работал, не разрешала сотрудникам публиковать результаты исследований, проведенных на производстве. Стремясь контролировать качество производимого пива, Стьюдент собирал небольшие выборки (из соображений экономии). Он обнаружил, что один из типов кривых Пирсона соответствовал распределению, удобному для проведения этих экспериментов в малом масштабе. Так, если Стьюдент хотел оценить среднюю кислотность пива, произведенного на заводе за определенный период, он вычислял среднее из уровней кислотности, измеренных в дюжине бочек, которые составляли выборку. Проблема заключалась в том, что Стьюдент не знал возможную ошибку, возникающую при оценке среднего генеральной совокупности по выборочному среднему, то есть число, необходимое для определения точности статистического вывода и допустимого предела кислотности. Чтобы определить это, Стьюденту нужно было знать распределение выборочного среднего. Было замечено, что если выборка достаточно большая — 30 или более объектов, — распределение выборочного среднего приближается к нормальному (благодаря центральной предельной теореме), но если выборка маленькая, то дело обстоит иначе.

Распределение *t* Стьюдента

Стьюдент вычислил верное распределение, известное сегодня — после того как в 1925 году его отметил Фишер — как распределение *t* Стьюдента. На самом деле это целое семейство распределений, зависящих от количества степеней свободы; в общем случае оно более плоское, чем нормальное распределение, с более длинными хвостами, что отражает большую неуверенность в выводах. Эта вероятностная модель оказалась неотъемлемой частью статистических методов наших дней ввиду своей надежности, поэтому она используется не только для выводов на основании малых выборок из нормальной генеральной совокупности (для

метод физики и таксономический подход естественных наук. Статистика стала новым способом организации мышления и взаимодействия с миром, находя себе применение в самых разных областях, в том числе в биометрических лабораторикоторой среднее и стандартное отклонение неизвестны), но также и для ненормально распределенных данных. Распределение t оказалось практически нечувствительным к гипотезе нормальности.



Как можно видеть, распределение t Стьюдента (серый цвет) обладает более широкими хвостами по отношению к нормальному (черный цвет).

Возвращение из забвения

Тем не менее Стьюдент оставался маргинальной фигурой до тех пор, пока Фишер не извлек его работу из забвения, хотя и обладал своеобразным чувством юмора (как можно заметить по мнемотехническому правилу, которое он придумал для куртозиса). Чтобы запомнить слово *платикуртический*, которое относится к кривым, более плоским, чем кривая нормального распределения, Стьюдент думал об утконосе (англ. *platypus*); а чтобы запомнить слово *лептокуртический*, которым называют более острые кривые с более очерченными хвостами, он представлял двух кенгуру, которые встречаются, прыгая на своих широких хвостах. Фишера и Стьюдента в 1912 году познакомил преподаватель Фишера в Кембридже, авторитетный астроном. Сложность математических доказательств, на которые Стьюдент ссылался в своих письмах, вдохновили Фишера на точное вычисление распределений различных выборочных статистик и последующую фиксацию своих первых успехов. Безразличие же Пирсона в этом отношении объяснялось его убежденностью в том, что выделение малых тенденций, наблюдаемых в биологических данных, требует использования больших выборок: «Только грязные торговцы пивом используют небольшие выборки!» — шутя говорил он своим ассистентам.

ях, на сельскохозяйственных предприятиях и в пивоваренной промышленности. Новая звезда возвестила о своем появлении на небосводе.

ГЛАВА 3

Математические основы статистического вывода

В 1920-е годы Фишер принял эстафету у поколения статистиков, сформировавшегося вокруг Пирсона. Его статья «*О математических основах теоретической статистики*» и последовавшие за ней книги «*Статистические методы для исследователей*» и «*Планирование экспериментов*» ознаменовали становление статистического вывода как математической дисциплины. В них Фишер изложил критерий значимости, дисперсионный анализ и рандомизацию в качестве основных принципов любой работы ученого-натуралиста с фактами.

Статистический вывод — это набор методов, которые позволяют формулировать суждения об общем (генеральная совокупность) на основании частного (выборка), предоставляя — и этим статистика отличается от гадания — меру уверенности в предсказании, вероятность ошибки.

Как мы видели в предыдущих главах, объединение рудиментарных статистических методов Лапласа и Гаусса, ограниченных астрономией, и науки о государстве, возникшей на базе демографии и нарождающейся социальной науки, произошло на рубеже XIX и XX веков на «территории» биологии, так как благодаря влиянию евгеники и биометрии теория эволюции была переформулирована как статистическая проблема.

До Фишера статистика, в которой доминировал титан Карл Пирсон, находилась в следующей ситуации. В описательной статистике, хотя и не существовало явного различия между выборкой и генеральной совокупностью, были известны простые графические представления (столбчатая диаграмма, гистограмма, диаграмма рассеяния) и вычислялись основные показатели центральной тенденции (среднее, медиана, мода), дисперсии (стандартное отклонение, хотя оно было и не единственной мерой), позиции (квартили и перцентили) и формы (асимметрия и куртозис). Переход от разведывательного анализа данных к математической теории вероятностей проходил через подгонку теоретических распределений — нормальной кривой или кривых Пирсона — к наблюдаемым распределениям частот методом наименьших квадратов или методом моментов. Качество подгонки оценивалось с помощью полезнейшего изобретения, вроде критерия хи-квадрат. Наконец, статистический вывод мог похвастаться только двумя быстрыми методами: предсказаниями, основанными на анализе регрессии и корреляции, и в особенности обратными вероятностными методами, имевшими большое влияние вплоть до эпохи Лапласа и базирующимися на теореме Байеса (байесовский, или субъективный вывод).

Фишер заполнил пробел в этом важнейшем секторе, заложив основы методов оценки и вывода, которые сегодня считаются классическими (вывод на основании частот, или объективный вывод). Если Пирсон учил, как извлекать интересующую информацию из путаницы данных, Фишер показал, как познать целое (генеральную совокупность), наблюдая часть (выборку). Он был конструктором, который окончательно связал описательную статистику и статистический вывод, соединив последний с математикой, что дало дисциплине новую жизнь.

И что еще важнее, Фишер упорядочил последовательность шагов статистического метода. Начальный разведывательный анализ имеющихся данных сменялся построением вспомогательной вероятностной модели, вслед за этим шла фаза оценки неизвестных параметров модели на основании данных выборки, и как итог — фаза подгонки модели к реальности с помощью критерия значимости и планирования эксперимента. Можно сказать, что Фишер довел до абсолюта создание методологического корпуса статистики: выбор теоретической модели на основании эмпирических данных, математическая дедукция свойств этой модели, оценка неизвестных параметров и заключительная проверка модели с помощью эксперимента. Подход, состоящий в сборе информации в ходе эксперимента и подготовке выводов на ее основе, составляет суть статистического вывода, и в отличие от вычисления вероятностей это не индуктивный, а дедуктивный процесс, сопровождающийся определенными ошибками, которые можно оценить количественно.

ПРОБЛЕМЫ И КРИТЕРИИ ВЫВОДА

В 1919 году Фишер принял должность консультанта по статистике на экспериментальной сельскохозяйственной станции в Ротамстеде, после того как отклонил предложение Карла Пирсона работать в лаборатории Гальтона под руководством последнего, ввиду их усугубившихся разногласий. Молодой ученый переехал вместе с семьей на старую ферму неподалеку от станции на севере Лондона. Ее владельцы, производители удобрений, пригласили его навести порядок в огромном количестве данных, которые накопились за годы работы. Время показало, что это решение было правильным. Заведующий станцией сэр Эдвард Джон Рассел (1872–1965) сохранял атмосферу свободы, которая стимулировала научный обмен между биологами, химиками и статистиками. Здесь Фишер превратился в неутомимого агрария-исследователя, и по дороге между фермой и станцией рождались его самые гениальные идеи, касающиеся всех областей статистики.

В своей программной статье *«О математических основах теоретической статистики»* (*On the mathematical foundations of theoretical statistics*), представленной в *Лондонском королевском обществе* в 1921 году и опубликованной в 1922-м, Фишер предложил номенклатуру, сегодня встречающуюся в каждом учебнике по статистическому выводу. Так, здесь впервые появился и был употреблен 57 раз термин «параметр» в его современном статистическом понимании. Ошибочное утверждение, сделанное астрофизиком Артуром Эддингтоном в книге *«Движения звезд и структура Вселенной»* (*Stellar movements and the structure of the universe*, 1914), а также ряд других вопросов, сформулированных Пирсоном в переписке, пока не прервалось их общение, задали отправную точку, которая подтолкнула Фишера к подробному изучению вопроса статистических оценок.

Статья начинается с утверждения, что целью статистических методов является «редукция» данных: выражение всей значимой информации о генеральной совокупности, содержащейся в выборке, через несколько числовых показателей. Вслед за этим Фишер описывает понятие «статистической модели», позволяющее ясно различать генеральную совокупность (реальную или гипотетическую) и выборку из нее — две связанные сущности, граница между которыми до тех пор была нечеткой. Имеющиеся данные должны рассматриваться как случайная выборка из генеральной совокупности, распределение которой по отношению к изучаемому явлению характеризуется набором параметров, обозначаемых греческими буквами (например, параметр θ). Действительно, для каждого возможного значения параметров существует отдельная генеральная совокупность, такая, что основной вопрос, который задает себе каждый статистик, выглядит следующим образом: к которой из бесконечного количества возможных генеральных совокупностей принадлежит выборка, находящаяся передо мной?

Кроме того, Фишер указал на три класса математических проблем, с которыми сталкивается статистический вывод. В первую очередь это проблемы «спецификации», состоящие в определении модели генеральной совокупности, то есть семейства распределений, зависящих от одного или нескольких параметров θ, из которого (предположительно) делаются выборки. Во вторую очередь это проблемы «оценки» — основного направления статистического вывода (мы еще рассмотрим их более подробно). И на третьем и последнем месте — проблемы «распределений», то есть точное выведение распределения выборочной статистики на основе распределения в генеральной совокупности, считающегося известным. Выборочные распределения задают вероятность, с которой какая-либо статистика принимает значение между двумя заданными границами (или, что то же самое, относительную частоту, с которой она находится в этих пределах при бесконечном повторении процесса выборки). Решение этого класса проблем является в каком-то смысле необходимым условием для статистического вывода, так как позволяет найти стандартную ошибку оценки, а также сравнить между собой качества различных оценок. Это позволяет вычислить точность оценки и измерить неопределенность в предсказании значения неизвестного параметра или параметров.

Сосредоточившись на проблемах теории оценки, Фишер утверждал, что речь шла о выборе наиболее подходящего значения параметра θ на основании данных выборки, или, точнее, на базе статистик — обозначаемых латинскими буквами (например, T), — которые вычисляются по наблюдаемым данным. Почему для оценки среднего в генеральной совокупности μ используется выборочное среднее? Или почему используют стандартное отклонение выборки S для оценки стандартного отклонения в генеральной совокупности σ? Теория статистической оценки, разработанная Фишером, отвечает на эти вопросы, определяя, каким критериям должна удовлетворять хорошая оценка.

Первый и самый очевидный критерий, который предъявляется к оценке параметра θ посредством статистики T, — состоятельность: то есть T должна приближаться к истинному значению θ по мере увеличения размера выборки. Следовательно, если выборка достаточно большая, значение T с большей вероятностью совпадает с θ. Состоятельными считаются оценки, которые при увеличении размера выборки стремятся к истинному значению параметра.

> Следует допустить, что любой вывод о целом на основании частного будет сопровождаться определенной степенью неуверенности, но это не значит, что данный вывод не может быть абсолютно строгим.
>
> **Р. Э. Фишер,** ***«Планирование экспериментов»*** **(1935)**

Вторым желательным качеством явилось не только стремление T к истинному значению параметра θ, но и «эффективность» такого стремления, то есть минимизация ошибок. Стандартная ошибка оценки должна быть как можно более малой (далее мы увидим, что для создания эффективных оценок Фишер предложил метод наибольшего правдоподобия).

СМЕЩЕНИЕ И ЭФФЕКТИВНОСТЬ

Сегодня три критерия, приведенные Фишером, претерпели небольшие изменения, хотя их смысл теперь включен в другие критерии.

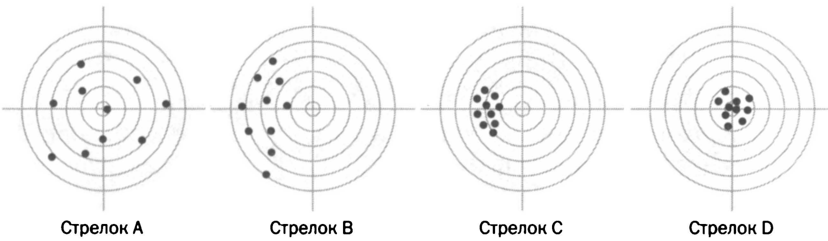
Смещение
Оценка T считается правильной и несмещенной для параметра θ, если для любого размера выборки среднее значение ее распределения равно θ, то есть если ожидаемое значение статистики T является истинным значением θ. Иначе оценка считается смещенной.

Эффективность
Эффективность, или *точность*, оценки определяется как величина, обратная дисперсии ее выборочного распределения, то есть обратная квадрату ее стандартного отклонения, стандартной ошибки (понятие *дисперсии*, квадрата стандартного отклонения, было введено Фишером в 1918 году как более простое в вычислительном отношении). Таким образом, эффективность, или точность, оценки оказывается связана с ее дисперсией (эти величины обратно пропорциональны друг другу): чем больше дисперсия оценки, тем менее точна эта оценка, и наоборот. Это понятие имеет особенное значение для сравнения несмещенных оценок, так как среди них предпочтение должно отдаваться более эффективным, то есть обладающим наименьшей дисперсией.

Средняя квадратичная ошибка
Несмотря на это, иногда возникает проблема выбора из двух оценок с противоположными свойствами: одна несмещенная, а другая смещенная, но с меньшей дисперсией. В этих случаях разумно выбрать оценку с меньшей средней ошибкой предсказания параметра (формально: с меньшей квадратичной ошибкой для заданного размера выборки, то есть суммой дисперсии оценки и квадрата ее смещения). Примером может служить оценка дисперсии σ^2 в генеральной совокупности. В соответствии с принципами, оптимальным выбором является использование не выборочной дисперсии S^2 (которая определяется как среднее ква-

И наконец, третье, еще более ограничивающее условие, — условие достаточности, которое предполагает, что статистика T не должна терять никакую информацию, содержащуюся в выборке, она должна объединять всю информацию, нужную для оценки соответствующего параметра. Когда статистика T является достаточной для θ, никакая другая оценка не даст больше информации о параметре на основании данных выборки.

дратов разниц данных по отношению к среднему), а «полудисперсии», или скорректированной выборочной дисперсии $\hat{S}^2$, при ее вычислении вместо деления на n (размер выборки) делят на $n - 1$. Причина в следующем: при использовании выборки вычисляется вариация по отношению к среднему выборки (а не к среднему генеральной совокупности, что нас на самом деле интересует), и это ведет к недооценке вариации в генеральной совокупности. При делении на $n - 1$ получается несколько большее значение, которое является лучшей оценкой дисперсии в генеральной совокупности, так как получаемый показатель является несмещенной оценкой. Тем не менее с точки зрения средней квадратичной ошибки правильнее использовать выборочную дисперсию S^2, а не «полудисперсию» $\hat{S}^2$. Оценка без коррекции является предпочтительной. И наконец, когда выборки большие и сложно получить корректные оценки с высокой эффективностью, необходимым минимумом, по Фишеру, является состоятельная оценка, то есть та, которая при увеличении размера выборки стремится к истинному значению параметра.



Если мы сравним статистические оценки с выстрелами нескольких стрелков, то сможем лучше понять, какими качествами должна обладать хорошая оценка. Выстрелы стрелка А не отклоняются в каком-то одном направлении, но видно, что они очень разбросаны (это соответствует несмещенной, но неэффективной оценке). Выстрелы стрелка В смещены влево и разбросаны (оценка смещенная и неэффективная). Выстрелы стрелка С кучные, но отклоняются (смещенная и эффективная оценка). И выстрелы стрелка D располагаются близко к центру (оценка несмещенная и эффективная), что является наилучшим вариантом.

Кроме того, можно доказать, что в этом случае T будет эффективной оценкой. Достаточность оказалась высшим критерием, так как подразумевала два других, более слабых (эффективность и состоятельность). Когда достаточная статистика была найдена, стало возможным утверждать, что проблема оценки полностью решена. Фишер признал, что, к сожалению, не всегда для конкретного параметра существует достаточная оценка.

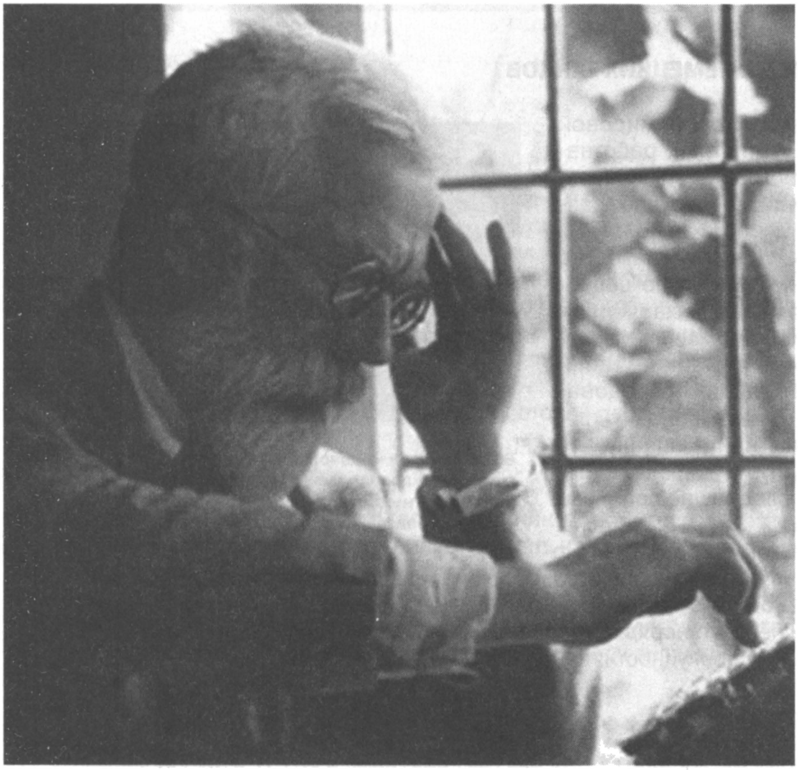
Первым методом, использовавшимся для построения оценок, был метод моментов, предложенный Карлом Пирсоном. Идея простая: взять в качестве оценки среднего генеральной совокупности выборочное среднее, в качестве оценки стандартного отклонения генеральной совокупности — выборочное стандартное отклонение, и так далее.

В общем случае моменты генеральной совокупности приравнивались к выборочным моментам, и находились неизвестные параметры. В своей статье Фишер пришел к выводу, что эффективность этого метода построения оценок недостаточна, так как многие оценки не удовлетворяют указанным критериям. Оценки, полученные методом моментов, были состоятельными, но обычно неэффективными (точными при минимальной вариации). Преимущество этих оценок — в их простоте, а недостатки — в том, что не принималось во внимание распределение генеральной совокупности, из которого производилась выборка, и не использовалась вся доступная информация.

Поэтому Фишер всегда ссылался на метод Пирсона как на «традиционный, но неэффективный». Карл Пирсон не признавал себя побежденным и даже в своей последней статье (опубликованной посмертно в 1936 году в *«Биометрике»*) всеми силами защищал достоинства своего метода, начав текст с риторического вопроса: «Теряете время, подгоняя кривые с помощью метода моментов?»

Хорошие оценки, особенно на больших выборках, можно было получить с помощью метода максимального правдоподобия, изобретенного Фишером и в своей зачаточной форме изложенного в его первой статье в 1912 году. Прямой предшественник метода максимального правдоподобия встречается у Гаусса и даже у Даниэля Бернулли, хотя байесовский вывод, вдохновивший Лапласа, затмил эту и другие работы. И все же в продвижении этого метода как универсального конструктора оценок Фишер пошел гораздо дальше, чем эти математики.

Чтобы понять функцию правдоподобия, заново введенную Фишером и одну из самых важных для вывода, следует четко различать два довольно близких понятия. Пусть θ — неизвестный параметр генеральной совокупности, а X — случайная вы-



ВВЕРХУ:
Портрет Рональда Э. Фишера за работой с арифмометром *Millionaire* (фото Энтони Баррингтона-Брауна из книги Дж. Ф. Бокса «*R.A. Fisher: The Life of a Scientist*», 1978).

ВНИЗУ:
***Rothamsted Research*, экспериментальная сельскохозяйственная станция в Ротамстеде, один из старейших в мире исследовательских центров в области сельского хозяйства, где Фишер проводил эксперименты, которые позволили ему разработать теорию статистики.**

ПРОБЛЕМА НЕМЕЦКИХ ТАНКОВ

В годы Второй мировой войны оценка масштабов производства немецких танков занимала умы статистиков стран Антигитлеровской коалиции.

Во время Второй мировой войны статистики, работавшие на стороне союзников, столкнулись с трудно разрешимой проблемой: как оценить общее количество немецких танков по серийным номерам захваченных у противника машин? Предположим, немецкие танки нумеровались последовательно от 1 до N. Значит, требовалось найти оценку для N. Допустим, у захваченных танков следующие серийные номера: 2, 3, 7, 16. На основании этой выборки нужно оценить N, то есть общий размер генеральной совокупности немецких танков. Согласно методу моментов, для нахождения оценки N приравняем момент первого порядка генеральной совокупности, то есть среднее генеральной совокупности:

$$\mu = \frac{N+1}{2},$$

к которому мы добавим 1, так как счет начинается не с 0, к выборочному моменту первого порядка, то есть выборочному среднему:

борка из этой совокупности. С одной стороны, существует вероятность получения выборки X при условии некоторого значения θ (предполагается, что оно известно), обозначаемая $p(X|\theta)$ (X — переменная, θ фиксирован) и определяющая вероятность появления всякой выборки. В проблеме оценки случай противоположный: наблюдается выборка X, но значение θ неизвестно. Тем не менее описанная функция остается полезной, так как, заменив X на наблюдаемые значения, из $P(X|\theta)$ для каждого θ у нас будет вероятность получить выборочное значение X. Изменяя θ при фиксированном X, можно получить функцию, называемую функцией правдоподобия, $L(\theta|X)$, где X фиксирован, а θ — переменная. Следует отметить, что после перемены

$$\bar{X} = \frac{2+3+7+16}{4} = 7.$$

Приравняв эти значения, найдем N равным 13. Но так как в выборке встретился танк с номером 16, логично сделать вывод, что наилучшей оценкой будет наибольший встретившийся номер, то есть 16. Немцы произвели как минимум 16 танков. Тем не менее, если рассматривать только максимальное число в выборке, мы будем недооценивать общий размер генеральной совокупности, так как выборочный максимум может быть равным или меньшим, но никогда — большим, чем N. В действительности наилучшей возможной оценкой будет эффективная оценка (несмещенная и с минимальной дисперсией), формула которой для N следующая:

$$m + \frac{m-n}{n},$$

где m — наибольший наблюдаемый серийный номер и n — размер выборки. Эту формулу можно интерпретировать как сумму выборочного максимума и «пустого среднего» выборки. К максимальному значению прибавляется среднее из промежутков между имеющимися наблюдениями, исходя из предположения, что за максимумом находится еще столько же элементов, сколько пропущено между имеющимися значениями. В нашем примере лучшей оценкой для N будет

$$16 + \frac{16-4}{4} = 19 \text{ танков всего.}$$

ролей X и θ в соответствии со сменой точки зрения, возникающей при выводе, функция правдоподобия не является распределением вероятности, следовательно (как всегда подчеркивал Фишер), не подчиняется правилам исчисления вероятностей (когда подставляются конкретные значения из выборки). Эта функция отражает наши знания о параметре генеральной совокупности. И вместо предположения известного θ и вычисления вероятности наблюдать различные выборки X допустим, что наблюдается конкретная выборка X, и оценим правдоподобие различных значений θ. Функция правдоподобия — ключевой инструмент оценки совместимости наблюдаемых выборочных значений и возможных значений параметра.

