ВВЕДЕНИЕ

Эта книга является учебным пособием для студентов-психологов по курсу «Математические методы в психологии». В ней описываются основные математические методы, которые предлагаются математической теорией и которые широко применяются на практике в современных психолого-педагогических исследованиях.

В предлагаемом учебном пособии автор поставил цель доступно познакомить читателей, не имеющих специального математического образования, с основами современного статистического анализа данных. Эта цель определяет характер изложения материала. Изложение ведется практически без строгих математических доказательств, но с подробными обсуждениями, объяснениями и иллюстрациями. Ограничиваясь простыми математическими средствами, мы стараемся по возможности не только описывать конкретные методы статистического анализа, но и разъяснять их сущность и границы применимости.

Имеющиеся в настоящее время руководства по статистическому анализу данных либо требуют достаточно хорошей математической подготовки, либо предназначены для технических приложений, либо не отражают современный уровень статистических методов обработки данных, либо являются для широкого круга читателей малопонятными из-за отсутствия точного описания излагаемых статистических моделей. Эти обстоятельства послужили одной из причин, побудивших автора написать эту книгу.

Книга написана на основе курса лекций «Математические методы в психологии», которые в течение ряда лет автор читал студентам Московского городского психолого-педагогического университета (МГППУ). Необходимость дать студентам сравнительно простое учебное пособие по прикладной статистике было второй причиной появления этой книги.

Поскольку книга адресована психологам и педагогам, то этим определяются не только отбор материала и характер его изложения, но и подбор иллюстрированных примеров. Эти примеры берутся из психологии и педагогики. Однако они носят учебный (гипотетический) характер и не являются результатами каких-либо реальных психолого-педагогических исследований. Приведено большое количество примеров для самостоятельной работы.

Для приведенных в учебном пособии примеров характерен небольшой объем исходных данных. Это сделано не только дня облегчения расчетов вручную, но и потому, что на практике большинство психолого-педагогических исследований проводится именно для малых объемов исходных данных. Решение простых примеров в учебном процессе позволяет наглядно ощутить, как работают на практике статистические методы и подготовить студентов к использованию компьютерных статистических программ. Решение задач вручную возможно во многих случаях лишь при наличии соответствующего набора таблиц математической статистики. Учебное пособие (см. приложение 1) снабжено достаточно полным набором таких таблиц, который включает малодоступные статистические таблицы для основных непараметрических критериев.

В главах 1—2 данного пособия излагаются основные понятия теории вероятностей. Знание психологами элементов теории вероятностей важно потому, что многие психологические признаки и связи между ними описываются вероятностными характеристиками. Далее, в главах 3—13 излагаются конкретные математические методы обработки экспериментальных данных в современной психологии.

Если психологические исследования базируются на некоторых конкретных измерениях того или другого психологического феномена, свойства, характеристики, черты и т. д., то для получения научных и практических выводов из таких измерений необходимо использовать математические методы современного статистического анализа данных. Эти методы позволяют психологу-практику обосновывать правильность используемых методик и приемов, обобщать данные эксперимента, находить зависимости между исследуемыми психологическими признаками, выявлять существенные различия между различными группами испытуемых с точки зрения исследуемого признака, строить прогнозы поведения испытуемых, избегать логических и содержательных ошибок при интерпретации полученных данных и т. д.

При этом необходимо помнить, что математические методы статистического анализа данных — лишь инструменты психологических исследований. Наиболее важным в каждом эксперименте является четкая постановка задачи, детальное планирование опыта, выбор непротиворечивых гипотез и метода исследования, содержательная интерпретация результатов обработки экспериментальных данных, полученных с помощью использования того или иного метода статистического анализа. С этой точки зрения психолог играет ведущую роль. Выработку профессионализма и хорошей интуиции психолога стимулирует его знакомство с основами современного статистического анализа данных. Это позволяет психологу либо самостоятельно, либо совместно с привлекаемыми математиками грамотно ставить задачу психологического эксперимента, выбирать конкретный метод статистического анализа, использовать статистические пакеты программ для персональных компьютеров при реализации выбранного метода анализа, давать содержательную интерпретацию полученных математических результатов анализа данных. Противоречия в психологических и математических выводах свидетельствуют о некорректности решения задачи или некорректности интерпретации результатов.

