М. И. ГРАБАРЬ
К. А. КРАСНЯНСКАЯ

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПЕДАГОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ

НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ
ИНСТИТУТ СОДЕРЖАНИЯ И МЕТОДОВ
ОБУЧЕНИЯ АПН СССР

М. И. ГРАБАРЬ
К. А. КРАСНЯНСКАЯ

ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПЕДАГОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

●

НЕПАРАМЕТРИЧЕСКИЕ МЕТОДЫ

МОСКВА · «ПЕДАГОГИКА» · 1977

371
Г 75

Печатается по решению
Редакционно-издательского совета АПН СССР

Грабарь Мстислав Игоревич, Краснянская Клара Алексеевна
ПРИМЕНЕНИЕ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПЕДАГОГИЧЕСКИХ ИССЛЕДОВАНИЯХ.
Непараметрические методы

Заведующая редакцией **И. Н. Цекова.** Редактор **В. Г. Иоффе.**
Художник **Н. И. Суслов.** Художественный редактор **Е. Э. Дятлова.**
Технический редактор **О. Н. Самойлова.**
Корректор **В. Н. Рейбекель.**

ИБ 215.

Сдано в набор 21.04.77 г. Подписано в печать 30.08.77 г. А-01652
Формат 84×108$^1/_{32}$ Бумага тип. № 2.
Печ. л. 4,25 Усл. печ. л. 7,14 Уч.-изд. л. 6,99 Тираж 3000 экз. (БЗ—29—13—77 г.)
Заказ 224 Цена 50 коп.

Издательство «Педагогика» Академии педагогических наук СССР и Государственного комитета Совета Министров СССР по делам издательств, полиграфии и книжной торговли
Москва, 107066, Лефортовский пер., 8

Подольский филиал ПО «Периодика» Союзполиграфпрома при Государственном комитете Совета Министров СССР по делам издательств, полиграфии и книжной торговли
г. Подольск, ул. Кирова, д. 25

Грабарь М. И., Краснянская К. А.

Г 75 Применение математической статистики в педагогических исследованиях. Непараметрические методы. М., «Педагогика», 1977.

136 с. с ил. (Науч.-исслед. ин-т содержания и методов обучения Акад. пед. наук СССР).

Книга посвящена проблеме проверки гипотез, возникающих в педагогических исследованиях, с помощью методов математической статистики.

Рассматриваются общие принципы, на основе которых осуществляется такая проверка. Изложение каждого конкретного метода сопровождается примерами его использования на материале педагогических исследований.

Книга адресуется научным работникам в области педагогики, психологии, социологии.

Г $\frac{60300-059}{005(01)-77}$ БЗ—29—13—77.

ВВЕДЕНИЕ

Педагогические явления и процессы обладают с точки зрения возможностей применения методов теории вероятностей и математической статистики для изучения этих явлений и процессов следующими особенностями. Во-первых, весьма слабо разработана до настоящего времени практика измерений случайных величин, характеризующих те или другие стороны (состояния) педагогических явлений и процессов, что главным образом и препятствует проведению объективного количественного анализа при изучении сложных закономерностей этих процессов. Во-вторых, неизвестны и не могут быть установлены с помощью качественного анализа типы вероятностных законов распределения упомянутых выше случайных величин.

Указанные особенности не позволяют исследователям-педагогам применять в своей работе методы математической статистики по аналогии с тем, как эти методы применяются в естественных или технических науках.

Поэтому одной из задач, возникающих в связи с «математизацией» педагогической науки, является задача классификации, с одной стороны, самих педагогических проблем, а с другой — вероятностных и статистических методов с целью определения, какие именно методы теории вероятностей и математической статистики пригодны для решения той или иной педагогической проблемы.

Частичному решению этой задачи и посвящена настоящая работа. При этом авторы ограничились лишь одной (правда, весьма важной) педагогической проблемой — проблемой проверки статистических гипотез, возникающих в педагогических исследованиях.

В настоящее время теория проверки статистических гипотез представляет собой глубоко разработанный раздел математической статистики, опирающийся на достаточно сложный аппарат многомерного статистического анализа, которому посвящена многочисленная научная и учебная литература (см., например, книги [10, 11, 13]). Любое изложение общей теории проверки статистических гипотез неизбежно должно предполагать у читателя очень серьезную математическую подготовку, каковой, к сожалению, не обладают большинство исследователей-педагогов. Поэтому авторы постарались прежде всего отобрать такие типы педагогических гипотез, которые допускают проверку средствами одномерного статистического анализа[1]. Вместе с тем в связи с упомянутым выше состоянием практики количественных измерений из методов одномерного статистического анализа были отобраны те методы, которые допускают использование данных педагогического эксперимента, полученных в результате не только количественных, но и качественных измерений с помощью дискретных шкал наименований и порядка (подробнее об этих шкалах и вообще о проблеме измерений см. раздел 1). Наконец, поскольку теоретические вероятностные законы для большинства педагогических явлений неизвестны, пришлось отказаться от рассмотрения тех методов математической статистики (в математической литературе они известны под названием «критерии согласия»[2]), которые позволяют проверять гипотезы о согласии данных эксперимента с тем или другим типом закона распределения.

В итоге авторы остановили свой выбор на так называемых непараметрических методах математической статистики (см. раздел 3).

Относительно конкретного содержания и структуры данной работы отметим следующее.

В работе рассматриваются типы статистических гипотез, характерных для педагогических исследований (раздел 2), излагается общая методика их проверки (раздел 3). В соответствии с названием работы выделяются характерные особенности непараметрических

[1] Одномерный статистический анализ — это изучение вероятностных свойств одной случайной величины на основе выборочных наблюдений над этой величиной.

[2] В настоящее время более употребительным стал термин «тест» (test). Однако, в связи с тем что в педагогике этот термин имеет свое значение, в работе используется термин «критерий».

критериев (раздел 3, 5). Так как возможность применения того или иного из этих критериев зависит от характера измерений свойств педагогических явлений, в разделе 1 рассматриваются основные типы измерений, возможные в педагогических исследованиях.

Для экспериментальной педагогики характерна постановка исследований, преследующих цель выявления эффективности педагогических средств путем сравнения достижений или свойств одной и той же группы учащихся в разные периоды времени (такие группы получили название зависимых выборок) или разных групп учащихся (независимые выборки). В связи с этим критерии, рассмотренные в работе, также разделены на две группы: для зависимых выборок (раздел 4) и независимых выборок (раздел 5).

В процессе исследования эффективности педагогического средства в большинстве случаев можно выделить два этапа: поисковый эксперимент и массовая проверка. При проведении поискового эксперимента для получения объективных данных достаточна выборка небольшого объема. Большая часть критериев, рассмотренных в данной работе, приспособлена именно для оценки результатов сравнения двух выборок небольшого объема.

При изложении каждого критерия принята следующая структура: указывается тип экспериментальных данных, к которым применим критерий; характер (дискретный или непрерывный) изучаемого свойства; шкала измерения результатов эксперимента; формулируется проверяемая гипотеза, и дается правило принятия решения об отклонении или принятии гипотезы.

В заключение дается сравнительный анализ возможностей использования каждого из рассмотренных критериев при проведении педагогических исследований.

Предполагается, что читатели знакомы с простейшими понятиями теории вероятностей, математической статистики и основными методами репрезентативного отбора в объеме таких книг, как, например, [4, 7, 8] из списка литературы, приведенного в конце работы. Тем не менее для облегчения понимания последующего материала рассмотрим некоторые понятия, относящиеся к законам распределения случайных величин.

Случайные величины и законы их распределения

Под случайной величиной в теории вероятностей понимается такая величина, которая с определенными вероятностями может принимать некоторое множество числовых значений. Иными словами, нельзя указать точно, какое (какие) именно числовое значение (значения) из множества всех возможных значений примет случайная величина. Можно лишь сказать, что такое-то значение (такие-то значения) величина принимает с такой-то вероятностью.

Употребляется следующая классификация случайных величин.

Дискретной называется такая случайная величина, которая принимает множество «изолированных друг от друга» значений, т. е. для каждого ее возможного значения можно указать на числовой оси окрестность, не содержащую других значений этой величины. Простейшим примером дискретной случайной величины может служить случайная величина, принимающая лишь целочисленные значения. Такой случайной величиной будет, в частности, число ошибок, допущенных учащимся в какой-либо контрольной работе.

Непрерывной случайной величиной в широком смысле называется случайная величина, возможные значения которой заполняют сплошные промежутки числовой оси. Примером такой случайной величины может служить время, затраченное учащимся на решение какой-либо задачи или обдумывание ответа на заданный вопрос[1]. В дальнейшем (см. с. 8) мы будем рассматривать понятие непрерывной случайной величины в более узком, однако наиболее важном с точки зрения практических приложений смысле.

Законом распределения случайной величины в самом общем смысле считается всякое соотношение, устанавливающее связь между возможными значениями случайной величины и вероятностями, с которыми принимаются эти значения. Для дискретных величин простейшей формой закона распределения служит так называемая таблица распределения, в которой указываются возможные значения величины и соответствующие вероятности, с которыми принимаются эти значения. Однако наиболее употребительная форма закона распределения, одинаково пригодная для величин любого типа, связана с понятием функции распределения.

[1] Существуют также случайные величины «смешанного типа», которые нельзя отнести к категории чисто дискретных или чисто непрерывных величин, однако в данной работе такие величины рассматриваться не будут.

При изложении основных вопросов, связанных с законом распределения случайных величин, будут использованы некоторые понятия дифференциального и интегрального исчисления. Однако в дальнейшем эти понятия применяться не будут, и читатели, не владеющие элементами дифференциального и интегрального исчисления, могут ограничиться приводимыми в тексте наглядными геометрическими иллюстрациями.

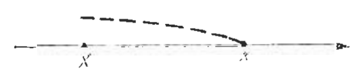
1. *Функция распределения случайной величины.* Будем обозначать большими буквами случайные величины, а их возможные значения — теми же, но малыми буквами.

Функцией распределения случайной величины X называется функция $F(x)$ аргумента x, определяемая равенством

$$F(x) = P(X < x), \quad (1)$$

где величина в правой части означает вероятность того, что случайная величина X принимает значение, меньшее чем заданное число x.

С геометрической точки зрения величину $F(x)$ можно истолковать как вероятность попадания «случайной точки» X на промежуток $(-\infty, x)$ числовой оси:



Перечислим основные свойства функции распределения, а также некоторые важные формулы и понятия, связанные с этими свойствами.

1°. Функция распределения любой случайной величины определена на всей числовой оси.

2°. $0 \leqslant F(x) \leqslant 1$ для любых x, причем $\lim\limits_{x \to -\infty} F(x) = 0$, $\lim\limits_{x \to +\infty} F(x) = 1$. Последние равенства означают, что попадание «случайной точки» X на числовую ось есть достоверное событие, а непопадание — невозможное событие.

3°. Функция распределения есть неубывающая функция, т. е. $F(x_1) \leqslant F(x_2)$ при условии, что $x_1 < x_2$.

4°. $P(x_1 \leqslant X < x_2) = F(x_2) - F(x_1)$, т. е. вероятность попадания случайной точки X в промежуток от x_1 до x_2 равна приращению функции распределения на концах этого промежутка.

5°. Функция распределения любой случайной величины непрерывна слева в каждой точке x_0 числовой оси, т. е. $\lim\limits_{x \to x_0} F(x) = F(x_0)$ при условии, что $x < x_0$.

Непрерывность и слева и справа, т. е. двусторонняя, может не иметь места. Так, можно показать, что функция распределения любой дискретной случайной величины имеет разрывы в точках, которые изображают на числовой оси возможные значения этой величины. Если функция распределения $F(x)$ непрерывна (и слева и справа) в каждой точке числовой оси, то случайную величину X мы будем называть непрерывной в узком смысле или просто непрерывной случайной величиной. Относительно применения подобных случайных величин в педагогике заметим следующее. Педагогические исследования, если они связаны с изучением личности, как правило, направлены на изучение и формирование ее психических свойств. Считается, что любое психическое свойство личности распределено непрерывно. Это положение используется многими психологами и педагогами, хотя оно еще не получило строгого обоснования.

6°. Если p—любое заданное число от 0 до 1, то значение x_p случайной величины X, определяемое равенством $F(x_p) = p$, называется квантилем, отвечающим вероятности p. Это такое значение случайной величины X, для которого справедливо равенство $P(X < x_p) = p$, т. е. вероятность попадания значения случайной величины в промежуток от $-\infty$ до x_p равна заданному значению p.

7°. Если рассматриваются две непрерывные случайные величины X и Y с функциями распределения $F(x)$ и $G(x)$, то неравенство $F(x) > G(x)$, справедливое для всех x, имеет следующий вероятностный смысл. Для любой заданной вероятности p, удовлетворяющей условию $0 < p < 1$, можно указать такие числа $x_1 < x_2$, что

$F(x_1) = P(X < x_1) = p = P(Y < x_2) = G(x_2)$ (см. рис. 1). Таким образом, с одной и той же вероятностью p величина X может принимать значения, меньшие чем величина Y. В этом случае говорят, что случайная величина X «стохастически» (т. е. в вероятностном смысле) меньше случайной величины Y. Наоборот, если $F(x) < G(x)$, то величина X «стохастически» больше Y.

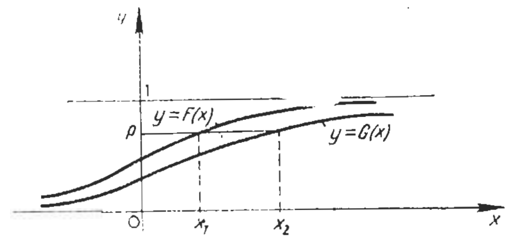


Рис. 1

2. *Плотность распределения случайной величины.* Предположим, что функция распределения $F(x)$ дифференцируема, т. е. имеет производную

$$F'(x) = p(x). \tag{2}$$

В этом случае функция $p(x)$ называется плотностью распределения случайной величины. Отметим, что существование плотности распределения возможно только для непрерывных случайных величин.