ОБМАННАЯ МОНЕТА

Пусть имеется монета, вероятность выпадения сторон которой, p, неизвестна, и при подбрасывании выпадает аверс. Монету подбрасывают четыре раза и получают следующую серию: ОРОО (орел-решка-орел-орел). Из исчисления вероятностей мы знаем, что

$$P(OPOO \mid p) = p^3(1-p).$$

Следовательно, функция правдоподобия выглядит так:

$$L(p \mid OPOO) = p^3(1-p).$$

Это выражение позволяет сделать вывод, например, что значение $p = 0{,}6$ более правдоподобно, чем 0,5, так как $L(0{,}6 \mid OPOO) = 0{,}0864$ и $L(0{,}5 \mid OPOO) = 0{,}0625$. Функция правдоподобия позволяет выделить более правдоподобные с точки зрения имеющихся данных значения параметра p. С помощью несложных вычислений можно доказать, что функция правдоподобия достигает максимума для значения 0,75. Таким образом, наша оценка, основанная на имеющейся выборке, будет $p = 0{,}75$. В сущности, это и есть основа метода оценки параметров с помощью максимального правдоподобия.

Фишер в качестве оценки θ предлагал выбирать то значение, которое соответствует максимальной вероятности появления наблюдаемых значений выборки. Другими словами, выбрать такое значение параметра, которое, оказавшись реальным, максимизирует вероятность иметь данные, наблюдаемые в реальности. Это ведет к необходимости найти максимум функции правдоподобия, чтобы выбрать значение θ, давшее бы максимальное значение функции $L(\theta|X)$. В заданных условиях оценки, полученные методом максимального правдоподобия, являются асимптотически корректными, эффективными (по мере увеличения размера выборки смещение стремится к нулю, а дисперсия — к минимуму) и достаточными (если существует другая подобная статистика для конкретного примера).

Этот метод Фишер использовал в статье 1915 года, которую Карл Пирсон критиковал особенно жестко. Метод не имел никакого отношения к теореме Байеса. Чтобы оценить коэффициент корреляции ρ в генеральной совокупности, Фишер выбрал значение, которое соответствовало максимальной вероятности получить коэффициент корреляции r, наблюдаемый в выборке, то есть максимуму функции правдоподобия. Понятие статистической модели, три типа проблем вывода (спецификация, оценка, распределение), три критерия оценок (состоятельность, эффективность, достаточность) и метод максимального правдоподобия определяют контекст исследовательской активности в теоретической или математической статистике на протяжении всего XX века, хотя туманный и слабо очерченный характер ряда доказательств, изложенных Фишером, доставил множество неприятностей математикам последующих десятилетий. Появление этого знаменитого труда Фишера открыло новую эру в статистике, введя целый ряд терминов (параметр, статистика, оценка и так далее), которые с тех пор считаются неотъемлемой частью статистической литературы.

«СТАТИСТИЧЕСКИЕ МЕТОДЫ ДЛЯ ИССЛЕДОВАТЕЛЕЙ»

В период с 1923 по 1924 год Фишер писал книгу *«Статистические методы для исследователей»* (*Statistical Methods for Research Workers*), которая увидела свет в 1925 году и на сегодняшний день выдержала 14 переизданий. Это самый влиятельный и известный труд Фишера. Он больше похож на учебник, чем на научную работу, благодаря убедительному стилю и характерному отсутствию математических доказательств. Возможно, этим и обусловлен его успех. Практические, технические, теоретические и философские проблемы представлены в книге наглядными числовыми примерами. Фишер был великим прикладным математиком, но понимал статистику как дисциплину, опирающуюся не только на дедуктивное рассуждение, свойственное математике, но и на индуктивное, к которому прибегают эксперты, отталкиваясь от имеющихся данных.

Во введении Фишер пишет, что статистика — не что иное, как математика, примененная к результатам наблюдений. Статистика занимается исследованием популяций, отдельных особей, молекул или мероприятий, обращая внимание на изменчивость и возможность редуцировать или упростить исходные данные, извлечь всю важную информацию, содержащуюся в рассматриваемой совокупности. При исследовании доступных выборок статистик делает выводы о полной совокупности, но они должны быть выражены (как безапелляционно подчеркивал Фишер) не языком вероятности (как считали сторонники теоремы Байеса, или обратных вероятностных методов), а скорее языком правдоподобия.

В разных главах Фишер разбирает вопросы, составляющие сегодня базовый курс статистического вывода. Интересно отметить, что автор начал с использования графиков. Само по себе рассматривание графиков ничего не дает, но часто подсказывает первоначальное направление анализа. После разбора фундаментальных вероятностных распределений (нормального, биномиального и распределения Пуассона) он приводит статистический рецепт, краеугольный камень всего произведения — «критерий значимости».

Все части книги, посвященные критерию значимости в его разных ипостасях (подгонки, гомогенности и независимости, для среднего, разницы средних и коэффициентов регрессии и корреляции), начинаются с набора данных, с которыми можно встретиться в ходе любого исследования. Комментируя их, Фишер подводит читателя к разным фазам статистического рассуждения, которые завершаются решением проблемы. Подход критерия основывается на знании выборочных распределений нормальных популяций, вычисленных самим Фишером и другими специалистами и описанных в математических статьях, которые не читают биологи и агрономы. В книге Фишер широко использует χ^2 Пирсона, t Стьюдента и новое распределение, которое с 1934 года будут называть F-распределением *Фишера — Снедекора*, по имени американского математика Джорджа Снедекора (1881–1974), уточнившего логарифмическую (логнормальную) аппроксимацию, использованную Фишером.

Но в чем состоял критерий значимости? В первую очередь в нулевой гипотезе H_0, которая устанавливает, например, что истинное значение неизвестного параметра таково, что $\theta = \theta_0$. Фишер дал это имя исходной гипотезе исследователя в 1935 году, так как в сельском хозяйстве исходной была гипотеза, что использование нового удобрения не приводит ни к каким изменениям, что оно неэффективно, то есть разница между средними цифрами роста при использовании и без использования удобрения близка к нулю.

Затем, после определения нулевой гипотезы, которую нужно проверить, выбирается статистика T и вычисляется ее значение по данным имеющейся выборки X, которое обозначается как $T(X)$. Так как распределение статистики T в выборке известно, определяется вероятность того, что статистика T примет значение, равное или большее наблюдаемому значению $T(X)$, при условии, что нулевая гипотеза верна (то есть предполагая, что истинное значение параметра θ равно θ_0). Математически это выглядит так: $P(T \geqslant T(X)|H_0)$. Это число было названо *p-значением*. Отсюда, если значение p достаточно мало — обычно меньше 0,05,— считается, что критерий оказался значимым и поэтому позволил опровергнуть нулевую гипотезу H_0. В противоположном случае тест оказывается незначимым для заранее заданного уровня значимости $\alpha = 0,05$, отклонить нулевую гипотезу H_0 оказывается невозможным, ее временно принимают.

> Все эксперименты проводятся для того, чтобы дать фактам возможность опровергнуть нулевую гипотезу.
>
> **Фишер, «*Планирование экспериментов*» (1935)**

Нулевая гипотеза отвергалась только в том случае, когда вероятность наблюдать выборку вроде имеющейся была очень низка. Статистическое рассуждение основывалось на следующей логической дизъюнкции: «Или произошло исключительное событие (очень маловероятное), или нулевая гипотеза неверна» (используя слова Фишера). P-значение, или вероятность значимости, — что в то время не всегда было легко вычислить — служило Фишеру своего рода мерой доказательств против нулевой гипотезы: чем оно меньше, тем сильнее имеющиеся доказательства против гипотезы. Очень малое значение указывало на то, что наблюдаемая выборка сильнее отличается от ожидаемой, чем это можно объяснить чистой случайностью, и поэтому исследователь имеет дело с неправдоподобной нулевой гипотезой, которую следует отвергнуть.

Закрепим эти понятия с помощью простой иллюстрации. Предположим, мы использовали новое удобрение на 20 растениях и наблюдали их рост в течение определенного периода времени, чтобы измерить влияние нового удобрения: увеличение (+) или уменьшение (−) скорости роста по отношению к росту без удобрения. Нашей нулевой гипотезой будет отсутствие всякого положительного эффекта от удобрения, то есть распределение ускорений (+) и замедлений (−) будет полностью случайным, как выпадение «орла» и «решки» при подбрасывании совершенно симметричной монеты. Поэтому, согласно нулевой гипотезе H_0, вероятность «+» будет равна вероятности «−», а именно $\theta = 0{,}5$. Представим, что после эксперимента мы видим 16 «+» и только 4 «−». Если мы выберем в качестве статистики T количество наблюдаемых «+», то выяснится, что вероятность получить 16 или более «+» исходя из предположения, что вероятность положительного эффекта равна 0,5, составляет, как легко вычислить (см. таблицу), только 0,006. То есть: $P(T \geqslant 16|H_0) = 0{,}006$. Так как значение p ниже порогового $\alpha = 0{,}05$, тест оказался значимым, и мы можем отвергнуть исходную нулевую гипотезу: есть эмпирические данные, противоречащие гипотезе, что удобрение не оказывает эффекта. Наоборот, все указывает на то, что оно стимулирует рост растений.

Количество «+»	Вероятность
16	0,004621
17	0,001087
18	0,000181
19	0,000019
20	0,000001
Всего	**0,006**

Таблица расчета вероятности получить 16 и более «+» согласно формуле биномиальной вероятности:

$$P(\text{количество «+» равно } k) = \binom{20}{k} \cdot 0{,}5^{20}.$$

В отличие от многих свои коллег, Фишер подчеркивал, что именно значение p, а не конкретное значение $T(X)$ статистического критерия, составляет меру доказательности рассуждений против нулевой гипотезы. Так, например, конкретное значение статистики χ^2, вычисленное для измерения расхождения между серией теоретических значений и наблюдаемых данных, не позволяло количественно оценить степень связи между двумя сериями значений (это можно было бы вычислить с помощью коэффициента корреляции), так как одно и то же значение статистики могло быть значимым для большой выборки, но незначимым для маленькой выборки. Более того, Фишер предупреждал, что уровень значимости α не должен быть фиксированным, жестким. Впрочем, его предупреждение забыли и уровень 0,05 широко приняли, вплоть до того, что значение $p = 0{,}051$ стали считать незначимым, а 0,049 — значимым. Выбор этого граничного значения — вопрос не математический, универсальный, а зависит от прагматического контекста: если проверяется новое лекарство, то уровень значимости 0,05 несет 5%-й риск того, что неэффективное лекарство будет признано эффективным (в этом случае, как и в некоторых других, уровень 0,01 или 0,001 может оказаться более подходящим).

В сущности, придуманные Фишером критерии значимости — разновидность статистических *modus tollens* («рассуждение от противного»). Традиционный *modus tollens* имеет следующую структуру.

Если A, то B.
Не B.
Следовательно, не A.

Новая статистическая версия:

Если нулевая гипотеза H_0 верна, то наблюдаемые данные статистически незначимы для $\alpha = 0{,}05$ *с высокой вероятностью* $1 - \alpha = 0{,}95$.

Наблюдаемая выборка X статистически значима на уровне $\alpha = 0{,}05$.

Следовательно, гипотеза H_0 неверна.

Тем не менее основная разница между логическим и статистическим рассуждением состоит в том, что последнее может быть ошибочным — в том смысле, что нельзя быть полностью уверенным и есть шанс допустить ошибку: вероятность 0,05 ошибочно отклонить нулевую гипотезу, которая на самом деле истинна. Из-за этой особенности критики критерия значимости считали, что ему недостает логической силы. Мы можем отвергнуть нулевую гипотезу, даже если она будет верной. Критерий значимости не может дать ответ на вопрос об истинности или ложности исходной гипотезы. Фишер здесь смешивал события невероятные с событиями невозможными. И все же статистика отличается от гадания именно способностью дать количественную оценку вероятности ошибки.

Фишер описывал критерий значимости как способ отвергнуть нулевую гипотезу, которая никаким образом не может быть доказана и установлена окончательно. Этот подход, связанный с опровержением, соответствовал направлению «фальсификации», инициированному философом Карлом Поппером (1902–1994). Для статистика и философа наука характеризуется постановкой экспериментальных доказательств, которые могут опровергнуть или фальсифицировать теории, описываемые учеными. Удивительно, что книга Фишера «*Планирование экспериментов*», углубляющая этот вопрос (о ней мы будем говорить далее), была опубликована в том же 1935 году, когда

Поппер отдал в печать свой главный труд — *«Логику научного открытия»* (хотя философ и не цитировал статистика). Методологически подход Фишера был разновидностью фальсификации в приложении к статистике: он состоял в опровержении гипотез, для которых наблюдения были относительно неправдоподобными (хотя решение об отклонении гипотезы, разумеется, могло быть пересмотрено на основании новых данных). Британского статистика от венского философа отличало то, что для нашего героя критерий значимости, будучи методологически дедуктивным (если А, то В; не А, следовательно, отвергаем H_0), был при этом индуктивным по содержанию, так как позволял обучаться на опыте, хотя всегда только на какое-то время. Нулевая гипотеза никогда не подтверждалась, но ее можно было опровергнуть. Если критерий оказывался значимым, гипотеза оказывалась неприемлемой в свете имеющихся данных; если нет, это говорило лишь о том, что гипотеза была совместима с данными. Не отклонить гипотезу не значило принять ее, если только речь не шла о серии последовательных незначимых тестов. Никакой отдельно взятый эксперимент не был для Фишера доказательством законов природы.

Как мы увидим в главе 5, подход Фишера имел некоторые недостатки. Во многих случаях имеющиеся доказательства против нулевой гипотезы были одновременно доказательствами в пользу некоторой альтернативной гипотезы, которую Фишер никогда не рассматривал в рамках критерия значимости. Таким же образом английский математик не настаивал на расчете и на важности вероятности ошибки. Другую трудность представлял технический вопрос выбора статистики для каждого теста. Выбор, разумеется, субъективный, хотя довольно стандартизованный. Фишер утверждал, что следует придерживаться принципа достаточности, выбирая достаточные статистики, то есть содержащие всю информацию, доступную в выборке. К сожалению, большинство статистик, приведенных Фишером в его книге, не удовлетворяли этому требованию (например, статистика χ^2).

В середине 1929 года Эгон Пирсон (1895–1980), сын Карла Пирсона и в то время уже видный статистик, опублико-

ANOVA

Кроме критерия значимости, книга Фишера описывала дисперсионный анализ — другую инновационную статистическую методику, известную во всем мире по английской аббревиатуре ANOVA (англ. ANalysis Of VAriance). С помощью критерия значимости можно было оценить эффективность удобрения по сравнению с отсутствием удобрений или другим удобрением. Это *критерий разницы средних* (в Приложении приведен числовой пример). Но как поступить, если мы хотим сравнить три или более удобрений? Первым вариантом (довольно неэффективным) будет попарное сравнение эффектов удобрений A, B и C: A с B, B с C, C с A. Но для уровня значимости $\alpha = 0{,}05$ проведение трех проверок увеличивает вероятность ошибки гораздо выше приемлемого порога: P (какие-либо ошибки в трех тестах) = 1 − P (ни одной ошибки в трех тестах) = $1-0{,}95^3 = 1-0{,}86 = 0{,}14$. Вероятность совершить ошибку, отвергнув нулевую гипотезу (о том, что нет различий), почти в три раза выше ожидаемой: 0,14 против 0,05. Если вместо трех удобрений у нас было бы четыре, пришлось бы сделать шесть сравнений, и это стало бы уже трудоемким делом. Чтобы преодолеть эти трудности, Фишер придумал дисперсионный анализ, который через сравнение выборочных дисперсий (отсюда его название) позволял делать некоторые выводы о соотношении средних значений в генеральной совокупности. Предположим, мы обработали шесть участков тремя разными удобрениями A, B и C (по два участка каждым удобрением). Измерим урожайность на каждом участке и вычислим среднюю эффективность каждого удобрения.

Удобрение A	Удобрение B	Удобрение C
11	6	1
9	5	3
Среднее A = 10	Среднее B = 5,5	Среднее C = 2

Затем вычисляется общее среднее:

$$\text{Среднее} = \frac{11+6+1+9+5+3}{6} = 5{,}83.$$

В этом эксперименте можно идентифицировать три вида вариабельности: общая вариация между шестью участками (у каждого своя урожайность), вариация между тремя удобрениями (A, B и C дали разный урожай) и вариация внутри каждой группы, связанная со случайностью, называемая также *внутренней*, или *остаточной вариацией* (не все участки, обработанные A, дали одинаковый результат). Сравнение этих трех источников вариации позволяет оценить соотношение эффекта между удобрениями A, В и С. Если вариация между удобрениями отличается по масштабу от вариации внутри каждой группы, разумно предположить, что эта разница связана с разницей эффектов А, В и С. И если эта разница статистически значима, нулевую гипотезу можно отвергнуть. Эта разница вариации «между» удобрениями и вариации «внутри» каждой группы и составляет предмет измерения при дисперсионном анализе через соотношение дисперсий, которое имеет F-распределение Фишера — Снедекора. Часть успеха дисперсионного анализа связана с его табличным представлением. Так как общая вариация равна сумме вариации между группами и вариации, связанной с ошибкой внутри каждой группы, общая сумма квадратов (SST) — каждого наблюдения по отношению к общему среднему — может быть выражена через сумму квадратов групп (SSA) — между средним в каждой группе и общим средним — плюс сумму квадратов ошибок (SSW) — каждого наблюдения по отношению к среднему в его группе —

$$SST = (11 - 5{,}83)^2 + (6 - 5{,}83)^2 + \ldots + (3 - 5{,}83)^2 = 68{,}83.$$
$$SSA = 2\cdot(10 - 5{,}83)^2 + 2\cdot(5{,}5 - 5{,}83)^2 + 2\cdot(2 - 5{,}83) = 64{,}33.$$
$$SSW = (11 - 10)^2 + (6 - 5{,}5)^2 + \ldots + (3 - 2)^2 = 4{,}5.$$
$$SST = 68{,}83 = 64{,}33 + 4{,}5 = SSA + SSW.$$

Далее вычисляем соответствующие средние, поделив каждое число на количество степеней свободы, то есть количество элементов минус 1. В нашем примере SST делится на 6–1 = 5 (так как есть шесть наблюдений); SSA — на 3–1 = 2 (так как есть три группы) и, наконец, SSW — на количество степеней свободы, получающееся при решении уравнения $SST = SSA + SSW$. Или: $SSW = SST - SSA = 5-2 = 3$, что совпадает с разностью между количеством наблюдений и количеством групп. Эти выкладки Фишер собрал в таблицу, где привел значение F для связи между квадратами средних в группах и квадратами ошибок.

Таблица ANOVA				
Источник вариации	**Сумма квадратов**	**Степени свободы**	**Среднее квадратов**	F
Между группами	64,33	2	32,17	21,44
Внутри групп	4,50	3	1,5	
Общая вариация	68,83	5		

Так как значение p или вероятность того, что распределение F с 2 и 3 степенями свободы примет значение, равное или большее 21,44, как приведено в таблице, равно 0,02, что меньше 0,05, можно отвергнуть гипотезу о том, что три удобрения действуют одинаково. Из данных следует, что удобрение А, несмотря на небольшой объем выборки, эффективнее.

вал в *Nature* анонимную рецензию на второе издание книги, и она привела Фишера в бешенство. Отношения между Фишером и младшим Пирсоном дружескими не стали. Возможно, Фишер полагал, что младшего Пирсона подстрекает отец. Основной претензией, сформулированной Эгоном, было то, что Фишер всегда предполагал, будто исследуемая генеральная совокупность распределена нормально, и точность тестов уменьшалась, если это предположение было неверным. Интересно отметить, что Стьюдент в своих письмах утверждал то же, но Фишер оставил эти доводы без внимания. И только Эгон Пирсон, под влиянием Стьюдента, с помощью симуляции на таблицах случайных чисел доказал, что многие тесты, основанные на знании выборочных распределений нормальных совокупностей, могут продолжать использоваться, так как вездесущее t-распределение было стабильным, даже если не выполнялось условие нормальности. Эта процедура сродни той, которую Стьюдент в свое время использовал для проверки эмпирической адекватности своего t-распределения, хотя в его распоряжении не было таблиц и ему пришлось взять колоду карт с цифрами роста и длины среднего пальца 3000 преступников. Современный вариант этого метода носит славное имя *метода Монте-Карло*.

Успех *«Статистических методов для исследователей»* обозначил конец эпохи корреляций и подгонки кривых. До Фишера статистики посвящали большую часть своих усилий вычислению коэффициентов, следуя примеру Карла Пирсона. Но все его исследования были пронизаны фундаментальным заблуждением, поскольку не делалось различия между проблемой оценки значения коэффициента, то есть степени ассоциации между двумя или более переменными, и проблемой, связанной с проверкой значимости этой ассоциации, ее возможного существования. Кроме того, после школы Пирсона Фишер снова ввел в оборот исследование выборок относительно небольшого размера, сделав статистические методы более живыми и достоверными.

«ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТОВ»

В последней главе *«Статистических методов для исследователей»* Фишер приводит примеры планирования экспериментов в сельском хозяйстве, с которым он познакомился в Ротамстеде. Позже, в статье 1926 года он окончательно усовершенствовал основные принципы любого эксперимента. В 1935 году поток идей нашел свое выражение в другом бестселлере, *«Планирование экспериментов»*, в котором Фишер собрал фундаментальные принципы планирования экспериментов, разработанные им в 1920-е годы. Этот инновационный труд выдержал восемь переизданий и был скорее сборником концепций, чем руководством по вычислению, он также оказал существенное влияние на экспериментальные исследования вообще и сельскохозяйственные в частности.