Множество математических методов современного статистического анализа данных можно разбить на две большие группы. К первой группе относятся методы, предназначенные для анализа данных, имеющих вероятностную (случайную) природу. Они предусматривают вероятностную интерпретацию обрабатываемых данных и полученных в результате обработки статистических выводов. Для понимания таких методов необходимо знание элементов теории вероятностей. Теория вероятностей изучает математические модели, имитирующие механизмы функционирования гипотетического (неконкретного) реального явления, результат которого зависит от влияния большого числа взаимодействующих случайных (не поддающихся строгому учету и контролю) факторов. Такого рода математические модели называют вероятностными моделями. Они используются в тех ситуациях, когда имеется хотя бы принципиально мыслимая возможность приближенного многократного воспроизведения всего комплекса условий, при которых производились измерения анализируемых данных.

Методы первой группы статистического анализа данных — это методы математической статистики. Они позволяют по результатам наблюдений конкретного явления или системы (исходным статистическим данным) строить статистическую модель явления или системы, т. е. такую вероятностную модель, которая в определенном смысле наилучшим образом соответствует исходным статистическим данным.

В большинстве руководств по математической статистике излагаются лишь так называемые параметрические методы, которые предполагают, что исследуемые психологические (случайные) признаки имеют нормальный закон распределения. В данном учебном пособии кроме параметрических методов излагаются и наиболее распространенные на практике непараметрические (ранговые) методы, т. е. методы, свободные от предположений о законе распределения признака и поэтому более универсальные.

Ко второй группе методов современного статистического анализа данных относятся методы, предназначенные для обработки данных, не имеющих вероятностной природы. Другими словами, эти методы используются в ситуации, когда изучаемое явление детерминировано (т. е. не зависит от мешающего влияния случайных факторов), или в ситуации, когда принципиально невозможно многократно повторить опыт хотя бы в приблизительно одинаковых условиях. В подобных ситуациях невозможна вероятностная интерпретация исходных статистических данных и получаемых в результате их обработки выводов. Методы анализа и моделирования данных из второй группы принято называть логико-алгебраическими, поскольку эти методы основаны на обычной логике и используют алгебраический и геометрический подходы. Ко второй группе методов современного статистического анализа данных относятся, например, методы кластерного анализа и многомерного шкалирования.

Если методы первой группы (назовем их вероятностно-статистическими) позволяют решать традиционные для статистики задачи (например, оценка неизвестных параметров, проверка гипотез, моделирование связи изучаемых признаков), то методы второй группы (логико-алгебраические) позволяют решать новые, специфические задачи, такие, как классификация объектов и признаков, сжатие информации, визуализация (наглядное представление) данных, отбор наиболее информативных показателей, включая выявление латентных факторов, моделирование сложных систем (латентно-структурный анализ) и т. д. Настоящее учебное пособие дает достаточно полное представление о современных логико-алгебраических методах. Отметим также, что логико-алгебраические методы на практике могут применяться в комплексе с вероятностно-статистическими методами.

Статистические программные пакеты для персональных компьютеров сделали методы статистического анализа данных более доступными, менее трудоемкими и наглядными.

В приложении 2 учебного пособия даются общие рекомендации по использованию статистических пакетов программ, а также подготовленные В. Т. Бордуковой и В. И. Бордуковой методические рекомендации по использованию пакета STATISTICA. Знакомство с данным учебным пособием позволяет успешно применять математические методы для обработки материалов психолого-педагогических экспериментальных измерений, начиная с постановки задач, выбора метода ее решения, включая использование статистического пакета программ, и кончая интерпретацией результатов анализа и практическими рекомендациями.

В приведенном списке литературы можно найти более глубокое описание приведенных в учебном пособии методов анализа данных, а также некоторые другие, более специальные методы.

Первый вариант этого учебного пособия был издан в 2006 году издательским центром МГППУ при всемерной поддержке ректора МГППУ, академика РАО В. В. Рубцова и первого проректора МГППУ, профессора А. А. Марголиса. Им автор выражает свою глубокую благодарность.

В настоящем издании учебного пособия были устранены замеченные опечатки и внесены небольшие дополнения при изложении некоторых вопросов.

Автор благодарен В. Т. Бордуковой и В. И. Бордуковой за их согласие включить в учебное пособие подготовленные ими методические указания по пакету STATISTICA. Автор также благодарен М. В. Ивановой за большую техническую помощь при подготовке настоящего издания учебного пособия.

