МИНИСТЕРСТВО ОБЩЕГО И ПРОФЕССИОНАЛЬНОГО
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ЧЕЛЯБИНСКИЙ ГОСУДАРСТВЕННЫЙ
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ

Е. В. Яковлев

Математические основы психологии

Учебное пособие

ЧЕЛЯБИНСК — 1999

УДК 15(021)
ББК 88.49
Я73

Яковлев Е. В. Математические основы психологии: Учеб. пособие: – Челябинск: Издательство ЧГПУ, 1999. – 61 с.

В пособии излагаются основные положения и факты теории вероятностей и математической статистики, необходимые психологу-практику.

Работа предназначена студентам, изучающим краткий курс «Математические основы психологии».

Рецензенты: В. А. Долгова, доктор психологических наук
В. М. Ситников, кандидат физико-математических наук

ISBN 5–85716–213–0

© Е. В. Яковлев, 1999
© Издательство ЧГПУ, 1999

Пояснительная записка

Основной целью курса «Математические основы психологии» является ознакомление студентов с методами теории вероятностей и математической статистики. Причём это знакомство не должно ограничиваться изучением отдельных статистических приёмов работы с накопленным психологическим материалом. Более важной является задача обучения принципам построения вероятностно-статистического языка. А именно, студент должен уметь отобразить результаты, полученные внутри абстрактных моделей, на психологическую реальность, и наоборот перевести психологическую проблему на язык математики. Понятно, что для этого требуется и обучение математической символике, и ознакомление с математическими структурами, и накопление определённых навыков построения выводов.

Всё вышесказанное относится в основном к практической психологии. Однако в теории психологии математика играет не меньшую роль. Без понимания математического языка невозможно глубокое изучение теоретической психологии. Наиболее развитые психологические теории существенно опираются на математику, и именно изложение их как математических теорий способствовало их бурному развитию. Это относится в частности к различным направлениям теорий обучения и поведения. Отметим также, что с 1964 года издаётся «Журнал математической психологии» (Journal of Mathematical Psychology), отражающий исследования в этой области.

Данное пособие предназначено в первую очередь студентам-психологам, изучающим курс «Математические основы психологии» по сокращённой программе. Как правило, такая программа предусматривает 10–14 лекционных часов и 10–14 часов семинарских занятий. Это накладывает определённые ограничения на выбор материала. Мы постарались дать представление о наиболее важных разделах данного курса, поделив его на четыре части: основы теории вероятностей, основы математической статистики, основы теории измерений и некоторые методы проверки статистических гипотез. В связи с ограниченным количеством лекционных часов последняя тема может быть вынесена на семинарские занятия или дана для самостоятельного изучения.

Введение

Неоспоримым является тот факт, что в процессе развития любой науки её основания приобретают всё более математическую форму. Предложения, лежащие в основании теории, становятся более чёткими, правила вывода новых фактов приобретают более строгую логическую структуру.

Изложение основ теории на языке математики даёт целый ряд преимуществ и зачастую определяет стратегические направления её развития.

В первую очередь, привлечение хорошо разработанного логического аппарата получения новых высказываний из уже имеющихся позволяет делать некоторые общие выводы чисто вербальными рассуждениями в рамках построенной теории. Даже из сравнительно небольшого набора естественных аксиом может вытекать огромное количество далеко не очевидных следствий, которые трудно получить без привлечения математики. С другой стороны, отдельные утверждения, кажущиеся справедливыми, но сложно проверяемыми, легко отбрасываются после несложного математического анализа.

Кроме того, использование математической теории дисциплинирует исследователя, заставляя оперировать конкретными понятиями, что, в свою очередь, позволяет избежать многочисленных ошибок, связанных с расплывчатостью определений.

Нельзя переоценить роль математики и при создании новых теорий. Даже само построение математической модели, описывающей некоторое явление, подтверждает правильность основ закладываемой теории. Если данная модель хорошо описывает поведение реальных объектов, значит, была правильно понята сущность явления.

С другой стороны, невозможно оценить справедливость какого-либо теоретического положения, если из него не следуют количественные предсказания. Например, теория утверждает, что при соблюдении определённых условий должно наблюдаться некоторое явление. Если в результате опыта наблюдалось данное явление, то это может служить косвенным подтверждением (но не доказательством!) данного теоретического положения. Однако если это явление не наблюдалось, то либо его действительно не было, либо точность измерений была недостаточной. Привлечение количественных теорий позволяет легко решить этот вопрос, поскольку в данном случае будет предсказываться не только проявление данного явления, но и сила его проявления. Своевременное отсечение неверных предсказаний и выводов позволит ещё на ранних этапах отбросить ложные направления развития теории.

Аналогичные преимущества даёт математическая трактовка проблемы выбора между двумя конкурирующими теориями. Математическое их представление позволяет выделить те предсказания одной теории, которые противоречат предсказаниям другой теории. После проведения эксперимента становится ясно, какие предсказания подтверждаются, что даёт возможность отбросить одну из теорий. С другой стороны, такой анализ позволяет выделить одинаковые предсказания конкурирующих теорий. Проверка таких предсказаний не даёт возможность выбрать между этими теориями. Это уберегает исследователя от проведения экспериментов, которые не дают нужной информации.

Более того, взвешенный математический подход позволяет не отбрасывать полностью неудачную теорию, а выделить те её положения, которые несовместимы с опытом. Если удаётся локализовать те положения теории, которые вызвали проблемы, то появляется возможность изменить их не затрагивая теорию в целом. Такой путь развития теории значительно экономнее по времени и даёт существенно более точный результат.

С другой стороны, никакая математика не может спасти несостоятельную психологическую теорию. Она либо не будет адекватно отображать действительность, либо в процессе постоянного устранения недочётов выродится в бесплодную схему, не имеющую прогностической силы.

Изложенные выше моменты относятся к одной стороне использования математики, а именно к части моделирования теорий. Другой стороной, является применение математических методов для описания следствий разрабатываемых теорий и, в частности, для применения достижений теории в конкретных ситуациях, для конкретных целей.

Развитие этих двух подходов применения идей математического моделирования в психологии привело к тому, что выделилось целое направление, называемое математической психологией.

Достижения этой области позволяют предположить, что со временем математическая психология будет составлять основное содержание теоретической психологии.

Практическая же психология изначально опиралась на математические методы и не сможет обойтись без них в дальнейшем. Даже наиболее распространённые в психологическом исследовании тестирование и анкетирование просто бессмысленны без математической обработки полученных результатов.

Однако, огромная роль математики в развитии психологии, а в особенности успехи математического моделирования не означают превращение психологии в отрасль прикладной математики. По целому ряду причин математика не может полностью смоделировать психические процессы. Даже восприятие информации при помощи зрения или слуха не удаётся адекватно представить математической моделью. Тем более, невозможно смоделировать деятельность мозга. Математическое мышление является лишь малой и весьма специфической частью человеческого мышления, и оно никогда не сможет адекватно представить все виды мышления. Причины этого кроятся в самой математике и достаточно сложны для неподготовленного читателя. В силу естественной ограниченности математического мышления принципиально неразрешима и проблема «искусственного интеллекта». «Машина» никогда полностью не заменит человека, как бы этого иногда нам не хотелось. В это трудно поверить, наблюдая гигантские достижения в области компьютерной техники, но, тем не менее, это так. Более того, компьютер не сможет полностью заменить человека даже в области математики. Поскольку даже математическое мышление не может быть реализовано «машиной».

§ 1. Основы теории вероятностей

1.1. Вероятностное пространство

Теория вероятностей — это раздел математики, «выясняющий закономерности, которые возникают при взаимодействии большого числа случайных факторов» (10, с. 113). То есть с точки зрения теории вероятностей нас интересует что произойдёт, или точнее, что может произойти при некотором случайном испытании.

Событие, которое может наступить в результате случайного испытания, называется *случайным событием*.

Рассмотрим простейший пример случайного испытания — подбрасывание монетки. Если отвлечься от чисто гипотетических возможностей — падения монеты на ребро, прилипания к потолку и т.д., то мы получим всего два исхода — выпадение «орла» или выпадение «решки». В условиях данного опыта эти два исхода нельзя разбить на более мелкие. Такие «неделимые» исходы называются *элементарными исходами* или *элементарными событиями*. При бросании игральной кости (кубика) такими исходами являются выпадение определённого числа очков от одного до шести. Выпадение чётного числа очков не является элементарным событием, т.к. охватывает несколько более мелких — выпадение двух, четырёх или шести очков. Но, безусловно, оно является случайным событием.

Совокупность Ω всех возможных элементарных исходов данного испытания называется *пространством элементарных событий (исходов)*.

Описав пространство элементарных событий, мы можем перейти к описанию *всех* возможных случайных событий в нашем испытании. Как мы отметили выше, случайное событие является объединением некоторого числа элементарных событий. То есть, используя терминологию теории множеств, мы можем сказать, что *случайное событие* — это подмножество пространства элементарных событий.

Например, для бросания игральной кости пространство элементарных событий — это $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$, где ω_k — выпадение k очков. Выпадение чётного числа очков — это событие $\omega = \{\omega_2, \omega_4, \omega_6\} \subseteq \Omega$.

Такое определение случайного события корректно, если Ω конечно или счётно, и совершенно не годится в общем случае. Для несчётных множеств нельзя, вообще говоря, построить логически непротиворечивую теорию, назвав случайным событием любое подмножество пространства элементарных исходов. Поэтому для построения общей теории приходится выделять некоторый класс подмножеств **A**, удовлетворяющий определённым условиям.

Отметим, однако, что в реальной практике такие случаи не встречаются, являясь математическими абстракциями. Поэтому в дальнейшем мы будем считать, что случайное событие — это подмножество пространства элементарных исходов.

Следующим важнейшим понятием является вероятность. В математическом энциклопедическом словаре *вероятность* определяется как «числовая характеристика степени возможности наступления какого-либо определённого события в тех или иных определённых, могущих повторяться неограниченное число раз условиях» (10, с. 118).

Рассмотрим сначала так называемое *классическое определение вероятности*. Пусть пространство элементарных исходов состоит из конечного числа равнозначных исходов, т.е.

$\Omega=\{\omega_1, \omega_2, \ldots, \omega_n\}$. Назовём исход благоприятным для интересующего нас события А, если это событие наступило. Например, выпадение двух очков является благоприятным исходом для события А = {выпадение чётного числа очков}.

Вероятность события А в этом случае определяется равенством $P(A) = m/n$, где m — число благоприятных для А исходов, n — число всех исходов.