Статистика, по Фишеру, необходима для получения ответа на вопросы вида «Какое удобрение лучше?», «Какое лекарство эффективнее?» и так далее. Иногда невозможно дать ответ на основе исследования воздействия удобрения или лекарства на метаболизм растения или другого организма, но можно собрать экспериментальные данные и сравнить результаты. Тем не менее сбор данных бывает очень сложным процессом, если эксперимент, который должен их дать, не спроектирован с должным вниманием или если ученый неспособен интерпретировать результаты. Первую проблему, как объяснял Фишер, можно решить с помощью планирования, вторую — с помощью логики научного вывода. Для статистика планирование и логика — две стороны одной медали.

Биологические исследования требуют проведения контролируемых экспериментов. Пассивного наблюдения недостаточно. Выборочная техника состоит в исследовании репрезентативной выборки из генеральной совокупности и измерении изучаемых показателей. Планирование экспериментов, наоборот, заключается в фиксации одних параметров и наблюдении за другими, с измерением возникающих изменений. Когда данные получают с помощью правильно спланированного эксперимента, создается более прочное эмпирическое обоснование для выводов о взаимосвязи между переменными. Объекты, получающие «обработку» (термин, связанный с удобрениями, стал применяться в широком значении), являются экспериментальными единицами. В случае сельскохозяйственного эксперимента такими единицами являются участки или разновидности растений. Переменная, воздействие которой на эти единицы измеряется, называется фактором. Любой хорошо спланированный эксперимент должен основываться, согласно Фишеру, не только на сравнении различных видов обработки, но также на проверке значимости наблюдаемых различий с помощью статистического теста. Следовательно, каждая обработка должна встречаться как минимум два раза, а лучше — несколько раз. Если мы хотим сравнить обработку А и В, в идеале следует применить их одновременно на множестве участков. Поставить все на одну карту, на единственный участок — слишком рискованно, так как это может привести к ошибочным выводам, ведь выборка не будет репрезентативной. Может случиться, что наблюдаемая разница между обработками А и В связана просто с разной плодородностью почвы на разных участках, а не с тем, например, что А эффективнее, чем В. Принцип повторения, сформулированный Фишером, помогал ограничить ошибку эксперимента, то есть случайную вариацию, не контролируемую экспериментатором (как, например, разная плодородность участков, на которых применяются обработки А и В).

Спросить мнение у статистика после того, как сделаны выводы по эксперименту, — зачастую все равно что просить его сделать посмертную экспертизу. Может, у него получится установить причину смерти эксперимента.

Выступление Фишера на Первом индийском статистическом конгрессе (1938)

Планируя эксперимент, ученый не знает о множестве факторов, которые могут повлиять на результат. Он не в состоянии управлять всеми причинами скрытого воздействия. Так, например, при изучении нового удобрения нет смысла сравнивать рост растений, получивших удобрение, в одной теплице с ростом растений в других теплицах или в другие годы, так как они могли расти или растут в других условиях. Наоборот, сравнивать следует рост в той же самой теплице у двух групп растений: у группы А, получающей удобрение, и другой группы В, называемой *контрольной*, которая не получает удобрение. Ученый может высадить две эти группы параллельными рядами — справа А, слева В, — но это не исключает вмешательства неизвестных факторов (угол падения солнечных лучей на каждый ряд или воздушный поток в теплице), маскирующих истинный эффект удобрения. Общий инструмент для работы с этими отклонениями Фишер назвал *принципом рандомизации*. Каждая пара растений из группы А и В должна распределяться по лункам случайным образом. Если подбрасывать монету, то при выпадении орла, например, нужно сажать растение А справа, а В — слева. И наоборот, при выпадении решки высаживать растение А слева, а растение В — справа, и так далее. Такой подход позволяет добиться того, что любая значимая разница в росте между двумя группами растений может быть объяснена использованием нового удобрения.

До Фишера распределение по группам производилось систематически, что могло искажать результаты. Рандомизация ничего не стоит и защищает от воздействия возможных известных и неизвестных факторов, устраняя случайную вариацию, которая делает доказательства менее явными. Без рандомизации могло случиться, что грядка, предназначенная для группы А, окажется более плодородной, чем грядка для группы В, и таким образом неоднородность почвы замаскировала бы истинный эффект нового удобрения. И в самом деле в Ротамстеде данные собирались таким образом, что дожди и метеоусловия маскировали возможное воздействие удобрений, с которыми проводились эксперименты; влияние двух факторов смешивалось. Фишер не только указал на несоответствия, но и объяснил, как это исправить. Неожиданно, в связи с вопросом о рандомизации как методе нейтрализации внешних факторов, Фишер чуть не порвал все отношения со своим старым другом Стьюдентом (хотя в 1937 году он написал ему хвалебный некролог). Этот принцип породил заметные разногласия, так как многие ученые считали, что раз они хорошо разбираются в исследуемом материале, то нужно проводить эксперименты в систематической манере, не отдавая себе при этом отчет, что тем самым они лишаются возможности использовать критерий значимости, так как для него требуются случайные выборки.

Иногда полностью рандомизированный план эксперимента сталкивается с трудно преодолимым препятствием — гетерогенностью экспериментальных объектов (например, почвы на участках). Случайное назначение экспериментальных воздействий предполагает, что все экспериментальные объекты однородны. Если это условие не выполняется, следует разделить их на классы (внутри которых, разумеется, все воздействия будут применяться рандомизированно). Причина для группировки в классы очевидна: чем более разнородны объекты, тем более масштабна ошибка эксперимента и тем меньше возможность получить значимую разницу, объяснимую различными воздействиями. Группировка «блокирует» этот внешний фактор, обусловливающий вариацию в результатах, которая не представляет интереса, так как зависит не от реакции на удобрение, а, например, от различных свойств почвы, где это удобрение применяется. Фишер называл такую конструкцию ***планом с блочной рандомизацией***.

Предположим, мы хотим исследовать пять видов обработки (A, B, C, D и E) на 20 участках. Допустим, у нас получился рандомизированный план: B, C, A, C, E, E, E, A, D, A, B, C, B, D, D, B, A, D, C, E, в котором каждая обработка встречается четыре раза. Несмотря на это, можно установить дополнительные ограничения для этого полностью рандомизированного плана, которые устранят часть эффектов, связанных с неоднородностью почвы — «градиентом плодородности», как говорил Фишер, — и поэтому увеличат чувствительность в отношении различий между экспериментальными воздействиями. Идея в том, чтобы разделить 20 участков на 4 блока и чтобы в каждом блоке каждая обработка встречалась один раз: AECBD, CBEDA, ADEBC, CEBAD. (Следует продолжать использо-

ПРЕДШЕСТВЕННИК СУДОКУ

Когда нужно устранить эффект более чем одного внешнего фактора, который может привести к ошибочным результатам, используется план «латинский квадрат». Если мы хотим исследовать эффективность пяти удобрений (A, B, C, D и E), но при этом их действие может быть ограничено типом почвы и используемыми инсектицидами (предположим, что каждый из этих факторов содержит пять типов), блочный план потребует $5 \cdot 5 \cdot 5 = 125$ объектов. Причины экономического характера препятствуют масштабным экспериментам. Так что можно прибегнуть к особому типу плана с неполными блоками — модели «латинский квадрат». Эта схема эксперимента состоит в назначении одного из внешних факторов столбцам, а другого — строкам квадрата, таким образом, чтобы каждый вариант экспериментального воздействия появился один раз в каждой строке и в каждом столбце. Соответственно, количество строк и столбцов должно быть одинаковым и быть равным количеству видов обработки. Перед нами квадрат, который называется латинским, так как математик Леонард Эйлер использовал для его заполнения латинские буквы. Популярная головоломка судоку — не что иное, как частный случай латинского квадрата, в котором вместо букв используются числа от 1 до 9. Этот хитроумный план позволяет исследователю получить больше информации из маленькой выборки, так как устраняет внешнюю вариацию с помощью блокировки одновременно двух внешних факторов, и возможность получить значимую разницу между обработками возрастает вдвое. В нашем примере пять удобрений изучаются с помощью 25 участков, расположенных в соответствии со следующим латинским квадратом.

D	E	C	B	A
B	D	E	A	C
C	A	B	D	E
E	B	A	C	D
A	C	D	E	B

Интересно отметить, что из 56 возможных латинских квадратов размером 5×5 так называемый *квадрат Кнута — Вика*, основанный на движении шахматного коня, позволяет достичь самых точных оценок среднего. Аналогично, диагональные латинские квадраты, которые содержат одинаковые элементы по диагонали, менее точны, что Фишер считал окончательным аргументом в пользу принципа рандомизации.

вать рандомизацию внутри каждого блока, чтобы избежать сюрпризов.) Таким образом, общая вариабельность данных эксперимента уменьшается, и становится возможным оценить с большей точностью ту ее часть, которая соответствует разнице между экспериментальными воздействиями.

И в полностью рандомизированных планах, и в блочных планах для изучения данных используется одна и та же статистическая техника — *дисперсионный анализ* или одна из его модификаций (однофакторный или двухфакторный дисперсионный анализ). Этот мощный инструмент, созданный Фишером, компенсировал недостатки некоторых трудоемких и часто приводящих к ошибкам методов, широко использовавшихся в то время, и позволял в одном эксперименте сравнить действие нескольких удобрений (фосфата, сульфата, хлората и их отсутствия), разделяя различные источники вариации с целью выделить вариацию, связанную с изучаемыми факторами, то есть с действием удобрений.

Фишер считал, что систематические планы не должны использоваться. Полностью рандомизированный план помогал избежать искажений, связанных с различной плодородностью участков, но общая ошибка эксперимента рисковала оказаться слишком большой. В хорошо спланированном эксперименте на рандомизацию могут накладываться дополнительные ограничения с целью исключить вариацию, связанную с различием в плодородности почвы, и облегчить оценку той части вариации, которая связана с различными удобрениями. С помощью блочного плана можно было в несколько раз увеличить ценность эксперимента, так как только последовательное повторение исходного эксперимента дает такой же уровень точности (в случае, когда повторение было возможно, так как в сельском хозяйстве сложно добиться повторения тех же метеорологических условий).

Еще одно достижение, связанное с именем Фишера,— возможность одновременно изучать несколько факторов в одном эксперименте благодаря искусному использованию статистики, что перевело эксперименты с одним фактором в разряд неэффективных и дорогих. На практике часто возникает необхо-

ZEA MAYS И НЕПАРАМЕТРИЧЕСКИЙ СТАТИСТИЧЕСКИЙ ВЫВОД

Третья глава *«Планирования экспериментов»* посвящена анализу знаменитого эксперимента, проведенного Чарльзом Дарвином с целью доказать, что растения, полученные путем перекрестного опыления, растут лучше, чем те, что возникли в результате самоопыления. С помощью Гальтона Дарвин сравнил рост 15 пар экземпляров вида *Zea mays*, то есть кукурузы. В каждой паре одно растение было продуктом перекрестного опыления, а другое — продуктом самоопыления. Пары высаживались одновременно в один сосуд таким образом, что условия внешней среды — вода, свет, температура и так далее — были идентичными для каждого из двух экземпляров. Эти меры предосторожности, предпринятые Дарвином, привели к появлению так называемой *парной выборки*, которая по сравнению с двумя выборками по 15 растений повышает чувствительность эксперимента, то есть способность получить значимую разницу, за счет уменьшения ошибки. Применив t-критерий Стьюдента (см. Приложение), Фишер изучил разницу средних значений роста и пришел к выводу, что Дарвин прав. При этом ученый не упустил возможность упрекнуть его в том, что он не рандомизировал распределение типов растений по половинам сосуда. Также он обвинил Гальтона в манипуляции данными выборки, в их переупорядочивании.

Непараметрический вывод

Ожидая критики, которую могли выразить некоторые статистики-теоретики, далекие от экспериментов (явный намек на Эгона Пирсона), в связи с тем, что использование критерия значимости предполагает использование выборки из нормально распределенной совокупности, Фишер придумал новый метод, который приводил к аналогичным выводам. Это было одним из первых случаев применения метода, впоследствии ставшего непараметрическим выводом, большим прорывом в области статистического вывода, который получил особенно широкое развитие после Второй мировой войны и который отличался от параметрического вывода тем, что не предъявлял требований к форме распределения генеральной совокупности и его параметрам. Непараметрические тесты обладают меньшей чувствительностью, чем параметрические, но не опираются на гипотезу нормальности, что делает их более общими.

димость одновременно измерить влияние нескольких факторов, а также их возможное взаимодействие. Многофакторный эксперимент имеет преимущество благодаря возможности изучать одновременно два и более фактора вместо проведения независимых экспериментов. Более того, использование многофакторного плана позволяет измерить возможное взаимодействие факторов, что недоступно в случае, когда эксперименты проводятся отдельно. Для двух факторов — одного с тремя уровнями и другого с двумя (например, три уровня азотного удобрения, соответствующих дозам от 0 до 2, и два уровня калийного, 0 и 1) — имеем многофакторный эксперимент с $3 \times 2 = 6$ вариантами экспериментального воздействия. Отклик будет наблюдаться в шести различных комбинациях.

Фишер яростно боролся с распространенным мнением, будто следует изменять только один фактор за раз. До тех пор пока он не устранил это убеждение, большинство исследователей полагали, что идеальным является исследование одного фактора в одном эксперименте. Однако природа, так сказать, лучше реагировала на хорошо спланированную анкету, а не на отдельный вопрос.

Книга *«Планирование экспериментов»* даже чаепитие превратила в статистическую задачу. Еще со времен Ротамстеда Фишер завел обыкновение пить чай со всеми сотрудниками своего отдела. Однажды, когда он передавал чашку доктору Мюриэль Бристоль, она отказалась от нее, сказав, что ей больше нравится, когда в чашку сначала наливают молоко. С ее точки зрения, вкус чая отличался в зависимости от того, в каком порядке наливали чай и молоко. Фишер ответил, что это не имеет значения. Уильям Роуч — другой сотрудник, который впоследствии женился на Мюриэль Бристоль, — предложил провести эксперимент: поставить перед Мюриэль ряд чашек, приготовленных разными способами, и проверить, сможет ли она их различить. Доктор Бристоль правильно определила чай в каждой чашке. Фишер использовал эту историю в своей книге как затравку к серии вопросов, которые должны были иллюстрировать линию подготовки к любому эксперименту: сколько чашек нужно использовать? В каком порядке? Сколько из них требовалось определить правильно?

Если бы была предложена только одна чашка каждого типа, вероятность случайного отгадывания составляла бы 1/2. Это слишком много для того, чтобы выяснить, действительно ли женщина могла отличить одну смесь от другой. Если верить в такую способность только в том случае, когда вероятность случайного правильного ответа достаточно низка (меньше 0,05, то есть эта неприятность происходит менее чем в 5% ситуаций), было недостаточно предложить по 2 чашки каждого типа, так как можно угадать случайно 1 из 6 (есть 6 различных способов выбрать 2 объекта из 4, и только один из них правильный), то есть 17%. Недостаточно было бы предложить и по три чашки каждого типа, так как случайная догадка здесь возможна в 1 из 20 ситуаций (есть 20 способов выбрать 3 объекта из 6), что соответствует вероятности равной, но не меньшей, чем установленный лимит 0,05. И наконец, предложив по четыре чашки каждого типа, можно достичь вероятности случайного угадывания 1/70 (существуют 70 способов выбрать 4 объекта из 8), то есть 0,014, и если в этих условиях доктор Бристоль сумеет правильно определить чашки, будут основания утверждать, что она способна отличить один состав от другого. Эта граница позволяла узнать, действительно ли результат был лишь догадкой или она и вправду могла почувствовать, каким образом был приготовлен чай.

Кроме того, Фишер подчеркивал, что чашки нужно предлагать доктору Бристоль в случайном порядке, тогда эксперимент будет правильно спланирован и можно будет применить критерий значимости. Этот великолепный пример психофизического эксперимента английского статистика разорвал шаблон традиционного проведения эксперимента.

СТАНОВЛЕНИЕ СТАТИСТИЧЕСКОЙ МЫСЛИ

Фишер произвел революцию в экспериментальных исследованиях, описав широко применяемые сегодня методы максимально полезного использования экспериментов с ограниченными выборками, избегая, насколько это возможно, воздействия внешних факторов. Этот болезненный ребенок, любящий учение и одаренный глубоким геометрическим видением, вырос в ученого, внесшего самый большой вклад в статистику — если не создавшего ее полностью. В 1929 году Фишер был принят в Королевское общество. И когда в 1933 году ушел на пенсию Карл Пирсон, его должность в Университетском колледже Лондона была распределена между двумя людьми: для его сына Эгона была создана кафедра статистики, и кафедра евгеники — для Фишера, который оставил Ротамстед, чтобы перейти на «кафедру Гальтона», хотя Карл Пирсон стремился не допустить этого. Эгон Пирсон унаследовал от своего отца антипатию к Фишеру, и их неприязнь не замедлила проявиться под общей крышей, в результате чего атмосфера в обеих лабораториях (биометрической и евгенической) сделалась невыносимой.

И все же для Фишера это были приятные и успешные годы. Ему оказывали почести, и со временем он получил международное признание. Джордж Снедекор со своим экстраординарным сборником *«Статистические методы»* (*Statistical Methods*, 1940) и Гарольд Хотеллинг много сделали для его продвижения в Америке. Европейское издание *«Статистических таблиц для биологических, сельскохозяйственных и медицинских исследований»* (*Statistical tables for biological, agricultural and medical research*, 1938) в соаторстве с Фрэнком Йейтсом (1902–1994), его самым способным учеником из Ротамстеда, также помогло распространить идеи Фишера. Но наибольший вклад в раскрытие его понимания статистики, связавшего британский статистический вывод с континентальной теорией вероятностей, внес учебник шведского математика Харальда Крамера *«Математические методы статистики»* (1946). Так, в этой книге появляется неравенство Крамера — Рао, выведенное или швейцарским математиком, или индийским статистиком Кальямпуди Радхакришна Рао (докторантом Фишера), которое устанавливает нижний лимит минимальной дисперсии оценки, дополняя теорию Фишера.

Результатом всего этого стала определенная автономия статистических методов, которые развивались вокруг концепции статистической модели, предложенной Фишером (хотя в некоторых областях статистики, таких как разведывательный анализ данных, описанный Джоном Тьюки в 1977 году, эта модель не используется, поэтому разведывательный анализ считается в статистической науке редкостью). С нашей точки зрения, хотя многие историки науки говорят о статистической революции XIX века, настоящая революция произошла в 1920–30-е годы, когда статистический вывод получил вероятностную и одновременно экспериментальную направленность. Если отбросить весь биологический, социологический и прочий материал, то статистика, как отмечал Фишер, станет вторичной дисциплиной. Именно приложения являются материалом, который делает эту науку чем-то большим, чем прикладная математика.

Эта сторона статистики — способность проливать свет на самые таинственные области, — ускорила ее оформление как науки. Одним из символов этого становления выступает статистическая лаборатория в Университете Айовы, основанная в 1933 году Снедекором (Фишер посещал ее несколько раз) и приведшая к настоящему буму во время и после Второй мировой войны, когда статистические лаборатории объединили усилия с университетами и промышленностью. Статистический анализ, который раньше казался эксцентричной затеей (вроде анализа эффективности молитвы или длины веревки виселицы, выполненного Гальтоном), стал ежедневной реальностью в эконометрике, метеорологии, эпидемиологии (биостатистике), промышленной инженерии (контроль качества) и других областях. Рост количества областей исследования, кафедр, книг и специальных журналов стимулировало также распространение компьютеров, облегчающих использование статистических методов (например, для генерации случайных чисел без участия таблиц).

Критерий значимости и принципы проведения эксперимента, выведенные Фишером, нашли множество применений — от изучения удобрений до исследования вакцин. Не говоря уже о том, что обнаружение знаменитого бозона Хиггса в июле 2012 года не обошлось без вычисления p-значения: физики сообщили, что вероятность получения наблюдаемого эффекта в ускорителе частиц исходя из предположения, что это фоновый шум (нулевая гипотеза), составляет менее 0,0000003, и это значение интерпретируется как серьезное доказательство существования данной частицы (потому что этот сигнал объяснить нельзя иначе). *P*-значение стало еще меньше после экспериментов в марте 2013-го, подтвердив мудрые слова Фишера в *«Планировании экспериментов»*:

> «Феномен можно считать экспериментально доказуемым, когда известен способ проведения эксперимента, редко дающий статистически незначимый результат».

Сельскохозяйственными экспериментами Фишер подвел итог основной части статистической теории, и чтобы распространить применение этих методов, отделив их от биометрии и других технических контекстов, закрепил возможность их гибкого использования, и в результате статистика нашла себе место на всех уровнях. В первых рядах этой революционной армии статистиков всегда будет стоять фигура Рональда Эйлмера Фишера, который заложил первый камень моста, соединившего эту математическую дисциплину с экспериментальной практикой.

ГЛАВА 4

Согласование идей Дарвина и Менделя

Еще со студенческой поры Фишер поставил перед собой задачу согласовать позиции Дарвина и Менделя, то есть естественный отбор и законы, управляющие наследственностью. Без сведений, содержащихся в «*Генетической теории естественного отбора*» (1930), дарвинизм оставался бы в забвении, и синтетической теории эволюции потребовались бы многие годы для становления.

Работая на опытной станции в Ротамстеде, Фишер не только развил статистику как математическую и экспериментальную науку, но и разработал целую серию биологических экспериментов с целью совместить эволюционную теорию Дарвина и теорию наследственности Менделя. Хотя станция официально не участвовала в исследованиях, Фишеру разрешили развести мышей, улиток и кур и даже выделили ему для этого участки (за колонией мышей ухаживали жена и дети).

Однако его увлечение этим вопросом началось гораздо раньше. В период с 1912 (когда была опубликована его первая статья) по 1919 годы (когда он обосновался в Ротамстеде) Фишер написал почти 100 текстов, причем более 90 из них были связаны с биологическими вопросами, а меньшая часть — со статистикой или математикой. Особое значение имела статья 1918 года о генетике: *«Корреляция между родственниками в свете гипотезы менделевского наследования»* (*The Correlation between Relatives on the Supposition of Mendelian Inheritance*).

Читая лекции подросткам, британский ученый начал обдумывать вопрос, поставленный Карлом Пирсоном: соответствует ли вариация в человеческих популяциях менделевской модели наследования? В Кембридже, где большинство были последователями Менделя, Фишер убедился, что законы Менделя объясняют наследственность, и старался разрешить споры между биометристами и менделианцами, доказав, что измерения первых соответствуют принципам вторых. Даже если каждая характеристика или наследственный фактор (с 1922 года вместо термина «фактор» Фишер использовал *ген*) по отдельности подчиняются дискретным законам Менделя, совокупность наследственных факторов, встречающаяся у индивидуумов и в популяции, подчиняется непрерывному закону естественного отбора Дарвина, точно так же как сумма ошибок астрономических наблюдений распределяется нормально, даже если отдельные ошибки не имеют нормального распределения.