ГЛАВА 1

СЛУЧАЙНЫЕ СОБЫТИЯ И ИХ ВЕРОЯТНОСТИ

§ 1. Случайные события и операции над ними

В природе часто наблюдаются явления и события, исходы которых практически однозначно определяются заданными условиями. Такие явления и события принято называть детерминированными или закономерными. Например, при равномерном и прямолинейном движении материальной точки пройденный путь S однозначно определяется формулой $S = v \cdot t$, если задать скорость v и время t. Другой пример дает второй закон Ньютона классической механики: сила, действующая на тело, равна произведению массы тела на сообщаемое этой силой ускорение.

Но бывают такие явления и события, для которых сохранение основных условий опыта не гарантирует однозначность исхода. При подбрасывании монеты нельзя предсказать исход: упадет монета гербом вверх или нет. При измерении одной и той же физической характеристики одним и тем же прибором, в одних и тех же условиях получаем различные результаты. Результаты таких опытов не определяются заранее однозначно в силу влияния большого числа разнообразных причин, не поддающихся строгому учету и контролю. Такого рода явления и события принято называть *недетерминированными* или *случайными*.

Перейдем к более детальному описанию событий с неопределенным исходом.

Будем считать, что проводится реальный или мысленный опыт при не изменяющемся во времени действии большого числа случайных (не поддающихся строгому учету и контролю) факторов, не позволяющих сделать однозначные выводы о том, произойдет или не произойдет интересующее нас событие. При этом предполагается, что имеется принципиальная возможность многократного повторения опыта при одних и тех же условиях. Неконтролируемый (случайный) разброс в результатах наблюдений объясняется многочисленностью, сложностью и неизученностью формирующих их факторов.

Наиболее простые примеры опытов, для которых полностью выполнены описанные условия, дают азартные игры. Действительно, опыты с подбрасыванием монеты, игральной кости или с вытягиванием наугад карты из колоды карт можно многократно повторять в одинаковых условиях, но они не позволяют делать полностью определенные заключения о том, произойдет или не произойдет в результате данного опыта интересующее нас событие, например, появление герба, шестерки или туза пик.

Соблюдение вышеуказанных условий проведения опытов в более серьезных и сложных сферах человеческой деятельности — в психологии, медицине, экономике, технике и т. д. — требует в каждом конкретном случае специального рассмотрения.

С каждым опытом связано множество всех возможных взаимно исключающих исходов опыта. Каждый из этих исходов будем называть *элементарным событием* (или *элементарным исходом*), а множество всех таких исходов — *пространством элементарных событий* (или исходов). Элементарные события будем обозначать ω, а пространство элементарных событий — Ω. Таким образом, например, в результате опыта с конечным числом исходов ω обязательно происходит одно из элементарных событий, причем одновременно с ним не может произойти никакое другое элементарное событие.

Если пространство Ω содержит конечное или счетное число элементарных событий, то Ω называется *дискретным* пространством элементарных событий. В противном случае Ω называется *непрерывным*.

Пусть сначала Ω — дискретное пространство. Тогда элементарные события можно занумеровать. Обозначив их через $\omega_1, \omega_2, \ldots$ $\ldots, \omega_n, \ldots$, получаем, что все пространство $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n, \ldots\}$.

Рассмотрим несколько примеров опытов с дискретными пространствами Ω. В этих примерах под монетой понимается идеально симметричная монета, очень тонкая и однородная по плотности, а под игральной костью понимается геометрически правильный куб с занумерованными гранями, однородный по плотности.

Здесь уместно отметить модельный характер вышеуказанных условий проведения опытов с монетой и игральной костью. Опыты проводятся не с реальными монетами и костями, а с их абстракциями, точнее с их математическими моделями, т. е. такими абстракциями, в которых отношения между реальными элементами заменены подходящими отношениями между математическими категориями. Точность математической модели описания реального явления проверяется практикой.

Как будет ясно из дальнейшего, теория вероятностей изучает вероятностные модели реальных явлений, т. е. математические модели, имитирующие механизмы функционирования таких явлений, на которые существенно влияет большое число случайных факторов.

Примеры.

1) Подбрасывание монеты. Если обозначить через ω_1 выпадения герба *Г*, а через ω_2 выпадение решки *Р*, то $\Omega = \{\omega_1, \omega_2\}$.

2) Выбрасывание игральной кости. Тогда $\Omega = \{\omega_1 = 1,\ \omega_2 = 2,\ \omega_3 = 3,\ \omega_4 = 4,\ \omega_5 = 5,\ \omega_6 = 6\}$.

3) Двукратное бросание монеты. Имеем: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$, где ω_1 — выпадение *ГГ*, ω_2 — выпадение *ГР*, ω_3 — выпадение *РГ*, ω_4 — выпадение *РР*.