Обратная связь $p(x)$ и $F(x)$ выражается формулой

$$F(x) = \int_{-\infty}^{x} p(x) \cdot dx. \tag{3}$$

Свойства 2°, 3° и 4° функции $F(x)$ в сочетании с формулами (2) и (3) определяют следующие основные свойства плотности распределения:

$$p(x) \geqslant 0, \tag{4}$$

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} p(x) \cdot dx, \tag{5}$$

$$\int_{-\infty}^{\infty} p(x) \cdot dx = 1 \text{ (вероятность достоверного события)}. \tag{6}$$

График плотности распределения $p(x)$ называется кривой распределения или кривой вероятностей. Такая кривая позволяет дать наглядную геометрическую иллюстрацию рассмотренным выше соотношениям (см. рис. 2). Заметим, что в силу (6) полная площадь под кривой распределения равна 1.

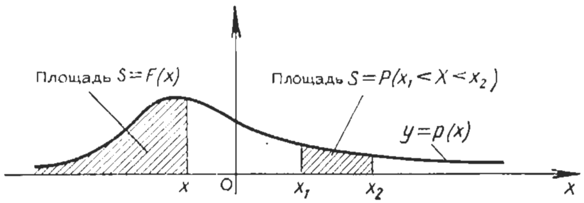


Рис. 2

Если случайная величина имеет плотность распределения $p(x)$, то закон распределения такой величины обычно указывается путем задания функции $p(x)$.

Если возможные значения случайной величины ограничены снизу (сверху) или лежат в конечном интервале, то плотность распределения полагается равной нулю вне промежутка возможных значений случайной величины.

3. *Математическое ожидание, или среднее значение случайной величины.* Законы распределения случайных величин (заданные при помощи таблицы распределения, функции распределения или плотности распределения) позволяют определить числовые характеристики случайных величин, важнейшей из которых служит так называемое математическое ожидание, или среднее значение случайной величины.

Рассмотрим сначала дискретную величину X, принимающую с вероятностями $p_1, p_2, \ldots, p_n$ конечное множество значений $x_1, x_2, \ldots, x_n$. Пусть производится N независимых наблюдений над величиной X, и пусть при этом значение x_1 наблюдалось m_1 раз, значение x_2 наблюдалось m_2 раз, ..., значение x_n наблюдалось m_n раз. Обозначим через $\widetilde{M}[X]$ среднее арифметическое всех результатов наблюдений. Тогда

$$\widetilde{M}[X] = \frac{m_1 x_1 + m_2 x_2 + \ldots + m_n x_n}{N} = x_1 \cdot \frac{m_1}{N} + x_2 \cdot \frac{m_2}{N} + \ldots + x_n \cdot \frac{m_n}{N}.$$

Числа $h_1 = \frac{m_1}{N}$, $h_2 = \frac{m_2}{N}$, ..., $h_n = \frac{m_n}{N}$ суть частоты, с которыми наблюдаются возможные значения $x_1, x_2, \ldots, x_n$ случайной величины X.

Согласно закону больших чисел, при возрастании числа независимых наблюдений N каждая частота h_i будет (в определенном вероятностном смысле) приближаться к вероятности p_i. Поэтому число $M[X]$, определяемое равенством

$$M[X] = x_1 p_1 + x_2 p_2 + \ldots + x_n p_n, \tag{7}$$

представляет собой аналог среднего арифметического результатов наблюдений над величиной X. Число $M[X]$ и называется математическим ожиданием или средним значением случайной величины X. Если дискретная величина X может принимать бесконечное множество значений, то конечная сумма в правой части (7) заменится суммой бесконечного числового ряда.

В том случае, когда X есть непрерывная величина с плотностью распределения $p(x)$, среднее значение (математическое ожидание) определяется по аналогии с формулой (7) с заменой суммы интегралом[1]:

$$M[X] = \int_{-\infty}^{\infty} x\, p(x) \cdot dx. \tag{8}$$

Среднее значение непрерывной случайной величины служит характеристикой центральной тенденции в распределении значений этой величины в том случае, когда плотность распределения $p(x)$ симметрична относительно некоторого значения аргумента или близка к симметричной.

Так, если график $p(x)$ имеет вертикальную ось симметрии $x = x_0$, то среднее значение $M[X] = x_0$. При значительной асимметрии среднее значение уже не будет показателем центральной тенденции.

[1] В том случае, когда среднее значение выражается суммой бесконечного числового ряда или интегралом с бесконечными границами, оно существует не для всякой случайной величины.

4. *Некоторые важные законы распределения.*

а) Биномиальный закон распределения. Рассмотрим так называемую схему Бернулли, т. е. последовательность n независимых испытаний, в каждом из которых некоторое событие A может наступить с вероятностью p (и, следовательно, не наступить с вероятностью $q=1-p$). Рассмотрим, далее, случайную величину, которая представляет собой число наступлений события A во всех n испытаниях. Очевидно, возможными значениями этой случайной величины будут числа: 0, 1, 2, 3, ..., $n-1$, n, и следовательно, величина будет дискретной. Вероятности возможных значений такой случайной величины определяются известной формулой Бернулли

$$P_n(m) = C_n^m \, p^m q^{n-m}, \qquad (9)$$

где $C_n^m = \dfrac{n(n-1)(n-2)\cdots(n-m+1)}{1\cdot 2\cdot 3\cdots m}$ — так называемый биномиальный коэффициент, а $m=0, 1, 2, ..., n$*. Совокупность возможных значений m и их вероятностей (9) и определяет биномиальный закон распределения, который обычно задается таблицей распределения:

m	0	1	2	3	. . .	$n-1$	n
$P_n(m)$	q^n	$n\,p\,q^{n-1}$	$C_n^2\,p^2\,q^{n-2}$	$C_n^3\,p^3\,q^{n-3}$	. . .	$n\,p^{n-1}\,q$	p^n

Доказано, что математическое ожидание (среднее значение) для случайной величины, распределенной по биномиальному закону, равно np.

б) Нормальный закон распределения (закон Гаусса). Практическое применение многих из рассматриваемых далее непараметрических методов статистики основано на так называемом стандартном нормальном распределении, плотность которого задается формулой

$$p(x) = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{x^2}{2}} \qquad (10)$$

* Следует иметь в виду соглашение, что $C_n^0 = C_n^n = 1$.

и описывает закон распределения случайной величины X с математическим ожиданием (средним значением), равным 0, и дисперсией 1.

Приведем график функции (10), называемый обычно кривой Гаусса, и укажем некоторые специальные квантили величины X, которые наиболее часто используются при проверке статистических гипотез. В соответствии с рис. 2 площади под кривой Гаусса равны вероятностям попадания значений случайной величины на те или другие промежутки числовой оси. Заметим, что кривая Гаусса симметрична относительно оси Oy.

Площади, заштрихованные на рис. 3, 4, 5, иллюстрируют вероятностные соотношения:

$$P(|X| > 1{,}64) = 0{,}10; \quad P(X > 1{,}64) = 0{,}05;$$
$$P(|X| > 1{,}96) = 0{,}05; \quad P(X > 1{,}96) = 0{,}025;$$
$$P(|X| > 2{,}58) = 0{,}01; \quad P(X > 2{,}58) = 0{,}005.$$

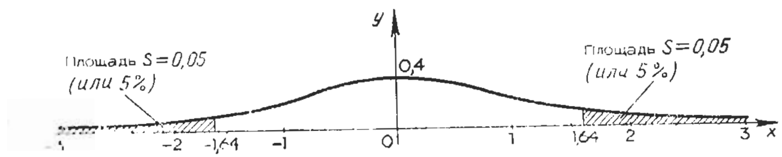


Рис. 3

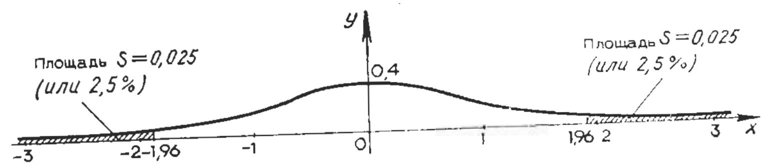


Рис. 4

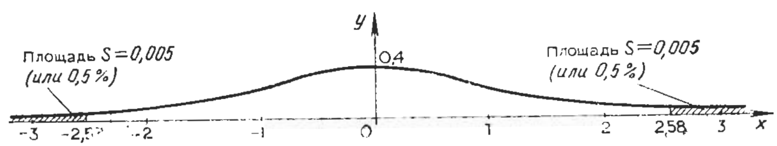


Рис. 5

Числа 1,64; 1,96; 2,58 являются квантилями нормального распределения, которые отвечают вероятностям $p=0{,}95$; $p=0{,}975$; $p=0{,}995$.

в) Распределение χ^2 (хи-квадрат). Такое название получил закон распределения неотрицательной случайной величины, плотность распределения которой определяется формулой[1]

$$p(x)=\frac{1}{2^{\frac{\nu}{2}}\Gamma\left(\frac{\nu}{2}\right)}\cdot x^{\frac{\nu}{2}-1}\cdot e^{-\frac{x}{2}} \quad \text{для} \quad x>0. \qquad (11)$$

Параметр ν, от которого зависит выражение $p(x)$, называется числом степеней свободы и может принимать целые значения 1, 2, 3, ... Можно показать, что математическое ожидание (среднее значение) случайной величины, подчиняющейся распределению χ^2, равно именно ν, в чем и заключается вероятностный смысл этого параметра.

Приведем графики (см. рис. 6, 7, 8) плотности распределения (11) для значений $\nu=1$, $\nu=2$, $\nu\geqslant3$, показывающие влияние ν на форму кривой вероятностей.

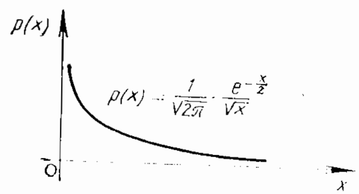


Рис. 6

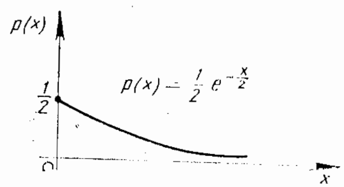


Рис. 7

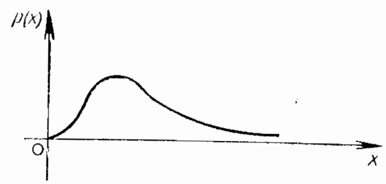


Рис. 8

5. *Эмпирическая функция распределения случайной величины.* Поскольку проверка статистических гипотез осуществляется на основе выборочных наблюдений над случайными величинами — величинами, характеризующими изучаемые генеральные совокупности, наряду с общим понятием функции распределения будет использоваться понятие выборочной, или эмпирической, функции распределения. Определяется это понятие следующим образом.

Пусть $x_1, x_2, ..., x_n$ — наблюдаемые в выборке объема n значения случайной величины X, расположенные в порядке возрастания. Будем предполагать наблюдения взаимно независимыми. Положим

$$S_n(x)=\frac{k(x)}{n}, \qquad (12)$$

где $k(x)$ — число наблюдаемых значений x_i, которые меньше или равны x. Величина $k(x)$ принимает значения 0, 1, 2, ..., n.

Приведем общий вид графика функции (12), считая для простоты, что все значения x_i различны (см. рис. 9).

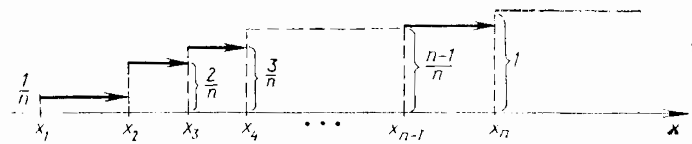


Рис. 9

Связь между теоретической функцией распределения, определяемой согласно (1), и эмпирической функцией распределения (12) устанавливается следующей теоремой.

Теорема (В. И. Гливенко, 1933). С вероятностью 1

$$\sup_{-\infty<x<\infty}|S_n(x)-F(x)|\to 0 \quad \text{при} \quad n\to\infty^*.$$

Иными словами, с увеличением объема выборочных наблюдений эмпирическая функция распределения приближается (равномерно относительно аргумента x) к теоретической функции распределения, причем этот вывод справедлив с вероятностью 1. Следует, однако, заметить, что теорема Гливенко справедлива только при том условии, что наблюдаемые значения случайной величины измерены с помощью шкал количественных измерений (см. раздел 1).

[1] См.: *Митропольский А. К.* Техника статистических вычислений. М., 1971, с. 247, 325.

* Символ $\sup A$ означает верхнюю грань числового множества A.

1 ОСНОВНЫЕ ТИПЫ ИЗМЕРЕНИЙ В ПЕДАГОГИКЕ

Измерение в самом широком смысле может быть определено как приписывание чисел объектам или событиям согласно определенным правилам. Эти правила должны устанавливать соответствие между некоторыми свойствами рассматриваемых объектов и некоторыми свойствами чисел.

Значение измерения состоит в том, что операции с числами, приписанными объектам, позволяют сравнивать между собой эти объекты по состоянию измеренного свойства. Правила, на основе которых объектам приписываются числа, отражают различные свойства этих объектов. Поэтому в зависимости от целей исследования и возможностей исследователя выбирается то или иное правило, использование которого позволяет измерить нужное свойство объектов. Каждое такое правило порождает свой тип измерения — свою «измерительную шкалу».

Можно выделить четыре основных вида измерительных шкал, получивших широкое распространение. Согласно предложенной С. С. Стивенсом [12, 14, 17] и принятой в настоящее время терминологии, они получили следующие названия: шкала наименований, шкала порядка (ранговая шкала), интервальная шкала, шкала отношений. Измерения, осуществляемые с помощью двух первых шкал, считаются качественными, а измерения, осуществляемые с помощью двух последних шкал, — количественными. Принято шкалы, приводящие к качественным измерениям, называть дискретными; шкалы, приводящие к количественным измерениям, — непрерывными.

В каждой шкале строго определены свойства чисел, которые приписываются объектам. В этом смысле упомянутые выше шкалы можно расположить в определенном порядке таким образом, что в каждой последующей шкале используются кроме свойств чисел предыдущей шкалы и другие свойства чисел. Поэтому чем больше порядок шкалы, тем больше арифметических действий разрешается проводить над числами, приписанными объектам.