Например, выпадение чётного числа очков при бросании игральной кости возможно с вероятностью ½, поскольку благоприятных исходов 3 (выпадение 2, 4 или 6 очков), а всего исходов — 6.

В общем случае, вероятность вводится следующим образом. Сопоставим каждому событию А из пространства элементарных исходов Ω число P(A). Мы получим при этом некоторую числовую функцию P(A). Эта функция называется *вероятностью*, если она удовлетворяет следующим условиям:

1) аксиома неотрицательности: $P(A) \geq 0$;
2) аксиома нормированности: $P(\Omega) = 1$;
3) аксиома сложения: $P(A + B) = P(A) + P(B)$, если события А и В *несовместны*, т.е. не могут произойти одновременно.

Отметим, что классическое определение вероятности вытекает из данного сейчас общего определения, предложенного А. Н. Колмогоровым. Действительно, т.к. все исходы равнозначны, то их вероятности должны быть равны, т.е. $P(\omega_1) = P(\omega_2)=\ldots= P(\omega_n)$. С другой стороны, $P(\omega_1+ \ldots + \omega_n) = P(\Omega) = 1$. Таким образом, вероятность каждого элементарного исхода ω_k равна $P(\omega_k) = 1/n$. Следовательно, если событие А состоит из m элементарных событий, то $P(A) = m/n$.

Набор (Ω, **A**, P), где Ω — пространство элементарных исходов, **A** — множество событий, P — вероятность называется *вероятностным пространством*.

Определив для каждого конкретного эксперимента (например, бросания монеты) вероятностное пространство, мы можем забыть о конкретной природе событий и исследовать их чисто математическими методами. Поскольку число элементарных исходов несущественно для рассуждений в рамках теории вероятностей, мы получаем необходимые нам сведения, исследовав одну математическую модель.

1.2. Случайная величина и её основные характеристики

Одним из важнейших понятий теории вероятностей является понятие случайной величины.

Под *случайной величиной* будем понимать некоторую переменную величину, принимающую в зависимости от случая те или иные числовые значения. То есть случайная величина — это функция, ставящая в соответствие каждому элементарному исходу ω некоторое число $\xi(\omega)$.

Например, число очков выпавших при бросании игральной кости является случайной величиной. Опишем её: $\xi(\omega_1) = 1, \ldots, \xi(\omega_6) = 6$. Для того же бросания кости можно определить другую случайную величину. Предположим, что двое играют в такую игру: если при бросании игральной кости выпадает чётное число очков, то первый игрок отдаёт один рубль второму игроку, а если нечётное, то второй отдаёт один рубль первому. Опишем случайную величину, равную выигрышу первого игрока: $\xi(\omega_1) = \xi(\omega_3) = \xi(\omega_5) = 1$, $\xi(\omega_2) = \xi(\omega_4) = \xi(\omega_6) = -1$.

Таким образом, в рамках одного вероятностного пространства можно определить бесконечно много различных случайных величин, описывающих разные случайные процессы.

Основной характеристикой случайной величины является её функция распределения.

Пусть ξ — некоторая случайная величина, принимающая действительные значения. Тогда *функцией распределения* случайной величины ξ называется функция $F(x)=P(\xi < x)$, т.е. значением функции распределения является вероятность того, что случайная величина ξ принимает значение меньшее x. Значение x, при котором $F(x) = \alpha$ для фиксированного значения вероятности α называется *квантилью*.

Обычно выделяют дискретные и непрерывные случайные величины. *Дискретные* величины характеризуются тем, что можно перенумеровать все их возможные значения. Число очков при бросании игральной кости является как раз примером дискретной случайной величины. *Непрерывные* же случайные величины могут принимать любые значения из некоторого промежутка на множестве действительных чисел. Так время, затраченное на выполнение какой-либо работы, является непрерывной случайной величиной. Строгое определение непрерывной случайной величины гласит: «Случайную величину, принимающую вещественные значения, называют непрерывной, если непрерывна её функция распределения» (18, с. 26). Отсюда следует, что для непрерывных распределений вероятность каждого отдельного значения случайной величины равна нулю. Для дискретных же распределений все такие вероятности принимают некоторые положительные значения. Именно в этом и кроется глубинный смысл противопоставления непрерывных и дискретных случайных величин.

Если функция распределения случайной величины непрерывна и дифференцируема, то закон распределения описывают также при помощи *плотности вероятности* $p(x)$, которая может быть определена как производная функции распределения.

В этом случае функция распределения имеет вид

$$F(x) = \int_{-\infty}^{x} p(y)dy.$$

Функция распределения даёт исчерпывающую информацию о поведении случайной величины. Однако для практических целей, как правило, достаточно знания нескольких числовых характеристик распределения, позволяющих оценить такие его свойства как центр группирования значений, разброс значений вокруг этого центра и т.д. Важнейшими из таких характеристик являются *моменты* распределения. Мы обсудим только два из них: математическое ожидание (первый момент) и дисперсию (второй момент).

Для дискретной случайной величины ξ, принимающей значения $x_1, x_2, x_3, \ldots, x_n$ с вероятностями $p_1, p_2, p_3, \ldots, p_n$, *математическое ожидание* определяется по формуле

$$M\xi = \sum_{k=1}^{n} x_k p_k .$$

То есть математическое ожидание — это некоторое «осреднение» случайной величины с учётом вероятности получения отдельных значений. Для непрерывной случайной величины такое осреднение производится при помощи интеграла. Более точно, математическое ожидание непрерывной случайной величины равно

$$M\xi = \int_{-\infty}^{+\infty} xp(x)dx,$$

где *p(x)* — плотность вероятности.

Величина разброса случайной величины вокруг среднего значения характеризуется дисперсией.

Дисперсия случайной величины ξ определяется по формуле

$$D\xi = M(\xi - M\xi)^2 \text{ или } D\xi = M\xi - (M\xi)^2.$$

Дисперсию часто обозначают через σ^2 (сигма в квадрате), где σ означает *среднеквадратичное отклонение*. То есть среднеквадратичное отклонение — это корень квадратный из дисперсии. Эту характеристику используют, когда нужно, чтобы показатель разброса случайной величины выражался в тех же единицах, что и сами значения случайной величины.

1.3. Примеры распределений

Вернёмся теперь от числовых характеристик распределений случайных величин к законам их распределения. Очевидно, что существует бесконечное число различных функций распределения, а значит и бесконечное число самих вероятностных распределений. Однако среди этого великого многообразия можно выделить такие распределения, которые встречаются на практике наиболее часто. Многие из них лежат в основе целых областей знания, таких как теория массового обслуживания, теория надёжности, теория измерений, теория игр и т.п. Мы бегло рассмотрим два распределения, которые играют особую роль в теории вероятностей.

Одним из самых распространённых дискретных распределений является *биномиальное распределение*. Оно возникает в тех случаях, когда нас интересует, сколько раз происходит некоторое событие в серии из определённого числа независимых наблюдений (опытов), выполняемых в одинаковых условиях. Последовательность независимых испытаний, в которых результатом каждого из испытаний может быть один из двух исходов, например, «успех» и «неуспех», и вероятность «успеха» в каждом из испытаний одна и та же, называется *схемой Бернулли*.

Итак, предположим, что случайная величина ξ равна числу «успехов» в n испытаниях Бернулли, причём вероятность успеха в каждом испытании равна p. Тогда ξ принимает значение k с вероятностью

$$P(\xi = k) = C_n^k p^k (1-p)^{n-k},$$

где C_n^k — число сочетаний из n элементов по k элементов. Эта формула и описывает закон *биномиального распределения* с параметрами n и p.

Математическое ожидание и дисперсия случайной величины, имеющей биномиальное распределение, равны $M\xi = np$ и $D\xi = np(1-n)$ соответственно.

На рисунке 1 показаны вероятности $P(\xi = k)$ при $n = 10$ для различных значений p.

Для биномиального распределения существуют два типа таблиц данных. В таблицах первого типа приводятся вероятности $P(\xi = k)$, а в таблицах второго типа вероятности $P(\xi \le k)$ при различных значениях n и p. Достаточно полные такие таблицы с подробными комментариями можно найти, например, в (8) или (11).

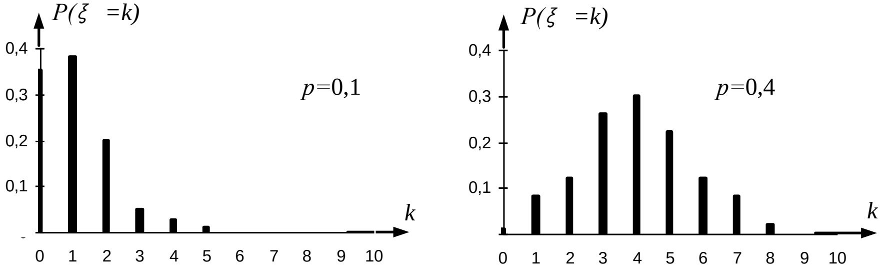

Рис. 1. Вид биномиального распределения
для p = 0,1 и p = 0,4 при n = 10

Важнейшим среди непрерывных распределений является *нормальное (гауссовское) распределение*. Это распределение возникает, когда «значение непрерывной случайной величины формируется под воздействием очень большого числа независимых случайных факторов, причём сила воздействия каждого отдельного фактора мала и не может превалировать среди остальных, а характер воздействия — аддитивный (т.е. при воздействии случайного фактора F на величину a получается величина $a + \Delta F$, где случайная «добавка» ΔF мала и равновероятна по знаку)» (7, с. 169).

Если функция распределения случайной величины ξ задаётся формулой

$$F(x) = \frac{1}{\sqrt{2\pi}\cdot\sigma}\int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}}\,\mathrm{d}t,$$

то говорят, что случайная величина ξ имеет *нормальное распределение* с параметрами a и σ^2. Функцию $F(x)$ часто называют *функцией Лапласа*. Понятно, что плотность нормального распределения определяется формулой

$$p(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(t-a)^2}{2\sigma^2}}.$$

Математическое ожидание и дисперсия в этом случае равны соответственно $M\xi = a$ и $D\xi = \sigma^2$. Если $a = 0$ и $\sigma = 1$, то распределение называют *стандартным нормальным распределением*.

Смысл параметров нормального распределения хорошо виден из рисунка 2.