Героями юности Фишера были Дарвин и Людвиг Больцман, вместе с Максвеллом создавший статистическую механику. По аналогии с бесконечной совокупностью молекул газа, которые изучала статистическая механика, Фишер предполагал — как в абстрактной области статистического вывода, так и в более практическом смысле эволюционной биологии — гипотетическую бесконечную популяцию, из которой извлекались наблюдаемые выборки. Последующая статья 1922 года о генетическом доминировании описывала эту революционную аналогию:

> «Эволюцию для естественного отбора можно сравнить с аналитическим подходом в теории газов: можно выдвигать самые разнообразные гипотезы о природе отдельных молекул и, тем не менее, определять общие законы поведения газов».

Модель, созданная Фишером для менделевских популяций, в сущности была адаптацией газовой модели из статистической механики. Непрерывная вариация, наблюдаемая в общей популяции, может быть объяснена произведением действия множества дискретных наследственных факторов.

В черновике 1916 года Фишер впервые упоминает статистический термин «дисперсия», которому дает определение на первой же странице. И вкратце упоминает дисперсионный анализ как способ выделить часть дисперсии, связанную с конкретной наследственной причиной. Основной текст черновика составлял тезис, что теория Менделя не опровергается биометрическими данными. В письме Карлу Пирсону, датированном 1916 годом, он писал:

> «Недавно закончил статью о менделизме и биометрии, которая может быть тебе интересна. Я столкнулся с фактом, что анализ человеческих данных не противоречит менделизму. Но тема эта запутанная».

Фишер отправил статью в Лондонское королевское общество, но рецензенты высказали сомнения по поводу ее содержания. Одним из них был не кто иной, как Карл Пирсон, который хотя и не высказывался открыто против результатов Фишера, но считал его проект малоубедительным и, возможно, не до конца понимал использовавшуюся математику. Другим рецензентом был биолог Р. К. Паннет, от которого впоследствии по неожиданному стечению обстоятельств Фишер унаследовал должность в Кембридже в 1943 году. Годы спустя Фишер сказал, что статью рецензировали статистик, не разбирающийся в биологии, и биолог, не знающий статистики. В защиту рецензентов следует отметить, что статьи Фишера были не всегда удобочитаемыми, и, как неоднократно замечал Стьюдент, фишеровское *evidently*[1] обычно означало несколько часов напряженного труда для остальных смертных.

В конце концов Фишер отозвал статью и в 1918 году отправил ее в Королевское общество Эдинбурга, где она была опубликована — не без борьбы и благодаря финансовой помощи Леонарда Дарвина (1850–1943), сына Чарльза Дарвина и его друга, который со времен Кембриджа покровительствовал ему и поддерживал во время экономических трудностей. Первый шаг на пути к унификации был сделан.

---

[1] Очевидно (англ.) — *Примеч. перев.*

ЗАТМЕНИЕ ДАРВИНИЗМА

Чарльз Дарвин сделал подвижной естественную классификацию Линнея. Несмотря на то что динамизм Дарвина, противопоставляемый фиксированным классам Линнея, витал в воздухе (он встречается также в трансформизме Ламарка), оригинальность натуралиста состоит в предоставлении поясняющего механизма: естественный отбор, задуманный как метафора, — так написано в *«Происхождении видов»* (1859). Теорема Дарвина об эволюции основывается в целом на методиках одомашнивания и разведения животных и растений («естественный отбор» как обобщение «искусственного отбора», проводимого человеком, но без действующего субъекта, демиурга, этот отбор осуществляющего, и потому — без какого-либо финала) и реализуется в виде эволюционных деревьев, которые упорядочивают живущие виды и останки вымерших видов (филогенетическая реконструкция морфологических таксономий).

После смерти Дарвина в 1882 году и до возрождения его идей в 1930-е годы наблюдалось «затмение дарвинизма» — эволюционная биология оказалась в состоянии прострации вследствие развития менделевской теории наследования. Работы Менделя стали широко известны в 1900-м, через 34 года после их публикации и через 16 лет после смерти автора, причем на континенте — благодаря ботанику Хуго де Фризу, а на британских островах — благодаря Уильяму Бэтсону (ему мы также обязаны термином *генетика*), который использовал их для опровержения недарвиновских теорий (ламаркизма и мутационизма), подразумевавших скачкообразные изменения видов. Бэтсон подчеркивал разницу между Менделем и Дарвином, представляя первого как врага теории эволюции, и второго — как ответственного за удручающее состояние, в котором оказалась теория Менделя.

Когда в 1906 году не стало Уэлдона, в 1911-м — Гальтона, Карл Пирсон остался практически единственным защитником ортодоксального подхода: *Natura non facit saltus*[2]. На самом

[2] Природа не делает скачков (лат.) — *Примеч. перев.*

СЛИШКОМ ХОРОШО, ЧТОБЫ БЫТЬ ПРАВДОЙ

Грегор Мендель.

Основным результатом экспериментов Менделя по скрещиванию растений было открытие, что определенные признаки передаются потомству без ослабления или изменения, будучи перенесенными какими-то отдельными единицами или частицами, которые Мендель называл *факторами*, а мы сегодня именуем *генами*. Монах-августинец провел также исчерпывающий анализ результатов экспериментов. Так, скрещивая желтый и зеленый горох, он получил урожай, в котором из 8023 горошин 6022 (≈75%) были желтыми (доминантный признак) и 2001 (≈25%) — зелеными (рецессивный признак). Речь шла о втором законе Менделя, или о законе расщепления признаков. В статье *«Работы Менделя открыты заново?»* (*Has Mendel's work been rediscovered?*), опубликованной в 1936 году, Фишер посредством критерия χ^2 обнаружил почти полное соответствие наблюдений Менделя в отношении знаменитых экспериментов с горохом и теоретическими результатами, которые должны были получиться. Самым неожиданным было то, что в некоторых экспериментах Мендель сделал ошибочные выводы. Тем не менее наблюдения имели значительное сходство с этими ошибочными заключениями. Фишер считал, что необязательно сам Мендель фальсифицировал данные: просто кто-то из его ассистентов, не проделав работу должным образом, знал, что именно Мендель хотел получить... Этот вопрос вызвал много споров, и до сих пор нет единого мнения о том, сам Мендель или его помощник изменили данные. Иногда даже небольшое расхождение вызывает подозрение.

деле первые биометрические работы Пирсона состояли в изучении закона наследования Гальтона и подкреплении гипотезы постепенности, с помощью которой биометристы заключали, что эволюция происходит не скачкообразно, как утверждали последователи мутационизма, а через непрерывный отбор благоприятных изменений для распространения через потомков.

Будучи оплотом дарвинизма, биометрическая школа вступила в жестокие споры с менделианским напором мысли. Это железное противостояние испытало воздействие принятой Пирсоном и почерпнутой им еще в Германии философии науки, которая заставляла его воспринимать биометрию лишь как описание без рассуждений, как исключительно количественную теорию естественной эволюции. Пирсон хотел делать вероятностные предсказания об эволюции в череде поколений, но не вступать в метафизическую дискуссию о наследственных механизмах. Эта цель соответствовала библии позитивизма Пирсона — *«Грамматике науки»*, сходство которой с философией идеалистов было отмечено и обличено Владимиром Лениным в *«Материализме и эмпириокритицизме»* (1909). Эта особенная философия, с одной стороны, привела его к развитию полностью математической теории наследственности, наделенной инструментами для описания наблюдаемых феноменов, а с другой — помешала оценить экстраординарный вклад, представленный Фишером в 1918 году. Для Пирсона бесконечные популяции и совокупности наследственных факторов, о которых говорил Фишер, были недоступными для наблюдения и, следовательно, нереальными. Неприязнь к образам, которые использовал Фишер, была огромна.

РЕВОЛЮЦИЯ НА ФЕРМЕ

Споры между биометристами и генетиками не стихали, пока Фишер не доказал, что эмпирические параметры организмов соответствуют законам наследственности. Британский статистик был инициатором объединения идей Дарвина и Менделя, так как доказал, что данные измерений оказываются результатом сложения большого количества менделевых факторов (генов) и значения коэффициентов корреляции, полученные в эксперименте, объясняются действием всей совокупности этих факторов.

Фишер выполнил двойную миссию. С одной стороны, в 1930-е годы он внес значительный вклад в рождение неодарвинизма — синтетической теории эволюции. В этом синтезе соединились многие научные области (биометрия, генетика, анатомия, эмбриология, палеонтология), как это видно из списка авторов: Феодосий Добржанский (генетик), Эрнст Майр (зоолог), Джордж Гейлорд Симпсон (палеонтолог) и так далее. С другой стороны, он создал популяционную генетику, которая является одним из столпов современной эволюционной теории — дисциплиной, в которой, образуя гармоничное, математически смоделированное единство, сходятся эволюционная биология и генетика.

Здесь стоит рассказать о революционном труде Фишера, который он надиктовал своей жене в Ротамстеде — *«Генетической теории естественного отбора»* (*The Genetical Theory of Natural Selection*, 1930), а также о работах других знаменитых генетиков: *«Эволюция в менделевских популяциях»* (*Evolution in Mendelian Populations*, 1931) Сьюэла Райта (1889–1988) и *«Причины эволюции»* (*The Causes of Evolution*, 1932) Дж. Б. С. Холдейна (1892–1964), который в 1937-м возглавил кафедру биометрики Университетского колледжа Лондона при финансовой поддержке вдовы Рафаэля Уэлдона. Фишер, Райт и Холдейн — «три тенора» популяционной генетики, закрепившие роль дарвиновского отбора в качестве главного эволюционного механизма в терминах статистического эффекта менделевской генетики.

«Естественный отбор — не эволюция». Этим категорическим утверждением начинается книга Фишера, впоследствии ставшая классикой популяционной генетики. Этот афоризм должен был привлечь внимание к другому необходимому компоненту теории эволюции: менделевской генетике.

Единицами эволюции были не отдельные организмы, а популяции, каждая со своим генетическим распределением. В отсутствии мутаций и при предположении постоянства среды эволюция популяции рано или поздно останавливается. Несмотря на то что количество возможных комбинаций вариантов генов (аллелей) очень велико, оно конечное, поэтому процесс остановился бы на комбинации, лучше всего адаптированной к условиям среды, даже если бы естественному отбору пришлось идти к этому в течение многих поколений на протяжении большого периода времени. Тем не менее мутации случаются, хотя и редко. История выживания нового мутантного гена зависит, как показал Фишер, от капризов судьбы и от пользы или ущерба, которые этот ген приносит в борьбе за выживание.

Математическое рассуждение Фишера в его книге начиналось с гипотезы о появлении одного мутантного гена в популяции из миллиона индивидуумов, с распределением Пуассона или «редких событий» со средним $1 + e (e \geqslant 0)$, где e обозначает «преимущество для селекции». Если в популяции присутствовали экземпляры фенотипически различные (например, белая и черная моль), каждый из которых мог соответствовать одному или нескольким генотипам (в зависимости от доминантности или рецессивности аллеля), и в конкретном поколении наблюдаемое соотношение двух фенотипов было равно r, а в следующем — $r(1 + e)$, то e — преимущество аллеля, связанного с этим фенотипом (как у черной моли по отношению к белой моли — она лучше маскировалась в дыму английских фабрик). Разумеется, это преимущество необязательно постоянно во времени и в разных условиях среды, ведь то, что благоприятно здесь и сейчас, может не являться таковым в другое время или в другом месте. В случае моли преимущество в 0,01 в пользу моли, маскирующейся в промышленной среде, означает, что если у белого варианта появилось 100 единиц потомства, то у черного — 101 (на 1% больше).

Знакомые генетики спрашивали меня: правда ли, что великий генетик Р. Э. Фишер был также известным статистиком?

Леонард Джимми Сэвидж (1976)

В этих условиях Фишер вычислил вероятность исчезновения мутантного гена в n-м поколении. При отсутствии преимущества ($e = 0$) вероятность исчезновения в 63-м поколении равна 0,9698, то есть почти 97% в пользу исчезновения.

Удивительно, для преимущества в 1% (e = 0,01) указанная вероятность составляет 0,9591, почти 96% в пользу исчезновения. Всего на 1% меньше. Дальнейшие вычисления показывают, что в 127-м поколении вероятность сохранить ген равна 0,0271 с преимуществом и 0,0153 без него, то есть мутантный ген имеет почти вдвое больший шанс выживания, хотя у обоих эта вероятность на самом деле небольшая. В пределе вероятность сохранения благоприятной мутации составила около 2% (с другой стороны, для нейтральной мутации она была почти нулевой). Тем не менее, если популяция велика и насчитывает миллионы индивидуумов, будет существовать значительное количество особей с благоприятной мутацией, что позволит возникнуть адаптивному изменению, хотя многие благоприятные мутации утратятся по пути.

С помощью этих вычислений Фишер хотел также показать, что направление и смысл эволюции имеют мало общего с мутациями, поскольку без эволюционного преимущества эффект мутации в биологическом виде незначителен и в пределе равен нулю (не говоря уже о том, что большинство мутаций приводят к летальным изменениям). Естественный отбор — это процесс, посредством которого маловероятное событие, мутация, с течением времени увеличивает свою вероятность. Таким образом, естественный отбор — главный двигатель эволюции. Это доказывало правоту Дарвина и возвращало к жизни дарвинизм, чему так сопротивлялись последователи Менделя. Биологические выводы из математических результатов, полученных Фишером, были чрезвычайно важны и подкреплены экспериментами с фруктовой мушкой (*Drosophila melanogaster*, сумасшедшая репродуктивная активность которой облегчает изучение мутаций и скрещиваний).

Кроме того, труд Фишера содержал «фундаментальную теорему естественного отбора», прославляющую союз учений Дарвина и Менделя, квинтэссенцию взглядов Фишера на естественный отбор. Ее основное положение: «Скорость роста биологической адаптации популяции в любой момент равна генетической вариабельности, присутствующей в этот момент

ТРУДНЫЙ ХАРАКТЕР

Рональд Эйлмер Фишер был наделен многими талантами, но имел также заметные недостатки. Мятежный дух заставлял его упорствовать и спорить по пустякам, часто довольно грубо — устно или на письме, — в разногласиях с коллегами, статистиками и генетиками. Один из примеров мы уже видели — столкновение с Карлом Пирсоном. А в следующей главе будет представлено его мнение в отношении статистического вывода или связи между курением табака и возникновением рака. Эта жесткость в научной идеологии распространялась на его религиозные и политические убеждения, происходившие из его консерватизма и почитания традиций, а также склонности отрицать коммунизм и любые формы прогресса. Ввиду евгенических убеждений он считал, что не все люди равны.

К этому добавлялись некоторые странности, типичные для математиков-гениев. Например, его склонность терять важные документы. С другой стороны, плохое зрение не стало препятствием для обретения отличной физической формы благодаря ежедневным пробежкам. Чтобы работать дома в тишине (это было затруднительно, учитывая количество детей, которые обременяли его и финансово), он требовал, чтобы между ним и детьми были как минимум две закрытые двери, что позволило бы ему сконцентрироваться.

Скверный преподаватель

По свидетельствам, Фишер был однозначно плохим преподавателем. Он опускал комментарии в ходе обучения и исследований. По этому поводу приведем анекдот, рассказанный шотландским статистиком У. Кокреном (1909–1980):

> «На одном из занятий он привел результат без доказательства. После нескольких безуспешных попыток справиться самостоятельно я зашел к нему в кабинет и попросил представить мне доказательство. Он ответил, что переложил его куда-то; открыл множество ящиков и в конце концов решил, что проще сделать его заново. Мы сели, и он записал выражение, с которого я начал свои попытки. «Очевидно, дальше рассуждение идет так», — сказал он и написал выражение в две строки. — «Теперь предположим, что нужно развить это», — и написал уравнение на три строки. Взглянул на выражение и прокомментировал: «Единственно возможный следующий шаг — такой», — и выдал выражение на четыре с половиной строки. Помолчал 45 секунд и сказал: «Результат должен следовать отсюда», — написав ниже выражение, о котором я его спрашивал. Урок был окончен».

в популяции». Это достаточно запутанная формулировка, и потребовалось несколько лет, чтобы оценить ее по достоинству.

Чтобы естественный отбор мог действовать на признак, нужно, чтобы было из чего выбирать, то есть несколько аллелей, или альтернативных форм гена, кодирующего признак. Фишер математически доказал, что чем больше генетическая вариабельность в популяции, тем выше скорость эволюции. Большой генетической вариации соответствуют большие эволюционные сдвиги. Фишер сравнивал свою теорему со вторым принципом термодинамики, или законом энтропии, которая может только увеличиваться. Естественный отбор работает медленно, но неизбежно постепенно увеличивая частоту благоприятных генов, которые встраиваются в геном вида, что увеличивает приспособляемость организмов. Следовательно, отбор превращает хорошо адаптированный аллель в доминантный, а губительные мутации — в рецессивный.

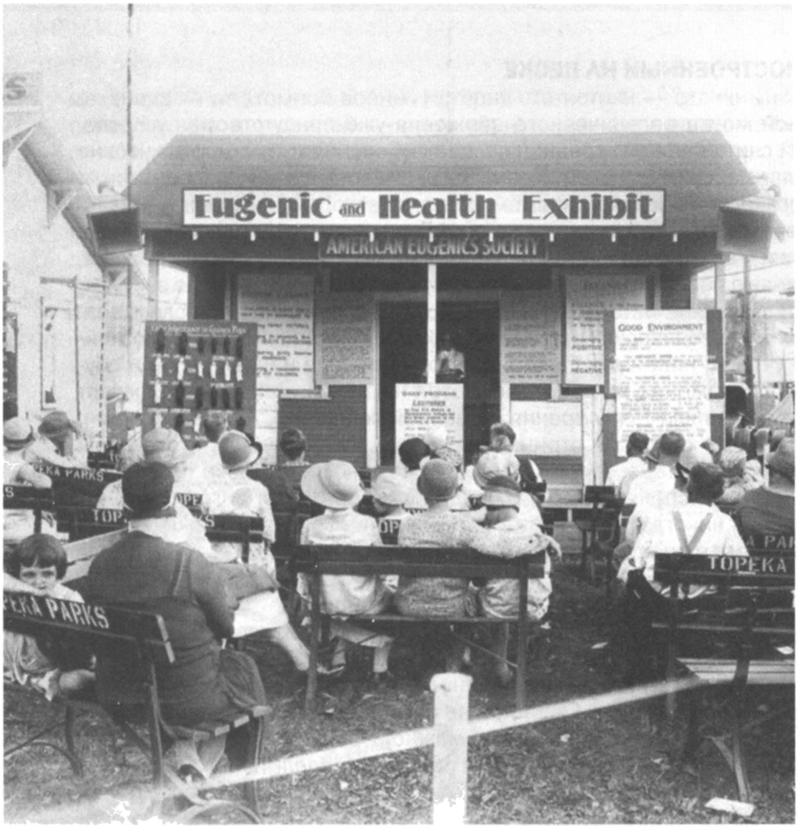
Популяционная генетика сообщила математическую стройность теории эволюции. Несмотря на это, Фишер и Холдейн придерживались двух гипотез, которые критиковал Райт. Во-первых, они рассматривали генетический материал организма как мешок фасоли, то есть набор не взаимодействующих друг с другом генов. Именно Райт обобщил их упрощенные модели. Во-вторых, они рассматривали популяции в целом, что приводило к тому, что естественный отбор выглядел практически однонаправленным процессом, без ответвлений. Но Райт обратил внимание на то обстоятельство, что большие популяции обычно разделены на меньшие локальные популяции, в которых процветала эндогамия, что делало естественный отбор более изменчивым, давая начало понятию *адаптивного ландшафта*.

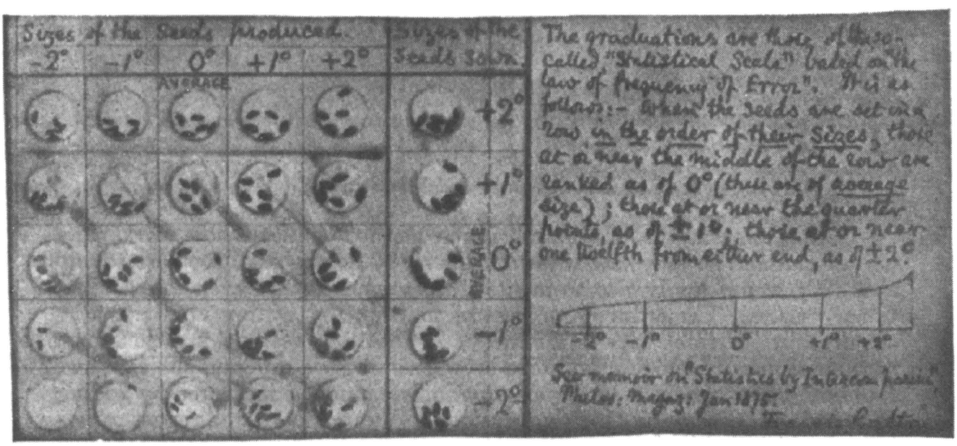
В настоящее время, несмотря на вклад Фишера и других генетиков, остаются сомнения по поводу соотношения в процессе эволюции ролей естественного отбора и мутаций и, в частности, по поводу их действия на молекулярном уровне. Для кого-то основной эволюционной силой на молекулярном уровне является простой «генетический дрейф», то есть изменение аллельных частот биологического вида вследствие стохастического эффекта, связанного с размножением (аллели детей являются случайной выборкой из аллелей родителей), и отбор оказывает свое действие на морфологическом уровне, на уровне организмов. Для других, наоборот, мутантные гены являются выборочно нейтральными, роль мутаций нельзя игнорировать, и отбор действует на молекулярном и морфологическом уровнях. Иными словами, окончательно не известно, геном или вид являются субъектом эволюции. Однако существуют разногласия по поводу непрерывности или дискретности эволюционных изменений (градуализм). Так, сторонники «пунктирного равновесия», в отличие от ортодоксальных неодарвинистов, считают, что в эволюции чередуются периоды быстрых и медленных изменений. Никто не оспаривает позицию Дарвина, но неодарвинисты не единодушны.

СТАТИСТИКА, ДАРВИНИЗМ И ЕВГЕНИКА

Список мотиваций не был бы полным, не упомяни мы, что Фишер усердно продвигал евгенику — дисциплину, которая стимулировала его интерес к работам в области человеческой генетики. В Кембридже Фишер и другие блестящие ученые (например, Джон Мейнард Кейнс) с энтузиазмом сотрудничали с *Евгеническим образовательным обществом*, основанным в 1907 году Гальтоном и до 1911 года возглавляемым Леонардом Дарвином (он также был председателем Первого международного конгресса по евгенике, проходившего в Лондоне в 1912 году и посвященного памяти Гальтона). Кроме того, Фишер создал евгеническое общество в стенах университета.