4) Двукратное бросание игральной кости. Тогда Ω представляет собой множество $\{\omega_k,\ k = 1, 2, 3, \ldots, 36\}$, где каждое ω_k обозначает некоторую пару чисел (m, n), где $m = 1, 2, 3, 4, 5, 6$, $n = 1, 2, 3, 4, 5, 6$ и m означает количество выпавших очков при первом бросании, а n означает количество выпавших очков при втором бросании.

Кроме элементарных (или неразложимых) событий приходится иметь дело и с составными (или разложимыми) событиями. Составное событие в случае дискретного Ω имеем в том случае, если происходит какое-либо из элементарных событий, причем таких элементарных событий можно указать, по меньшей мере, два. В таком случае говорят, что составное событие состоит, по меньшей мере, из двух элементарных событий.

Случайным событием называется любое подмножество дискретного пространства Ω элементарных событий.

Случайные события будем обозначать буквами $A, B, C, \ldots$. Приведем примеры случайных событий.

Примеры.

1) В опыте с выбрасыванием игральной кости можно говорить о событии A, состоящем в том, что выпало четное число очков, или о событии B, состоящем в том, что выпало нечетное число очков. Тогда $A = \{\omega_2, \omega_4, \omega_6\}$, $B = \{\omega_1, \omega_3, \omega_5\}$.

2) При двукратном бросании монеты можно говорить о событии A, состоящем в том, что выпал хотя бы один раз герб, или о событии B, состоящем в том, что дважды выпал герб или дважды выпала решка. Тогда $A = \{\omega_1, \omega_2, \omega_3\}$, $B = \{\omega_1, \omega_4\}$.

3) При двукратном бросании игральной кости можно говорить о событии A, состоящем в том, что сумма выпавших очков не менее десяти. В этом случае событие A содержит 6 элементарных событий, означающих соответственно выпадение пар чисел (4, 6); (5, 5); (5, 6); (6, 4); (6, 5); (6, 6).

До сих пор мы рассматривали задачи, в которых пространство Ω состояло из не более чем счетного числа элементарных событий.

[. . .]

Введенные операции над событиями допускают геометрическую интерпретацию. Пусть событиями A и B являются попадания соответственно в большой и малый круг (рис. 1).

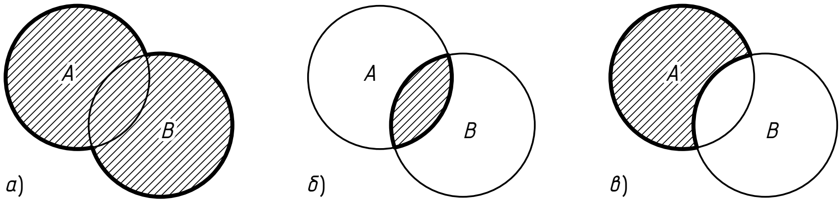

Рис. 1

Тогда событием $A+B$ является заштрихованная область на рис. 1, *а*, событием AB является заштрихованная область на рис. 1, *б*, событием $A-B$ является заштрихованная область на рис. 1, *в*.

Дадим еще некоторые определения.

Все пространство Ω элементарных событий называется *достоверным событием*, а пустое множество $\varnothing$, т. е. множество, не содержащее ни одного элементарного события, называется *невозможным событием*.

Событие $\overline{A}=\Omega-A$ называется *дополнением к A* или *противоположным событию A*. Например, если при двукратном бросании монеты событие A состоит в том, что герб выпал хотя бы один раз, то событие $\overline{A}$ означает, что герб не выпал ни разу. В событии $\overline{A}$ содержатся всевозможные элементарные события, не входящие в A, если Ω — дискретное пространство.

Два события A и B называются *равными* ($A=B$), если A происходит только в том случае, когда происходит B. События A и B в этом случае содержат одни и те же элементарные события для дискретного Ω. События A и B называются *несовместными*, если $AB=\varnothing$. Это значит, что A и B не могут произойти одновременно. В случае дискретного Ω у событий A и B нет ни одного общего элементарного события. Например, если при двукратном бросании игральной кости событие A означает, что сумма выпавших очков является нечетным числом, а событие B означает, что выпали (6,6), то A и B — несовместные события.

Несколько событий $A_1, A_2, \ldots, A_n$ называются *попарно несовместными*, если несовместной является каждая пара из этих событий. Например, все элементарные события являются несовместными.