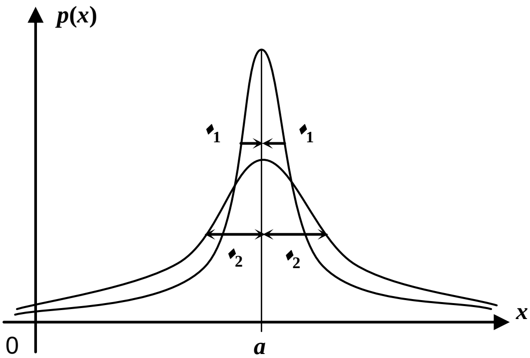

Рис. 2. Плотность нормального распределения со средним *a* и различными значениями дисперсии σ^2

Для функции распределения и для плотности вероятности нормального распределения существует множество таблиц разной степени точности (см., например, (8; 11; 20)). В практических вычислениях наиболее удобны таблицы для плотности и таблицы квантилей.

При операциях со случайными величинами возникает несколько новых видов распределений. В первую очередь это распределение Стьюдента, распределение хи-квадрат и F –распределение. Эти распределения играют важную роль в статистическом анализе и широко используются. Мы не будем приводить здесь описание перечисленных распределений, так как для практического использования требуются лишь конкретные значения соответствующих функций распределения, которые можно найти в многочисленных таблицах (см., например, 8; 9; 11).

При использовании нормального закона распределения следует учитывать следующее важное обстоятельство. Нормальное распределение не является универсальным распределением, описывающим любые явления. Это лишь один из многих типов распределения, имеющихся в природе, хотя и наиболее распространённый. Однако нормальный закон распределения имеет понятные математические свойства и хорошо изучен, что делает его очень популярным. Кроме того, с помощью нормального закона несложно описываются многие другие распределения. В одних случаях точно, а в других — с некоторой точностью приближения.

Использованию нормального распределения для приближённого описания распределений случайных величин не препятствует то обстоятельство, что эти величины могут принимать значения из некоторого ограниченного интервала (например, возраст от 0 до 100 лет), а нормальное распределение не может быть сосредоточено ни на каком конкретном интервале. Дело в том, что вероятность больших отклонений нормальной случайной величины от центра распределения практически равна нулю, и ею можно пренебречь. Этот факт хорошо иллюстрирует график плотности нормального распределения (см. рис. 2).

1.4. Предельные теоремы

Предположим, что мы произвели достаточно большое число n испытаний Бернулли. Подсчитаем число тех опытов, в которых произошло интересующее нас событие, пусть это будет *m.* Тогда частота $\nu = m/n$ нашего события в *n* повторениях приблизительно равна вероятности наступления этого события. Причём эта оценка тем точнее, чем больше опытов мы провели. Это несложное, на первый взгляд, замечание составляет основное содержание исторически одной из первых теорем теории вероятностей — *теоремы Бернулли*, доказанной в конце семнадцатого века.

В более точной формулировке она звучит так: вероятность того, что частота успеха ν отличается от вероятности успеха *p* более чем на ε, стремится к нулю с ростом числа испытаний, т.е. $\mathrm{P}\{|\nu - p| > \varepsilon\} \to 0$.

Теорему Бернулли называют также *слабым законом больших чисел*.

Сам *закон больших чисел* носит более общий характер и утверждает следующее.

Если последовательность $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимых одинаково распределённых случайных величин такова, что существует их математическое ожидание *a*, то для любого $\varepsilon > 0$ вероятность того, что среднее арифметическое значение случайных величин отличается от математического ожидания *a* более чем на ε, стремится к нулю с ростом *n*, т.е.

$$P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}\xi_i - a\right| > \varepsilon\right\} \to 0.$$

Другими словами, при достаточно большом количестве случайных величин их математическое ожидание приблизительно равно среднему арифметическому.

Следующей важнейшей теоремой теории вероятностей является так называемая *центральная предельная теорема*.

Если последовательность $\xi_1, \xi_2, \ldots, \xi_n, \ldots$ независимых одинаково распределённых случайных величин такова, что существует их математическое ожидание a и дисперсия σ^2, то для любого $\varepsilon > 0$

$$P\left\{ \frac{\sum_{i=1}^{n} \xi_i - na}{\sqrt{n\sigma^2}} < x \right\} \xrightarrow[n\to\infty]{} F(x),$$

где $F(x)$ — функция стандартного нормального распределения.

То есть после соответствующей нормировки сумма достаточно большого числа одинаково распределённых независимых случайных величин приближённо описывается нормальным законом распределения. Таким образом, центральная предельная теорема выявляет ту роль, которую играет нормальное распределение. Оно обычно возникает в явлениях, подверженных большому количеству малых случайных воздействий.

§2. Основы математической статистики

2.1. Общие вопросы

Математическая статистика, как научная дисциплина, представляет собой «раздел математики, посвящённый математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов» (10, с. 344). При этом под *статистическими данными* понимается любая числовая информация, характеризующая некоторую совокупность объектов, обладающих теми или иными общими признаками. Например, количество учащихся в классе, имеющих оценку «5» по геометрии, или процент студентов, получающих стипендию и т.д.

Статистический подход к изучению каких-либо процессов или явлений состоит в мысленном расчленении происходящих изменений на две части — закономерную и случайную и выявлении закономерной изменчивости на фоне случайной. Изучая, как и теория вероятностей, случайные явления, математическая статистика использует те же определения, понятия и методы, однако, решает другие задачи. Теория вероятностей исследует явления, *полностью* заданные их моделью и выявляет ещё до опыта те статистические закономерности, которые будут иметь место после его проведения. В математической статистике вероятностная модель явления определена с точностью до неизвестных параметров. Отсутствие сведений об этих параметрах компенсируется возможностью проводить необходимые испытания и на их основе восстанавливать недостающую информацию. То есть в некотором смысле задачи математической статистики являются обратными к задачам теории вероятностей.

В математической статистике принято выделять два основных направления исследований.

Первое направление связано с оценкой неизвестных параметров. Изучение данных, полученных опытным путём, позволяет дать приближённые оценки неизвестных характеристик изучаемого объекта.

Второе направление связано с проверкой гипотез, выдвинутых до опыта. В рамках этого направления разрабатываются критерии, на основании которых по результатам проведённого эксперимента гипотеза либо принимается, либо отвергается.

Основная ценность методов математической статистики для психолога-исследователя состоит в возможности делать некоторые выводы обо всей совокупности объектов, используя данные о сравнительно небольшой группе из этой совокупности.

Основными понятиями математической статистики являются понятия генеральной совокупности, выборки и теоретической функции распределения.

Генеральной совокупностью называется множество объектов, каждому из которых присуще определенное значение некоторой числовой характеристики. Поскольку для статистических исследований важны лишь численные значения этой характеристики, математическая статистика абстрагируется от физической природы самих объектов. Поэтому с точки зрения теории генеральная совокупность представляет собой набор чисел.

Для установления параметров генеральной совокупности проводится некоторое число испытаний. Каждое испытание состоит в том, что случайным образом выбирается один объект генеральной совокупности и определяется значение интересующей характеристики. В результате получается набор чисел, который называется *выборкой* из данной генеральной совокупности.

Рассматривая значение выборки как случайную величину, мы можем рассмотреть её функцию распределения, которая в этом случае называется *теоретической функцией распределения*. Всё это даёт возможность привлечь для исследования аппарат теории вероятностей.

2.2. Оценивание неизвестных параметров

Поскольку в статистических расчётах мы имеем дело с данными, подверженными случайной изменчивости, невозможно точно определить характеристики исследуемых случайных процессов. Это вынуждает прибегать к некоторым их оценкам, т.е. к приближённым значениям. Поэтому для статистики большую роль играют так называемые *выборочные* или *эмпирические* значения характеристик.

Как следует из теоремы Бернулли, вероятность наступления события приближенно равна частоте наступления этого события в некоторой конечной серии опытов. Следовательно, в качестве оценки для вероятности некоторого события естественно взять частоту наступления этого события.

Оставив в стороне вопросы точности таких оценок, опишем наиболее распространённые выборочные характеристики.

Учитывая то, что вероятность оценивается при помощи частоты наступления события, *выборочную функцию распределения* можно найти по формуле

$$\hat{F}(x) = \frac{f(x)}{n},$$

где $f(x)$ — число наблюдённых значений случайной величины ξ меньших x и n — объём выборки (т.е. число испытаний).

Известная теорема Гливенко-Кантелли утверждает, что при увеличении объёма выборки выборочная функция распределения приближается к теоретической функции распределения. Заменяя теоретическую функцию распределения её выборочным аналогом, мы получим выборочные значения остальных характеристик.

Так *выборочное математическое ожидание* определяется по формуле

$$\hat{M}\xi = \bar{x} = \frac{1}{n}\sum_{k=1}^{n} x_k.$$

Таким образом, математическое ожидание можно оценить при помощи среднего арифметического значений случайной величины ξ .

Выборочная дисперсия может быть найдена по формуле

$$\hat{D}\xi = \frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2,$$

где $\bar{x}$ — среднее значение наблюдений.

Следовательно, *выборочное среднеквадратичное отклонение* вычисляется по формуле

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$

Несколько более точные оценки дисперсии и среднеквадратичного отклонения получаются по формулам

$$\hat{\sigma}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2, \qquad \hat{\sigma} = \sqrt{\frac{1}{n-1}\sum_{k=1}^{n}(x_k - \bar{x})^2}.$$

Такие оценки являются *несмещёнными* в том смысле, что математическое ожидание от полученных по ним значений равно дисперсии и среднеквадратичному отклонению соответственно.

Аналогично можно оценить и другие параметры распределения изучаемой случайной величины.

Полученные таким образом оценки являются *точечными* в том смысле, что они оценивают неизвестный параметр одним числом или точкой. Однако, как мы знаем, точечная оценка не совпадает с оцениваемым параметром, поэтому бóльший интерес представляет указание области, в которой этот параметр находится с вероятностью не меньше заданной. Строится такая область следующим образом. Сначала выбирается число α, лежащее в пределах от нуля до единицы. Это будет вероятность, с которой параметр должен попасть в построенную нами область. Как правило, в качестве такой вероятности берут значения 0,9; 0,95 или 0,99 (т.е. вероятности 90%, 95% или 99%). Используя выборочную функцию распределения полученной точечной оценки параметра, мы можем найти интервал, в который этот параметр попадёт с заданной вероятностью. Такой интервал называется *доверительным интервалом* с уровнем доверия α.