Рассмотрим каждую из четырех шкал.

Шкала наименований. Построение шкалы наименований и ее использование при измерении некоторого свойства рассматриваемых объектов возможно, если установлен критерий, позволяющий по состоянию измеряемого свойства распределить эти объекты на несколько классов, причем каждый объект должен при этом попасть только в один класс.

Всем объектам, попавшим в один и тот же класс, приписывается одно и то же число. Объектам другого класса приписывается другое число. В этих условиях числа, приписанные объектам, обладают только свойствами равенства и различия. Так, если нескольким объектам приписано одно и то же число, то эти объекты одинаковы (равны) по состоянию измеряемого свойства. Если объектам приписаны разные числа, то они различны по состоянию измеряемого свойства.

Измерением по шкале наименований является, например, классификация учащихся по полу и приписывание числа 1 каждой девочке и числа 2 каждому мальчику рассматриваемого коллектива. В данном примере классификация учащихся проводится по одному признаку. Такая классификация некоторой совокупности объектов позволяет выявить структуру этой совокупности по рассматриваемому признаку.

Другим примером измерения по шкале наименований является классификация учащихся по состоянию двух признаков: пол и успешность выполнения контрольного задания. В состоянии каждого признака выделим в данном случае по две градации: девочка, мальчик и верный ответ, неверный ответ. В результате по состоянию двух признаков коллектив учащихся разделится на четыре класса: девочки, верно выполнившие задание; девочки, неверно выполнившие задание; мальчики, верно выполнившие задание; мальчики неверно выполнившие задание. Далее всем объектам первого класса припишем, например, число 1, объектам второго класса — число 2, третьего класса — число 3, четвертого — число 4. Классификация некоторой совокупности объектов по нескольким признакам позволяет не только выявить структуру совокупности, но и получить данные, необходимые для выяснения связи двух и более признаков. В приведенном примере такими признаками являются пол учащихся и выполнение контрольного задания.

В шкале наименований числа, приписанные объектам, используются в качестве ярлыков — их можно заменить любыми символами: буквами, значками. В связи с этим количественная обработка экспериментальных данных проводится не с самими этими числами, а с числами, характеризующими количества объектов, попавших в каждый класс. Методы, применяемые для такой обработки, называют иногда статистикой качественных признаков.

Шкала наименований допускает несколько статистических операций. Прежде всего возможно подсчитать число объектов в каждом классе и найти простое или процентное отношение этого числа к общему числу рассматриваемых объектов, т. е. подсчитать абсолютную, относительную, или процентную, частоты класса. На основе полученных результатов можно выделить класс с наибольшим числом объектов (наибольшей абсолютной частотой), который принято называть м о д о й. Если продолжить наблюдения, не изменяя условий, при которых была определена мода, то мода будет представлять объекты, появление которых можно ожидать с большей вероятностью по сравнению с появлением других объектов. Например, при выяснении возраста учащихся, посещавших автомобильный кружок, оказалось, что больше всего членов кружка попало в группу учащихся 14—15 лет. Это значение является модой, и вероятность появления новых членов кружка 14—15 лет больше, чем вероятность появления учащихся старше 15 и младше 14 лет.

Несмотря на кажущуюся примитивность шкалы наименований, измерения по этой шкале могут быть использованы для проверки некоторых статистических гипотез (см. разделы 4 и 5) и для вычисления показателей корреляции качественных признаков[1].

---

[1] Вопросы стохастической связи и корреляции в данной работе не рассматриваются.

Шкала порядка. Измерения по шкале порядка обладают всеми свойствами и возможностями измерений по шкале наименований и некоторыми новыми свойствами. Измерение по этой шкале возможно, если существует разумный критерий, задающий на множестве объектов, обладающих изучаемым свойством, отношения равенства и порядка применительно к состоянию данного свойства, т. е. для любых двух объектов A и B такой критерий позволяет установить, во-первых, истинность одного из следующих утверждений: $A=B$ или $A\neq B$; во-вторых, в случае $A\neq B$ установить истинность одного из следующих утверждений: $A>B$ или $A<B$. Кроме того, для любых трех объектов A, B, C, таких, что $A>B$ и $B>C$, согласно этому критерию должно быть верным неравенство $A>C$*; для любых трех объектов A, B, C, таких, что $A=B$ и $B=C$, согласно критерию должно быть верным равенство $A=C$.

Пусть, например, установлен некоторый критерий уровня усвоения понятия «функция». Согласно этому критерию ученик A усвоил понятие лучше, чем ученик $B (A>B)$, а B — лучше ученика $C (B>C)$. Тогда измерение усвоения по шкале порядка возможно, если согласно этому критерию ученик A усвоил данное понятие лучше, чем C.

Свойство транзитивности отношения порядка, которое кажется столь «естественным», вовсе не обязательно выполняется автоматически при попытках упорядочить объекты. Допустим, например, что мы хотим упорядочить группу шахматистов по результатам сыгранных ими партий (каждая пара шахматистов играет одну партию), условимся считать, что игрок A лучше игрока $B (A>B)$, если A выиграл партию у B. Свойство транзитивности при таком упорядочении не имеет места. Действительно, если A выиграл у B, а B выиграл у C, то это еще не означает, что A обязательно выиграет у C.

В педагогических исследованиях постоянно приходится иметь дело с измерением уровня знаний, умений и навыков, способностей, мнений, склонностей, интересов и др.

---

* Это свойство отношения порядка называется транзитивностью.

При измерении результатов обучения учащихся шкала порядка может быть получена всякий раз, когда имеется критерий, позволяющий расположить учащихся по степени увеличения (уменьшения) измеряемого признака, если при этом невозможно определить, на сколько равных единиц по состоянию признака один объект больше (меньше) другого.

Количество различных чисел — балловых оценок или рангов, которые приписываются объектам, зависит от особенностей критерия, используемого для определения различного состояния измеряемого свойства в этих объектах. Если критерий позволяет установить, например, 6 различных состояний, то шкала порядка должна быть составлена из 6 чисел, представляющих монотонно возрастающую (убывающую) последовательность[1]. Поскольку шкала порядка устанавливает только отношения равенства и порядка, то для приписывания объектам могут быть использованы любые шесть чисел, которые можно расположить в порядке возрастания (убывания), например: 1, 2, 3, 4, 5, 6, или 2, 3, 8, 12, 20, 24, или 0, 4, 6, 8, 11, 13.

В связи с этим свойством шкалы вывод, сделанный на основе измерений некоторого признака в группе учащихся по n-балльной шкале порядка, будет объективным только в том случае, если он всякий раз следует из экспериментальных данных, которым в установленном порядке приписываются любые n чисел, образующие монотонную (возрастающую или убывающую) последовательность.

Особенности шкалы порядка таковы, что выводы, полученные на основе результатов арифметических действий с измерениями, сделанными по этой шкале, существенно зависят от выбора балловых оценок, т. е. они (выводы) в некоторых случаях могут отражать не объективные свойства рассматриваемых явлений, а случайные свойства, которые являются следствием особенностей принятой системы балловых оценок. Поэтому основное ограничение измерений по шкале порядка состоит в том, что с числами (баллами, рангами), приписанными объектам, нельзя производить арифметических действий: вычислять суммы, находить средние значения, дисперсии и т. п.[1]

Однако существуют некоторые характеристики группы объектов, которые остаются неизменными при любой n-балльной системе оценок состояния измеряемого свойства. Кроме моды такой характеристикой также является медианный объект, который превосходит по состоянию измеряемого свойства 50% объектов рассматриваемой группы и меньше которого остальные 50% объектов. Число, приписываемое данному объекту, называется м е д и а н о й и принимается за меру центральной тенденции группы объектов.

Дисперсия (мера рассеивания значений измеряемого признака в выборке) описывается в рамках измерений по шкале порядка посредством чисел, получивших название процентилей, из которых наиболее часто используются центили, децили, квартили. Центили представляют ряд из 99 чисел, приписанных таким 99 объектам рассматриваемой выборки, что первый объект превышает по состоянию измеряемого свойства 1% объектов выборки, второй объект — 2%, ..., последний — 99%. Децили представляют ряд из 10 чисел, приписанных таким 10 объектам выборки, что первый из них превышает 10% объектов выборки, второй — 20%, ..., последний — 90%. Определение центилей и децилей имеет смысл для больших выборок.

Квартили — три числа, приписанные таким объектам выборки, что первый из них превышает 25% объектов выборки, второй (медиана) — 50%, третий — 75%.

Выше было указано, с какой осторожностью следует пользоваться шкалами порядка, т. е. балловыми (ранговыми) оценками. Однако это вовсе не означает, что балловые оценки не могут служить основой для достаточно объективных и достоверных педагогических выводов. Наоборот, в математической статистике разработано несколько специальных критериев (так называемые ранговые критерии), которые позволяют проверять статистические гипотезы именно на базе балло-

[1] Т. е. каждый последующий член последовательности больше (меньше) предыдущего.

[1] Примеры необъективных выводов, сделанных на основе сравнения средних баллов двух групп учащихся, можно найти в статье Л. М. Фридмана «О корректном применении статистических методов в психолого-педагогических исследованиях». — «Советская педагогика», 1971, № 3.

вых (ранговых) оценок. Описание некоторых ранговых критериев и примеры их использования в педагогических исследованиях будут даны в разделах 4 и 5.

Интервальная шкала. Интервальную шкалу, или шкалу равных единиц, мы можем получить, если с помощью некоторого критерия сумеем измерить интервал между объектами в состоянии изучаемого свойства, т. е. установить, на сколько единиц один объект больше (меньше) другого. В такой шкале возможны арифметические операции над числами, приписанными объектам. Например, можно говорить, что 15—12=9—6, чего нельзя было утверждать, если измерения получены по шкале порядка. Ограниченность интервальной шкалы заключается в произвольном выборе нулевой точки (начала отсчета), и поэтому мы не можем определить, во сколько раз один объект больше другого. Здесь уместна аналогия с измерением температуры по шкале Цельсия. Например, показание термометра в 20° не означает вдвое большую температуру по сравнению с показанием в 10°. Тот же недочет характерен и для шкалы времени, которая также является интервальной шкалой, не имеющей строго фиксированного начала отсчета. С аналогичной ситуацией мы сталкиваемся в социальных науках. Абсолютный нуль (полное отсутствие изучаемого качества) при измерении знаний, умственного развития, социальных установок и т. п. неизвестен и, по-видимому, вообще не имеет смысла, подобно абсолютному нулю времени[1]. Поэтому в зависимости от характера рассматриваемой данной социальной наукой проблемы приходится принимать тот или иной условный нулевой уровень. Например, когда мы оцениваем выполнение учащимся какого-либо задания по числу верных (неверных) ответов на вопросы данного задания, то нуль верных ответов не означает полного отсутствия знаний, проверяемых заданием, у этого учащегося.

В отличие от естественных наук в социальных науках (в том числе и в педагогике), за исключением шкалы времени, которая хотя и используется достаточно широко, но играет в теории измерений особую роль,

[1] Что касается температуры, то, как известно, физикам удалось проникнуть более глубоко в сущность понятия теплоты и определить абсолютный нуль температуры (шкала Кельвина).

в настоящее время нет измерительных шкал интервального типа[1].

Ни одна из предлагавшихся многими психологами и социологами шкал для количественных измерений (шкала Терстоуна, шкала Векслера, шкала Ликерта, шкалограммный анализ Гутмана, латентный анализ Лазарсфельда[2], шкалы z-оценок, стенайнов и др.) математически не обоснована. Более того, многие из этих шкал строятся так, чтобы «любыми средствами» получить нормальное распределение результатов измерения. Такой подход является следствием неправильных представлений о гауссовском нормальном законе как «универсальном» законе для стохастических явлений.

Заканчивая краткое описание интервальной шкалы, отметим, что она позволяет применять к результатам измерений почти все статистические операции. Исключение составляет, например, вычисление часто употребляемого коэффициента вариации, который определяется по формуле $V = \frac{\sigma}{\bar{x}} \cdot 100\%$, где $\bar{x}$ — среднее значение выборки, σ^2 — дисперсия выборки. Действительно, если начало отсчета на шкале выбрать так, чтобы $\bar{x}=0$, то выражение для V теряет смысл.

Шкала отношений. Шкала отношений позволяет определить не только, на сколько больше (меньше) один объект другого в отношении измеряемого качества, но и во сколько раз больше (меньше). В такой шкале можно устанавливать равенство отношений чисел, приписываемых объектам. Например, можно говорить, что $\frac{6}{9} = \frac{2}{3}$. Это свойство шкалы вытекает из фиксированного положения нуля (факт, который можно строго математически доказать). Примерами таких шкал являются шкалы длины, веса, шкала температуры Кельвина и т. п.

Любая интервальная шкала превращается в шкалу отношений, если строго фиксировать начало отсчета (см., например, об измерении некоторого свойства числом

[1] Имеются в виду собственно социальные шкалы, ибо в ряде случаев социальные науки, например экономика и психология, используют физические единицы измерения.

[2] См.: Математические методы в современной буржуазной социологии. М., 1966.

верных ответов на с. 24—25). В шкале отношений применимы к полученным измерениям все арифметические операции и, следовательно, все понятия и методы математической статистики.

Таким образом, становится очевидным, что переход от одной шкалы измерения к другой (начиная со шкалы наименований) сопровождается расширением класса допустимых для полученных измерений математических операций и методов. В этом смысле наилучшей является шкала отношений. Однако изложенные выше соображения позволяют считать, что для решения проблем, стоящих перед социальными науками, вполне достаточна интервальная шкала[1].

Поскольку глубокое изучение социальных явлений и процессов невозможно сегодня без применения мощного математического аппарата, и в первую очередь методов математической статистики, проблема количественных измерений в социальных науках (и, в частности, в педагогике) заключается в переходе от дискретных шкал порядка к непрерывным интервальным шкалам. Интересное с теоретической точки зрения решение этой проблемы, основанное на отыскании «объективных весов» балловых оценок методами «распознавания образов», найдено Б. Битинасом[2]. К сожалению, предлагаемая в его работе методика измерения математически не обоснована и настолько сложна, что требует использования ЭВМ.