В 1911 году Фишер провел конференцию для группы студентов и изложил идею Гальтона, согласно которой нормальная кривая применима также к нравственным и интеллектуальным качествам человека, и люди подразделяются на множество классов — от слабоумных до гениев. Интеллектуальные и нравственные достоинства считались очевидно наследуемыми признаками, и поэтому браки должны были заключаться между лицами одного класса. Для Фишера книга Гальтона *«Наследственность таланта»* (*Hereditary Genius*) была одной





ВВЕРХУ: **Выставка, посвященная евгенике, на ярмарке в Канзасе в 1929 году.**

ВНИЗУ: **Эксперименты Гальтона по изучению наследования у гороха. Справа графически представлена функция распределения результатов, которую английский ученый сравнивал с законом ошибок.**

ДОМ, ПОСТРОЕННЫЙ НА ПЕСКЕ

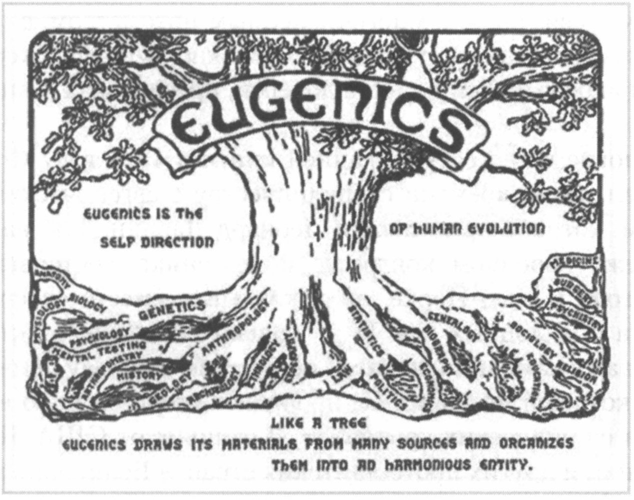
Основной мотив евгенического движения уже присутствовал у Кетле, который считал, что его средний человек объединяет в себе физические и моральные качества расы. Кроме того, существовала идея, связанная с биологическим эволюционизмом, согласно которой благодаря социальным мерам и селекции можно сохранить или изменить расовые характеристики (Гальтон). Тем не менее историки науки до сих пор не решили, какую роль в развитии статистики сыграла евгеника. Некоторые считают, что статистические методы были разработаны для решения проблем, возникающих в евгенических исследованиях. Эта доктрина не только мотивировала ученых, например Гальтона, Карла Пирсона и Фишера, но и обусловила содержание их работ (несмотря на то что, например, Эджуорт и Юл не разделяли точку зрения о расовой селекции). Другая фракция, наоборот, категорически отрицает эту связь, подчеркивая, что методы биометрической лаборатории в Университетском колледже полностью отличались от используемых в соседней лаборатории евгеники и что Карл Пирсон никогда не вступал в евгенические общества (даже если он не делал этого из неприятия менделизма).

Окончательное разделение статистики и евгеники

Возможно, биометрия и евгеника не были полностью изолированы. И хотя некоторые методы, такие как хи-квадрат, нашли множество применений (в агрономии, генетике, промышленности и так далее), другие методы, вроде карты родословных Гальтона, не были так распространены. Общественное значение статистики невозможно отрицать: ее формирование проходило в контакте с биометрией и попытками сделать евгенику царицей наук (что видно по плакату Второго международного конгресса по евгенике). Тем не менее повсеместность статистических методов, то есть их распространение во множестве естественнонаучных и социальных областей, делает возможным их независимость от идеологии, так же как сегодня классическая механика не зависит от баллистики или квантовая

из великих работа XIX века, сравнимой с *«Происхождением видов»* Дарвина, и в какой-то степени ее дополняла.

Одна из первых статей Фишера увидела свет в 1914 году на страницах *Eugenics Review*, журнала, бывшего символом евгенического движения, где он впоследствии опубликовал более 200 статей, не считая рецензий на книги и комментарии. Она называлась *«Некоторые надежды последователя евгеники»* (*Some hopes of a Eugenist*). В статье, уже прочитанной в универмеханика — от атомной войны. История статистики — отличный пример переплетения «внутренней» и «внешней» истории науки: евгеника была мостиком, который позволил связать эволюционную биологию и статистику и как следствие — стимулировать создание первых современных статистических представлений. Благодаря этому взаимодействию статистика получила достаточно силы, чтобы укорениться в других научных областях, это дало ей независимость от евгенической идеологии и одновременно достаточно уважения в связи с эволюционной биологией, определив становление популяционной генетики и создание синтетической теории эволюции.



«Евгеника и естественное движение эволюции человека» — девиз Второго международного конгресса по евгенике (1921); эта наука представлена как дерево, объединяющее разнообразные социальные и биологические дисциплины во главе с генетикой и статистикой.

ситетском обществе в Кембридже, он защищал евгенику как путь к прогрессу человечества. Три года спустя он опубликовал редакторскую статью, в которой продвигал принятие политических мер, способных увеличить рождаемость в профессиональных классах, а в низких — контролировать ее. На эту тему он рассуждал также в последних главах *«Генетической теории естественного отбора»*.

Фишер связывал упадок цивилизации с историческим моментом, когда рождаемость в высших классах начинала снижаться за счет увеличения рождаемости в низших классах (по его мнению, менее приспособленных интеллектуально). Используя данные переписи населения Великобритании, Фишер доказал обратную связь между рождаемостью и социальным положением: высшие классы имели низкую рождаемость, тогда как в низших классах наблюдалась рождаемость высокая. Семьи с высоким социальным статусом не могли позволить себе иметь много детей, так как небольшое количество детей было экономическим преимуществом. Чтобы преодолеть это бедствие, британский евгеник предлагал ослабить финансовое бремя, связанное с многочисленным потомством, с помощью государственных субсидий. Возможно, Фишер, у которого были двое сыновей и шесть дочерей, преследовал и личные цели.

Одновременно с публикацией книги в 1930 году Фишер посвящал много времени сотрудничеству с евгеническим обществом, которое представлял Леонард Дарвин. Так, на Третьем международном конгрессе по евгенике, состоявшемся в 1932 году в Нью-Йорке, он докладывал вместо своего наставника преклонных лет. И не только это: Фишер активно участвовал в кампании, начатой обществом в пользу принятия закона, который позволил бы проводить стерилизацию на основании евгенических критериев. В отличие от США, Германии, Дании и других протестантских стран, в Великобритании так и не были приняты законы о стерилизации — ни насильственной, ни добровольной. Несмотря на это, стоит уточнить, что британские евгеники постоянно подчеркивали большее влияние социального класса, чем расы, на наследование умственных качеств, в отличие от их североамериканских или немецких коллег.

После своего перехода из Ротамстеда в Университетский колледж в 1933 году Фишер продолжил евгенические исследования в лаборатории Гальтона. Вместе с другими коллегами он углубил сбор данных о человеческих родословных и изучение групп крови и резус-фактора (Rh). В 1950 году он выступал против Декларации о расе и расовых предрассудках ЮНЕСКО, которая устанавливала, что понятие расы — лишь инструмент классификации, не связанный с культурой, этносом или результатами теста на уровень интеллекта.

Сегодня слово *евгеника* уже несколько старомодно. То, что казалось руководящей идеей, оказалось инертным. Тем не менее, чтобы дать контекст убеждению Фишера в положительных качествах евгеники, следует отметить, что сегодня многие ученые выступают в пользу генетической инженерии, применяемой не только в отношении патологий, но и физических характеристик, таких как цвет волос или глаз.

ГЛАВА 5

Борьба с индукцией и научный метод

Параллельно своим открытиям в математике и биологии часть своего времени Фишер посвящал размышлениям о применимости статистических методов, особенно о байесовском выводе по сравнению с выводом, основанным на частотном подходе, который считал более правильным. Нет ни одной нити в полотне статистики, которую Фишер не рассмотрел бы подробно, что привело его к дискуссии с Ежи Нейманом и Эгоном Пирсоном по поводу подтверждения гипотез и, уже в последние годы жизни, с врачами по поводу связи табака и рака.

Преодолев тяжелый финансовый и духовный кризис, в 1943 году Фишер вернулся в Кембридж, в свою *alma mater*, и возглавил кафедру генетики, сменив на этом посту Р. Паннета. Жизнь с Фишером была нелегкой из-за противоречивости его характера: яркий и невзрачный, жестокий и дружелюбный, скупой и щедрый. Все одновременно. К денежным проблемам добавлялись тяжелая работа и уход за детьми. Невнимание к здоровью жены привело к необратимому семейному кризису в 1942 году. В этом же году старший сын, пилот истребителя, погиб в авиакатастрофе над Сицилией, оставив обоих родителей опустошенными. Брак распался, когда Фишер переехал в Кембридж... один.

Математическая статистика, созданная ученым в бурные 1920-е годы, внезапно вызвала споры (личные и концептуальные). Это обстоятельство заставило Фишера глубоко задуматься о внутренней логике новых методов научного вывода, о статистическом выводе, который сегодня называется классическим. Еще в 1935 году он опубликовал предварительную статью под заголовком *«Логика индуктивного вывода»* (*The Logic of Inductive Inference*), в конце предыдущего года вызвавшую многочисленные возражения в *Королевском статистическом обществе*. Но в 1950-е годы он посвятил больше времени этому вопросу. За спорной статьей *«Статистические методы и на-*

учная индукция» (*Statistical Methods and Scientific Induction*), представленной в *Королевском статистическом обществе* в 1955 году, последовала книга *«Статистические методы и научный вывод»* (*Statistical Methods and Scientific Inference*), опубликованная в 1956 году, в ней Фишер углубляет философские аспекты статистического вывода.

В этом труде Фишер попытался предложить унифицированное видение вывода, используя три подхода к существу проблемы: метод максимального правдоподобия, критерий значимости и фидуциальную вероятность (ее определение будет дано ниже). Книга представляет собой панораму статистического вывода от Байеса до современности. Кроме всего прочего, в ней Фишер осуждал Байеса, Карла Пирсона и других «ложных пророков». Британский статистик не упустил случая высказать свою неприязнь коллегам из США, в понимании которых статистика является исключительно математикой, безо всякой связи с экспериментальными науками. По мнению некоторых коллег, например Мориса Кендалла, эта книга — как и брошюра о раке и курении на 47 страницах, увидевшая свет в 1959 году, — вообще не должна была быть написана.

В любом случае Фишер уделял внимание трем спорным моментам: смыслу вероятности, недостаткам байесовского вывода и логике проверки гипотез.

ДАТЬ ОПРЕДЕЛЕНИЕ ВЕРОЯТНОСТИ

Хотя слово *вероятность* широко используется в языках, родственных латинскому (в котором *probabile* значит что-то вроде «достойный одобрения»), математическое понятие вероятности впервые появилось около 1660 года. С самого рождения этому понятию был свойствен дуализм. Идея напоминала двуликого Януса, который символизировал перемены в идее знаков из эпохи Возрождения. Утверждение считалось вероятным, если было хорошо доказанным. В эпоху Возрождения мир начал предоставлять доказательства посредством знаков.

Не только книги ученых мужей являлись убедительными свидетельствами, но также, говоря словами Галилея, книга природы. Таким образом, вероятным знаком был сигнал частый, повторяющийся, с помощью которого мир обеспечивал доказательства (так же, как дым является признаком огня).

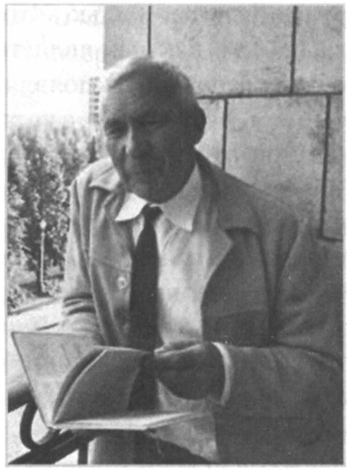
Поэтому вероятность появляется в связи, с одной стороны, с убеждением, а с другой — с частотой. Как схоластическая модальность возможности, вероятность могла объясняться *de dicto* (через суждения и их доказательства) или *de re* (через предметы и тенденцию, продемонстрированную в нескольких эпизодах, давать стабильные относительные частоты). Слово *вероятность* впервые было использовано для чего-то, что можно измерить, в *«Логике Пор-Рояля»* — учебнике мыслительного искусства, напечатанном приблизительно в 1662 году несколькими коллегами Паскаля, обосновавшимися в этом анклаве янсенизма.

И Пуассон в своем труде 1837 года о законе больших чисел, и Курно в книге о моральной науке, опубликованной в 1843-м, разъясняют, что вероятность смешивает два понятия, которые следует различать с хирургической точностью. Во-первых, это *шанс*, или физическая вероятность, которая определяет легкость, или склонность, как говорят сегодня, к возникновению события; во-вторых, это *эпистемологическая вероятность*, отражающая уверенность в возникновении события. В то время как первый отсылает к объективному свойству события (возможность его наступления, полезно при моделировании), вторая является субъективной (относится к нашему знанию, полезна для выводов).

Любопытно, что преподобный Томас Байес за 100 лет до этого писал: «Под *шансом* я понимаю то же, что под *вероятностью*». Но в 1850-е годы мир был уже не таким, каким он был во времена Байеса и Лапласа. Объективный аспект вероятности стал гораздо более важным, чем субъективный, просто потому что реальность была наполнена частотами. Количество цифр, наводнивших мир, сместило стрелку весов. Джон Венн в *«Логике случая»* (*Logic of Chance*) в 1866 году склонялся к частотному, а не персональному подходу к вероятности.

АКСИОМАТИЧЕСКИЕ РЕШЕНИЯ

Эти две интерпретации вероятности разделяют один и тот же математический формализм: аксиомы Колмогорова (1903–1987), сформулированные советским математиком в 1933 году. Любая интерпретация вероятности, которая удовлетворяет этим аксиомам (а их несколько), является хорошей реализацией понятия. Предложенные аксиомы учитывают рассуждения, формализованные в классическом («правиле Лапласа», применимом только к равновероятным случаям) и в частотном определении вероятности (теореме Бернулли, применимой только к повторяющимся феноменам), а также связывают теорию вероятностей с теорией множеств и теорией мер, превращая ее в стабильную математическую теорию: она быстро распространилась по всей центральной Европе, позволив решить многие проблемы. С другой стороны, субъективная интерпретация вероятности (как степень уверенности в высказывании или согласия с происхождением события, разная у всех людей, хотя и подчиненная очень строгим правилам внутренней непротиворечивости) была независимо формализована итальянским статистиком Бруно де Финетти (1906–1985) в 1937 году и распространена Леонардом Сэвиджем (1917–1971) в 1954 году, что вернуло к жизни байесовский вывод и этот подход к вероятности, связанный с полезностью (понятие, введенное Даниилом Бернулли, племянником Якоба, в 1737 году и впоследствии Фрэнком Рамсеем в 1931).



Андрей Николаевич Колмогоров.

Тем не менее статистический вывод в XIX веке оставался отчетливо байесовским (для оценки неопределенности использовались методы обратной вероятности Байеса и Лапласа). Только когда огромное количество цифр, доступных зарегистрированных частот, стало реальностью не только в астрономии (с накоплением выборочных данных в социологии, биологии и агрономии), благодаря Фишеру развился, как мы видели в главе 3, объективный статистический вывод и перевесил вывод байесовский, или субъективный. С продолжающимся наблюдением регулярных событий в других областях естественных наук, отличных от звездного неба, субъективная интерпретация вероятности как степени уверенности, по Лапласу, была замещена объективной, или частотной интерпретацией: вероятность основывается не на убеждениях, а на эмпирических частотах. С самого начала Фишер понимал, что каждая интерпретация вероятности определяет свою теорию вывода, так как вероятностные концепции являются кирпичиками, из которых складывается статистический вывод.

«БАЙЕС — ЭТО ВСЕ, ЧТО ТЕБЕ НУЖНО...»

Для многих ученых статистика призвана ответить на фундаментальный вопрос: когда будет корректным утверждать, что множество наблюдений дает доказательства в пользу гипотезы или против нее? Самая древняя попытка решить этот вопрос восходит к 1763 году: теорема Байеса, опубликованная в *«Эссе о решении проблем в теории случайных событий»* (*An essay towards solving a problem in the doctrine of chances*), подписанном преподобным Томасом Байесом. Эта теорема, предшественница методов обратной вероятности и байесовского вывода, была центральным результатом теста, задуманного с целью опровергнуть скептическую критику, поднятую против индукции шотландским философом Дэвидом Юмом в сочинении *«О чудесах»* (*Of Miracles*), так как открывала математическую дискуссию о вероятности, понимаемой как уверенность.

Только в этом теологическом контексте, испытавшем влияние Ньютона, можно понять, как, например, доктор Джон Арбетнот в 1710 году произвел вычисления, которые стали первым критерием значимости статистической гипотезы: если вероятность рождения мальчика равна вероятности рождения девочки (то есть 1/2), вероятность того, что будут зарегистрированы — как это было показано — 82 следующих друг за другом года, в которых мальчиков родится больше, чем девочек, равна $(1/2)^{82}$, то есть практически нулю. Поэтому гипотеза о равном количестве рождающихся мальчиков и девочек была опровергнута, и Арбетнот интерпретировал эту закономерность как аргумент (индуктивный) в пользу божественного вмешательства. Поэтому формула Байеса позволяла выражать вероятностные суждения о верности гипотезы (вероятность *a posteriori*), основываясь на данных (правдоподобии), а также на субъективной оценке того, что гипотеза верна (вероятность *a priori*).

Причины, которые привели Байеса к его теореме, были скорее теологическими и социологическими, чем чисто математическими.

Карл Пирсон (1926)

Тем не менее проблема обратной вероятности обрела свою форму в работах Якоба Бернулли в 1713 году. В 1704 году швейцарский математик сообщил в письме Лейбницу, что он открыл теорему, которая позволила ему вычислить *a posteriori*, с определенной точностью, неизвестные вероятности событий, известных эмпирически, как если бы они были известны *a priori*, изначально. Однако золотая теорема Бернулли не была в точности примером обратной вероятности, так как она утверждала, что для «известной» вероятности происхождения события относительная частота происхождения этого события стремится к этому числу (слабый закон больших чисел). Речь идет о теореме, исключительно и бесспорно относящейся к теории вероятностей. Так, Бернулли смог вычислить количество подбрасываний симметричной кости, необходимых для того, чтобы с «моральной уверенностью» (то есть с вероятностью большей или равной 0,999 — стандартом, аналогичным используемому современными статистиками 95- или 99-процентному уровню уверенности) относительная частота выпадения «шестерки» отличалась от $p = 1/6$ (вероятность, считающаяся известной) не более чем на 0,01: 1388 889 раз. В теореме вероятность p была фиксирована и вычислялась вероятность наблюдать какие-либо данные при условии, что относительная частота благоприятных исходов f_n стремится к p при увеличении количества экспериментов. Бернулли делал свои наблюдения на материале, в то время называвшемся ***прямыми проблемами вероятности***, на задачах, в которых предполагалась известной вероятность благоприятного исхода и вычислялась вероятность некоторой последовательности благоприятных и неблагоприятных исходов.

Но если p неизвестно, то как можно пользоваться этой теоремой? Парадоксально, Бернулли создал свою теорему как раз для случаев, когда не было заранее известно значение p. Но он сопротивлялся желанию обратить теорему, удовлетворившись включением всех возможных значений p в определенном интервале (забегая вперед, скажем, что он создал оценку для интервала p для определенного уровня доверия, моральной уверенности; вслед за этим в астрономии появилась теория возможных ошибок, с помощью которой были созданы оценки для интервала с уровнем доверия 50%). Другими словами, Бернулли нашел способ вычисления следующей вероятности (при известном p): $P(p$ в интервале $f_n \pm \varepsilon \mid p)$. У него был соблазн считать вычисленные значения значением вероятности $P(p$ в интервале $f_n \pm \varepsilon \mid f_n)$, где известное p заменено на известное f_n. Разумеется, это рассуждение ошибочно, так как второе выражение не следует из первого. Кажется, только Лаплас поддался соблазну «обратить» теорему и вывести вероятность I из наблюдаемой частоты f_n, несмотря на то что эта мысль уже присутствовала у Бернулли, который в случае успеха смог бы решить проблему индукции — перехода от частного к общему, от выборки к генеральной совокупности (индуктивный вывод).

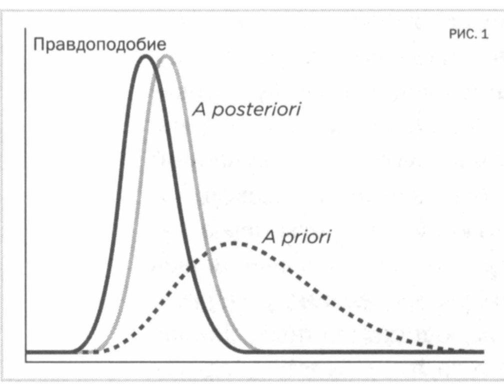
Решение этой проблемы, канонизировав эпистемологическую интерпретацию вероятности, Лаплас вывел из теоремы Байеса, которая считает неизвестную вероятность p случайной величиной. Работа Байеса была первой систематической попыткой вычислить вторую указанную вероятность: через назначение вероятности *a priori* и формулу Байеса вычислялась искомая вероятность. Предполагая распределение p *a priori* на интервале [0,1], основываясь на имеющихся данных, Лаплас вычислял вероятность (*a posteriori*) того, что p будет находить-

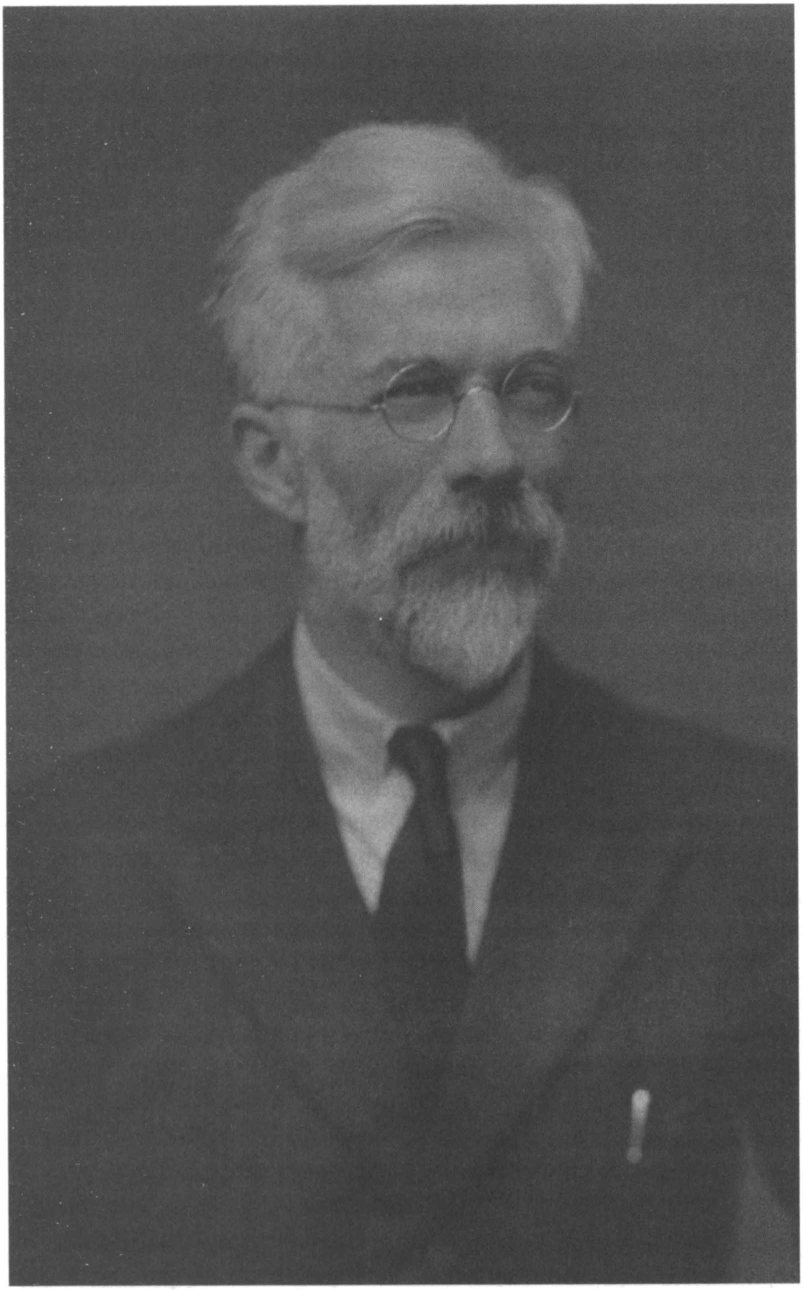
Распределение *a priori* и правдоподобие представлены пунктирной и сплошной линиями соответственно. Распределение *a posteriori*, вычисленное с помощью теоремы Байеса, изображено линией серого цвета (на горизонтальной оси располагаются возможные значения параметра θ, который требуется оценить). Распределение *a posteriori* находится посередине, между распределением *a priori* и правдоподобием. В этом конкретном примере оно гораздо ближе к правдоподобию, чем к вероятности *a priori*, что отражает количество информации, почерпнутой из данных.