2.3. Теория статистического вывода

Как мы уже отмечали в начале параграфа, одной из задач статистики является проверка гипотез о свойствах всей совокупности объектов на основании данных о некоторой её части. Решению этой важной задачи посвящён целый раздел математической статистики, называемый *теорией статистического вывода*.

Основная идея этой теории заключается в том, что некоторое множество объектов, выбранных случайным образом из общей совокупности объектов, имеет те же свойства, что и вся исходная совокупность. При этом естественно такое рассуждение имеет вероятностный характер. А, значит, оно неверно с некоторой вероятностью. Если вероятность справедливости некоторого утверждения близка к единице, то оно практически *достоверно*, и мы можем принять данное утверждение. Например, вероятность 0,99 означает, что для одного объекта из ста наш вывод может оказаться ошибочным. Следовательно, определив для себя приемлемый уровень ошибки, мы можем делать какие-либо выводы относительно исследуемых объектов.

Понятно, что для проверки некоторой гипотезы о реальных объектах средствами математической статистики мы должны, прежде всего, перевести наши знания об объекте на язык математики, т.е. другими словами, построить математическую модель. После этого гипотезу о свойствах объекта следует сформулировать как гипотезу о свойствах распределения некоторой случайной величины. Например, гипотеза об одинаковом уровне подготовленности учащихся двух классов может быть сформулирована как гипотеза о принадлежности оценок учащихся этих классов одному закону распределения.

Опишем общую схему применения статистических критериев для проверки гипотез.

Итак, предположим, что мы построили математическую модель нашего объекта.

Прежде всего, необходимо перевести проверяемую гипотезу на язык математики, т.е. выдвинуть *статистическую гипотезу*. Это может быть гипотеза о законе распределения некоторой случайной величины, или гипотеза о числовых значениях параметров исследуемой совокупности объектов, или гипотеза об однородности нескольких групп данных и т.д. Если мы можем непосредственно проверить выдвигаемую гипотезу, то не возникает никаких проблем. Однако чаще всего приходится проверять не саму гипотезу, а некоторые обязательные следствия из неё. При этом если следствия не выполняются, мы делаем вывод о ложности гипотезы. Однако, если следствия выполняются, мы не можем утверждать, что гипотеза справедлива. То есть косвенным образом гипотезу *доказать* нельзя, а можно лишь *опровергнуть*. Поэтому, чтобы расширить круг применяемых критериев, принято формулировать так называемую «*нулевую*» гипотезу, ложность которой означала бы истинность нашего предположения. Например, вместо проверки неоднородности групп по какому-либо показателю, мы будем проверять их однородность. Тогда отвержение второй гипотезы будет означать справедливость первой. Нулевая гипотеза, как правило, носит более общий характер и поэтому отвергнуть её бывает легче, чем подтвердить альтернативную гипотезу.

Итак, будем считать, что мы выдвинули статистическую гипотезу. Традиционно проверяемую гипотезу обозначают через H_0, а конкурирующую с ней альтернативную гипотезу — через H_1. Поскольку при проверке любой статической гипотезы решение всегда принимается с некоторой долей вероятности, мы должны определить для себя приемлемую точность вывода, то есть задать степень риска получить неверный вывод. Этот риск, представленный как вероятность, называется *уровнем значимости* и обычно обозначается через α. Другими словами, уровень значимости — это вероятность отвержения выдвинутой гипотезы в случае, когда она на самом деле верна. Выбор величины уровня значимости α зависит от величины потерь, которые мы понесём в результате неверно принятого решения. Однако, поскольку такое сопоставление в практических задачах бывает весьма затруднительным, прибегают к некоторым стандартным уровням значимости. Например, α берут равным 0,1; 0,05; 0,025 и т.д. Так уровень значимости 0,01 означает, что в одном случае из ста мы будем ошибочно отвергать высказанную гипотезу при использовании данного статистического критерия. Наиболее распространённым является уровень значимости 0,05, так как, с одной стороны, его достаточно для большинства задач, а с другой стороны, такая точность согласуется с потерями при моделировании реальных объектов.

После выбора уровня значимости мы должны составить некоторую функцию от результатов наблюдения, которая сама является случайной величиной и при справедливости гипотезы H_0

должна иметь известное нам распределение. Например, если мы предполагаем, что средние значения в двух группах данных равны, то отношение оценок дисперсий, вычисленных двумя способами, имеет распределение Стьюдента. На этом факте основаны некоторые методы дисперсионного анализа.

Для выбранного уровня значимости мы находим так называемое *критическое* значение выбранной функции. Эти значения приведены в таблицах соответствующих распределений и их нахождение, как правило, не требует никаких дополнительных вычислений. Найденное критическое значение выделяет *критическую область* (или *область отбрасывания*).

Наконец, вычисляем значение выбранной функции для полученных экспериментальных данных и сравниваем его с критическим. Если полученное значение попадает в критическую область, то мы считаем результаты наблюдений несовместимыми с гипотезой и поэтому отвергаем её и принимаем альтернативную гипотезу с вероятностью $1 - \alpha$. При этом мы предполагаем, что альтернативная гипотеза H_1 является отрицанием основной гипотезы H_0. Этот случай достаточен для целей психологического эксперимента.

Поясним более общую постановку проблемы на следующем примере. Предположим, что гипотеза H_0 состоит в том, что средний возраст студентов вуза равен 19 годам. Отрицанием этой гипотезы будет гипотеза о том, что средний возраст не равен 19 годам. Однако в качестве альтернативной гипотезы можно выдвинуть предположение о том, что средний возраст равен 20 годам. Никак не сказываясь на общей схеме применения критерия, это различие выявится лишь при вычислении вероятности справедливости альтернативной гипотезы. Примеры вычисления таких вероятностей можно найти, скажем, в (14, с. 258). Отметим лишь, что вероятность зависит от конкретного вида альтернативной гипотезы. Так, если альтернативная гипотеза состоит в том, что средний возраст равен 19,5 лет, вероятность будет уже другой.

Из приведённой схемы применения статистического критерия следует возможность совершения ошибок двух видов:

- отвержение гипотезы в случае, когда она на самом деле верна;
- принятие гипотезы в случае, когда она неверна.

Эти возможности называются соответственно ошибками *первого* и *второго рода.* Вероятность ошибки первого рода равна выбранному уровню значимости α. Вероятность ошибки второго рода обозначают через β, и в общем случае $1 - \beta$ не обязательно равно α. Величина $1 - \beta$ называется *мощностью критерия* и зависит от альтернативной гипотезы. Понятно, что чем ближе к единице мощность критерия, тем он эффективнее.

Все статистические критерии проверки гипотез строятся по изложенной выше схеме и различаются только конкретным видом выбираемой функции от значений наблюдения.

§ 3. Основы теории измерений

Теория измерений занимается вопросами численного описания характеристик объекта. Одним из её основных понятий является понятие шкалы измерения. Дадим, прежде всего, общее определение шкалы с математической точки зрения.

Пусть **A** — эмпирическая система с отношениями, **R** — полная числовая система с отношениями, **f** — функция, изоморфно отображающая **A** в подсистему **R**, **G** — группа допустимых преобразований на **R**. Тогда упорядоченный набор **<A, R, G>** называется **шкалой**, а набор **<A, R, G, f>** — **реализацией шкалы**.

Зафиксируем следующие группы преобразований на множестве действительных чисел **R**:

$\mathbf{G_1}$ — группа всевозможных взаимно однозначных отображений;

$\mathbf{G_2}$ — группа отображений, сохраняющих отношение порядка;

$\mathbf{G_3}$ — группа линейных отображений вида $\varphi(x) = \alpha x + \beta$, где $\alpha > 0$;

$\mathbf{G_4}$ — группа растяжений вида $\varphi(x) = \alpha x$, где $\alpha > 0$;

$\mathbf{G_5}$ — группа сдвигов вида $\varphi(x) = x + \beta$.

Легко увидеть, что введённые группы образуют следующую цепочку включений:

$$\mathbf{G_1 > G_2 > G_3 > G_4} \text{ и } \mathbf{G_3 > G_5}.$$

Из теории измерений известна следующая основная теорема классификации шкал:

Теорема. 1) Имеется пять типов шкал измерений признаков: шкала наименований $\mathbf{S_1}$, шкала порядка $\mathbf{S_2}$, шкала интервалов $\mathbf{S_3}$, шкала отношений $\mathbf{S_4}$ и шкала разностей $\mathbf{S_5}$; 2) каждая шкала $\mathbf{S}_i$ имеет в качестве группы допустимых преобразований группу $\mathbf{G}_i$ и полностью ею определяется; 3) группы $\mathbf{G}_i$ частично упорядочены по включению, причём, если $\mathbf{G}_i > \mathbf{G}_j$, то шкала $\mathbf{S}_j$ реализует более точный способ измерений, чем шкала $\mathbf{S}_i$.

Рассмотрим подробнее перечисленные выше шкалы.

Шкала наименований — это любое взаимно однозначное отображение изучаемой эмпирической системы во множество чисел.

Данная шкала позволяет лишь различать объекты. Например, эмпирическому признаку пол ученика можно сопоставить одно из чисел 0 или 1. Скажем, девочка — 0, мальчик — 1. Поскольку группа допустимых преобразований $\mathbf{G_1}$ состоит из любых взаимно однозначных отображений чисел друг в друга, то можно было бы выбрать другие два различных числа для нашей шкалы, например, девочка — 100, мальчик — 50. Числа, сопоставленные полу ученика, не несут в себе никакого смысла присущего собственно числам. Их нельзя, скажем, складывать или сравнивать. Так для второй шкалы $50 + 50 = 100$ не означает, что два мальчика равны одной девочке. А из сравнения $100 > 50$, не следует, что девочка (100) лучше, чем мальчик (50).

Шкала наименований не предоставляет никакой количественной информации, она лишь даёт объектам имена, отсюда и название — шкала наименований.

При использовании шкалы наименований объекты измерения распределяются по непересекающимся классам, охватывающим все изучаемые объекты. Каждому классу даётся наименование, числовое обозначение которого является одним из шкальных значений. Один класс — это совокупность объектов, имеющих одно и то же шкальное значение. Другими словами, при осуществлении измерения моделируются только отношения равенства и неравенства.

В более общей трактовке можно использовать для шкалы наименований любые объекты, а не только числа, скажем, мальчики — «+», девочки — «–», или мальчики — «А», девочки — «Б», или мальчики — ♂, девочки — ♀. Более того, для практических целей бывает удобнее использовать именно буквенные или знаковые обозначения, чтобы не возникало ненужных ассоциаций и соблазнов делать поспешные выводы (два мальчика равны одной девочке, девочки лучше мальчиков и т.д.). Отметим попутно, что шкала наименований, состоящая из двух значений, активно используется в различного рода исследованиях под названием дихотомическая шкала.