В 1973 г. одним из авторов была предложена вероятностная модель процесса решения субъектом (учащимся, испытуемым) любой интеллектуальной задачи. Эта модель позволяет, в частности, рассматривать шкалу оценок некоторого свойства учащихся по числу верных (неверных) ответов на ряд предлагаемых вопросов как дискретное извлечение из некоторой непрерывной интервальной шкалы[3].

[1] По этому поводу см.: *Фресс П., Пиаже Ж. (ред.).* Экспериментальная психология. М., 1966.

[2] См.: *Битинас Б.* Многомерный анализ в педагогике и педагогической психологии. Вильнюс, 1971.

[3] См.: *Грабарь М. И.* Об одной вероятностной модели построения шкалы количественных измерений в психолого-педагогических исследованиях. VI сессия семинара по методологии педагогики и методике педагогических исследований (тезисы докладов). М., 1973.

Таким образом, применение статистических методов, допускаемых шкалами интервального типа к оценкам, определяемым числом верных (неверных) ответов учащихся на несколько контрольных вопросов, можно считать в математическом отношении достаточно обоснованным. Эти соображения будут неоднократно использоваться в последующих разделах данной работы. При этом необходимо иметь в виду, что такой вывод справедлив только в том случае, когда характер вопросов отвечает двум условиям.

Во-первых, все вопросы должны быть направлены на изучение состояния одного и того же психического свойства учащихся (например, какого-либо умения или навыка). Во-вторых, каждый вопрос должен представлять некоторую интеллектуальную задачу, для решения которой учащемуся недостаточно только воспроизвести изученный материал, а требуется рассуждать, сравнивать, делать выводы по готовому алгоритму или в соответствии с им (учащимся) самим найденным методом решения поставленной задачи.

В качестве примера контрольных работ, которые не отвечают указанным выше требованиям (но весьма характерны для педагогических исследований по изучению знаний учащихся), приведем следующий.

Контрольная работа, проверяющая усвоение материала IV класса, которая состоит из:

а) задачи, решаемой составлением уравнения;

б) примера на вычисление значения выражения;

в) примера на тождественные преобразования;

г) вопроса, проверяющего усвоение некоторого геометрического понятия;

д) вопроса, проверяющего только память учащегося (типа «Дайте определение...»).

В этом случае объединение в одно целое заданий, проверяющих совершенно различные, в некоторых случаях никак не связанные между собой знания и умения (пример на вычисление и знание геометрического понятия и др.), нарушает первое требование. Включение вопроса д) нарушает выполнение второго условия, так как по своему характеру этот вопрос не является интеллектуальной задачей.

2 СТАТИСТИЧЕСКИЕ ГИПОТЕЗЫ В ПЕДАГОГИЧЕСКИХ ИССЛЕДОВАНИЯХ

Под статистической гипотезой понимается любое предположение о свойствах случайных величин или событий.

Не претендуя на полную классификацию гипотез[1], возникающих в педагогических исследованиях, остановимся на некоторых основных типах.

1) Гипотезы о типах вероятностных законов распределения случайных величин, характеризующих изучаемое свойство явления или процесса.

Применительно к педагогическим исследованиям примерами гипотез данного типа являются следующие:

а) оценки учащихся по некоторому тесту, проверяющему состояние вычислительных навыков, имеют нормальное распределение в совокупности учащихся пятых классов Москвы;

б) процент верных ответов учащихся при решении 10 логических задач имеет экспоненциальное распределение в совокупности учащихся восьмых классов экспериментальных школ.

В общем виде содержание гипотез первого типа, возникающих в педагогических экспериментах, можно формулировать так:

Некоторое свойство педагогического явления имеет определенный закон распределения (нормальный, экспоненциальный, Пуассона и т. д.).

2) Гипотезы о свойствах тех или других числовых параметров (средних значений, медиан, дисперсий и др.), характеризующих изучаемые случайные величины.

Примерами данного типа гипотез в педагогике могут служить:

[1] С различными типами и возможным содержанием гипотез в экспериментальной педагогике можно познакомиться в работе М. Н. Скаткина «Беседа с приступающими к работе над диссертацией» (М., 1967).

а) среднее число верных ответов на 12 контрольных заданий, проверяющих усвоение понятия «производная» в девятых экспериментальных классах, не меньше 9;

б) процент положительных оценок («3», «4», «5») за выполнение контрольной работы, проверяющей навык тождественных преобразований, не меньше 75% в совокупности учащихся пятых классов школ РСФСР;

в) среднее квадратическое отклонение (σ) числа верных ответов учащихся на вопросы определенного теста находится в пределах от 2 до 5 верных ответов в совокупности учащихся седьмых классов Первомайского района Москвы.

В общем виде применительно к педагогическим исследованиям гипотезы данного типа можно сформулировать так:

Значение параметра, характеризующего некоторое свойство изучаемого педагогического явления, не меньше (не больше) некоторого заданного значения или заключается в заданных пределах.

3) Гипотезы о стохастической (вероятностной) зависимости[1] двух и более признаков (факторов), характеризующих различные стороны рассматриваемого явления или процесса.

В педагогических экспериментах примерами гипотез данного типа могут быть следующие:

а) успеваемость класса стохастически зависит от уровня обучаемости учащихся;

б) усвоение разделов курса алгебры VII класса, изложенных на более современной математической основе, стохастически зависит от квалификации учителя.

В общем виде содержание гипотез данного типа применительно к педагогике можно сформулировать так:

Два или более свойств рассматриваемого педагогического явления стохастически зависимы, и зависимость эта подчиняется определенному закону (например, линейному); некоторый фактор или факторы оказывают

[1] Стохастическая зависимость двух (или более) факторов отличается от обычной функциональной зависимости тем, что изменение состояния одного из факторов не определяет однозначно изменения состояний другого фактора (других факторов), а определяет лишь вероятностный закон распределения этих состояний.

влияние на изучаемое свойство педагогического явления, и эта стохастическая зависимость подчиняется определенному закону.

4) Гипотезы о равенстве или различии законов распределения случайных величин, характеризующих изучаемое свойство в двух и более совокупностях рассматриваемых явлений.

Примерами такого типа гипотез могут служить:

а) усвоение начального курса математики (I—III классы) не имеет существенных различий у учащихся, начавших обучение с 6 лет или с 7 лет;

б) проблемное обучение в I классе эффективнее в в отношении общего развития учащихся по сравнению с традиционной методикой обучения;

в) учащиеся VIII класса усваивают раздел «Функция» по новой программе не хуже, чем учащиеся VIII—IX классов, обучавшиеся по старой программе.

Применительно к ситуациям, имеющим место в педагогических исследованиях, в общем виде содержание гипотез четвертого типа можно сформулировать следующим образом:

Состояние одного и того же свойства имеет одинаковое или различное распределение в каждой из двух или более совокупностей учащихся, отличающихся содержанием, методом или организацией обучения или социальной средой.

Проверка гипотез первого типа проводится с помощью методов математической статистики, получивших название «критерии согласия». Использование этих критериев возможно только на основе количественных измерений изучаемого свойства педагогического явления в выборке.

Проверка гипотез второго типа в основном осуществляется с помощью методов, получивших название параметрических (критерий Стьюдента, критерий Снедекора — Фишера и др.) и опирающихся только на количественные измерения.

Установление факта наличия или отсутствия связи двух и более признаков (проверка гипотез третьего типа) возможно даже при качественных измерениях изучаемых признаков с помощью определения значения коэффициентов Юла, Кендалла, Спирмена и др. Однако более глубокий анализ стохастических связей, а именно установление тесноты и типа этой связи, осуществляется методами корреляционного, регрессионного и дисперсионного анализов, которые применяются только на основе количественных измерений.

Проверка гипотез четвертого типа проводится с помощью методов, получивших название критериев значимости. Часть из них применима только при количественных измерениях изучаемого свойства, остальные можно использовать также и при качественных измерениях.

В соответствии со своим названием настоящая работа посвящена рассмотрению так называемых непараметрических методов математической статистики. Использование этих методов возможно при проверке гипотез двух последних типов. В настоящей работе авторы ограничились рассмотрением непараметрических методов, применимых в основном при проверке гипотез четвертого типа, т. е. гипотез о принадлежности наблюдений двух выборок к одной и той же совокупности или к различным совокупностям, имеющим одинаковый (или различный) закон распределения изучаемого свойства. Следует отметить, что в настоящее время для большинства педагогических исследований характерна постановка именно таких гипотез.

Проверка статистических гипотез такого типа осуществляется на основе сравнения результатов, полученных в выборках, каждая из которых составлена из членов одной из рассматриваемых совокупностей.

3 ОБЩИЕ ПРИНЦИПЫ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ

3.1. Нулевая гипотеза

Рассмотрим какую-либо подлежащую проверке статистическую гипотезу, которую часто называют также нулевой гипотезой, и обозначим ее через H_0. Примеры различного рода статистических гипотез в педагогике были приведены в разделе 2.

Общая теория проверки статистических гипотез составляет весьма сложный раздел математической статистики, изложение которого выходит за рамки данной работы. Поэтому мы ограничимся некоторыми общими положениями, необходимыми для понимания последующих практических рекомендаций. В частности, в качестве нулевой гипотезы H_0 мы будем рассматривать гипотезу, утверждающую, что изучаемые выборки взяты из генеральных совокупностей с одинаковым законом распределения, а различие в результатах выборок объясняется чисто случайными причинами. Примером нулевой гипотезы такого типа в педагогических исследованиях является утверждение о том, что различие в результатах выполнения двумя группами учащихся одной и той же контрольной работы вызвано случайными причинами, а на самом деле уровень выполнения этой работы для той и другой группы учащихся одинаков.

Проверка нулевой гипотезы H_0 обычно осуществляется путем сравнения ее с другой гипотезой H_1, которая называется альтернативной (в общей теории проверки гипотез рассматривается, как правило, множество альтернатив, принадлежащих тому или другому классу статистических гипотез).

Так, для упомянутого выше примера гипотезы H_0 в педагогике одна из возможных альтернатив H_1 означает, что уровни выполнения работы в двух группах учащихся различны и это различие определяется влиянием неслучайных факторов, например тех или других методов обучения.

На основе определенных правил (о чем речь пойдет ниже) надлежит сделать выбор: либо отвергнуть H_0 и тем самым принять H_1, либо принять H_0 и тем самым отвергнуть H_1.

3.2. Принцип практической невозможности. Уровни значимости и уровни достоверности

В математической статистике проверка тех или иных гипотез о случайных величинах и событиях основана на принципе так называемой практической невозможности. Заключается этот принцип в следующем. Задается заранее некоторая, обычно весьма малая вероятность α (например, $\alpha=0{,}1$; $\alpha=0{,}05$; $\alpha=0{,}01$ и т. д.), именуемая уровнем значимости. При этом случайные события, вероятность которых меньше или равна α, считаются практически невозможными. Следовательно, если некоторое событие с вероятностью, меньшей или равной α, все-таки произошло, то наступление такого события не может вызываться только случайными причинами. Это событие следует рассматривать как «неслучайное». Оно привлекает наше внимание, становится для нас, так сказать, «значимым». При этом чем меньше вероятность осуществившегося события, тем больше его «неслучайность» и тем важнее раскрыть причины этой неслучайности. Для того чтобы нагляднее представить роль уровня значимости в оценке неслучайности события, можно считать, что уровень значимости, выраженный в процентах ($\alpha\cdot 100\%$), показывает, сколько раз в среднем в 100 случаях мы рискуем ошибиться, объявив изучаемое событие неслучайным. Так, уровень значимости $\alpha=0{,}05$, т. е. 5%-й уровень значимости, допускает ошибку в среднем в 5 случаях из 100.

Рассмотрим величину $\theta=1-\alpha$. Очевидно, что $\theta\times\times 100\%$ показывает, сколько раз в среднем в 100 случаях будет справедлив вывод о неслучайности события, поэтому величина θ называется уровнем достоверности

При более высоком, чем 5%-й, уровне значимости (например, 10%-м) большее число событий нельзя рассматривать как случайные. Однако достоверность такого вывода ниже (90% против 95%). Наоборот, более низкий уровень значимости (например, 1%-й) приводит к более осторожным, но и более достоверным (уровень достоверности 99%) выводам. Роль уровня значимости для проверки гипотез раскрывается в пунктах 3.3 и 3.4.

3.3. Критерии для проверки нулевой гипотезы

Всякое правило, на основе которого нулевая гипотеза отклоняется или принимается, называется критерием для проверки этой гипотезы. В упрощенном виде схема построения критериев для проверки нулевой гипотезы указанного в пункте 3.1 типа такова.

Пусть рассматриваются две или более выборки любых объемов. В той или иной математической форме определяется величина, отражающая различие результатов этих выборок. Обозначим такую величину через S. Допустим, что предположение о справедливости нулевой гипотезы H_0 позволяет установить такой теоретический закон распределения величины S, который зависит от объемов выборок, но не зависит от самих результатов этих выборок. Тогда величина S может быть использована для построения критерия проверки H_0 и называется в этом случае статистикой критерия. Строится критерий следующим образом. Если закон распределения статистики S известен, то можно выбрать такую часть множества всех возможных значений S, что вероятность попадания конкретного значения S в эту часть не превосходит принятого уровня значимости. Такая часть множества всех возможных значений статистики называется критической областью критерия, отвечающей данному уровню значимости. Из предыдущего следует, что выбор критической области не зависит от результатов изучаемых выборок. Пусть критическая область для принятого уровня значимости выбрана. По результатам конкретных выборок находим числовое значение статистики S и рассматриваем две возможности.

1. Найденное значение S попало в критическую область.

Если предположить, что нулевая гипотеза H_0 справедлива, то отсюда, согласно сказанному выше, следует, что осуществилось событие (попадание S в критическую область), вероятность которого не превосходит уровня значимости, т. е. осуществилось маловероятное событие, которое на принятом уровне значимости мы должны считать неслучайным, или значимым, событием.