ся на определенном расстоянии (символ в конце с. 133) от наблюдаемой относительной частоты f_n. Основываясь на количестве выпадений «шестерки», он вычислял вероятность того, что вероятность выпадения «шестерки» будет в окрестности наблюдаемой относительной частоты.

Байесовские статистики пытались узнать вероятность того, что некоторый неизвестный параметр θ находится между двумя заданными значениями. Чтобы это сделать, нужно знать две вещи: во-первых, правдоподобие P (X|θ), то есть вероятность наблюдать выборку, извлеченную из генеральной совокупности, согласно предполагаемому значению параметра; во-вторых, вероятность *a priori* параметра θ, или распределение θ *a priori*, которая отражает вероятность того, что неизвестный параметр находится в заданных пределах. Распределение *a posteriori* (θ|X), вычисленное по правилу Байеса, есть не что иное, как компромисс между распределением *a priori* и правдоподобием, между тем, что мы знаем и что мы оцениваем на основе данных наблюдений (см. рисунок 1).

Любовь к числам и объективности, характерная для XIX века, заставила математиков искать альтернативу этому сомнительному рассуждению. Фишер сделал борьбу с байесовским выводом одним из смыслов своей научной деятельности. По его мнению, статистические методы привели к более полному пониманию индуктивной логики, создав базу для научного вывода, так как индуктивный вывод приводил, в отличие от дедуктивного, к увеличению знаний (потому что позволял обучаться на основании опыта, хотя и всегда с некоторой неопределенностью, которую, впрочем, можно было измерить, что делало вывод достаточно строгим). Хотя основная роль в дедуктивном или прямом выводе (от общего к частному, от популяции к выборке) принадлежит вероятности, ин-





Рональд Э. Фишер в 1943 году, когда он вернулся в Кембридж, чтобы возглавить кафедру генетики. Тогда же семейные трудности завершились распадом его брака.

дуктивный, или обратный вывод (от частного к общему, от выборки к популяции) основан на правдоподобии и в некоторых случаях — на фидуциальной вероятности. Ни в одном из них байесовская вероятность не используется.

Помимо прочих недостатков, Фишер отмечал, что сторонники байесовского подхода незаметно подменяют обратный, или индуктивный, вывод прямым выводом, или вероятностной дедукцией, постулируя исходное знание: распределение *a priori* параметра θ. Являясь математическим выражением, формула Байеса была неоспоримой (даже если для Фишера это было совсем не очевидно), но ее использование требовало назначения вероятности *a priori* правдивости оцениваемой гипотезы, весьма спорной. В ситуации полного незнания неприемлемо давать всем возможным значениям θ одинаковую вероятность (равномерное распределение) или вероятность, которая зависит только от состояния осведомленности (субъективная вероятность), так как исследователи могут использовать несовместимые между собой значения *a priori*, впадая в недопустимый субъективизм. (Действительно, сегодня известны несколько парадоксов, таких как «парадокс Линдли», которые показывают, как байесовский вывод терпит полную неудачу, если выбираются неадекватные данные *a priori*: вся вероятность приписывается *a posteriori* некоторым значениям параметра, независимо от наблюдений.) Кроме того, факт, что с увеличением размера выборки точная форма распределения *a priori* теряет связь с правдоподобием (как на рисунке 1, с. 134), заставил Фишера утверждать, что наиболее естественно делать выводы без каких-либо предположений *a priori*.

Несмотря на это, для Фишера индуктивный вывод был возможным, если производился не через байесовский способ. В отличие от философа Карла Поппера, Фишер не верил в то, что наука должна вернуться к простой доказательной модели, удалившись от экспериментальной практики. Большинство математиков, будучи чуждыми искусству дедукции, путали неуверенный вывод (в котором неуверенность можно измерить) с нестрогим выводом. Обучение через опыт проводилось с помощью критериев значимости, которые, как мы узнали в главе 3, использовались для выводов на основании наблюдаемых данных без каких-либо отсылок к предшествующим убеждениям (*a priori*). Правдоподобие является рациональной мерой убеждения, так что, в отличие от вероятности (которая позволяет только дедуктивные рассуждения, поэтому формула Байеса основывается на данных *a priori*), допускает индуктивные рассуждения, будучи именно тем, что проверяется в ходе тестирования.

> На первой странице есть логическая ошибка, которая делает ошибочными остальные 395 страниц, — принятие постулата Байеса.
>
> **Фишер — о книге «*Теория вероятности*» (1939) астронома Гарольда Джеффриса**

Примерно в 1930 году Фишер понял, что в некоторых ситуациях можно трансформировать полученную информацию о параметре в вероятностные высказывания, не прибегая к теореме Байеса. С помощью туманной аргументации Фишер определил распределение вероятности параметра θ на основе данных и не принимая во внимание какое-либо распределение *a priori*. Это была так называемая *фидуциальная вероятность*. Фишер рассматривал $P(X|\theta)$ как функцию двух переменных и, подставив наблюдаемое выборочное значение X и вычислив θ как функцию от X, использовал $P(\theta|X)$ как распределение вероятности θ с практической точки зрения. Он нашел способ обратить вероятностные утверждения о наблюдениях, получив значение параметра в вероятностных утверждениях о параметре на основании наблюдений.

В рассуждении о фидуциальности присутствует перенос вероятности с X на θ, с выборочной статистики на параметр, что является интуитивным, но запутанным, так как меняет состояние параметра, который из неизвестного, но постоянного значения превращается в случайную величину. Для Леонарда Сэвиджа «фидуциальный подход Фишера был попыткой приготовить байесовскую яичницу, не разбив ни одного байесовского яйца», так как единственным отличием доверительного

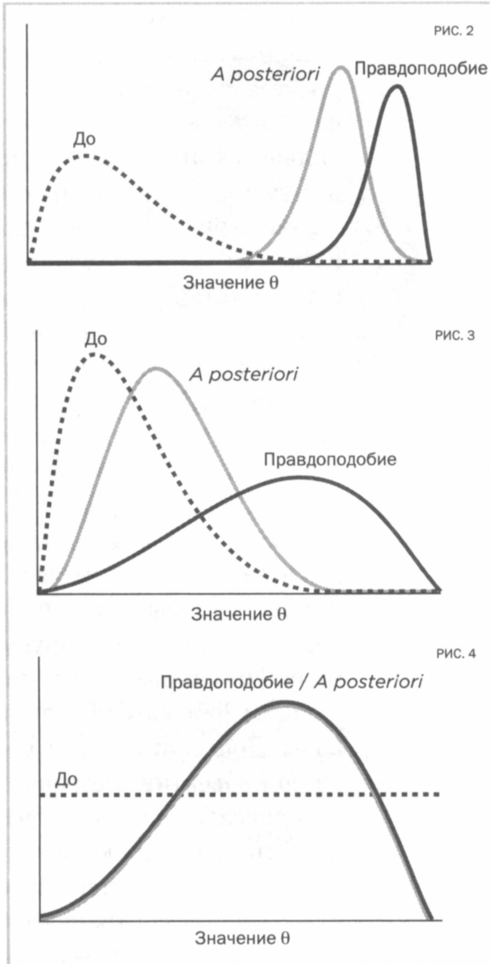
ВЗОЙДЕТ ЛИ ЗАВТРА СОЛНЦЕ?

Идеализированный портрет Томаса Байеса.

Надеясь опровергнуть Юма, который писал, что восход Солнца на следующий день есть лишь вероятное событие, Ричард Прайс (1723–1791), философ, занимавшийся посмертной публикацией документов Байеса, применил теорему своего коллеги, чтобы вычислить вероятность восхода Солнца. Принимая во внимание количество дней, в которые Солнце восходило, Лаплас улучшил эти вычисления, получив «правило последовательности»: если событие повторяется подряд любое количество раз, вероятность его очередного повторения равна этому числу плюс 1, деленному на то же самое число плюс 2. Так, если предположить, что Солнце восходит непрерывно в течение 5000 лет, то есть 1826 213 дней (Лаплас считал, что Земля очень молода — ей всего 5000 лет), вероятность восхода завтра равна 1826 214/ 1826 215 (99,9999 %). Но Лаплас подчеркивал, что в этом случае речь идет скорее о проблеме небесной механики, а не о проблеме вероятности, так как, согласно этому правилу, чем старше мы становимся, тем больше становится вероятность прожить еще больше. Таким образом, 80-летний человек имеет большую вероятность прожить еще день, чем 20-летний. Это противоречит здравому смыслу.

метода от метода Байеса было отсутствие знаний *a priori*. Действительно, фидуциальное распределение можно вычислить как распределение *a posteriori* по отношению к неинформативному (нейтральному, равномерному) распределению *a priori*. То, что сделал Фишер, ослабило его позиции, и даже в книге 1956 года он выступает в поддержку байесовского подхода при условии достаточной информации о параметре на основании выборки, так как при вычислении распределения *a posteriori* с помощью теоремы Байеса правдоподобие будет определяющим (как на рисунке 1, с. 134). В остальных случаях он выступал за фидуциальный подход.

Усилия по вытеснению теоремы Байеса, поддержанные такой важной персоной, как Фишер, не возымели успеха, и во второй половине XX века произошло возвращение байесовского вывода — несомненно, более древнего подхода к статистическому выводу, — в связи с теорией принятий решений. Байесовский подход является формальным, алгоритмическим подходом к туманной идее «обучения через опыт для принятия лучших решений». Он обеспечивает механизм для комбинирования наших знаний *a priori* с данными выборки, предоставляя вывод, объединяющий всю доступную информацию.

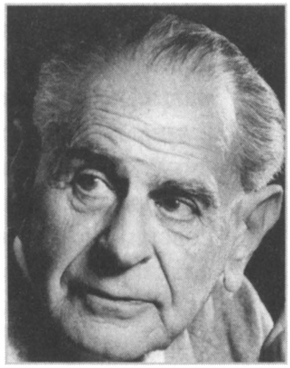


Сегодня некоторые статистики считают, что выводом будущего будет байесовский, — если будет вообще, так как классические методы часто имеют недостаточную точность, не принимают во внимание информацию, происходящую из предшествующих исследований, и увеличивают доверие к гипотезе. Пока классический вывод предполагает, что параметр θ фиксирован, и пытается его оценить, байесовский вывод интерпретирует его как случайную величину и делает объектом исследования вероятность $P(\theta|X)$. Если размер выборки X большой, оба метода дают в общем одинаковые результаты, поскольку информация из выборки намного превосходит информацию *a priori* (как можно видеть на рисунке 2, распределение *a posteriori* больше похоже на кривую правдоподобия, чем распределение *a priori*). Тогда как если выборка маленькая, эти методы могут приводить к разным результатам, поскольку информация *a priori* значительно превосходит информацию из выборки (на рисунке 3 распределение *a posteriori* сильно отличается от кривой правдоподобия). Тем не менее в ситуациях максимальной неуверенности использование в качестве начального нейтрального (неинформативного, равномерного) распределения ведет к результатам, совпадающим с результатами классического подхода (на рисунке 4 распределение *a posteriori* и кривая правдоподобия совпадают, так как кривая *a priori* является равномерной). Несмотря на это, часто байесовский метод труден в применении, так как предполагает числовые вычисления и использование метода Монте-Карло. Возможно, его возвращение в наши дни связано с распространением компьютеров.

В отличие от субъективного байесовского подхода, сегодня реабилитирован подход объективный, в котором вероятность *a priori* основывается не на личных убеждениях статистика, а на некоторых регулируемых опорных начальных распределениях. Некоторые статистики считают, что такой путь является лучшим способом объединить байесовский и классический выводы. Действительно, и Байес, и Лаплас использовали данные *a priori* объективные: равномерные распределения. Тем не менее ортодоксальные последователи Байеса считают такой подход нечестным и требуют, как де Финетти и Сэвидж, использования личных вероятностей, надеясь на силу эмпирических данных в отношении нейтрализации разницы в назначении начальной вероятности между различными субъектами, без необходимости вводить дополнительные ограничения кроме состоятельности или соответствия аксиомам математической теории вероятностей. Препятствием является то, что если один человек считает какую-либо гипотезу невозможной, а другой *a priori* присваивает ей какую-либо положительную вероятность, теорема Байеса никогда не сможет привести их к согласию, независимо от того, сколько экспериментов они проведут вместе.

Конечно, сторонники объективного байесовского метода пытаются нейтрализовать этот первоначальный релятивизм

КОГДА КУН ПОЗНАКОМИЛСЯ С БАЙЕСОМ



Сэр Карл Поппер.

Возрождение байесовских методов было во многом связано с модными течениями в философии науки. Различают два класса недедуктивного рассуждения: «индукцию», то есть вывод в условиях неопределенности, и «абдукцию», то есть спекулятивное создание теоретических гипотез для объяснения феноменов. И индукцию, и абдукцию эпистемологи второй половины XX века пытались снабдить вероятностным аппаратом. Первый шаг сделал Рудольф Карнап, философ, принадлежавший к Венскому кружку, учрежденному в США. Он попытался излечить смертельную рану индуктивной логики: тот факт, что надежность индуктивного рассуждения бледнеет перед надежностью дедуктивного. Для этого Карнап создал аксиоматическую теорию доказательств, основанную на наборе правил, которые должны были количественно охарактеризовать индуктивную вероятность или логику гипотезы, то есть вероятность гипотезы H при условии доступных доказательств e. Если $P(H|e) = 1$, он говорил, что e обусловливает H. И наоборот, если $P(H|e) = 0$, то e обусловливает отрицание H. Наконец, если $0 < P(H|e) < 1$, это число отражает степень частичного объяснения H логической структурой e. Эта идея присутствовала в работах Кейнса и байесовского астронома Гарольда Джеффриса, для которых любая индуктивная вероятность являлась условной, связанной с доступными доказательствами. В основном для Карнапа индуктивное доказательство было сродни дедуктивной импликации, но его логика вскоре встретила тяжелые технические и концептуальные препятствия.

Правдоподобие по Попперу

Карл Поппер тоже яростно боролся против этого индуктивизма, разжигая споры. Как и Фишер, он решительно отвергал индуктивное использование вероятности, предлагая понятие *правдоподобия* в качестве замены (хотя определение правдоподобия по Попперу отличается от определения по Фишеру). Поппер по отношению к философам-индуктивистам был тем же, кем Фишер был по отношению к байесовским статистикам. Попытки некоторых философов найти подходящую вероятностную логику для теорий и гипотез провалились; но признание, что наука предполагает субъективные суждения и оценки, как выразился Томас Сэмюэл Кун в работе *«Структура научных революций»* (*The Structure of Scientific Revolutions*, 1962), возродило надежды многих эпистемологов на байесовский вывод.

(который субъективисты устраняют, надеясь на гипотетический будущий общий лимит), вводя различные правила для назначения начальной вероятности, такие как «принцип недостаточного основания» Лапласа (или *«безразличия»*, как его назвал экономист Джон Мейнард Кейнс), который назначает одинаковую вероятность всем неизвестным событиям. Тем не менее, используя только равномерные или полуравномерные распределения, чтобы оставаться объективным, байесовский статистик получает те же самые результаты, что дает классический подход, тогда как для лучших результатов — например, оценки параметров с наименьшей ошибкой — требуется введение *a priori* различных распределений, и в этом случае спор между классическими и байесовскими статистиками возвращается в исходную точку.

ИНДУКЦИЯ, ДЕДУКЦИЯ И РЕШЕНИЕ

Фишер вступал в борьбу не только с байесовской школой. Внутри объективного вывода, благословленного британским статистиком, росла школа, основанная на работах Эгона Пирсона и Ежи Неймана (1894–1981). Этот математик польского происхождения с молодости интересовался применением статистики в сельском хозяйстве. Благодаря гранту он провел год (1925–1926) в лаборатории Карла Пирсона, хотя и был разочарован тем, что английский гигант не был знаком с континентальной абстрактной математикой. Следующий учебный год он решил провести в Париже, ассистируя на лекциях Анри Леона Лебега. И если бы не переписка с Эгоном Пирсоном, Нейман променял бы статистику на интегралы, вернувшись в Варшаву.

Когда Карл Пирсон передал эстафету своему сыну Эгону, тот почти сразу же пригласил Неймана в Университетский колледж. Вместе они создали тандем, выработав новую статистическую парадигму, основанную на критериях значимости, разработанных Фишером: на проверке гипотез, определение которой они совершенствовали в разных статьях с 1928 по 1933 годы, когда они представили фундаментальную лемму, занявшую центральное место в новой теории. На следующий год Нейман переформулировал индуктивную статистику, укрепив оценку с помощью доверительных интервалов (в некотором смысле бывших улучшением фидуциальных интервалов Фишера) и дав начало современной теории выборочного метода — случайной выборке в ее различных формах как основному принципу применения статистики.

> Иногда Фишер публиковал такие обвинения, что только святой мог бы их стерпеть.
>
> **Леонард Джимми Сэвидж (1976)**

Вначале Фишер назвал работы Неймана блестящими и отметил, что Нейман определял вывод в небайесовских терминах (чтение трактата о вероятности, написанного Рихардом фон Мизесом, не сделало его радикальным последователем частотного метода). Но, одновременно со вступлением Неймана в *Королевское статистическое общество* в 1935 году, Фишер резко прервал связь с ним, напав на его сельскохозяйственные исследования и назвав его чистым математиком, без связи с экспериментальной наукой (на это обвинение Нейман ответил, очевидно, не слишком сдерживаясь). В то время Фишер писал, что если нетерпимость к новым идеям является признаком старости, то Карл Пирсон обладал ею еще с молодых лет. Стоит отметить, что сам Фишер страдал от той же проблемы, очень быстро превращаясь в эгоцентричного динозавра от статистики. Он никогда не был особенно щедр к Нейману, несмотря на то что Нейман обожал его, и его теория доверительных интервалов и проверки гипотез разъяснила таинственную фидуциальную вероятность и банальные критерии значимости. Трения между Фишером и Нейманом продолжались в течение их совместной работы под крышей *Университетского колледжа* и не прекратились, когда в 1938 году Нейман отправился в Беркли, в Соединенные Штаты.

Хотя исторически Нейман опубликовал свою теорию доверительных интервалов после теории проверки гипотез, с логической точки зрения они идут в обратном порядке. В 1930 году у Неймана уже был зачаток идеи, возможно появившийся под воздействием фидуциального подхода, который Фишер разрабатывал параллельно (хотя и не сослался на него). Так, в 1934 году он указал, что гораздо более интересной, чем точечная оценка, является оценка с интервалом, внутри которого с определенной уверенностью можно найти значение оцениваемого параметра. Доверительный интервал должен сопровождать точечную оценку, отражая вариабельность оценки. Давать оценку без указания интервала было довольно рискованно. Но, в отличие от традиции давать точечную оценку и возможную ошибку (которая определялась интервалом с уверенностью 50%), Нейман предлагал возможность построить с помощью «осевых» переменных интервалы с любым желаемым уровнем уверенности (например, привычные 95 и 99%). Для любого определенного уровня уверенности можно было вычислить погрешность. Разумеется, при увеличении уверенности увеличивалась погрешность, точно так же как способом увеличить уверенность было увеличение размера выборки.

Например, можно показать, что если мы делаем выборки объемом 16 из нормально распределенной генеральной совокупности с неизвестным средним μ и стандартным отклонением 4, с вероятностью 0,95 выборочное среднее будет отличаться от среднего в генеральной совокупности μ не более чем на 1,96 единицы. Следовательно, если выборочное среднее окажется равным 40, можно ожидать, что μ находится в интервале $40 \pm 1{,}96$ (с 95-процентной уверенностью).

«С 95-процентной уверенностью» — значит, что оценка интервала проводится с помощью метода, который при большом числе повторений дает правильный ответ в 95% случаев. Как если бы этот интервал нам сообщил кто-то, кто говорит правду в 95% случаев; мы можем быть достаточно уверены, но не абсолютно уверены. Следует отметить, как настаивал Нейман, что если I — конкретный 95-процентный доверительный интервал, то нельзя сказать, что вероятность того, что I содержит истинное значение параметра θ, равна 0,95, потому что параметр θ либо содержится, либо не содержится в I, нет других вариантов, так как он является константой с определенным, хотя и неизвестным значением. Другими словами, вероятность того, что I содержит θ, может принимать два значения: 1 или 0, в зависимости от того, содержится θ в I или нет. Формула, с помощью которой был вычислен интервал I при подстановке данных выборки, имеет вероятность 0,95, что интерпретируется, согласно объективному или частотному определению вероятности, как то, что 95% выборок дадут интервал, который в действительности будет содержать значение параметра. Тем не менее невозможно узнать, является ли наш конкретный интервал I одним из них, но мы ожидаем, что это так с 95-процентной уверенностью.

> Вытаскиваю из мешка с черными и белыми фасолинами горсть фасолин и считаю, сколько черных и белых фасолин оказалось у меня в руке; отсюда предполагаю, что в мешке черные и белые фасолины находятся приблизительно в такой же пропорции.
>
> **Чарльз Сандерс Пирс о выборке как основе индукции**

Когда в 1955 году Фишер и Нейман снова скрестили шпаги по поводу провокационной статьи, которую Фишер передал в Королевское статистическое общество, Фишер понял, что концепция Неймана представляет опасность для его фидуциального метода, являясь не просто его вырожденной копией (и это несмотря на то, что фидуциальные интервалы Фишера не совпадали с доверительными интервалами Неймана, когда применялись к непараметрическим проблемам, таким как проблема Беренса — Фишера). Известно, что в ходе необычного рассуждения Фишер изменил состояние параметра θ, сделав для него возможным иметь распределение вероятности. Из константы он превратил его в случайную величину, и это заставило его выйти из парадигмы классической статистики и погрузило в недра статистики байесовской. Потому что для сторонников байесовского подхода было возможным интерпретировать 95-процентный доверительный интервал I — в том смысле, что параметр θ находится там с вероятностью (субъективной) 0,95.