Несмотря на кажущуюся простоту шкалы наименований, она при грамотном использовании позволяет получить важные данные.

Шкала порядка — это любое отображение упорядоченной эмпирической системы в числовую систему с отношениями, сохраняющее порядок. Примером такой шкалы служит хорошо известная шкала балльных оценок: 2, 3, 4, 5. Выбранные числа сохраняют в себе упорядоченность всей числовой системы, т.е. их всегда можно сравнить, например, $2 < 3$, $3 < 5$, $3 = 3$. Такая шкала не только позволяет различать объекты, но и сравнивать их. Например, 2 и 3: ученик, оцененный двойкой, знает или умеет меньше, чем ученик, оцененный тройкой. Но такая шкала также не позволяет выполнять арифметические операции. Хотя $2 + 2 = 4$, две «двойки» не равны одной «четвёрке», так же как сумма знаний «двоечника» и «троечника» не равна знаниям одного «отличника».

Группа допустимых преобразований данной шкалы состоит из преобразований, сохраняющих порядок. Следовательно, мы могли бы вместо шкалы 2, 3, 4, 5 взять шкалу 1, 15, 16, 40. Она также позволяет сравнивать объекты. Заметим здесь, что «расстояние» между оценками не несёт в данном случае никакой смысловой нагрузки. Нельзя сказать, что ученик, получивший 16 баллов, знает на 1 балл больше ученика, получившего 15 баллов, а ученик, имеющий 15 баллов, знает на 14 баллов больше ученика, имеющего 1 балл. Шкала порядка указывает только на порядок расположения объектов, но ничего не говорит о расстояниях между ними.

Таким образом, порядковая шкала моделирует не только отношения равенства и неравенства, но и отношения порядка между ними.

Если опять вернуться к более общей ситуации, забыв про числа, то в качестве примера порядковой шкалы можно взять шкалу «неудовлетворительно», «удовлетворительно», «хорошо», «отлично», или шкалу «высший сорт», «первый сорт», «второй сорт», «третий сорт», или же шкалу «верно», «неверно». Важно, чтобы была возможность упорядочить эти оценки также как и числа.

Порядковые шкалы несут в себе, очевидно, гораздо больше информации, чем шкалы наименований, и в то же время не требуют от исследователя точности количественных оценок. Оценивая качество объектов, исследователь расставляет их в некотором порядке, не обращая внимания на величину различий между ними.

Шкала интервалов — это шкала порядка, для которой задана операция, позволяющая сравнивать интервалы между делениями на шкале. Типичным примером является температурная шкала Цельсия. Расстояния между точками на шкале всегда можно сравнить между собой. Используя допустимые преобразования, можно получить другие температурные шкалы. Например, если **C** и **F** — температуры по Цельсию и Фаренгейту соответственно, то $\mathbf{F} = 1{,}8\mathbf{C} + 32$.

Отметим, что допустимое в данной шкале преобразование $\varphi(x) = \alpha x + \beta$ действует следующим образом: оно растягивает шкалу в α раз и сдвигает её на β единиц. Отсюда легко следует, что отношения длин одних и тех же интервалов в разных реализациях одной шкалы равны. Основной характерной чертой шкалы интервалов является наличие масштабной единицы, что позволяет выяснить не только, в каком из сопоставляемых объектов признак выражен сильнее, но и насколько сильнее выражен. Как мы видим, основную нагрузку в данном случае несут на себе величины интервалов, а числовые отметки можно задавать произвольным образом (0° по Цельсию равны 32° по Фаренгейту и т.д.). Отсюда и название шкалы.

Другим примером шкалы интервалов может служить процентное выражение каких-либо данных, скажем, качественная успеваемость — процент учащихся, не имеющих «двоек» и «троек» и т.д.

Таким образом, интервальные шкалы помимо отношений равенства и порядка изучаемых объектов моделируют отношения равенства «интервалов» между объектами, имея фиксированную масштабную единицу.

Понятно, что любую порядковую шкалу можно превратить в интервальную, описав правило сравнения интервалов. Но именно в обосновании равенства и разности интервалов между объектами и состоит основная трудность.

Шкала отношений — это шкала интервалов, в которой естественным образом задана точка отсчёта. Например, шкала измерений роста или веса. Допустимыми преобразованиями этой шкалы являются только растяжения $\varphi(x) = \alpha x$, где $\alpha > 0$. Например, вес можно мерить в сантиметрах, метрах и т.д. Основным требованием для шкалы отношений является отображение в ноль одного и того же объекта. Отсюда следует, что в различных реализациях одной шкалы сохраняются отношения не только между интервалами, но и между самими шкальными значениями. А это означает, что можно ответить на вопрос, во сколько раз сильнее выражен признак в одном из сравниваемых объектов по отношению к другому.

В виду достаточно жестких требований к системе шкала интервалов редко используется в психологических исследованиях. В качестве примера можно привести разве что затраченное время на выполнение некоторого задания.

Переход от шкалы интервалов к шкале отношений сопряжён с большими трудностями при незначительном увеличении числа способов статистической обработки, поэтому с точки зрения практики такой переход малоэффективен.

Шкала разностей — это шкала интервалов с жёстко закреплённым масштабом. Допустимые преобразования для этой шкалы — только сдвиги $\varphi(x) = x + \beta$. В качестве примера такой шкалы можно привести возраст. В силу достаточно серьёзных ограничений на систему шкала разностей практически не получила распространения в психологических исследованиях. Она используется, пожалуй, лишь в методе парных сравнений.

Отметим, в заключение, что по традиции шкалы наименования и порядка относят к *качественным* шкалам, а шкалы интервалов, отношений и разностей — к *количественным*. Такое деление естественным образом вытекает из свойств перечисленных шкал. С позиций статистической обработки данных чаще выделяют три группы шкал: номинальные, порядковые и количественные. Это объясняется тем, что разработанные статистические методы недостаточно точны для учёта особенностей, например, шкалы интервалов по отношению к шкале разностей.

Для большей наглядности сведём всё вышеизложенное в таблицу, добавив для полноты картины сведения о возможности вычисления статистических характеристик, часть из которых появится в следующем параграфе.

Таблица 1

Типы шкал и их основные характеристики

Шкала	Свойства	Группа допустимых преобразований	Примеры	Статистический аппарат
Шкала наименований **S_1**	Различает предметы по наличию свойства. Не различает уровней проявления свойства	**G_1** — группа всевозможных взаимно однозначных отображений	Порядковый номер, пол, результат сдачи зачёта	Частота Мода
Шкала порядка **S_2**	Различает уровень проявления свойств объекта. Не определяет величину различия в проявлении свойств	**G_2** — группа отображений, сохраняющих отношение порядка	Школьные оценки, год обучения, стаж	Частота Мода Медиана Коэффициент Кендэлла Размах Коэффициент Спирмена
Шкала интервалов **S_3** Определяет величину различия проявления свойства.	Имеет масштабную единицу. Сравнивает, на сколько больше проявляется свойство.	Температура, рейтинг, семестры Частота Относительная частота Квантили	Мода Медиана Среднее	Дисперсия Коэффициент корреляции

	Не определяет уровень исчезновения свойства (естественный ноль шкалы) **G_3** — группа линейных отображений вида $\varphi(x) = \alpha x + \beta$, где $\alpha > 0$			
Шкала отношений **S_4**	Определяет любые отношения между уровнями проявления свойств. Имеет масштабную единицу и фиксированное начало шкалы	**G_4** — группа растяжений вида $\varphi(x) = \alpha x$, где $\alpha > 0$	Масса, длина, скорость выполнения задания, процент учащихся, объём часов по предметам	Частота Относительная частота Квантили. Мода Медиана Среднее Дисперсия Коэффициент корреляции
Шкала разностей **S_5**	Определяет накопление свойства. Имеет масштабную единицу. Не имеет фиксированного начала шкалы	**G_5** — группа сдвигов вида $\varphi(x) = x + \beta$	Номера учебных недель, хронометрия	Частота Относительная частота Квантили Мода Медиана Среднее Дисперсия Коэффициент корреляции

При измерении качественных показателей, скажем, успеваемости, исследователь, естественно использует качественные шкалы, а именно шкалу наименований или шкалу порядка. Пользуясь шкалой наименований, можно лишь отметить различия, но невозможно сравнить проявление описываемых качеств. Следовательно, с точки зрения дальнейшей обработки информации предпочтительнее шкала порядка. Использование количественных шкал для первоначальных измерений в психологии физически невозможно в силу объективных причин. Не существует прибора, который позволил бы измерить, скажем, уровень интеллектуального развития, также как мы можем измерить вес или температуру. Следовательно, с точки зрения психологии проблема измерения и оценки объекта исследования состоит в количественной оценке информации, заключённой в качественных первичных оценках.

Для использования более точных шкал требуется больше информации, а, следовательно, лучшее понимание качественных закономерностей свойств исследуемой системы. Попытки построить сразу интервальную шкалу или шкалу отношений при недостатке информации приводит к использованию субъективных, а зачастую и неверных оценок. Хотя как было отмечено выше, это не даёт значительных преимуществ для дальнейшей статистической обработки.

В порядковых шкалах основную нагрузку несут на себе не сами числа, сопоставляемые конкретным проявлениям признака, а их взаимное расположение. Поэтому удобнее использовать не сами значения этих чисел, а их ранги.

Рангом наблюдения называется тот номер, который получит это наблюдение после упорядочения всех данных по возрастанию (или убыванию) силы проявления признака.

Процедура перехода от совокупности наблюдений к последовательности их рангов называется *ранжированием*.

Если все данные наблюдений различны, не возникает никакой трудности при их упорядочении. Если же есть повторяющиеся данные, то переходят к так называемым *средним рангам*. Они вычисляются следующим образом. Предположим, что мы имеем n одинаковых наблюдений, называемых *связкой*. Если бы они были различными, то занимали бы после ранжирования места, допустим, с i-того по $i+n$-тое. Чтобы посчитать средний ранг связки, надо сложить все эти ранги и разделить их на число элементов в связке, т.е. вычислить выражение $[i+(i+1)+(i+2)+\ldots+(i+n)]/n$. Каждому элементу в связке сопоставляется этот средний ранг. Например, после ранжирования выборки 3, 4, 5, 9, 10, 17, 42 мы получим последовательность 1, 2, 3, 4, 5, 6, 7, а после ранжирования выборки 2, 3, 3, 7, 15, 15, 15 — последовательность 1, 2½, 2½, 4, 6, 6, 6.