Вместе с тем справедливость H_0 означает, что наблюдаемое различие результатов выборок (отраженное в статистике S) определяется чисто случайными причинами. Таким образом, H_0 вступает в противоречие с данными наблюдений, и потому H_0 отвергается на принятом уровне значимости. Тем самым мы принимаем альтернативу H_1, которая означает, что наблюдаемое различие результатов выборок не случайно, или, как говорят, это различие статистически значимо.

2. Значение S не попало в критическую область.

В этом случае нет оснований для отклонения H_0 (нулевая гипотеза не противоречит данным наблюдений), и она принимается на данном уровне значимости.

В дальнейшем будут рассматриваться только критические области, определяемые одним или двумя неравенствами следующих типов:

$$1)\ S > S';\quad 2)\ S < S'';\quad 3)\ S < S'',\ \ S > S'\ \ (\text{при } S'' < S').$$

Первые две критические области называются односторонними, а третья — двусторонней критической областью. Числа S', S'' называются критическими значениями статистики или границами значимости. Каждому уровню значимости α отвечают свои критические значения.

На рис. 10, 11, 12 показаны критические области одностороннего и двустороннего типа. Возможные значения статистики располагаются на числовой оси S. Закон распределения статистики изображен кривой — графиком плотности распределения. Площади под кривой равны вероятностям попадания S на соответствующие промежутки числовой оси.

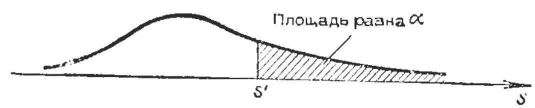


Рис. 10

2 Зак. 224

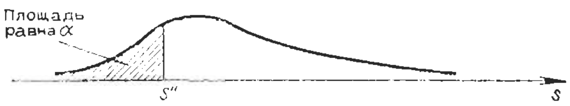


Рис. 11

Рис. 12

Из рис. 12 ясно, что двусторонняя критическая область для уровня значимости α может быть получена объединением односторонних критических областей, отвечающих уровню значимости $\frac{\alpha}{2}$.

При использовании специальных критических областей одностороннего и двустороннего типов правило для отклонения нулевой гипотезы формулируется особенно просто. Если используется односторонняя область, изображенная на рис. 10, то нулевая гипотеза отклоняется всякий раз, когда $S > S'$. Для односторонней области второго типа H_0 отклоняется, когда $S < S''$. Наконец, в случае двусторонней критической области H_0 отклоняется как в случае, когда $S < S''$, так и в случае $S > S'$[1].

3.4. Ошибки первого и второго рода

Поскольку принятие того или другого решения при проверке статистических гипотез основано на вероятностных суждениях, иными словами, любой вывод справедлив лишь с определенной вероятностью, решение об отклонении или принятии нулевой гипотезы может оказаться ошибочным. Логически возможны следующие случаи:

[1] Выбор той или другой критической области связан с рассматриваемыми альтернативами (см. далее разделы 4 и 5).

(I) H_0 верна, но мы отклоняем H_0;
(II) H_0 верна, и мы принимаем H_0;
(III) H_0 неверна, но мы принимаем H_0;
(IV) H_0 неверна, и мы отклоняем H_0.

Случаи (II) и (IV) отвечают верному решению, а случаи (I) и (III) — ошибочному. Говорят, что в случае (I) мы допускаем ошибку первого рода, а в случае (III) — ошибку второго рода.

Важно отметить, что, согласно изложенной выше схеме построения критериев для проверки нулевой гипотезы, вероятность ошибки первого рода всегда не превосходит принятого уровня значимости.

В самом деле, мы отклоняем H_0 всякий раз, когда наблюдаемое значение статистики S попадает в критическую область. Если H_0 на самом деле верна, то вероятность попадания S в критическую область, т. е. вероятность нашего ошибочного решения, не превосходит уровня значимости.

Таким образом, роль уровня значимости, по существу, заключается в том, что он обеспечивает заранее малую вероятность (а значит, относительную редкость) ошибочных решений при отклонении нулевой гипотезы. Например, при 5%-м уровне значимости, отклоняя нулевую гипотезу, мы совершаем ошибку в среднем не более чем в 5 случаях из 100. Из сказанного следует также, что вероятность верного решения в случае (II) всегда не меньше уровня достоверности.

Выбор того или другого уровня значимости (и следовательно, уровня достоверности) не может быть осуществлен на основании каких-либо общих теоретических соображений. Он в значительной степени определяется целями и особенностями исследования применительно к конкретным областям науки. Суть дела в том, насколько нежелателен (или допустим) риск ошибки при отклонении верной нулевой гипотезы. Например, крайне нежелателен такой риск при проверке статистических гипотез с помощью дорогостоящих или трудно осуществляемых экспериментов. В большинстве современных научных исследований используется либо 5%-й, либо 1%-й уровень значимости. В педагогических исследованиях, не нуждающихся в очень высоком уровне достоверности, представляется разумным принять 5%-й уровень значимости. Именно такой уровень и принят в настоящей работе. Однако в ряде случаев для сопоставления выводов будут рассматриваться и другие уровни значимости. В дальнейшем для простоты мы будем обозначать вероятность ошибки первого рода той же буквой α, что и уровень значимости. Вероятность ошибки второго рода будет обозначаться буквой β. Величина $\Phi = 1 - \beta$ называется мощностью критерия и представляет собой вероятность отклонения неверной нулевой гипотезы, т. е. вероятность правильного решения в случае (IV).

В отличие от ошибки первого рода вероятность ошибки второго рода β далеко не всегда удается ограничить заранее малым числом. Поэтому с целью проверки гипотез на данном уровне значимости α стараются выбрать такой критерий, для которого величина β была бы возможно меньшей, т. е. выбрать возможно более мощный критерий. Классические результаты в этом направлении принадлежат Ю. Нейману и Е. Пирсону [10, 11, 13]. Поскольку изложение общей теории проверки статистических гипотез не входит в задачу данной работы, мы ограничимся следующим определением, позволяющим выделить такой класс критериев, для которых при некоторых условиях можно гарантировать малую вероятность ошибки второго рода. Критерий для проверки нулевой гипотезы называется **состоятельным**, если при неограниченном возрастании объемов рассматриваемых выборок вероятность ошибки второго рода $\beta \longrightarrow 0$ (стремится к нулю). Таким образом, используя для проверки статистических гипотез состоятельные критерии, мы можем заранее обеспечить малую вероятность β ошибки второго рода, т. е. ошибочного принятия неверной нулевой гипотезы, путем изучения выборок достаточно большого объема.

3.5. Параметрические и непараметрические критерии

В пункте 3.3 было отмечено, что теоретический закон распределения статистики критерия не зависит от результатов изучаемых выборок. Однако этот закон распределения может зависеть от законов распределения рассматриваемых случайных величин в тех генеральных совокупностях, из которых взяты выборки, и, в частности, он может зависеть от параметров генеральных совокупностей (средних значений, дисперсий и т. д.). Если критерий основан на том или другом конкретном типе распределения генеральных совокупностей или использует параметры этих совокупностей, то такой критерий называется параметрическим. Например, параметрическими критериями являются широко используемые на практике критерии, основанные на t-распределении Стьюдента и F-распределении Фишера — Снедекора, поскольку эти критерии исходят из предположения о нормальном законе распределения случайных величин в генеральных совокупностях. Если критерий не опирается на предположение о конкретном типе распределения генеральных совокупностей и не использует параметры этих совокупностей, то он называется *непараметрическим критерием*. Следует заметить, что термин «непараметрический критерий», введенный впервые И. Вольфовицем[1], не отражает в достаточной мере особенностей таких критериев. Более удачным надо признать другое используемое в научной литературе название — «критерий, свободный от распределения» (говорят также о статистиках, свободных от распределения). Однако в настоящей работе мы сохраним традиционный и более компактный термин «непараметрический критерий». Возможности практического применения параметрических и непараметрических критериев существенно зависят от характера тех измерений, которые может осуществить исследователь. Параметрические критерии требуют для своего применения количественных измерений, т. е. измерений по шкале интервалов или шкале отношений. В то же время большинство непараметрических критериев может применяться и в тех случаях, когда результаты выборок измерены с помощью шкал наименования и порядка.

Поскольку при современном состоянии педагогической науки мы располагаем весьма ограниченным набором количественных показателей, характеризующих отдельные стороны процесса обучения и его результатов,

---

[1] *Wolfowitz I.* Additive Partition Functions and a Class of Statistical Hypotheses. The Annals of Mathematical Statistics, 13, 1942, p. 247—279.

именно непараметрические критерии служат в большинстве случаев единственным средством для проверки педагогических гипотез. Кроме того, несомненным достоинством непараметрических критериев является относительная простота вычислительных процедур, связанных с практическим применением этих критериев.

В последующих разделах (4 и 5) будут рассмотрены некоторые, наиболее важные непараметрические критерии и даны рекомендации по их применению в педагогических исследованиях. При этом мы ограничимся довольно типичной для педагогики задачей, когда сравниваются результаты двух выборок.

4 СРАВНЕНИЕ РЕЗУЛЬТАТОВ ДВУХ ЗАВИСИМЫХ ВЫБОРОК

В настоящей работе рассматривается несколько способов проверки статистических гипотез относительно совокупностей учащихся на основе измерений состояния некоторого свойства или признака у каждого члена различных выборок, составленных из этих совокупностей. При отборе этих способов учитывалось, что свойства учащихся, изучаемые в педагогических экспериментах, могут быть измерены по разным шкалам. В то же время использование того или иного способа проверки статистических гипотез зависит не только от характера измерений, но и от особенностей выборок. А именно существует два вида выборок: зависимые и независимые. Если результаты измерения некоторого свойства у объектов первой выборки не оказывают влияния на результаты измерения этого свойства у объектов второй выборки, то такие выборки считаются независимыми. В тех случаях, когда результаты одной выборки влияют на результаты другой выборки, выборки считаются зависимыми. Классический способ получения зависимых измерений — это двукратное измерение одного и того же или различных свойств у членов одной и той же группы учащихся. В тех случаях, когда каждому объекту первой выборки соответствует эквивалентный (родственный) по состоянию изучаемого свойства до начала эксперимента объект во второй выборке, выборки иногда также считаются зависимыми. Последний случай зависимых выборок практически не встречается в педагогических экспериментах в связи с тем, что чрезвычайно трудно бывает подобрать для учащихся, составляющих одну выборку, эквивалентных (родственных) в том или другом смысле учащихся в другой выборке. Поэтому в данной работе в качестве примеров зависимых выборок всегда будут рассматриваться те или другие результаты двукратного обследования одной и той же группы учащихся.

Для проверки статистических гипотез на основе результатов измерений некоторых свойств объектов зависимых и независимых выборок в математической статистике разработаны специальные методы. В данном разделе будут рассмотрены три метода, основанные на результатах сравнения измерений свойств объектов двух зависимых выборок: критерий Макнамары, знаковый критерий, критерий Вилкоксона.

4.1. Критерий Макнамары

Критерий предназначен для сравнения распределений объектов двух совокупностей по состоянию некоторого свойства на основе измерений (хотя бы по шкале наименований) этого свойства в двух зависимых выборках из рассматриваемых совокупностей.

Данные. В педагогических исследованиях нередко возникает проблема сравнения состояния некоторого свойства у членов двух зависимых выборок, когда данное свойство может быть измерено только по шкале наименований. Например, отношение группы учащихся к некоторой профессии до и после беседы по профориентации измерено по шкале наименований, имеющей следующие категории: совсем не нравится — не нравится — безразлична — нравится — очень нравится. В этом случае возникает необходимость сравнения ответов одних и тех же учащихся до и после беседы, так как полученные результаты позволят судить об эффективности данной беседы в отношении изменения мнения учащихся о профессии в положительную или отрицательную сторону.

Для тех случаев, когда измерение состояния изучаемого свойства проводится по шкале наименований, имеющей только две категории, разработан специальный критерий для сравнения результатов двух зависимых выборок. Этот критерий называется критерием Макнамары. Он может быть использован в исследовании, о котором говорилось выше, если отношение учащихся к профессии будет измерено по шкале наименований, имеющей только две категории, например: нравится — не нравится. Принято обозначать одну из двух категорий шкалы значком «0», вторую — значком «1».

Будем считать, что случайная переменная X характеризует состояние некоторого свойства в рассматриваемой совокупности объектов при первичном измерении данного свойства. А случайная переменная Y характеризует состояние этого же свойства в той же совокупности объектов при вторичном измерении.

Пусть имеется две серии наблюдений

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_N;$$

$$y_1,\ y_2,\ \ldots,\ y_i,\ \ldots,\ y_N$$

над случайными переменными X и Y, полученные при рассмотрении двух зависимых выборок. Составлено N пар вида (x_i, y_i), где x_i, y_i — результаты двукратного измерения одного и того же свойства у одного и того же объекта.

В педагогических исследованиях пары (x_i, y_i) могут быть результатами измерения состояния одного и того же свойства у одного и того же ученика до и после применения некоторого педагогического средства, причем x_i — состояние свойства до применения этого средства, а y_i — после его применения.

x_i, y_i — измерения по шкале наименований, имеющей две категории, обозначенные «0» и «1». В связи с этим пары (x_i, y_i) могут быть только четырех видов (0, 0), (0, 1), (1, 0), (1, 1).

Для использования критерия данные суммируются в виде четырехклеточной таблицы, которая называется «таблица 2×2».

Таблица 1

		Классификация y_i: $y_i=0$	Классификация y_i: $y_i=1$	
Классификация x_i	$x_i=0$	a (число пар, у которых $x_i=0$, $y_i=0$)	b (число пар, у которых $x_i=0$, $y_i=1$)	$a+b$
	$x_i=1$	c (число пар, у которых $x_i=1$, $y_i=0$)	d (число пар, у которых $x_i=1$, $y_i=1$)	$c+d$
		$a+c$	$b+d$	N

Допущения. Для применения критерия Макнамары необходимо выполнение следующих требований: 1) выборки случайные[1]; 2) выборки зависимые; 3) пары (x_i, y_i) взаимно независимы, т. е. члены выборки никак не влияют друг на друга (в педагогических исследованиях выполнение этого требования равносильно, например, исключению возможности консультаций и списывания членами выборок ответов друг у друга); 4) шкала измерений — шкала наименований с двумя категориями (например, выше — ниже, хуже — лучше, больше — меньше, да — нет, согласен — не согласен, есть — нет и т. п.).