Тогда как сторонники байесовского подхода отвечают на вопрос, почему используется конкретный интервал I, сторонники частотного подхода отвечают на вопрос, почему доверительные интервалы используются в целом, утверждая, что метод Неймана является дедуктивным процессом, который достигает 95% успеха при длительном применении. Уверенность — не окончательная мера точности (связанная с вычисленным числовым интервалом), а начальная.

Проверка гипотез, как мы увидим в дальнейшем, обладает тесной фундаментальной связью с доверительными интервалами. Пытаясь улучшить логическое обоснование критериям значимости Фишера, Пирсон и Нейман придумали множество улучшений. Лейтмотивом их исследований был следующий вопрос: что делать при получении значимого результата статистического теста? Это значило отвергнуть нулевую гипотезу. Но какие гипотезы еще могли быть? В этом смысле тест значимости был хуже, чем бесполезен. Он не давал никакого ответа.

Теория Неймана — Пирсона предполагала реальный выбор между двумя противоположными гипотезами. Проверка гипотез — это алгоритм для выбора одного из двух утверждений о параметре на основании информации, содержащейся в выборке. Одно будет принято, другое будет отвергнуто. После формулировки нулевой гипотезы H_0 формулируется альтернативная гипотеза H_1, которая отличается от исходной. Вслед за этим выбирается желаемый уровень значимости, обозначающий границы, вне которых расхождения считаются «слишком» большими. Как правило, выбирается $\alpha = 0{,}05$ (комплементарным значением обычно является 0,95). Это число определяет допустимый риск, то есть долю выборок, которые мы будем считать значимыми, то есть несовместимыми с нашей нулевой гипотезой (в этом случае — 5%). Таким же образом выбирается статистика T, распределение которой в выборке должно быть известно и которая служит мерой расхождения между нулевой гипотезой, альтернативной гипотезой и данными выборки. С помощью α и T строится «критическая область», или «область отказа», и дополняющая ее «область принятия нулевой гипотезы» (последняя имеет доверительный интервал $1-\alpha$). Попадание значения $T(X)$, наблюдаемого в выборке, в ту или иную область определяет, значима ли наблюдаемая разница, следует ли отклонить нулевую гипотезу и принять альтернативную.

Вся проверка гипотез сводится, таким образом, к принятию или отклонению предложенной нулевой гипотезы (с принятием, в последнем случае, альтернативной гипотезы). Тем не менее могут возникнуть следующие ситуации (схематично представленные в таблице).

а) Принимается верная нулевая гипотеза. Это правильное решение.

б) Отвергается неверная нулевая гипотеза. Это другое правильное решение.

в) Отвергается верная нулевая гипотеза. Очевидно, совершается ошибка, которая называется *ошибкой I рода*. Вероятность совершить эту ошибку определяется уровнем значимости α, заранее выбранным.

г) Принимается неверная нулевая гипотеза. В этом случае тоже совершается ошибка, называемая *ошибкой II рода*. Вероятность сделать такую ошибку обозначается β, и вероятность $1 - \beta$ называется *мощностью теста*, так как выражает вероятность отклонить нулевую гипотезу в том случае, когда она неверна.

	Гипотеза H_0	
Решение	Верна	Неверна
Отклонить H_0	*Ошибка I рода*	Правильно
Не отклонить H_0	Правильно	*Ошибка II рода*

Нейман и Пирсон доказали, что во многих случаях при фиксированной вероятности ошибки I рода α (то есть, предполагая частотную интерпретацию повторяющейся выборки, при фиксированной доле случаев, когда принимается ошибочное решение об отклонении истинной нулевой гипотезы) можно сконструировать и использовать тесты максимальной мощности, то есть тесты, которые уменьшают до минимума вероятность β ошибки II рода, максимизируя одновременно мощность теста и его чувствительность, то есть способность определить истинность нулевой гипотезы. В знаменитой лемме, опубликованной в 1933 году, Нейман и Пирсон доказывают, что в случае простых конкурирующих гипотез (которые присваивают неизвестному параметру конкретное значение) автоматически существует класс оптимальных проверок, небольшого размера и максимальной мощности — основанных на связи правдоподобия (см. Приложение). В 1933 году они писали:

> «Не ожидая, что сможем узнать про каждую гипотезу, истинная она или ложная, мы пытаемся найти правила, которые помогут обращаться с ними таким образом, чтобы ошибаться не слишком часто».

Согласно подходу Неймана и Пирсона, проверка гипотезы — не что иное, как правило принятия решения. Если поступать в соответствии с правилами, то в долгосрочной перспективе истинная нулевая гипотеза будет отклоняться не чаще, чем, скажем, в 5 случаях из 100, и также будет доказательство того, что ложная гипотеза отклоняется достаточно часто. Статистические тесты, однако, не являются правилами индуктивного вывода, а лишь правилами индуктивного поведения. Их цель — не обосновать наши убеждения, но подогнать наши решения к наблюдаемым данным. Невозможно определить, истинна или ложна гипотеза, но можно поступать с ней таким образом, что в долгосрочной перспективе будет сделано не слишком много ошибок. В отличие от Фишера, Нейман и Пирсон считали, что не рассуждение, а действие может быть индуктивным. Объектом статистики является такое использование опыта, которое ведет к корректным действиям. Не больше и не меньше.

Процессы контроля качества в промышленности следуют этому представлению. Так, во время Второй мировой войны проверка гипотез использовалась для отбора продовольствия для американской армии, так как делая выборки из каждой партии, можно было иметь уверенность, что как минимум 95 % партий не были дефектными. Эгон Пирсон написал целую книгу на эту тему, она сгорела во время одного из первых авианалетов на Лондон. Эмиграция Неймана в США в 1938 году облегчила закрепление этого созвездия идей на другой стороне Атлантики; в конце концов оно превратилось в математическую теорию решений, набросанную приблизительно в 1950-м печально известным Абрахамом Вальдом (он погиб в авиакатастрофе).

Во многих случаях Нейман поддерживал привлекательную доктрину, согласно которой индуктивный вывод невозможен и следует довольствоваться индуктивным поведением. Убедительное мнение, которое сделало его неугодным участником философских диспутов о статистике. С его точки зрения, математическая статистика не воздавала должное предполагаемому индуктивному характеру научного поиска, так как ее структура исключительно дедуктивна. Так же как последователи байесовского подхода принимают в качестве исходного пункта априорное распределение вероятности, Фишер всегда исходил из математической функции правдоподобия или выборочного распределения, выведенного заранее. Доверительные интервалы, напротив, получались из рассуждений о свойствах определенных случайных величин. Проверка гипотез — простые правила, в них нет места выводу, ни индуктивному, ни дедуктивному, так как в них присутствует вероятность ошибки. Логика, таким образом, опирается на решение.

Согласно Фишеру, Нейман и Пирсон полностью исказили его изобретение, так как целью критерия значимости, как мы объяснили в главе 3, был не выбор между двумя альтернативными гипотезами, а проверка того, что наблюдение соответствует нулевой гипотезе. Его любимые критерии были

ЖИВОТРЕПЕЩУЩИЕ ВОПРОСЫ ТЕОРЕМЫ НЕЙМАНА — ПИРСОНА

Несмотря на то что проверка гипотез была широко принята, она представляла некоторые технические трудности, которыми не следовало пренебрегать. Во-первых, многие исследователи считали, что для фиксированного значения α отклонение нулевой гипотезы, в тех случаях, когда оно происходит, будет более весомым при увеличении объема выборки n. Однако это неверно. Если мы попытаемся подтвердить, что среднесуточная производительность машины равна 5000 единиц в день, и создадим большую выборку (длинную серию ежедневных наблюдений), довольно вероятно, что мы обнаружим статистически значимую разницу и отклоним нулевую гипотезу о среднем, равном 5000.

Ежи Нейман.

Но заключение может быть и таким, что среднее равно 5000 + 0,00001, совершенно незначимая с практической точки зрения разница. Так как критическая область зависит от размера выборки, значение, за пределами которого отвергается нулевая гипотеза о том, что среднее равно 5000, приближается к 5000, по мере того как увеличивается n (наблюдаемое среднее должно быть очень близко к теоретическому среднему, если выборка большая). Небольшой эффект на большой выборке может быть настолько же весомым, как большой эффект на маленькой выборке. Чтобы избежать этой неприятности, некоторые советуют приводить в со-

превращены в вульгарные рецепты принятия и отклонения. В то время как критерий значимости строился с отправной точкой в виде единственной гипотезы и его целью было подтверждение статистической модели, проверка гипотез рассматривает две противоположные гипотезы, и ее главная цель — выбрать одну из двух.

Кроме того, согласно Фишеру, Нейман и Пирсон формализовали критерий значимости в (предположительно) неясной области, так как результат теста был связан со значением p

ответствие величину критерия и размер выборки. Во-вторых, учитывая привилегированное положение нулевой гипотезы (так как α фиксировано заранее), иногда нулевая гипотеза принимается даже тогда, когда она не слишком соответствует данным. Кроме того, обязательность выбора между нулевой и альтернативной гипотезами иногда толкает к решениям, когда обе гипотезы плохо подходят к данным выборки, что никогда не происходит при байесовском подходе (в Приложении в конце книги мы остановимся на этом вопросе).

Мощность критерия

Нейман подчеркивал, что незначимость критерия в отношении отклонения гипотезы не означает автоматически ее подтверждение, так как это зависит от мощности критерия, от того, насколько она высока. Некоторые статистики считают, что сила, с которой нулевая гипотеза подтверждается выборкой, может быть оценена с помощью величины, называемой «строгость», которая играет роль, аналогичную p-значению. В то время как p-значение определяется, как мы увидели в главе 3, из вероятности $P(T \geqslant T(X) \mid H_0)$, строгость критерия определяется как $P(T \geqslant T(X) \mid H_1)$. Чем выше эта вероятность, тем более «жестким», более строгим является критерий в смысле способности отличить истинную гипотезу от ложной. Эксперимент может подтвердить гипотезу, только когда предпринималась серьезная попытка ее опровергнуть. Наконец, для сложных гипотез фундаментальная лемма неверна и поиск более мощного критерия невозможен, поэтому трудно одновременно контролировать обе вероятности ошибки. Еще Фишер в свое время показал, что, если быть совсем точным, вычисление ошибки второго рода и, соответственно, мощности критерия не всегда возможно, так как альтернативная гипотеза не всегда определяется однозначно.

(которое показывало, до какой степени данные не противоречат нулевой гипотезе), а не с решением принять нулевую или альтернативную гипотезу. Не одно и то же — сообщить значение p, которое отражает доказательства, собранные из выборки, и сообщить о принятии или отклонении нулевой гипотезы, с последующим (ложным) убеждением, что эта гипотеза истинная или ложная просто вследствие того, что противоречит или не противоречит наблюдаемым данным. В самом деле, использование значения p приводит к тому, что все статистики из одной и той же выборки получат одинаковый результат. Наоборот, два статистика, проводящих проверку гипотез на одной и той же выборке, могут прийти к различным результатам, используя разные значения α. Значение p является свойством выборки, тогда как величина α является свойством проверки.

В связи с этим Фишер постоянно утверждал, что интерпретация уровня значимости α как частоты ошибочного решения в повторных выборках из одной и той же совокупности разрушает внутреннюю логику критерия значимости, так как в естественных науках редко есть возможность провести повторную выборку. Аналогия повторных выборок, использованная Нейманом и Пирсоном, работала только в случае проверки гипотез в промышленных условиях с партиями выборок. Кроме того, выражение «ошибка II рода» намекало на возможность ошибочно принять нулевую гипотезу как истинную, когда проведение критерия значимости исключало бы такой шаг.

Разница между двумя теориями была не столько математической, сколько логической и философской. В провокационной статье, представленной Фишером в 1955 году Королевскому статистическому обществу, британский статистик яростно обвинял Неймана в том, что тот поддался «североамериканскому прагматизму», так как казался более озабоченным ускорением производства, а не корректными статистическими выводами. Польский математик неправильно понял статистический вывод, сводя его, как писал Фишер, к сфере рабов Уолл-стрит, а не к свободным ученым в поиске истины. Нейман разрубил гордиев узел логики индуктивного вывода, о котором Фишер говорил как об иллюзорном. Но он перепутал контроль качества с научным выводом, ученого с коммерсантом. «Индуктивное поведение» казалось Фишеру попыткой уйти от проблемы, на самом деле существующей в «индуктивном рассуждении». Фишер хотел не зарабатывать деньги, а, скорее, учиться на экспериментах.

Ответ, который Нейман не замедлил дать, начинался с защиты несчастного Вальда от выпадов Фишера: соотношение между статистическим выводом и теорией принятия решений, придуманной Вальдом, такое же, как между тактикой и страте-

УПОТРЕБЛЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ И ЗЛОУПОТРЕБЛЕНИЕ ИМИ

Существующий методологический синкретизм приводит к многочисленным ошибкам, допускаемым при использовании статистических инструментов. Среди самых частых — следующие.

1. В разведывательном анализе данных часто используется среднее как каноническая мера центральной тенденции, которая группирует наблюдения, хотя медиана в общем случае больше подходит для этой цели, так как обладает меньшей волатильностью, то есть меньше подвержена действию крайних значений.
2. При регрессионном анализе достаточно надежным значением коэффициента корреляции является 0,6, хотя доказуемо, что в этом случае математическая модель объясняет всего 36 % наблюдений.
3. Перманентная иллюзия, плод *пастиша*, будто статистические тесты основываются на силлогизме: «Если нулевая гипотеза верна, то выборка X не может наблюдаться. Мы наблюдаем X, соответственно, исходная гипотеза неверна». Однако тесты основываются на гораздо более вероятном силлогизме: «Если нулевая гипотеза верна, то выборка X в высокой степени маловероятна. Мы наблюдаем X, следовательно, гипотеза в высокой степени неправдоподобна».
4. Священный статус статистической проверки как способа принятия дихотомических решений иногда приводит к тому, что на основании $\alpha = 0{,}05$ принимается нулевая гипотеза при $p = 0{,}051$ и, наоборот, отвергается при 0,049. Так же статистически значимый результат на уровне, скажем, 0,001 обычно интерпретируется, как если бы альтернативная гипотеза имела поддержку 0,999. Но то, что нет доказательств против гипотезы, не значит, что есть доказательства в ее пользу.
5. Путаница между значением p, то есть вероятностью наблюдать существующую выборку при условии истинности нулевой гипотезы, и вероятностью того, что нулевая гипотеза истинна в свете имеющейся выборки (эта вероятность может быть вычислена только с помощью теоремы Байеса). Эта неправильная инверсия терминов известна как *ошибка обвинителя*: если ты виновен, то логично, что все доказательства свидетельствуют против тебя; но тот факт, что все доказательства свидетельствуют против тебя, не значит, что ты виновен, как часто ошибочно делают вывод общественные обвинители.
6. Наконец, часто встречается пренебрежение мощностью критерия. Среди исследователей распространено убеждение, что если критерий оказался незначимым, то нулевая гипотеза является подтвержденной; но это нельзя подтвердить, не вычислив сначала функцию мощности критерия, которая отражает способность выделять различия.

гией. Кроме того, Нейман защищал свой подход с альтернативными гипотезами. Он указывал, что знаменитый эксперимент с женщиной, пробующей чай с молоком (описанный выше), был плохо спроектирован, раз не указывает, по отношению к чему проверялась нулевая гипотеза (то есть не указывает численное значение способности женщины отгадывать, если таковая у нее есть, в виде альтернативной гипотезы). В отношении центральной темы дискуссии Нейман повторял, что индуктивное поведение разрешает раз и навсегда извечную проблему индуктивного вывода.

Со временем польский математик начал говорить об индуктивном поведении — даже в присутствии философа Карнапа — как о важной концепции современной философии науки, находя его корни у Гаусса и Лапласа. В каком-то смысле голоса Неймана и Поппера перекликались, так как оба утверждали, что не существует индуктивного метода рассуждения. Если для Поппера возможный результат эксперимента — опровержение или, наоборот, подкрепление научной теории, то для Неймана это отклонение или принятие нулевой гипотезы (даже если, как в случае Фишера, Поппер едва ли цитировал Неймана).

Даже Эгону Пирсону пришлось вмешаться в спор. Правда, в отличие от Неймана, ему удалось не уйти в область философии, ограничившись тем, что лексикон теории принятия решений принадлежал скорее Нейману, чем ему самому. Гармония между двумя математиками практически исчезла, когда второй отправился в США.

Диспут между Фишером и Нейманом в 1955 году породил целый ряд споров, в которых участвовали не только статистики, но и философы, заинтересованные в научном выводе. Они подчеркивали, что теория проверки гипотез пригодна для проверки, но не для оценки уровня поддержки, который гипотеза получает после проведения эксперимента. Другими словами, классический вывод больше подходит для того, чтобы подвергнуть гипотезу опыту; но после того как заговорит природа, байесовский вывод должен подхватывать эстафету (так как позволяет сравнивать альтернативы посредством их вероятности *a posteriori*).

Тем не менее основной целью проверки гипотез является не измерение степени поддержки, получаемой гипотезами на основании наблюдаемых данных, а оценка расхождений между данными и гипотезами. В классической схеме вероятности вступают в игру как вероятности ошибки, а не как вероятности гипотез. Как и доверительный уровень, вероятность ошибки является начальной мерой точности, а не финальной. Критерии, придуманные Фишером, Нейманом и Пирсоном, не могут стать тем, чем они не являются. Нельзя требовать от них того, что они не могут дать.

И все же сегодня торжествует методологический эклектизм, особенно в социальных науках, где критерий значимости Фишера и проверка гипотез Неймана — Пирсона, и даже иногда байесовские модели, существуют в виде сплава, реализуемого на техническом, но невозможного на концептуальном уровне. В 1960-е годы теории Фишера и Неймана — Пирсона начали молчаливо смешивать в бесформенный гибрид, использование которого превратилось в механический ритуал. Согласно слабой мысли «все подходит!» («любой статистический метод является эффективным инструментом», «не нужно ввязываться в логические изыскания»), скрывается проблема философской глубины с последствиями в виде формирования и интерпретации результатов, потому что упомянуть значение p распределения *a posteriori* или величину критерия, мощность проверки и принятое решение — не одно и то же.

КУРЕНИЕ НАНОСИТ БОЛЬШОЙ ВРЕД ЗДОРОВЬЮ

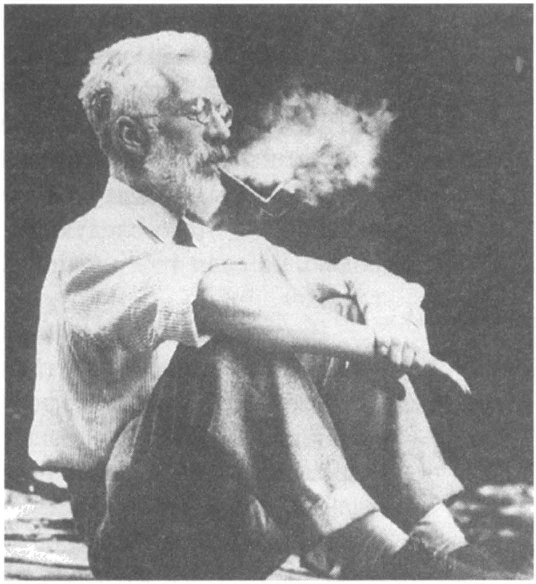
К 1920 году значительно увеличилось количество смертей от рака легких. Хотя уже существовали работы о возможной связи между этим типом рака и курением табака, в 1950-е годы, благодаря работам Ричарда Долла (1912–2005) и Остина Брэдфорда Хилла (1897–1991), этот вопрос пробудил истинный интерес и стимулировал резкие дискуссии в общественности. Эти эпидемиологи стояли во главе распространения принципов экспериментального планирования Фишера в клинических исследованиях.

Долл и Хилл опубликовали статистическое исследование, в котором случаями являлись пациенты, лечившиеся от рака легких, а контрольной группой — пациенты, лечившиеся от других болезней. С помощью анализа клинических записей пациентов, которым уже диагностировали этот вид рака, было установлено, что распространенность рака среди курильщиков в 11–20 раз выше, чем среди некурящих. Этот вывод был статистически значим на уровне 0,001.

Тем не менее эти работы получили множество возражений, в том числе от таких уважаемых людей, как Ежи Нейман. Но может быть, главным паладином критики был сам Фишер (который на множестве фотографий держит в руках трубку). Этот заядлый курильщик работал также консультантом у ряда производителей табака, опубликовал множество статей и одну брошюру, ставя под сомнение взаимосвязь между раком, сигаретами и статистикой.

Одним из препятствий, которым воспользовался Фишер, было то, что исследование показывало больший риск для курильщиков, но это не значило, что причиной является именно табак. Если А и Б прямо коррелируют, это не означает, что А является причиной Б, потому что может быть и наоборот: Б является причиной А, то есть рак легких вызывает привычку курить. Или может существовать третий фактор В, который является общей причиной для А и Б: люди, приобретающие привычку курить, имеют какой-то генетический признак, который делает их склонными впадать в табачную зависимость и одновременно получать рак; эту возможность Фишер обсуждал на основании данных, полученных от близнецов. Британский статистик сравнивал корреляцию, обнаруженную Доллом и Хиллом, с ложной корреляцией, существующей между ростом числа разводов и импортом яблок.

Фишер добавлял, что, в отличие от сельскохозяйственных экспериментов или испытаний вакцин, исследование Долла и Хилла не подходит для планирования эксперимента и является простым проспективным исследованием, так как раз-



ВВЕРХУ СЛЕВА:
Фишер в ходе визита в Индию, в компании К. Р. Рао, индийского статистика, работавшего в Кембридже.

ВВЕРХУ СПРАВА:
Один из споров, в которых участвовал заядлый курильщик Фишер, касался связи между курением и раком легких.

ВНИЗУ:
В 2002 году была открыта мемориальная доска, посвященная Фишеру, в присутствии троих его детей — Джун, Маргарет и Гарри.

деление на две группы (случаи и контроль) не было проведено случайным образом и было подвержено действию трудноустранимых внешних факторов. Кроме того, он подчеркивал, что разделение курильщиков на две группы — вдыхающих и не вдыхающих дым — привело к тому, что неожиданно те, кто не вдыхал дым, были более подвержены раку легких. Фишер выводил из этого следующее заключение: «Дым вредит здоровью, но если тебе нужно курить, то лучше отправлять дым глубже».