Как мы уже выяснили, измерение каких-либо показателей представляет собой выбор шкалы, которая отражала бы имеющиеся в исследуемой системе отношения. Этот выбор всегда ложится на исследователя. Правильнее было бы говорить не о выборе, а о построении шкалы, поскольку недостаточно решить использовать, к примеру, пятибалльную шкалу. Самое трудное — определиться, за что выставлять конкретную оценку.

Построенная шкала должна быть надёжной и валидной, т.е. точно измерять именно то, для чего она предназначена. Разберём подробнее эти две характеристики.

Прежде всего, разберёмся с понятием *надёжности*.

Целью любого измерения является установление истинного значения измеряемой величины, т.е. существующей в действительности неискажённой величины признака, присущего данному индивиду. Этот изучаемый признак проявляется довольно устойчиво в тестах, подготовленных для его измерения.

Однако в философии хорошо известен постулат о неизбежности погрешности измерения: результат эксперимента всегда содержит ошибку, как бы тщательно не проводились измерения. Принятие этого постулата неизбежно приводит к одному из основных положений теории измерения — к тезису о невозможности знания абсолютного значения измеряемой величины. Следовательно, измерениям подвергается только наблюдаемый результат измерения, искажённый под влиянием различных факторов. Можно выделить множество факторов, искажающих истинное значение измеряемой величины, например, влияние цели, задач и характера исследования, условия тестового опроса и т.д.

Итак, всякое измерение всегда содержит ошибку. Значит, задача исследователя — понизить эту ошибку до приемлемого уровня.

Но точность измерения составляет только одну сторону понятия надёжности. Другой важной характеристикой надежности является возможность повторить измерения с тем же результатом.

Валидность шкалы означает её способность измерять именно тот признак, для измерения которого она предназначена.

Соотношение между надёжностью и валидностью проще всего представить аналогией с часами: часы могут иметь точный (надёжный) ход, но, будучи поставленными на неверный час, они непригодны (невалидны) для получения ответа на главный вопрос — который час?

§4. Некоторые методы проверки статистических гипотез

4.1. Методы определения связи признаков

Во многих практических задачах при исследовании объектов, обладающих несколькими признаками, необходимо бывает выяснить, насколько эти признаки связаны между собой. Например, как связаны оценки по математике и оценки по физике, или успеваемость в вузе с типом темперамента и т.д. Поскольку мы ведём речь о статистических критериях, прежде всего мы должны выразить экспериментальные данные в числовой форме, т.е. измерить их. Так как значения в различных шкалах несут в себе различный запас информации, методы определения связи признаков существенно зависят от шкалы измерения этих признаков. Поэтому первым этапом анализа является классификация типа данных, т.е. отнесение их к той или иной шкале измерений. Вторым этапом является проверка гипотезы об отсутствии связи признаков. На третьем этапе производится оценка силы связи признаков, если гипотеза об их независимости была отвергнута.

Как было отмечено выше, для определения связи признаков, измеренных в разных шкалах, применяются разные критерии. Для данных, измеренных в шкале наименований, чаще всего используются таблицы сопряжённости и статистика Фишера-Пирсона хи-квадрат (χ^2). Для данных, измеренных в шкале порядка, используют ранжирование и коэффициенты корреляции Спирмена и Кендэлла. Для данных, измеренных в количественных шкалах, применяют коэффициент корреляции Пирсона и средства регрессионного анализа. Остановимся подробнее на этих критериях, ограничившись случаем двух признаков.

До конца этого пункта под «нулевой» гипотезой H_0 будем понимать гипотезу о независимости двух признаков A и B. При этом признаки A и B будем называть *независимыми*, если значение, принятое признаком A, не влияет на вероятности возможных значений признака B и наоборот.

Таблицы сопряжённости

Предположим, что в нашем распоряжении имеется выборка, состоящая из n объектов. Каждый объект характеризуется двумя признаками, скажем, A и B. Пусть признак A принимает r значений $A_1, A_2,\ldots, A_r$, а признак B принимает s значений $B_1, B_2,\ldots, B_S$. Составим таблицу 2 сопряжённости данных признаков.

Таблица 2

Таблица сопряжённости признаков *A* и *B*

	B_1		B_2		…		B_s		Сумма по столбцам
A_1	n_{11}	ν_{11}	n_{12}	ν_{12}	…	…	n_{1s}	ν_{1s}	P_1
A_2	n_{21}	ν_{21}	n_{22}	ν_{22}	…	…	n_{2s}	ν_{2s}	P_2
…	…	…	…	…	…	…	…	…	…
A_r	n_{r1}	ν_{r1}	n_{r2}	ν_{r2}	…	…	n_{rs}	ν_{rs}	P_r
Сумма по строкам	Q_1		Q_2		…		Q_s		n

Здесь через n_{ij} обозначено количество объектов выборки, обладающих комбинацией уровней A_i и B_j, через $P_i = n_{i1} + n_{i2} + \ldots + n_{is}$ — число появлений значений признака A_i и через $Q_j = n_{j1} + n_{j2} + \ldots + n_{jr}$ — число появлений значений признака B_j. Через ν_{ij} в таблице обозначена ожидаемая частота совместного появления значений A_i и B_j, найденная по формуле

$$\nu_{ij} = \frac{P_i Q_j}{n}.$$

Статистика Пирсона-Фишера

Критерий Пирсона-Фишера позволяет выяснить, существует ли какая-либо связь между признаками A и B, или же они ведут себя независимо друг от друга? Этот критерий основан на свойствах распределения χ^2.

Вычислим значение статистики Пирсона-Фишера, пользуясь формулой

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{(n_{ij} - \nu_{ij})^2}{\nu_{ij}}.$$

Теперь нужно сравнить полученное значение с критическим для выбранного уровня значимости значением, найденным по таблицам распределения χ^2 с $(r - 1)(s - 1)$ степенями свободы. Если полученное значение превосходит табличное, то гипотеза о независимости признаков A и B отвергается на выбранном уровне значимости.

Следует отметить, что критерий Пирсона-Фишера не следует использовать при малом числе наблюдений. Для его корректного применения считается достаточным, чтобы все ожидаемые частоты ν_{ij} были бы не меньше 5.

Проиллюстрируем применение критерия Пирсона-Фишера на следующем примере. У каждого члена группы из 195 человек по методике Чижковской был определён тип нервной деятельности (признак B). Затем каждому испытуемому было предложено выбрать инструкцию, регламентирующую его дальнейшую деятельность (признак A). Каждый из признаков имеет два уровня. Для признака A — это высокая реактивность (+) и низкая реактивность

(–) нервной деятельности. Для признака B — детальная инструкция и краткая инструкция. Результаты измерений с уже вычисленными значениями P_i, Q_j и V_{ij} приведены в таблице 3.

Таблица 3

Предпочтение различных видов инструкции высокореактивными и низкореактивными индивидами

Тип инструкции	+		–		Сумма по столбцам
Детальная	63	52,2	42	52,7	105
Краткая	34	44,8	56	45,2	90
Сумма по строкам	97		98		195

Вычисление статистики хи-квадрат по приведённой выше формуле даёт значение $\chi^2 = 9{,}58$. Число степеней свободы в нашем случае равно $(2 - 1)(2 - 1) = 1$. По таблице распределения χ^2 с одной степенью свободы находим, что вероятность получить значение 7,88 равна 0,005, а вероятность получить значение 10,83 равна 0,001. Следовательно, мы можем сделать вывод: вероятность случайно получить для независимых признаков наше значение 9,58 не превышает 0,005. Можно говорить поэтому, что гипотеза о независимости признаков A и B противоречит результатам эксперимента и мы должны её отвергнуть, т.е. связь между этими признаками присутствует.

Как показывает приведённый выше пример, практическое применение критерия Пирсона-Фишера не представляет никакой сложности и не содержит больших вычислений. Излагаемые дальше методики также просты в работе, поэтому мы не будем иллюстрировать их конкретными вычислениями. Достаточное количество примеров по использованию методик можно найти в (14).

Как мы уже отмечали, в шкалах порядка основную нагрузку несут на себе не сами числа, сопоставленные проявлению признака, а их взаимное расположение. Следовательно, для нас важны не значения этих чисел, а их порядок. Поэтому для дальнейшей обработки данных, измеренных в порядковой шкале, полезнее осуществить ранжирование данных. В результате каждому объекту будет сопоставлена пара чисел (r_i, s_i), где r_i — ранг объекта относительно признака A и s_i — ранг объекта относительно признака B.

Критерий Спирмена

Вычислим статистику S, используя формулу

$$S = \sum_{i=1}^{n}(r_i - s_i)^2.$$

Затем, чтобы сгладить действие численности группы n, перейдём к *коэффициенту ранговой корреляции Спирмена*:

$$\rho = 1 - \frac{6S}{n^3 - n}.$$

Коэффициент ρ лежит на отрезке от –1 до 1. Свои крайние значения он принимает в случаях, когда признаки A и B полностью противоположны ($\rho = -1$) или когда они полностью совпадают ($\rho = 1$). Чем ближе значение ρ к нулю, тем меньше связь между признаками. Следует отметить, что если связь между признаками более сложная, чем просто монотонная, то коэффициент Спирмена не всегда может отличить зависимость от независимости.

Для проверки гипотезы об отсутствии связи признаков надо воспользоваться таблицами распределения коэффициента Спирмена. Если полученное значение превосходит табличное по модулю, то предположение о независимости признаков отвергается на выбранном уровне значимости.

Отметим в заключение, что наличие связок, т.е. повторяющихся данных, снижает точность изложенного метода.

Коэффициент корреляция Пирсона

Предположим, что нам дана выборка из n объектов, каждому из которых сопоставлены два значения, характеризующих признаки A и B. Т.е. с каждым объектом связаны две последовательности оценок $a_1, a_2, \ldots, a_n$ и $b_1, b_2, \ldots, b_n$. Вычислим значение коэффициента корреляции по формуле

$$r = \frac{\sum_{i=1}^{n}(a_i - \overline{a})(b_i - \overline{b})}{\sqrt{\sum_{i=1}^{n}(a_i - \overline{a})^2}\sqrt{\sum_{i=1}^{n}(b_i - \overline{b})^2}}, \text{ где } \overline{a} = \frac{1}{n}\sum_{i=1}^{n} a_i, \overline{b} = \frac{1}{n}\sum_{i=1}^{n} b_i.$$

Коэффициент r принимает значения от –1 до 1. Причём, он равен –1 или 1, если связь между признаками линейная. Если признаки независимы и коэффициент корреляции существует, то он равен 0. С другой стороны, равенство нулю коэффициента r не означает статистической независимости исследуемых признаков.