Гипотезы. Предположим, что законы распределения случайных величин X и Y одинаковы. Тогда выполняется и такое равенство[2]:

$$P\ (x_i = 0,\ y_i = 1) = P\ (x_i = 1,\ y_i = 0) \qquad (1)$$

для всех N пар (x_i, y_i).

Критерий Макнамары и предназначен для проверки справедливости данного равенства. Нулевая гипотеза имеет вид

$$H_0: P\ (x_i = 0,\ y_i = 1) = P\ (x_i = 1,\ y_i = 0)$$

для всех i. В качестве альтернативной гипотезы выбирается гипотеза

$$H_1: P\ (x_i = 0,\ y_i = 1) \neq P\ (x_i = 1,\ y_i = 0)$$

для всех i. Если гипотеза H_1 справедлива, то это означает, что законы распределения переменных X и Y различны, т. е. состояния изучаемого свойства существенно (значимо) различны в одной и той же совокупности при первичном измерении этого свойства (например, до применения нового метода обучения) и при вторичном его измерении (например, после применения нового метода обучения). Справедливость нулевой гипотезы приводит к выводу об отсутствии значимых различий в

[1] Т. е. выборка получена из рассматриваемой совокупности методом случайного отбора.

[2] Равенство (1) означает, что появление пар (0,1) и (1,0) равновероятно.

состоянии изучаемого свойства при первичном и вторичном измерениях его состояния у объектов рассматриваемой совокупности.

Гипотезы могут быть записаны в другой форме, которая позволяет их проще интерпретировать в соответствии с содержанием и особенностями проводимого эксперимента:

$$H_0: P\ (x_i = 0) = P\ (y_i = 0) \quad \text{для всех } i, \qquad (2)$$

$$H_1: P\ (x_i = 0) \neq P\ (y_i = 0) \quad \text{для всех } i, \qquad (3)$$

$$H_0: P\ (x_i = 1) = P\ (y_i = 1) \quad \text{для всех } i, \qquad (4)$$

$$H_1: P\ (x_i = 1) \neq P\ (y_i = 1) \quad \text{для всех } i. \qquad (5)$$

Например, при проверке эффективности беседы по профориентации равенство (1) можно интерпретировать так: вероятность изменения после беседы отрицательного отношения к профессии на положительное равна вероятности изменения положительного отношения на отрицательное. Равенство (2) можно интерпретировать так: вероятность положительного отношения к профессии одинакова до и после проведения беседы, равенство (4) — вероятность отрицательного отношения к профессии одинакова до и после проведения беседы.

Статистика критерия. Для проверки статистических гипотез с помощью критерия Макнамары подсчитывается значение величины, называемой статистикой критерия.

Допустим, что N пар (x_i, y_i) распределились следующим образом: число пар вида $(x_i=0,\ y_i=1)$ равно b, число пар вида $(x_i=1,\ y_i=0)$ равно c. Тогда, если $b+c>20$, то в качестве статистики выбирается величина

$$T_1 = \frac{(b-c)^2}{b+c}. \qquad (4.1.1)$$

Если $b+c \leqslant 20$, то используется величина T_2, равная наименьшему из значений b и c, т. е.

$$T_2 = \min\ (b,\ c). \qquad (4.1.2)$$

Значения статистик T_1 и T_2 не зависят от значений a и d — числа пар вида $(x_i=0,\ y_i=0)$ и $(x_i=1,\ y_i=1)$, так как эти пары представляют измерения объектов, индифферентных к воздействию средства, эффективность которого проверяется в проводимом эксперименте, и, естественно, не учитываются при рассматриваемом способе оценки результатов эксперимента.

Правило принятия решения. Пусть $b+c=n$ и α — принятый уровень значимости. Рассмотрим правила принятия решений в случае применения критерия Макнамары для проверки разного вида гипотез.

Проводится проверка гипотезы H_0: $P(x_i=0,\ y_i=1)=P(x_i=1,\ y_i=0)$ — при альтернативе H_1: $P(x_i=0,\ y_i=1)\neq P(x_i=1,\ y_i=0)$.

Если справедлива нулевая гипотеза, то статистика критерия $T_2=\min(b,\ c)$ распределена по биномиальному закону с $p=0{,}5$. Поэтому для $n\leqslant20$ по табл. A* по значению n и величине статистики критерия T_2 находим $P(T_2\leqslant T_2$ наблюдаемое), т. е. вероятность появления значения статистики, меньшего или равного наблюдаемому значению T_2 при данном значении n. Если эта вероятность меньше половины заданного уровня значимости α, т. е. $\frac{\alpha}{2}$, то H_0 отклоняется на уровне значимости α. При этом в случае, когда $b<c$, принимается гипотеза H_1: $P(x_i=0,\ y_i=1)<P(x_i=1,\ y_i=0)$, а в случае $b>c$ — гипотеза H_1: $P(x_i=0,\ y_i=1)>P(x_i=1,\ y_i=0)$.

Таблицы биномиального распределения, удобные для применения критерия Макнамары, составлены для $n\leqslant25$. Однако для $n>20$ при предположении о справедливости нулевой гипотезы распределение статистики критерия T_1 аппроксимируется распределением χ^2 (хи-квадрат) с одной степенью свободы ($\nu=1$). H_0 отклоняется на уровне значимости α, если наблюдаемое значение T_1 превосходит критическое значение статистики критерия, отвечающее данному уровню значимости α, которое определяется по таблице $Г$ распределения χ^2 для одной степени свободы. Для $\alpha=0{,}05$ $T_{1\ \text{критич}}\ T_1=3{,}84$; $\alpha=0{,}025\ T_{1\ \text{критич}}=5{,}02$; $\alpha=0{,}01$ $T_{1\ \text{критич}}=6{,}63$.

При отклонении H_0 принимается гипотеза H_1: $P(x_i=0,\ y_i=1)<P(x_i=1,\ y_i=0)$, если $b<c$, и гипотеза H_1: $P(x_i=0,\ y_i=1)>P(x_i=1,\ y_i=0)$, если $b>c$.

---

* Таблицы *А, Б, В, Г, Д, Е, Ж* даны в конце книги.

В случае, когда $b=c$, применение статистики критерия T_2 при $n\leqslant20$ и статистики T_1 при $n>20$ заведомо не позволяет отвергнуть нулевую гипотезу на любом уровне значимости α. Поэтому при $b=c$ результаты эксперимента не позволяют использовать критерий Макнамары для проверки статистических гипотез.

Примеры использования критерия Макнамары. Пример 1. Проводился эксперимент по выявлению эффективных форм профориентационной работы среди учащихся выпускных классов. С этой целью было разработано несколько различных по содержанию и методике проведения циклов лекций, бесед и экскурсий учащихся на предприятия. Эффективность каждого цикла выявлялась в отношении формирования у учащихся определенного мнения о некоторой профессии.

Один из циклов был направлен на формирование положительного отношения учащихся к профессии экономиста. Отношение учащихся к данной профессии выяснялось до и после проведения цикла на основе ответов учащихся на вопрос: «Ваше отношение к профессии экономиста?» (Ответы: «нравится» — «не нравится».)

Из 150 учащихся, которые прослушали данный профориентационный цикл, методом случайного отбора была составлена выборка из 20 учащихся. Результаты двукратного ответа на вопрос об отношении к профессии этих 20 учащихся представляют измерения по шкале наименований, имеющей две категории («нравится» обозначим значком «0», «не нравится» — значком «1»), такого свойства учащихся, как отношение к профессии экономиста. В этих условиях возможно применение критерия Макнамары для выявления тенденции в изменении отношения учащихся к данной профессии после посещения профориентационного цикла, так как выполнены все допущения данного критерия (см. с. 42).

Результаты двукратного опроса 20 учащихся запишем в форме таблицы 2×2. В условиях данного эксперимента значение a равно числу учащихся, которые оба раза дали ответ «нравится»; значение b — числу учащихся, которые первый раз дали ответ «нравится», второй раз — «не нравится»; значение c — числу учащихся, которые первый раз дали ответ «не нравится», второй раз — «нравится»; значение d — числу учащихся, которые оба раза дали ответ «не нравится».

Второй опрос

Первый опрос:	нравится	не нравится	
нравится	$a=2$	$b=2$	4
не нравится	$c=11$	$d=5$	16
	13	7	20

Проверяется гипотеза H_0: посещение данного профориентационного цикла лекций и экскурсий не оказывает влияния на отношение учащихся к профессии экономиста. В соответствии с полученными результатами ($b<$ $<c$) альтернативная гипотеза H_1 формулируется следующим образом: посещение данного цикла бесед и лекций положительно влияет на отношение учащихся к профессии.

В этих условиях для проверки гипотезы применяется критерий Макнамары для $n\leqslant 20$ ($n=b+c=11+2=$ $=13$; $13<20$), т. е. подсчитывается значение статистики T_2, равное наименьшему из значений b и c (см. формулу (4.1.2). В данном случае $T_2=2$. По табл. A вероятность появления значения $T_2\leqslant 2$ при $n=13$ равна 0,011. Если уровень значимости проверки гипотез $\alpha=$ $=0{,}05$, то $\frac{\alpha}{2}=0{,}025$ и верно неравенство $0{,}011<\frac{\alpha}{2}$. Следовательно, гипотеза H_0 отклоняется на уровне значимости $\alpha=0{,}05$ и принимается альтернативная гипотеза H_1. Таким образом, на основе результатов проведенного эксперимента можно сделать вывод о том, что разработанный цикл лекций и экскурсий способствует формированию положительного отношения к профессии экономиста.

П р и м е р 2. Проверялось влияние формы контроля знаний учащихся по некоторому разделу программы на результаты контрольного опроса. На одном и том же содержательном материале были составлены: письменная работа обычного типа из 3 заданий и тест из 20 вопросов. На основе результатов выполнения каждой из форм в отдельности учащиеся распределялись на две категории: усвоил — не усвоил. При выполнении письменной работы в первую группу относили учащихся, получивших оценки «3», «4», «5», выставленные в соответствии с нормами, разработанными экспериментаторами. При выполнении теста в первую группу относили учащихся, верно ответивших на 13 и более вопросов. Остальные учащиеся были отнесены ко второй группе.

Из разных школ было выбрано методом случайного отбора 100 учащихся. Каждый из них выполнял обе формы контрольных работ одну за другой. Результаты двукратного контроля знаний этих учащихся представляют измерения по шкале наименований с двумя категориями (усвоил — не усвоил) состояния знаний учащихся по определенному разделу. В этих условиях возможно применение критерия Макнамары для выявления значимости различия в распределении учащихся по состоянию знаний при различных формах контроля: письменная контрольная работа и тестовая форма, так как выполнены все допущения данного критерия.

Результаты двукратного выполнения работы запишем в форме таблицы 2×2. В условиях проведенного эксперимента она имеет вид:

Результаты по тесту:

Результаты по контрольной работе:	усвоил	не усвоил	
усвоил	$a=63$	$b=21$	84
не усвоил	$c=4$	$d=12$	16
	67	33	100

Проверяется гипотеза H_0: форма контроля за усвоением данного раздела программы не оказывает влияния на распределение учащихся по состоянию знаний. В связи с задачами эксперимента альтернативная гипотеза H_1 формулируется следующим образом: распределение учащихся по состоянию знаний различно при разных формах контроля.

В этих условиях для проверки гипотезы применяется двусторонний критерий Макнамары для $n>20$ ($n=$ $=b+c=4+21=25$), т. е. подсчитывается значение статистики T_1 по формуле (4.1.1) (см. с. 43). В данном случае

$$T_1=\frac{(b-c)^2}{b+c}=\frac{(21-4)^2}{21+4}=11{,}56.$$

Для уровня значимости $\alpha=0,05$ критическое значение $T_{1_{\text{критич}}}=3,84$ (см. с. 44). Следовательно, верно неравенство $T_{1_{\text{наблюд}}}>T_{1_{\text{критич}}}$. Поэтому нулевая гипотеза отвергается на уровне значимости $\alpha=0,05$ и принимается альтернативная гипотеза. Таким образом, на основе результатов проведенного эксперимента можно сделать вывод о том, что форма контроля за усвоением раздела программы существенно влияет на распределение учащихся по состоянию знаний.

Пример 3. В большом числе экспериментов принято оценивать выполнение учащимися задания двумя категориями: верно и неверно. Затем выполнение этого задания группой учащихся оценивается числом верных ответов или процентом верных ответов. В связи с задачами эксперимента нередко необходимо сравнить между собой проценты верных ответов одной и той же группы учащихся на один и тот же или разные вопросы. Критерий Макнамары может быть использован для сравнения значимости различия двух процентных отношений. Однако это сравнение невозможно провести, если мы располагаем только этими двумя отношениями. Например, известно, что процент верных ответов группы в 120 учащихся на один вопрос равен 57, на другой — 68. Если мы располагаем только этими данными, то применить критерий Макнамары невозможно. Для его применения нужно по выполнению двух вопросов разбить учащихся на такие четыре подгруппы, на основе численности которых заполняется таблица вида 2×2, необходимая для подсчета статистики критерия. Покажем процесс сравнения процентов верных ответов на конкретном примере.

Группа из 120 учащихся отвечала на два вопроса. Процент верных ответов на первый вопрос — 57, на второй — 68. При этом из 120 учащихся 50 верно ответили на два вопроса, 19 — верно ответили на первый вопрос и неверно на второй, 31 — неверно ответил на первый вопрос и верно на второй, 20 — неверно ответили на оба вопроса. В этих условиях возможно применение критерия Макнамары для сравнения уровней выполнения двух вопросов, так как выполнены все допущения критерия.

Таблица 2×2 будет иметь вид:

Результаты выполнения первого вопроса:

		верно	неверно	
Результаты выполнения второго вопроса:	верно	$a=50$	$b=31$	81
	неверно	$c=19$	$d=20$	39
		69	51	120

Проверяется гипотеза H_0: не существует различия в уровне выполнения обоих заданий данной группой учащихся — при альтернативе H_1: уровни выполнения заданий существенно различны.