Несколько несовместных событий $A_1, A_2, \ldots, A_n$ *образуют полную систему событий*, если их сумма является достоверным событием, т. е. $A_1+A_2+\ldots+A_n=\Omega$.

Множество всех элементарных событий, входящих в дискретное пространство Ω, образует полную систему событий. Еще один пример полной системы событий дают события A и B в опыте с подбрасыванием игральной кости, если под событием A понимать выпадение нечетного числа очков, а под событием B — выпадение четного числа очков.

В дальнейшем тот факт, что элементарное событие ω принадлежит пространству Ω элементарных событий, условимся обозначать следующим образом: $\omega \in \Omega$.

§ 2. Вероятности случайных событий

Пусть сначала Ω — дискретное пространство элементарных событий. Чтобы определить вероятность любого случайного события, аксиоматически вводится понятие вероятности элементарного события $\omega \in \Omega$.

Аксиома. Каждому элементарному событию $\omega \in \Omega$ ставится в соответствие некоторая числовая характеристика $\mathsf{P}(\omega) \geqslant 0$ шансов его появления, называемая вероятностью события ω, причем выполняется следующее условие нормировки:

$$\sum \mathsf{P}(\omega) = 1,$$

где знак $\sum$ означает конечную сумму вероятностей всех элементарных событий, если Ω содержит конечное число элементарных событий ω, и знак $\sum$ означает сумму ряда из вероятностей всех элементарных событий, если Ω содержит счетное число исходов ω.

Из приведенной аксиомы сразу следует, что для любого элементарного события $\omega \in \Omega$ вероятность $\mathsf{P}(\omega)$ удовлетворяет двойному неравенству $0 \leqslant \mathsf{P}(\omega) \leqslant 1$.

Вероятностью любого случайного события A называется сумма вероятностей всех элементарных событий, составляющих событие A.

Из этого определения и из аксиомы немедленно следует, что всегда $0 \leqslant \mathsf{P}(A) \leqslant 1$, $\mathsf{P}(\Omega) = 1$, $\mathsf{P}(\varnothing) = 0$ (здесь Ω — достоверное событие, $\varnothing$ — невозможное событие). Кроме того, если A и B — несовместные события, то $\mathsf{P}(A + B) = \mathsf{P}(A) + \mathsf{P}(B)$.

В общем случае из конкретных условий решаемой задачи обосновать определенное введение вероятностей элементарных событий не так-то просто. Это нетрудно сделать в том частном случае, когда

[. . .]

Зная матрицу перехода P, можно найти матрицу перехода $P(n)$ за n шагов по формуле $P(n) = P^n$.

Матрица P является матрицей перехода за один шаг, т. е. $P = P(1)$. Например, если

$$P = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix},$$

то матрица перехода за два шага

$$P_2 = P^2 = \begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix}\begin{pmatrix} 0{,}4 & 0{,}6 \\ 0{,}3 & 0{,}7 \end{pmatrix} = \begin{pmatrix} 0{,}34 & 0{,}66 \\ 0{,}33 & 0{,}67 \end{pmatrix}.$$

Понятие цепи Маркова обобщается на тот случай, когда вместо двух возможных состояний A_1 и A_2 рассматриваются $k \geqslant 2$ несовместных состояний $A_1, A_2, \ldots, A_k$, образующих полную систему состояний.

Задачи к главе 1

1. Брошены две игральные кости. Найти вероятность того, что: а) сумма выпавших очков нечетная, б) произведение выпавших очков больше 15.

2. Из колоды в 36 карт взяты наугад три карты. Найти вероятность того, что это будут король пик и два туза.

3. В коробке 20 одинаковых конфет, причем 15 из них изготовлены на фабрике № 1, остальные — на фабрике № 2. Наугад берут три конфеты. Найти вероятность того, что все эти конфеты изготовлены: а) на фабрике № 1, б) на фабрике № 2.

4. В конверте 7 мужских и 3 женские фотографии, которые неразличимы на ощупь. Из конверта вынимают наугад 4 фотографии. Найти вероятность того, что из них: а) две женские фотографии, б) есть хотя бы одна женская фотография.

5. На полке в случайном порядке расставлено 9 различных книг. Найти вероятность того, что определенные 4 книги стоят рядом в определенном порядке.

6. Найти вероятность того, что дни рождения 12 человек придутся на различные месяцы года.

7. В лифт на первом этаже девятиэтажного здания входят восемь человек. Найти вероятность того, что на одном из этажей из лифта не выйдет ни один человек.

[. . .]