Для коэффициента Пирсона построены таблицы распределений, пользуясь которыми мы можем сделать вывод о справедливости нашей гипотезы. Если полученное значение r по модулю превосходит критическое значение для выбранного уровня значимости, то гипотеза о независимости исследуемых признаков отвергается на данном уровне значимости.

Если гипотеза о независимости признаков отвергается, то обычно выясняют степень силы связи. Для этого используют различные *меры связи*, т.е. некоторые числовые характеристики, указывающие на эту связь. Наиболее распространёнными среди них являются уже знакомые нам коэффициенты корреляции Спирмена и Пирсона. Поскольку описанные выше методы определения связи признаков позволяют определять и силу этой связи, мы не будем приводить здесь другие коэффициенты. Более подробную информацию по этой теме можно найти в (12; 17).

Существует большое количество статистических критериев для определения связи признаков и установления силы этой связи. Мы рассмотрели наиболее универсальные из них. Эта универсальность означает минимум предположений о характере изучаемых данных, а, значит, минимум дополнительных проверок, которые, как правило, и представляют основную трудность. Применение же какого-либо критерия в несвойственной ему ситуации в лучшем случае даст очень неточный, а в худшем — просто неверный результат. Кроме того, рассмотренные критерии не требуют каких-либо специальных знаний и осуществляются с минимальными вычислениями. Для применения вышеизложенных критериев достаточно иметь элементарные вычислительные навыки и набор статистических таблиц в рамках любого справочника.

4.2. Критерии согласия.

Во многих статистических задачах мы предполагаем, что некоторые случайные величины имеют заданное распределение (нормальное, экспоненциальное и т.д.) и, исходя из этого предположения, применяем те или иные методы дальнейшей обработки. При этом естественно возникает вопрос: насколько наши предположения отвечают экспериментальным данным? Для ответа на этот вопрос применяются специально разработанные критерии, которые называются *критериями согласия*.

Наиболее распространённые из них это: критерий Колмогорова, критерии омега-квадрат и хи-квадрат, критерии асимметрии, эксцесса и т.д.

При проверке справедливости предположения о законе распределения следует выделять *простые* и *сложные* гипотезы:

- простая гипотеза указывает определённый закон распределения с конкретными параметрами, например, нормальное распределение с математическим ожиданием 0 и дисперсией 1 (так называемое, стандартное нормальное распределение)
- сложная гипотеза указывает семейство распределений, например, нормальное распределение с неизвестными математическим ожиданием и дисперсией.

Мы рассмотрим примеры критериев согласия, предназначенных для проверки простых гипотез, т.к. это наиболее распространённая ситуация для психологических исследований.

Критерий Колмогорова

Пусть мы имеем n значений $x_1,\dots, x_n$ некоторой характеристики, измеренной в количественной шкале. Прежде всего, перейдём к *вариационному ряду* $x_{(1)}, \dots, x_{(n)}$, т.е. расположим значения в нашей выборке по возрастанию. Обозначим через $F(x)$ предполагаемую функцию распределения. Необходимым условием применимости критерия Колмогорова является непрерывность функции $F(x)$, т.е. использовать его можно лишь для непрерывных распределений.

Предположим, что нам известны все параметры предполагаемого распределения. Например, если речь идёт о нормальном распределении, то мы должны знать математическое ожидание a и дисперсию σ.

Для проверки гипотезы о совпадении эмпирической и гипотетической функций распределения надо вычислить значение статистики Колмогорова D_n, пользуясь следующей формулой:

$$D_n = \max_{1 \le i \le n} \left\{ \frac{i}{n} - F(x_{(i)}), F(x_{(i)}) - \frac{i-1}{n} \right\}.$$

Значения функции распределения *F(x)* можно найти в соответствующих таблицах. Полученную величину D_n следует сравнить с найденными по таблицам критическими значениями. Гипотеза отвергается на выбранном уровне значимости, если полученное значение превосходит критическое значение для данного уровня значимости. Таблицы для статистики Колмогорова можно найти, например, в (8).

Критерий омега-квадрат

В тех же предположениях, что и критерий Колмогорова применяется критерий омега-квадрат.

Статистика омега-квадрат (ω^2) вычисляется по следующей формуле:

$$\omega^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^{n} \left(F(x_{(i)}) - \frac{2i-1}{2n} \right)^2 .$$

Если полученное значение превосходит найденное по таблице критическое значение, то гипотеза отвергается на данном уровне значимости. Таблицы для статистики омега-квадрат также можно найти в (8).

Отметим, что существуют модификации критериев Колмогорова и омега-квадрат, предназначенные для проверки сложных гипотез.

Критерий хи-квадрат К. Пирсона

Данный критерий относится к независимым испытаниям с конечным числом исходов и предназначен только для простых гипотез. Пусть мы имеем n независимых повторений некоего опыта, который заканчивается одним из r исходов. Критерий хи-квадрат предназначен для проверки гипотезы о том, что вероятности этих исходов равны p_1, ... p_r. Обозначим через m_1, m_2, ... m_n количества опытов, заканчивающихся, соответственно, первым, вторым, …, n-ным исходами и вычислим значение статистики хи-квадрат (χ^2) с $(r-1)$ степенью свободы по следующей формуле:

$$\chi^2 = \sum_{i=1}^{r} \frac{(m_i - np_i)^2}{np_i}.$$

Если полученное значение превосходит табличное, то гипотеза должна быть отвергнута на выбранном уровне значимости. Значение распределения χ^2 можно найти в соответствующих таблицах.

Следует отметить, что критерий хи-квадрат достаточно точен только при больших значениях n. Его рекомендуется применять, если все значения произведений np_i не меньше 10.

В заключение сделаем несколько общих замечаний по поводу использования критериев согласия.

1. Теоретическое представление о законе распределения, которому должна подчиняться выборка, всегда имеет характер математической модели, т.е. является в какой-то мере приближённым. Следовательно, точность статистических проверок должна быть сопоставима с точностью, которую мы ожидаем от математической модели в целом.

2. Критерии согласия очень чувствительны к «выбросам»,

т.е. к значениям резко отличающимся от остальных. А, значит, учитывая первое замечание, мы можем отбросить такие значения, чтобы выровнять распределение для дальнейших исследований.

4.3. Факторный анализ

Если можно явно выделить несколько факторов, влияющих на конечный результат, причём эти факторы принимают конечное число значений, то возникает задача *факторного анализа.* Типичный пример — сравнение по достигаемым результатам нескольких различных способов действия, например, методик преподавания, стимулирования и т.д. То, что, по нашему мнению, оказывает влияние на конечный результат, называется *фактором*, конкретную реализацию фактора (например, конкретную методику) называют *уровнем фактора* или *способом обработки*, а значение измеряемого признака (т.е. величину результата) называют *откликом*.

Если мы выделяем несколько факторов, то получаем задачу *многофакторного анализа*, а если один — то *однофакторного анализа.*

Бывает, что в рамках однофакторной модели влияние интересующего нас фактора не проявляется, хотя достаточно убедительные соображения указывают, что такое влияние должно быть. Причиной этого может быть большой внутригрупповой разброс, на фоне которого действие фактора становится малозаметным, либо наличие других факторов, сглаживающих это влияние. В первом случае дело можно поправить более тщательным отбором экспериментальных данных, во втором же надо вводить в рассмотрение дополнительные факторы. Один из методов борьбы с нежелательными воздействиями «мешающих» факторов основан на специальном планировании эксперимента. Цель такого планирования —

свести к нулю влияние «мешающих» факторов. А именно, при фиксированном уровне фактора проводят испытания на такой группе, внутри которой действия мешающих факторов уравновешивают друг друга. Однако, если информация о характере влияния таких факторов отсутствует, такой подбор становится невозможным. Тогда можно поступить следующим образом. Из большого числа потенциально пригодных объектов случайным образом выбираются те, которые и образуют требуемую группу. Методики такого выбора можно найти в (16). Если же этими методами не удаётся избавиться от мешающего фактора, надо переходить к схемам многофакторного анализа. Желательно выделить один главный «мешающий» фактор и перейти к методам двухфакторного анализа, которые наиболее эффективны и просты. Мы не будем рассматривать эти методы, а остановимся лишь на задачах однофакторного анализа. Более подробно эти вопросы изложены в (19).

Для сравнения влияния факторов на результат необходимо собрать определённый статистический материал. Обычно его получают следующим образом: каждый из k способов обработки применяют несколько раз (не обязательно одно и то же число раз) к исследуемому объекту и регистрируют результаты. Наиболее удобным способом представления полученных данных является таблица.

Таблица 4

Таблица однофакторного анализа

	Способ обработки			
	1	2	…	k
Результаты измерений	x_{11}	x_{12}	…	x_{1k}
	…	…	…	
	$x_{n(1)\,1}$	$x_{n(2)\,2}$	…	$x_{n(k)\,k}$

Здесь $n(i)$ — количество измерений, соответствующих i-тому способу обработки.

Первый вопрос, на который необходимо ответить при изучении экспериментальных данных следующий: Есть ли влияние, выделенного фактора на полученный результат? Другими словами может ли различие наблюдённых в опыте значений быть вызвано случайными причинами? То есть надлежит проверить справедливость следующего предположения: все данные таблицы принадлежат одному и тому же распределению, а, значит, отсутствуют эффекты обработки. Такое предположение обычно называют *нулевой гипотезой.*

Если гипотеза об отсутствии эффектов обработки отвергается, то проводится оценка действия этих эффектов или различий между ними и строятся доверительные интервалы для этих характеристик.

Ранговый однофакторный анализ

В том случае, когда мы ничего не знаем о законе распределения наблюдений, то для проверки нулевой гипотезы лучше использовать не сами значения наблюдений x_{ij}, а их ранги r_{ij} во всей совокупности данных. Используемые для этого критерии называются *ранговыми.* Их целесообразно использовать также для данных, измеренных в порядковой шкале.

Итак, проведём ранжирование данных таблицы 4 и перейдём к таблице 5.