В этих условиях для проверки гипотез применяется критерий Макнамары для $n>20$ ($n=b+c=31+19=$ $=50>20$). Подсчитаем значение статистики критерия T_1 по формуле (4.1.1):

$$T_1=\frac{(31-19)^2}{31+19}=2,88.$$

Для уровня значимости $\alpha=0,05$ $T_{1\text{критич}}=3,84$. Следовательно, верно неравенство $T_{1\text{наблюд}}<T_{1\text{критич}}$. Поэтому на уровне значимости $\alpha=0,05$ у нас нет достаточных оснований для отклонения нулевой гипотезы. Иначе говоря, нет достаточных оснований считать, что проценты верных ответов группы учащихся на вопросы 1 и 2 различаются статистически значимо.

4.2. Критерий знаков

Критерий предназначен для сравнения состояния некоторого свойства у членов двух зависимых выборок на основе измерений, сделанных по шкале не ниже порядковой.

Данные. Будем считать, что случайная переменная X характеризует состояние некоторого свойства в рассматриваемой совокупности объектов при первичном измерении данного свойства, случайная переменная Y характеризует состояние этого же свойства в той же совокупности объектов при вторичном измерении.

Имеется две серии наблюдений

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_N;$$
$$y_1,\ y_2,\ \ldots,\ y_i,\ \ldots,\ y_N$$

над случайными переменными X и Y, полученные при рассмотрении двух зависимых выборок. На их основе составлено N пар вида (x_i, y_i), где x_i, y_i — результаты двукратного измерения одного и того же свойства у одного и того же объекта.

В педагогических исследованиях объектами изучения могут служить учащиеся, учителя, администрация школ. При этом x_i, y_i могут быть, например, балловыми оценками, выставленными учителем за двукратное выполнение одной и той же или различных работ одной и той же группой учащихся до и после применения некоторого педагогического средства.

Элементы каждой пары x_i, y_i сравниваются между собой по величине, и паре присваивается знак «+», если $x_i<y_i$, знак «—», если $x_i>y_i$, и «0», если $x_i=y_i$. Установление соотношения «больше» («меньше») между двумя измерениями возможно, если эти измерения сделаны хотя бы по шкале порядка. Следовательно, знаковый критерий непригоден в случае измерений по шкале наименований.

Допущения. Для применения знакового критерия необходимо выполнение следующих требований: 1) выборки случайные; 2) выборки зависимые; 3) пары (x_i, y_i) взаимно независимы; 4) изучаемое свойство объектов распределено непрерывно[1] в обеих совокупностях, из которых сделаны выборки; 5) шкала измерений должна быть не ниже порядковой.

Гипотезы. Предположим, что законы распределения случайных величин X и Y одинаковы. Тогда выполняется также и такое равенство

$$P\ (x_i < y_i) = P\ (x_i > y_i) \quad \text{для всех пар} \quad (x_i,\ y_i),$$

которое означает, что вероятность того, что первое измерение (x_i) в паре (x_i, y_i) меньше второго измерения (y_i), равна вероятности того, что первое измерение в паре больше второго, для всех N пар. Справедливость этого равенства и проверяется с помощью знакового критерия. Таким образом, нулевая гипотеза будет иметь вид

$$H_0:\ P\ (x_i < y_i) = P\ (x_i > y_i) \quad \text{для всех } i.$$

[1] См. введение, с. 8.

При использовании знакового критерия в качестве альтернативной гипотезы выбирается гипотеза

$$H_1:\ P\ (x_i < y_i) \neq P\ (x_i > y_i) \quad \text{для всех } i.$$

Если гипотеза H_1 справедлива, то отсюда следует, что законы распределения величин X и Y различны, т. е. состояния изучаемого свойства существенно различны в одной и той же совокупности при первичном и вторичном измерениях этого свойства. Справедливость нулевой гипотезы интерпретируется следующим образом: в состоянии изучаемого свойства нет значимых различий при первичном и вторичном измерениях.

Пусть $F(x)$ и $G(x)$ — функции распределения случайных величин X и Y. Иногда знаковый критерий применяется для проверки нулевой гипотезы H_0: $F(x)=G(x)$ — при альтернативе H_1: $F(x)\neq G(x)$ более общего типа, чем указанная выше альтернатива $P\ (X<Y)\neq P(X>Y)$. Однако для таких альтернатив критерий не является состоятельным (см. [6]).

Статистика критерия. Для проверки гипотез с помощью знакового критерия на основе наблюдений подсчитывается значение величины T, называемой статистикой критерия. Значение T определяется следующим образом.

Допустим, что из N пар (x_i, y_i) нашлось несколько пар, в которых значения x_i и y_i равны. Такие пары обозначаются знаком «0» и при подсчете значения величины T не учитываются. Предположим, что за вычетом из числа N числа пар, обозначенных знаком «0», осталось всего n пар. Среди оставшихся пар подсчитаем число пар, обозначенных знаком «+» (т. е. те пары, в которых $x_i<y_i$). Значение величины T и равно числу пар со знаком «+».

Правило принятия решения. Пусть число пар, в которых $x_i\neq y_i$, равно n и α — принятый уровень значимости. Рассмотрим правила принятия решений при проверке разного вида гипотез.

I. Двусторонний критерий. Проводится проверка гипотезы

$$H_0:\ P\ (x_i < y_i) = P\ (x_i > y_i)$$

при альтернативе

$$H_1:\ P\ (x_i < y_i) \neq P\ (x_i > y_i).$$

Для $n \leqslant 100$ составлена специальная таблица (см. таблицу *Б*), в которой для каждого значения n даны критические значения t_α и $n-t_\alpha$ статистики T для разных уровней значимости α. В табл. *Б* эти значения даны для $\alpha=0{,}05$; $\alpha=0{,}02$; $\alpha=0{,}01$. При данном значении n гипотеза H_0 отклоняется на уровне значимости α, если для наблюдаемого значения T справедливо одно из неравенств $T<t_\alpha$ или $T>n-t_\alpha$.

2. Односторонний критерий. В тех случаях, когда имеются достаточные основания предполагать, что результаты второго измерения изучаемого свойства у одних и тех же объектов — y_i имеют тенденцию превышать[1] (или, наоборот, быть меньше) результаты первичного измерения — x_i, вместо двустороннего критерия используется односторонний критерий.

А. Для случая, когда x_i имеют тенденцию превышать y_i, проводится проверка гипотезы

$$H_0: P(x_i < y_i) \geqslant P(x_i > y_i)$$

— при альтернативе

$$H_1: P(x_i < y_i) < P(x_i > y_i).$$

Для $n \leqslant 100$ может быть использована та же табл. *Б*, что и для двустороннего критерия. H_0 отклоняется на уровне значимости α, указанном для одностороннего критерия, если наблюдаемое значение $T<t_\alpha$. Табл. *Б* предлагает значения t_α для уровней значимости $\alpha=0{,}025$; $\alpha=0{,}01$; $\alpha=0{,}005$.

Б. В том случае, когда y_i имеет тенденцию превышать по значению x_i, проводится проверка гипотезы

$$H_0: P(x_i < y_i) \leqslant P(x_i > y_i)$$

— при альтернативе

$$H_1: P(x_i < y_i) > P(x_i > y_i).$$

H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T>n-t_\alpha$, где значение $n-t_\alpha$ определяется из табл. *Б*.

Таблица критических значений знакового критерия основана на биномиальном распределении. Для достаточно больших значений n биномиальное распределение можно приближенно заменить (аппроксимировать) нормальным. В случае двустороннего критерия t_α определяется по формуле

$$t_\alpha = 0{,}5\left(n + W_{\frac{\alpha}{2}}\sqrt{n}\right), \tag{4.2.1}$$

где $W_{\frac{\alpha}{2}}$ — квантиль нормального распределения[1], определяемый для вероятности $\frac{\alpha}{2}$. Приведем некоторые значения $W_{\frac{\alpha}{2}}$: для $\alpha=0{,}05$ $W_{\frac{\alpha}{2}}=-1{,}96$; для $\alpha=0{,}02$ $W_{\frac{\alpha}{2}}=-2{,}58$; для $\alpha=0{,}01$ $W_{\frac{\alpha}{2}}=-3{,}29$. H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T<t_\alpha$ или $T>n-t_\alpha$.

В случае одностороннего критерия t_α определяется по формуле

$$t_\alpha = 0{,}5\left(n + W_\alpha\sqrt{n}\right), \tag{4.2.2}$$

где W_α — квантиль нормального распределения, определяемый для вероятности α. При $\alpha=0{,}05$ $W_\alpha=-1{,}64$, при $\alpha=0{,}02$ $W_\alpha=-2{,}05$; при $\alpha=0{,}01$ $W_\alpha=-2{,}58$.

При проверке гипотезы H_0: $P(x_i<y_i) \geqslant P(x_i>y_i)$ — и альтернативы H_1: $P(x_i<y_i)<P(x_i>y_i)$ H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T<t_\alpha$ (значение t_α определяется по формуле (4.2.2)).

При проверке гипотезы H_0: $P(x_i<y_i) \leqslant P(x_i>y_i)$ — и альтернативы H_1: $P(x_i<y_i)>P(x_i>y_i)$ H_0 отклоняется на уровне значимости α, если $T>n-t_\alpha$ (значение t_α определяется по формуле (4.2.2)).

Примеры использования знакового критерия.

Пример 1. Учащиеся выполняли контрольную работу, направленную на проверку усвоения некоторого понятия. Пятнадцати учащимся, 7 из которых получили отметку «2» и 8 — отметку «3», было затем предложено программированное пособие, составленное с целью формирования данного понятия у учащихся с низким уровнем обучаемости. После изучения пособия учащиеся снова выполняли ту же контрольную работу, которая оценивалась по пятибалльной системе.

---

[1] Т. е. в среднем чаще значение вторичного измерения больше значения первичного измерения изучаемого свойства одного и того же объекта.

[1] Определение квантиля см. во введении, с. 8.

Данный эксперимент проводился с целью проверки эффективности программированного пособия как средства повышения знаний слабых учащихся путем самообразования. Результаты двукратного выполнения работы учащимися представляют измерения по шкале порядка (пятибалльная шкала) такого качества, как усвоение некоторого понятия. В этих условиях возможно применение знакового критерия для выявления тенденции изменения состояния знаний учащихся после изучения пособия, так как выполняются все допущения этого критерия (см. с. 50).

Результаты двукратного выполнения работы (в баллах) 15 учащимися запишем в форме таблицы (см. табл. 2).

Таблица 2

Учащиеся (№)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Первое выполнение	2	2	2	2	2	3	3	3	3	3	2	2	3	3	3
Второе выполнение	2	3	3	4	3	2	3	4	4	3	4	3	2	4	4
Знак разности отметок	0	+	+	+	+	—	0	+	+	0	+	+	—	+	+

Проверяется гипотеза H_0: состояние знаний учащихся не повысилось после изучения пособия — при альтернативе H_1: состояние знаний учащихся повысилось после изучения пособия.

В соответствии с содержанием гипотез следует применить односторонний знаковый критерий (см. случай Б на с. 52). Подсчитаем значение статистики критерия T, равное числу положительных разностей отметок, полученных учащимися. Согласно данным табл. 2, $T=10$. Из 15 пар в 3 случаях разность измерений равна нулю, следовательно, остается только 12 (15—3=12) пар, т. е. $n=12$.

Для определения критических значений статистики критерия $n-t_\alpha$ используем табл. *Б*, так как $n<100$. Для уровня значимости $\alpha=0{,}05$ при $n=12$ значение $n-t_\alpha=9$. Следовательно, выполняется неравенство

$T_{наблюд}>n-t_\alpha$ (10>9). Поэтому в соответствии с правилом принятия решения (с. 52) нулевая гипотеза отклоняется на уровне значимости $\alpha=0{,}05$ и принимается альтернативная гипотеза, что позволяет сделать вывод об улучшении знаний учащихся после самостоятельного изучения пособия.

Пример 2. Проводился эксперимент по проверке целесообразности системы упражнений некоторого раздела учебника. С этой целью выяснялось мнение учителей относительно числа данных упражнений в начале и в конце учебного года. В связи с тем что мнения отдельных учителей резко расходились в отношении числа упражнений (одни требовали значительного увеличения числа упражнений, другие — уменьшения), было решено результаты опроса каждого учителя распределять только по двум категориям: увеличить число таких упражнений и уменьшить число таких упражнений. Всего было дважды опрошено 200 одних и тех же учителей. Результаты двукратного измерения мнения каждого учителя представляют измерения по шкале порядка (двухбалльная шкала) такого качества, как отношение учителя к системе упражнений учебника. В этих условиях возможно применение знакового критерия для выявления тенденции изменения мнения учителей относительно необходимого числа упражнений.

В условиях принятой шкалы измерения изучаемого свойства знак «+» присваивается паре вида: уменьшить число упражнений (результат первого опроса), увеличить число упражнений (результат второго опроса); знак «—» присваивается паре вида: увеличить (результат первого опроса), уменьшить (результат второго опроса); во всех остальных случаях присваивался знак «0».

В связи с большим объемом выборки и особенностью шкалы измерения результаты двукратного опроса 200 учителей удобно записать в форме таблицы (см. табл. 3).

При составлении гипотез необходимо учесть, что мнения отдельных учителей, полученные до начала эксперимента, не позволили выявить тенденцию в отношении к числу упражнений. Поэтому в условиях проводимого эксперимента нулевая гипотеза будет иметь вид: H_0 — в течение учебного года мнение учителей в отно-

Таблица 3

Первое мнение	Увеличить	Увеличить	Уменьшить	Уменьшить	
Второе мнение	Увеличить	Уменьшить	Увеличить	Уменьшить	
Знак разницы мнений	0	—	+	0	
Число учителей	52	74	36	38	200

шении числа упражнений учебника не изменится. Следовательно, альтернативная гипотеза должна быть сформулирована следующим образом: H_1 — мнение учителей в отношении числа упражнений изменится в течение учебного года.