В последующие годы было проведено множество проспективных исследований, а также экспериментов на животных, которые подтвердили без тени сомнений тезисы Долла и Хилла (и доказали, что, вопреки ошибочному заключению первого исследования, вдыхание дыма более опасно). По мере накопления доказательств Нейман изменил свое мнение, а позиция Фишера осталась твердой.

СТАТИСТИКА В XXI ВЕКЕ

Рональд Эйлмер Фишер никогда не был статистиком в университете. В 1957 году он решил оставить кафедру генетики в Кембридже, а через два года присоединился в качестве почетного исследователя к научно-промышленному комплексу, связанному с университетом Аделаиды (Австралия). Этот гений духа, которому в 1952 году королева Елизавета II пожаловала титул *сэра*, умер 29 июля 1962 года, в возрасте 72 лет, из-за рака кишечника.

Прогресс, стимулированный Фишером, обеспечил ему почетное место в пантеоне статистиков. Благодаря ему статистика стала основой многих экспериментальных наук. Пока эксперименты дают изменчивые данные, требуется статистика. Каждый научный факт обладает неизбежно статистической природой: речь идет о наборе повторяющихся наблюдений, подверженных факторам и ошибкам случайной природы. Статистика участвует в описании, моделировании, объяснении и предсказании этих данных. И этот процесс обычно состоит из следующих этапов: определение адекватной модели с использованием исчисления вероятностей, планирование эксперимента, описание и анализ собранных выборочных данных, оценка неизвестных параметров модели генеральной совокупности, проверка гипотез о модели, подгонка модели и принятие решений.

> Самое интересное в работе статистика — то, что ты можешь играть в любом саду.
>
> **Джон Тьюки**

Наравне с другими стилями научного рассуждения (геометрический стиль математических наук, гипотетико-дедуктивный — физических наук, экспериментальный — лабораторных наук, таксономический — естественных наук и историко-генетический — гуманитарных наук), существует свой стиль работы и мысли, связанный со статистической наукой, характеризующийся плодородной диалектикой рассуждения и экспериментирования.

Применение статистических методов распространено в самых разных областях, таких как инженерия, экономика, медицина и психология. Сегодня статистические техники используются при работе почтовых фильтров наших компьютеров, при наблюдении за звездными скоплениями, при выявлении налоговых обманов или при анализе инцидентов, таких как авария *Челленджера* в 1986 году, и так далее.

Распространение статистики — Рональд Эйлмер Фишер был важным участником этого процесса — привело не только к тому, что теперь карта лучше передает территорию, но и вследствие этого территория — наш глобализованный мир — трансформируется до неожиданных пределов благодаря применению карты. Мы живем в мире статистики, в котором карта неотличима от реальности.

Приложение

ПРОВЕРЯЯ ФИШЕРА, НЕЙМАНА И БАЙЕСА

Цель приложения — математически показать, что все три статистические школы обладают разными подходами к анализу одного и того же примера. С помощью простого числового примера читатель сможет убедиться, что каждая из этих философий статистики интерпретирует вероятностные вычисления по-своему. Предположим, неизвестный параметр генеральной совокупности θ может принимать только два значения: 0 и 1. Предположим также, что наблюдаемые данные выборки X могут иметь только четыре возможных результата: 1, 2, 3 или 4. Следующая таблица показывает вероятность $P(X|\theta)$ наблюдать каждый результат выборки в зависимости от значения параметра.

$P(X\mid\theta)$	$X=1$	$X=2$	$X=3$	$X=4$
$\theta=0$	0,980	0,010	0,005	0,005
$\theta=1$	0,098	0,900	0,001	0,001

ТЕСТ ЗНАЧИМОСТИ ФИШЕРА

Проверим нулевую гипотезу о том, что $\theta = 0$. По Фишеру, не будем упоминать никаких альтернативных гипотез ($\theta = 1$), так как нашей целью является не выбор между двумя противоположными гипотезами, а обоснование статистической модели, которая предполагает такое значение неизвестного параметра. Как мы помним из главы 3, значение p определялось как вероятность $P(T \geqslant T(X)|H_0)$, что в нашем дискретном случае означает вероятность наблюдать значение, равное или более редкое, чем значение, в действительности наблюдаемое при $\theta = 0$. Держа это в уме, какой вывод мы сделаем, если увидим, что $X = 2$? Согласно логике, глядя на таблицу выше, так как вероятность наблюдать такой результат при предположении $\theta = 0$ очень низка (всего 0,010), значение p должно быть маленьким. И действительно, 0,010 + 0,005 + 0,005 = 0,02, что ниже, чем обычный лимит 0,05, и показывает, что нулевая гипотеза не соответствует наблюдаемым данным и поэтому должна быть отклонена. А если $X = 3$? Тогда значение p равно 0,005 + 0,005 = = 0,01, что приводит к отклонению нулевой гипотезы $\theta = 0$ при большом уровне значимости. Наконец, если $X = 1$ (результат, лучше всего сочетающийся с нулевой гипотезой, так как он наблюдается с вероятностью 0,980), значение p равно 0,980 + + 0,010 + 0,005 + 0,005 = 1, что никоим образом не противоречит нулевой гипотезе. Таким образом, значение p — это математическая мера, которая указывает в критерии значимости, до какой степени выборка опровергает исходную гипотезу, но ничего не говорит о том, в какой степени это позволяет сделать выводы или подтвердить ее.

ПРОВЕРКА ГИПОТЕЗ НЕЙМАНА — ПИРСОНА

Рассмотрим нулевую гипотезу H_0: $\theta = 0$ и противоположную альтернативную гипотезу H_1: $\theta = 1$. Целью проверки является выбор между двумя гипотезами. Интуитивно, глядя на таблицу, если мы видим $X = 1$, мы примем нулевую гипотезу. Наоборот, если мы наблюдаем $X = 2$, мы ее отклоним, приняв альтернативную. Когда $X = 3$ или 4, решение не будет таким простым.

Как объясняется в главе 5, теория Неймана — Пирсона начинается с уравновешивания двух вероятностей ошибки.

Прежде всего фиксируется величина или уровень значимости критерия α, который ограничивает вероятность ошибки I рода (то есть частоту, с которой мы будем принимать ошибочное решение об отклонении истинной нулевой гипотезы). Затем находится критерий с наименьшей вероятностью ошибки II рода (принять нулевую гипотезу при ее ложности) или, что то же самое, самый мощный критерий, то есть с наибольшей вероятностью отклонить нулевую гипотезу, когда она ложная. Как доказали Нейман и Пирсон в знаменитой лемме, оптимальные критерии (с небольшими величинами и максимальной мощностью) основываются на отношении правдоподобия, то есть на отношении $P(X|\theta = 1) / P(X|\theta = 0)$, которое получается делением вероятности (правдоподобия) из таблицы.

	$X=1$	$X=2$	$X=3$	$X=4$
$\frac{P(X\mid\theta=1)}{P(X\mid\theta=0)}$	0,1	90	0,2	0,2

Легко заметить, что отношение правдоподобия приведет к отклонению нулевой гипотезы и принятию альтернативной гипотезы при $X = 2$ (как мы и предсказывали), так как частное принимает очень большое значение (правдоподобие альтернативной гипотезы в 90 раз больше, чем у нулевой). При $X = 1$ мы выберем нулевую гипотезу, так как частное имеет небольшое значение (0,1). И если $X = 3$ или 4, решение будет зависеть от величины, выбранной для критерия, так как выборочные данные плохо подходят к обеим гипотезам (вероятность наблюдать 3 или 4 низкая для обеих гипотез). Можно доказать, что для $\alpha = 0{,}01$ критическая область для H_0: $\theta = 0$ содержит только $X = 2$. Следовательно, для $X = 3$ или 4 оставим нулевую гипотезу. Мощность этого критерия вычисляется из вероятности $P(X = 2|\theta = 1)$ отклонить нулевую гипотезу при верной альтернативной и равна (по данным таблицы) 0,900. Значит, этот критерий обладает большой мощностью — другими словами, большой способностью отличить истинную гипотезу от ложной. В частности, если наблюдается $X = 1$ (незначимый результат), «строгость» критерия вычисляется как $P(T \geqslant T(X)|\theta = 1) = 0{,}900 + 0{,}001 + 0{,}001 + 0{,}098 = 1$, что дает прекрасное доказательство для вывода нулевой гипотезы против альтернативной.

Тем не менее с $\alpha = 0{,}02$ критическая область включает $X = 2$, 3 и 4, поэтому отклоним исходную гипотезу при любых обстоятельствах, даже если нулевая гипотеза будет более правдоподобна, чем альтернативная при $X = 3$ или 4. Как уже сказано, данные выборки 3 и 4 являются редкими событиями при любой из противоположных гипотез, но необходимость выбора в теории Неймана — Пирсона обязывает принять решение. Этот довод последователи байесовского вывода обычно выдвигают против защитников частотного подхода, так как при байесовском подходе, как мы увидим дальше, такое происходит не всегда.

Однако линией защиты классических статистик является апелляция к понятию *строгости*. Таким образом, например, решение о принятии альтернативной гипотезы при $X = 3$ (значимый результат) не указывает на возможность вывода этой гипотезы вне разумного сомнения, так как строгость критерия при H_1 равна (хотя доказательство этой формулы выходит за рамки этой книги) $P(T \leqslant T(X)|\theta = 1) = 0{,}098 + 0{,}001 + 0{,}001 = 0{,}1$, что слишком мало). Строгость критерия низкая, так как мощность очень высокая, с точностью 0,902. Рассмотрим иллюстративный пример, чтобы объяснить, почему существует такая связь: если мы будем использовать сеть, очень хорошо подходящую для рыбной ловли, у нас будет хороший шанс поймать рыбу и, следовательно, отклонить нулевую гипотезу о том, что в озере нет рыбы (высокая мощность); но если ячейки сети позволяют выловить почти все, мы не сможем узнать, велика рыба или мала, и поэтому подтвердить альтернативную гипотезу о размере рыбы в озере (низкая строгость). В сущности, наблюдение выборочного результата 3 приводит к отклонению H_0 (потому что для $\theta = 0$ такое наблюдение маловероятно), но из этого не следует истинность H_1 (при $\theta = 1$, так как и для этого значения подобное наблюдение маловероятно). Внимательный читатель спросит, почему мы не используем типичное значение $\alpha = 0{,}05$. Причина в том, что дискретный пример потребовал бы в этом случае введения «случайного критерия», что усложнило бы рассуждение.

БАЙЕСОВСКИЙ ВЫВОД

Байесовский анализ требует *a priori* установить распределение θ. Следовательно, применяя теорему Байеса (описанную в главе 1), можно скомбинировать эту вероятность *a priori* и правдоподобие и получить таким образом вероятность *a posteriori*, которая сможет сделать выбор между H_0 и H_1. Рассмотрим два различных варианта *a priori*. Первый будет равномерным, то есть нейтральным, неинформативным, присваивая одинаковую вероятность двум возможным значениям θ: $P(\theta = 0) = P(\theta = 1) = 1/2$. Второй, напротив, даст в пять раз большую вероятность значению $\theta = 1$: $P(\theta = 0) = 1/6$; $P(\theta = 1) = 5/6$. Таким образом, для каждого из возможных значений θ вероятность *a posteriori* вычисляется с помощью формулы Байеса, приведенной ниже:

$$P(\theta|X) = \frac{P(\theta) \cdot P(X|\theta)}{P(0) \cdot P(X|0) + P(1) \cdot P(X|1)}.$$

Как можно вычислить, в первом случае, если мы берем равномерное распределение и наблюдаем $X = 1$, вероятность *a posteriori* очевидно склоняется в пользу нулевой гипотезы: $P(\theta = 0|X = 1) = 0{,}91$, тогда как $P(\theta = 1|X = 1) = 0{,}09$. Если мы наблюдаем $X = 2$, вероятность *a posteriori* склоняется, как предсказано, в пользу альтернативной гипотезы: $P(\theta = 0|X = 1) = 0{,}01$ против $P(\theta = 0|X = 1) = 0{,}99$. Но что будет при $X = 3$ или 4 (выборочные значения, представлявшие проблему для классической теории)? При $X = 3$ видно, что правило Байеса склоняется к нулевой гипотезе против альтернативной: $P(\theta = 0|X = 3) = 0{,}83$ и $P(\theta = 1|X = 3) = 0{,}17$. Однако при использовании второго варианта данных *a priori* (который дает больше веса *a priori* значению $\theta = 1$, чем $\theta = 0$) ситуация радикально меняется:

$P(\theta = 0|X = 3) = 0{,}50$ и $P(\theta = 1|X = 3) = 0{,}50$. Равновесие! Как мы видим, выбор данных *a priori* имеет решающее значение в байесовском подходе и смещает равновесие в ту или иную сторону.

КЛАССИЧЕСКИЙ ВЫВОД

Наконец, на другом примере покажем, как классический вывод работает в реальной жизни. Вдохновимся примером, который Фишер взял из знаменитой статьи Стьюдента 1908 года. Изучается эффективность нового снотворного средства и измеряется количество часов сна, которое десять пациентов получают или теряют при использовании этого лекарства по сравнению с отсутствием лекарства. Это называется ***выборкой с парными наблюдениями***, так как сравнения производятся с теми же самыми десятью людьми (если бы речь шла о десяти разных людях в каждом случае, это были бы независимые выборки, которые требовали бы применения более сложных статистических тестов; с помощью парных выборок можно обнаружить эффекты, невидимые для ***независимых выборок***). При использовании средства наблюдаются следующие разницы: +1,2; +2,4; +1,3; +1,3; +0; +1; +1,8; +0,8; +4,6; +1,4. На первый взгляд лекарство является эффективным, но может оказаться, что эффект связан со случаем, а не с дозой. Выборочное среднее X равно +1,58 (что подтверждает наше мнение), но попробуем проверить нулевую гипотезу о том, что среднее в генеральной совокупности μ равно 0, против альтернативной гипотезы $\mu \neq 0$, другими словами, гипотезу о том, что если лекарство будет применено во всей генеральной совокупности, оно не возымеет никакого эффекта, против гипотезы о наличии эффекта.

Предположим, что число дополнительных или потерянных часов сна под действием лекарства следует нормальному распределению с неизвестными средним μ и стандартным отклонением σ. Основываясь на данных выборки, мы хотим оценить среднюю эффективность средства μ во всей генеральной совокупности. Из центральной предельной теоремы мы знаем, что для больших выборок ($n > 30$), в самых общих условиях,

$$\frac{\text{оценка параметра}}{\text{стандартное отклонение оценки}} \sim \begin{matrix}\text{стандартное нормальное} \\ \text{распределение}\end{matrix}.$$

В случае оценки среднего генеральной совокупности μ с помощью маленьких выборок в нормальной совокупности, если бы мы знали стандартное распределение генеральной совокупности σ, мы могли бы использовать нормальное приближение. С уверенностью 95% среднее генеральной совокупности μ отличается от выборочного среднего $\bar{X}$ менее чем на 1,96 стандартного распределения генеральной совокупности σ, поделенные на квадратный корень из размера выборки n. Или, как любил говорить Фишер, только 1 раз из 20 выходит за эти границы, установленные для классического уровня значимости 5%.

Для случаев, когда σ неизвестно (что происходит чаще), астроном Ф.В. Бессель предположил, что σ можно заменить скорректированным стандартным отклонением выборки $\hat{S}$ (квадратным корнем из полудисперсии выборки, определенной в главе 3, и который в нашем случае равен 1,23) и поддаться искушению сказать, что приемлемые значения не превышают

$$\pm 1{,}96 \cdot \frac{\hat{S}}{\sqrt{n}}.$$

Тем не менее эта оценка, будучи успешной в XIX веке, пыталась скорректировать тот факт, что $\hat{S}$ подвержено случайным вариациям выборки, и в некоторых случаях будет больше, а в некоторых меньше σ. Стьюдент первым догадался, что это упущение влияет на выводы на маленьких выборках, отметив, что нормальное распределение (откуда возникает $\pm 1{,}96$) нельзя использовать. Вместо него следовало использовать новое распределение — t Стьюдента, хвосты которого в области крайних значений уменьшаются гораздо медленнее. Следовательно, точность вывода повышалась при использовании значения $\pm 2{,}262$ (на 5-процентном уровне значимости). Интересно отметить, что Стьюдент отправил таблицы своего распределения Фишеру с комментарием: «Возможно, вы единственный человек, который никогда не воспользуется ими». Время показало, опровергнув мнение Карла Пирсона, повсеместность t-распределения Стьюдента, так как его использование обычно эффективно вне зависимости от нормальности первоначального распределения.

Подводя итог: если σ неизвестно, следует использовать приближение, открытое Стьюдентом:

$$\frac{\text{выборочное среднее} - \text{параметр}}{\text{стандартное отклонение выборочного среднего}} \sim t_{n-1} \text{ Стьюдента.}$$

Критерий t касается точности выборочного среднего и позволяет проверить значимость гипотезы о среднем генеральной совокупности. Если наш препарат не обладает никаким эффектом ($\mu = 0$), следует ожидать, что выборочное среднее $\bar{X}$ окажется в интервале

$$\mu \pm 2{,}262 \cdot \frac{\hat{S}}{\sqrt{n}} = 0 \pm 2{,}262 \cdot \frac{1{,}23}{\sqrt{10}} = (-0{,}88, +0{,}88).$$

Так как выборочное среднее равно +1,58, мы можем отклонить нулевую гипотезу: новый препарат является эффективным.

Список рекомендуемой литературы

Bell, E.T., *Los grandes matemáticos*, Buenos Aires, Losada, 2010.

Boyer, C., *Historia de la matemática*, Madrid, Alianza Editorial, 2007.

Fisher, R.A., *Statistical methods, experimental design and scientific inference*, Oxford University Press, 2003.

Grima, P., *La certeza absoluta y otras ficciones*, Barcelona, RBA, 2010.

Hacking, I., *La domesticación del azar*, Barcelona, Gedisa, 1995.

Hald, A., *A History of Mathematical Statistics from 1750 to 1930*, Nueva York, Wiley, 1998.

Peña, D., *Fundamentos de estadística*, Madrid, Alianza, 2008.

Porter, T., *The Rise in Statistical Thinking, 1820-1900*, Princeton University Press, 1986.

Rivadulla, A., *Probabilidad e inferencia científica*, Barcelona, Anthropos, 1991.

Stewart, I., *Historia de las matemáticas*, Barcelona, Crítica, 2008.

Stigler, S., *The History of Statistics*, Harvard University Press, 1986.

Указатель

Наука. Величайшие теории
Выпуск № 47, 2015
Еженедельное издание

РОССИЯ
Издатель, учредитель, редакция:
ООО «Де Агостини», Россия
Юридический адрес: Россия, 105066, г. Москва, ул. Александра Лукьянова, д. 3, стр. 1
Письма читателей по данному адресу не принимаются.

Генеральный директор: Николаос Скилакис
Главный редактор: Анастасия Жаркова
Финансовый директор: Полина Быстрова
Коммерческий директор: Александр Якутов
Менеджер по маркетингу: Михаил Ткачук
Младший менеджер по продукту: Елизавета Чижикова

Для заказа пропущенных выпусков и по всем вопросам, касающимся информации о коллекции, обращайтесь по телефону «горячей линии» в Москве:
☎ 8-495-660-02-02
Телефон бесплатной «горячей линии» для читателей России:
☎ 8-800-200-02-01

Адрес для писем читателей:
Россия, 150961, г. Ярославль, а/я 51, «Де Агостини», «Наука. Величайшие теории»
Пожалуйста, указывайте в письмах свои контактные данные для обратной связи (телефон или e-mail).

Распространение: ООО «Бурда Дистрибьюшен Сервисиз»

Свидетельство о регистрации СМИ в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (Роскомнадзор) ПИ № ФС77-56146 от 15.11.2013

УКРАИНА
Издатель и учредитель:
ООО «Де Агостини Паблишинг», Украина
Юридический адрес:
01032, Украина, г. Киев, ул. Саксаганского, 119
Генеральный директор: Екатерина Клименко

Для заказа пропущенных выпусков и по всем вопросам, касающимся информации о коллекции, обращайтесь по телефону бесплатной горячей линии в Украине:
☎ 0-800-500-8-40
Адрес для писем читателей:
Украина, 01033, г. Киев, а/я «Де Агостини», «Наука. Величайшие теории»
Україна, 01033, м. Київ, а/с «Де Агостіні»

Свидетельство о регистрации печатного СМИ Государственной регистрационной службой Украины
КВ № 20525-10325Р от 13.02.2014

БЕЛАРУСЬ
Импортер и дистрибьютор в РБ:
ООО «Росчерк», 220037, г. Минск, ул. Авангардная, 48а, литер 8/к, тел./факс: + 375 (17) 331 94 41
Телефон «горячей линии» в РБ:
☎ + 375 17 279-87-87
(пн-пт, 9.00–21.00)
Адрес для писем читателей:
Республика Беларусь, 220040, г. Минск, а/я 224, ООО «Росчерк», «Де Агостини», «Наука. Величайшие теории»

КАЗАХСТАН
Распространение:
ТОО «Казахско-Германское предприятие БУРДА-АЛАТАУ ПРЕСС»
Казахстан, г. Алматы, ул. Зенкова, 22 (уг. ул. Гоголя), 7 этаж.
☎ +7 727 311 12 86, +7 727 311 12 41 (вн. 109), факс: +7 727 311 12 65

Издатель оставляет за собой право изменять розничную цену выпусков. Издатель оставляет за собой право изменять последовательность выпусков и их содержание.

Отпечатано в полном соответствии с качеством предоставленного электронного оригинал-макета в ООО «Ярославский полиграфический комбинат»
150049, Ярославль, ул. Свободы, 97
Формат 70 x 100 / 16.
Гарнитура Petersburg
Печать офсетная. Бумага офсетная.
Печ. л. 5,5. Усл. печ. л. 7,128.
Тираж: 20 000 экз.
Заказ № 1514760.

ISSN 2409-0069

12+ Данный знак информационной продукции размещен в соответствии с требованиями Федерального закона от 29 декабря 2010 г. № 436-ФЗ «О защите детей от информации, причиняющей вред их здоровью и развитию».

Коллекция для взрослых, не подлежит обязательному подтверждению соответствия единым требованиям установленным Техническим регламентом Таможенного союза «О безопасности продукции, предназначенной для детей и подростков» ТР ТС 007/2011 от 23 сентября 2011 г. № 797

Дата выхода в России 28.11.2015

Рональд Фишер — ученый, снабдивший статистику инструментами, благодаря которым она обрела то огромное значение, которое имеет сегодня. Его основной вклад — статистический вывод, инновационный подход, связанный с понятием вероятности, который дал статистике, состоявшей прежде на службе других дисциплин, необходимый импульс для того, чтобы она стала полноправной наукой. Этому британскому математику и биологу мы обязаны статистическим методом, который применяется в планировании научных экспериментов. Он был ярым сторонником евгеники, зародившейся в первой половине XX века, и в этом контексте его исследования касались также генетики и современной эволюционной теории.

ISSN 2409-0069

00047

9 772409 006778

12+

Рекомендуемая розничная цена: 289 руб.