Таблица 5

Ранговая таблица однофакторного анализа

	Способ обработки			
	1	2	…	k
Ранги результатов измерений	r_{11}	r_{12}	…	r_{1k}
	…	…	…	
	$r_{n(1)\,1}$	$r_{n(2)\,2}$	…	$r_{n(k)\,k}$

Критерий Краскела-Уоллиса

Одним из наиболее универсальных ранговых критериев является критерий Краскела-Уоллиса. Он применяется, если нельзя сказать ничего определённого об альтернативах к нашей нулевой гипотезе.

Итак, пусть мы имеем $N = n(1) + n(2) + \ldots + n(k)$ наблюдений, ранги которых занесены в таблицу 5. Статистика Краскела-Уоллиса вычисляется по формуле

$$H = \frac{12}{N(N+1)} \sum_{j=1}^{k} n(j) \left(\frac{1}{n(j)} \sum_{i=1}^{n(j)} r_{ij} - \frac{N+1}{2} \right)^2 .$$

Небольшие таблицы распределения данной статистики можно найти в сборниках статистических таблиц. При больших объёмах N, находящихся за пределами таблиц, распределение H приблизительно равно распределению хи-квадрат с $(k - 1)$ степенями свободы.

Итак, мы отвергаем нулевую гипотезу на данном уровне значимости, если наблюдаемое значение H превышает соответствующее значение для распределения хи-квадрат.

Если в исходной таблице 4 много совпадающих значений, то выводы будут носить приближённый характер. Для повышения точности можно использовать модифицированную форму статистики. Подробные сведения по этому поводу можно найти, например, в (20). Там же можно ознакомиться с другими ранговыми критериями, предназначенными для более частных ситуаций. Эти критерии являются более мощными в каждом конкретном случае, но, применённые некорректно, дают гораздо худшие результаты. Мы не будем касаться этих методик. Отметим лишь, что для целей психологических исследований вполне достаточны универсальные методики.

Дисперсионный анализ

Если мы можем считать, что ошибки наблюдений носят случайный характер и имеют нормальное распределение с нулевым средним и общей дисперсией, то можно использовать методы однофакторного дисперсионного анализа. Опишем один из таких методов.

F–тест для дисперсий

Данный метод проверяет справедливость гипотезы о равенстве средних значений при различных способах обработки и основан на сопоставлении двух оценок для дисперсии.

Осуществляется эта проверка следующим образом.

Предположим, что данные, полученные в эксперименте, измерены в количественной шкале и помещены в таблицу 4. Отметим ещё раз, что данный метод основан на предположении о том, что ошибки наблюдений имеют стандартное нормальное распределение. Сначала необходимо вычислить средние значения X_j по каждому столбцу и общее среднее X. Для этого воспользуемся формулами

$$X_j = \frac{1}{n(j)} \sum_{i=1}^{n(j)} x_{ij}, \text{ где } j = 1,\dots,k, \text{ и } X = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n(j)} x_{ij}.$$

Затем найдём две оценки σ^2 и $\sigma_*{}^2$ для дисперсии:

$$\sigma^2 = \frac{1}{N-k}\sum_{j=1}^{k}\sum_{i=1}^{n(j)}(x_{ij} - X_j)^2 \quad \text{и} \quad \sigma_*^2 = \frac{1}{k-1}\sum_{j=1}^{k} n(j)(X_j - X)^2.$$

И в заключение вычислим отношение

$$F = \frac{\sigma_*^2}{\sigma^2}.$$

Это отношение должно иметь F-распределение с (k –1, N–k) степенями свободы. Если полученное значение F превышает табличное значение, то гипотеза отвергается на выбранном уровне значимости.

Если гипотеза оказалась несовместимой с результатами наблюдений, то можно вычислить доверительные интервалы для средних значений в каждой группе наблюдений, т.е. для результатов действия каждого способа обработки. Для коэффициента доверия (1 – α) доверительный интервал в i-той группе равен

$$(X_j - \frac{\sigma_*}{\sqrt{n(j)}} t_{1-\alpha},\ X_j + \frac{\sigma_*}{\sqrt{n(j)}} t_{1-\alpha}),$$

где $t_{1-\alpha}$ — квантиль уровня (1 – α) соответствующего распределению Стьюдента с (N – k) степенями свободы (т.е. такое значение t, при котором функция распределения равна 1 – α).

Практическое применение методов факторного анализа и критериев согласия также не составляет большого труда, но, тем не менее, принесёт несомненную пользу исследователю. Строгое доказательство выдвигаемых гипотез позволяет двигаться вперёд, сохраняя уверенность в том, что на предыдущих этапах не было допущено ошибок, способных перечеркнуть всю дальнейшую работу.

Задачи и упражнения

1. Описать пространство элементарных исходов Ω для следующих испытаний:

а) оценка за экзамен;

б) одновременное бросание двух монет;

в) бросание игральной кости (кубика) и монеты.

2. Для каждого испытания из задачи 1 привести пример случайного события не являющегося элементарным.

3. Для каждого испытания из задачи 1 привести пример случайного события, вероятность наступления которого равна 1/2. В каком из перечисленных испытаний нет событий, наступающих с вероятностью 1/3 и почему?

4. В ящике находится 4 белых, 6 красных и 10 синих шаров. Вычислить вероятности следующих случайных событий:

а) взят белый шар;

б) взят синий шар;

в) взят белый или красный шар.

5. Студент выучил 6 экзаменационных билетов из 30. Какова вероятность того, что он возьмёт со стола один из выученных билетов? Какова вероятность того, что преподаватель предложит студенту один из этих билетов?

6. Из пачки талонов, занумерованных всеми двузначными числами, наугад вытягивается один. Какова вероятность того, что номер этого талона состоит из одинаковых цифр?

7. На каждом пространстве элементарных исходов Ω из задачи 1 определить две различных случайных величины и полностью описать их.

8. Какие из перечисленных случайных величин являются

дискретными, а какие — непрерывными?

а) число «орлов» при 5 бросаниях монеты;

б) время, требующееся для подготовки к экзамену;

в) результаты теста по определению типа темперамента.

9. Описать функцию распределения случайной величины:

а) оценка, полученная студентом на экзамене;

б) число очков, выпавших при бросании игральной кости (кубика).

10. Вычислить математическое ожидание и дисперсию для случайных величин, описанных в задаче 9.

11. Вычислить выборочные значения математического ожидание, дисперсии и среднеквадратичного отклонения для следующей группы данных:

а) 1, 3, 3, 0, 4, 1;

б) 102, 106, 111, 112, 112, 114, 115, 115, 116 (Для упрощения вычислений воспользуйтесь следующим фактом: если из всех данных вычесть одно и тоже число x, то выборочное математическое ожидание уменьшится на это же число x, а выборочные значения дисперсии и среднеквадратичного отклонения не изменятся.).

12. В какой шкале измерений представлены следующие данные:

а) оценка студента за экзамен;

б) тип нервной деятельности;

в) номер зачётной книжки;

г) ранжирование по степени проявления познавательной самостоятельности;

д) время, затраченное на запоминание набора слов;

е) количество решённых задач из контрольной работы.

13. Для каждого типа шкалы измерений привести примеры, не использованные в данном пособии.

14. Осуществить ранжирование следующих данных:

а) 3, 4, 5, 25, 14.

б) 2, 31, 1, 5, 6, 6, 1,1.

Библиография

Учебная литература

1. Агапов Г. И. Задачник по теории вероятностей. – М.: Высшая школа, 1986.
2. Артемьева Е. Ю., Е. М. Мартынов. Вероятностные методы в психологии. – М.: Изд-во МГУ, 1975. – 206 с.
3. Бочаров П. П., Печенкин А. В. Теория вероятностей. Математическая статистика. – М.: Гардарика, 1998. – 328 с.
4. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 1979.
5. Емельянов Г. В., Скитович В. П. Задачник по теории вероятностей и математической статистики. – Л.: Изд-во ЛГУ, 1967.
6. Суходольский Г. В. Основы математической статистики для психологов. – Л.: Изд-во ЛГУ, 1972.

Справочная литература

7. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочное изд. Под ред. С. А. Айвазяна. – М.: Финансы и статистика, 1983. – 471 с.
8. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
9. Ликеш И., Ляга И. Основные таблицы математической стати-

стики. – М.: Финансы и статистика, 1985. – 356 с.
10. Математический энциклопедический словарь. – М.: Советская энциклопедия, 1988. – 847 с.
11. Мюллер П., Нойман П., Шторм Р. Таблицы по математической статистике. – М.: Финансы и статистика, 1982. – 278 с.

Монографии

12. Аптон Г. Анализ таблиц сопряжённости. – М.: Финансы и статистика, 1982. – 144 с.
13. Аткинсон Р., Бауэр Г., Кроттерс Э. Введение в математическую теорию обучения. – М.: Мир, 1969. – 486 с.
14. Гласс Дж., Стэнли Дж. Статистические методы в педагогике и психологии. – М.: Прогресс, 1976. – 495 с.
15. Готтсданкер Р. Основы психологического эксперимента. – М.: МГУ, 1982. – 463 с.
16. Джонсон Н., Лион Ф. Статистика и планирование эксперимента в технике и науке. – М.: Мир. Т. 1, 1980, – 610 с., Т. 2, 1981, – 520 с.
17. Кендэлл М. Ранговые корреляции. – М.: Статистика, 1975. – 212 с.
18. Тюрин Ю. Н., Макаров А. А. Статистический анализ данных на компьютере / Под ред. Б. Э. Фигурнова. – М.: ИНФРА-М, 1998. – 528 с.
19. Хартман Г. Современный факторный анализ. – М.: Статистика, 1972.
20. Холлендер М., Вульф Д. Непараметрические методы статистики. – М.: Финансы и статистика, 1983. – 518 с.
21. Яковлев Е. В. Педагогический эксперимент: квалиметрический аспект. – Челябинск: Издательство ЧГПУ, 1998. – 136 с.

Оглавление

Евгений Владимирович Яковлев

Математические основы психологии

Редактор Ю. В. Тихонова

ISBN 5–85716–213–0

Лицензия ЛР № 040277 от 17 августа 1997 г.

Формат 60x84/16 Заказ 548

Тираж 100 экз. Объём 3,5 уч. изд. л.

454080 г. Челябинск, пр. им. В. И. Ленина, 69

Издательство ЧГПУ

Отпечатано на ризографе типографии ЧГПУ

454080 г. Челябинск, пр. им. В. И. Ленина, 69