В соответствии с содержанием гипотез следует использовать двусторонний знаковый критерий для выборок большого объема (см. с. 53). Значение статистики критерия T равно числу пар со знаком «+», т. е. $T=36$ (см. табл. 3). Из 200 пар 90 пар имеют знак «0» ($52+38=90$), значит, $n=200-90=110$.

Критическое значение статистики критерия t_α определяем по формуле (4.2.1) (см. с. 53): $t_\alpha=0{,}5\left(n+W_{\frac{\alpha}{2}}\sqrt{n}\right)$. Для $\alpha=0{,}05$ $W_{\frac{\alpha}{2}}=-1{,}96$, поэтому при $n=110$ $t_\alpha=0{,}5(110-1{,}96\cdot\sqrt{110})=44{,}72$; $n-t_\alpha=110-44{,}72=65{,}28$.

Значение $T_{\text{наблюд}}=36$; следовательно, верно неравенство $T_{\text{наблюд}}<t_\alpha$. Поэтому нулевая гипотеза отклоняется на уровне значимости $\alpha=0{,}05$ и принимается альтернативная гипотеза. Результаты эксперимента позволяют не только утверждать, что мнение учителей изменилось в течение года, но и выявить тенденцию изменения в направлении уменьшения числа упражнений (число пар со знаком «—» больше числа пар со знаком «+» в табл. 3).

Пример 3. Предполагается, что изучение курса математики способствует формированию у учащихся одного из приемов логического мышления (например, приема обобщения) даже в том случае, если его формирование не проводится целенаправленно. Для проверки этого предположения был проведен следующий эксперимент.

Учащимся VII класса было предложено 5 задач, решение которых основано на использовании данного приема мышления. Считалось, что учащийся владеет этим приемом, если он дает верный ответ на 3 и более задачи.

Была разработана следующая шкала измерений: верно решена 1 или 2 задачи — оценка «0»; верно решено 3 задачи — оценка «1»; верно решено 4 задачи — оценка «2»; верно решено 5 задач — оценка «3».

Работа проводилась дважды: в конце сентября и конце мая следующего года. Ее писали 35 одних и тех же учащихся, отобранных методом случайного отбора из 7 разных школ. Результаты двукратного выполнения работы запишем в форме таблицы (см. табл. 4).

Таблица 4

Ученики, № п/п	Первая отметка	Вторая отметка	Знак разности	Ученики, № п/п	Первая отметка	Вторая отметка	Знак разности
1	1	1	0	19	1	1	0
2	1	0	—	20	0	1	+
3	1	0	—	21	0	1	+
4	2	2	0	22	1	2	+
5	0	1	+	23	2	1	—
6	0	0	0	24	0	1	+
7	0	1	+	25	0	1	+
8	0	0	0	26	1	2	+
9	0	0	0	27	0	0	0
10	1	1	0	28	1	0	—
11	0	1	+	29	1	0	—
12	3	3	0	30	0	1	+
13	2	2	0	31	0	0	0
14	1	0	—	32	1	2	+
15	0	1	+	33	1	0	—
16	0	1	+	34	3	2	—
17	1	1	0	35	0	1	+
18	0	1	+				

В соответствии с целями эксперимента формулируем нулевую гипотезу следующим образом: H_0 — изучение математики не способствует формированию изучаемого приема мышления. Тогда альтернативная гипотеза будет иметь вид: H_1 — изучение математики способствует овладению этим приемом мышления.

Характер гипотез приводит к необходимости применения одностороннего знакового критерия для их проверки (см. случай Б, с. 52). Согласно данным табл. 4, значение статистики $T=15$ — число разностей со знаком «+». Из 35 пар 12 имеют знак «0»; значит, $n=35-12=23$. По таблице Б для $n=23$ и уровня значимости $\alpha=0{,}025$ находим критическое значение статистики критерия $n-t_\alpha$, равное 16. Следовательно, верно неравенство $T_{\text{паблюд}}<n-t_\alpha\,(15<16)$. Поэтому в соответствии с правилом принятия решений (см. с. 52, случай Б) приходится сделать вывод о том, что полученные результаты не дают достаточных оснований для отклонения нулевой гипотезы, т. е. мы не располагаем достаточными основаниями для отклонения утверждения о том, что изучение математики само по себе не способствует овладению выделенным приемом мышления.

4.3. Критерий Вилкоксона

Психологам удалось разработать методику измерения большого числа психических свойств личности по интервальной шкале, а также методику измерения времени различных реакций по шкале отношений. Поэтому в психологических исследованиях нередко возникает проблема сравнения членов двух зависимых выборок по состоянию некоторого свойства, измеренного хотя бы по интервальной шкале. Эта проблема иногда может быть решена с помощью критерия Вилкоксона.

В педагогических исследованиях проблемы измерения времени реакций ученика рассматриваются редко, и в основном по отношению к скорости вычислений. При этом техника измерений разработана слабо и обычно сводится к использованию секундомера. Однако вполне возможна постановка экспериментов, связанных с определением времени реакций учащихся не только на вычислительные, но и на другие задания. Кроме того, измерение состояния некоторого свойства личности числом верных ответов на контрольные задания, проверяющие данное свойство, можно считать измерением по интервальной шкале. Правда, в этом случае использование критерия Вилкоксона приводит к объективным заключениям, если число вопросов достаточно велико, а сами вопросы составлены так, что позволяют проверить всесторонне изучаемое свойство личности. Дело в том, что для состоятельности критерия Вилкоксона очень важно отсутствие одинаковых измерений, т. е. результаты однократного измерения некоторого свойства у различных учащихся не должны совпадать. При небольшом числе заданий невозможно выполнить это условие. С увеличением числа заданий вероятность совпадения числа верных ответов у отдельных учащихся выборки снижается. Если же число заданий невелико, то вместо критерия Вилкоксона целесообразно использовать знаковый критерий.

В педагогических исследованиях критерий Вилкоксона может быть использован для проверки следующего предположения: медиана результатов измерения одного и того же свойства в некоторой совокупности учащихся до проведения эксперимента равна медиане результатов измерения того же свойства у учащихся той же совокупности после данного эксперимента.

Критерий Вилкоксона может применяться к результатам, полученным на основе однократного измерения некоторого свойства у учащихся одной выборки. В данном случае он используется для проверки предположения о том, что данная выборка принадлежит совокупности учащихся, медиана результатов которых равна определенному числовому значению.

Рассмотрим применение критерия при проверке различных статистических гипотез.

1. Случай двух зависимых выборок

Данные. Будем считать, что случайная переменная X характеризует состояние некоторого свойства в рассматриваемой совокупности объектов при первом измерении, а случайная переменная Y характеризует состояние этого же свойства в той же совокупности при втором измерении.

Пусть имеется две серии наблюдений

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_N;$$

$$y_1,\ y_2,\ \ldots,\ y_i,\ \ldots,\ y_N$$

над случайными переменными X и Y, полученные при рассмотрении двух зависимых выборок. Составим N пар вида (x_i, y_i), где x_i, y_i — результаты двукратного измерения некоторого свойства у одного и того же объекта.

Для каждой пары (x_i, y_i) находим $|D_i|=|y_i-x_i|$ — абсолютное значение разности измерений x_i и y_i. Пары, у которых $D_i=0$ — значения x_i и y_i совпадают $(x_i=y_i)$, не учитываются при подсчете значения статистики критерия. Пусть за вычетом числа таких пар из общего числа пар N остается n пар $(n \leqslant N)$. Выписываем для оставшихся n пар значения $|D_i|$ в ряд по возрастанию значений. Значению $|D_i|$, стоящему на левом фланге этого ряда, приписываем ранг 1, следующему за ним — ранг 2 и т. д., последнему — n. Если несколько последующих значений $|D_i|$, стоящих в этом ряду, совпадают, то каждому из них приписывается один и тот же ранг, равный среднему арифметическому их рангов. Правда, приписывание одинаковых рангов значениям $|D_i|$ снижает точность выводов, полученных при использовании критерия.

Далее каждому рангу приписывается знак «+», если он соответствует положительной разности $(D_i>0)$, и знак «—», если он соответствует отрицательной разности $(D_i<0)$.

Покажем процесс приписывания рангов на примере. Пусть имеется шесть пар наблюдений вида (x_i, y_i): (2,8); (12,14); (7,7); (10,4); (4,1); (12,8). Тогда $D_1=6$, $D_2=2$, $D_3=0$, $D_4=-6$, $D_5=-3$, $D_6=-4$. $|D_1|=6$, $|D_2|=2$, $|D_4|=6$, $|D_5|=3$, $|D_6|=4$.

За вычетом пары № 3 $(D_3=0)$ остается 5 пар. Запишем значения $|D_i|$ в ряд по возрастанию значений:

$$2,\ 3,\ 4,\ 6,\ 6. \qquad (1)$$

Числу 2 припишем ранг 1, числу 3 — ранг 2, числу 4 — ранг 3. Тогда числам 6 и 6 будут соответствовать ранги 4 и 5, т. е. каждому из них припишем ранг, равный среднему арифметическому их рангов: $\frac{4+5}{2}=4{,}5$. Далее каждому рангу припишем знак разности, которой этот ранг соответствует. В ряду (1) числу 2 будет соответственно приписан ранг (+1) $(D_2>0)$, числу 3 — ранг (—2) $(D_5<0)$, числу 4 — ранг (—3) $(D_6<0)$, числу 6 — ранг (+4,5) $(D_1>0)$, числу 6 — ранг (—4,5) $(D_4<0)$.

Запишем процесс приписывания рангов R_i в данном примере в форме таблицы (см. табл. 5).

Таблица 5

Номера пар (x_i, y_i)	1	2	3	4	5	6
Значение x_i Значение y_i	2 8	12 14	7 7	10 4	4 1	12 8
$x_i-y_i=D_i$ $\|D_i\|$	+6 6	+2 2	0	—6 6	—3 3	—4 4
Ранг $\|D_i\|$	4,5	1		4,5	2	3
R_i	+4,5	+1		—4,5	—2	—3

Допущения. Для применения критерия Вилкоксона необходимо выполнение следующих требований:

1) выборки случайные и зависимые;

2) изучаемое свойство объектов распределено непрерывно[1] в обеих совокупностях, из которых сделаны выборки;

3) распределение каждой из переменных X и Y симметричное (распределение случайной переменной X симметрично относительно прямой $x=c$, если $P(X \leqslant c-x)=P(X \geqslant c+x)$ для всех возможных значений x);

4) пары значений (x_i, y_i) взаимно независимы;

5) шкала измерений x_i и y_i должна быть хотя бы интервальной.

Критерий Вилкоксона более чувствителен к улавливанию особенностей измерений по сравнению со знаковым критерием, так как его применение основано не только на учете знаков разностей измерений x_i и y_i, но и на учете абсолютных значений этих разностей.

Гипотезы. Предположим, что законы распределения случайных переменных X и Y одинаковы. Тогда выполняется и такое равенство:

$$P\ (D_i > 0) = P\ (D_i < 0),$$

которое означает, что медиана всех разностей D_i равна нулю. Критерий Вилкоксона и предназначен для проверки справедливости данного равенства. Нулевая гипотеза имеет вид H_0: медиана $D_i=0$ — медиана разностей измерений x_i и y_i равна нулю.

[1] См. введение, с. 8.

При использовании критерия Вилкоксона в качестве альтернативной гипотезы выбирается гипотеза

$$H_1: \text{медиана} \quad D_i \neq 0.$$

Если H_1 справедлива, то отсюда следует справедливость утверждения: законы распределения переменных X и Y различны, т. е. состояние изучаемого свойства в совокупностях X и Y существенно различно. Из предположения о симметричности распределений переменных X и Y следует равенство медианы и среднего значения переменной X и равенство медианы и среднего значения переменной Y. Поэтому все заключения, справедливые относительно медиан этих переменных, должны быть верными относительно средних значений этих переменных. Для проверки альтернатив более общего вида критерий Вилкоксона не является состоятельным (см. [6]).

Статистика критерия. Для проверки гипотез с помощью критерия Вилкоксона на основе наблюдений подсчитывается величина T — статистика критерия. Допустим, что из N пар наблюдений (x_i, y_i) имеется несколько пар, в которых $x_i = y_i$ $(D_i = 0)$. Такие пары при нахождении значения T не учитываются, т. е. из общего числа N пар остается $n \leqslant N$ пар, у которых $x_i \neq y_i$. Значение T равно сумме положительных рангов $(R_i > > 0)$, присвоенных разностям D_i.

Правило принятия решения. Рассмотрим правило принятия решения при проверке различного вида гипотез.

1. Двусторонний критерий. Проводится проверка гипотезы H_0: медиана $D_i = 0$ — при альтернативе H_1: медиана $D_i \neq 0$.

Для $n \leqslant 20$ составлена таблица (см. табл. B), в которой для каждого значения n даются критические значения $W_{\frac{\alpha}{2}}$ и $W_{1-\frac{\alpha}{2}}$ статистики T для уровней значимости $\alpha = 0{,}10$; $\alpha = 0{,}05$; $\alpha = 0{,}02$; $\alpha = 0{,}01$. При данном n гипотеза H_0 отклоняется на уровне значимости α, указанном для двустороннего критерия, если наблюдаемое значение $T_{\text{наблюд}} < W_{\frac{\alpha}{2}}$ или $T_{\text{наблюд}} > W_{1-\frac{\alpha}{2}}$. Если верно неравенство $W_{\frac{\alpha}{2}} \leqslant T \leqslant W_{1-\frac{\alpha}{2}}$, то приходится сделать вывод о том, что полученные наблюдения не дают достаточных оснований для отклонения нулевой гипотезы.

2. Односторонний критерий. В ряде случаев полученные наблюдения или результаты проведенных ранее исследований приводят к выводу о том, что существует некоторая тенденция во взаимоотношениях между значениями переменных X и Y, а именно значение x_i имеет тенденцию превышать (или быть меньше) значения y_i. В этих условиях целесообразно вместо двустороннего критерия воспользоваться односторонним.

Если большая часть значений x_i превышает соответствующие значения y_i, то проводится проверка гипотезы H_0: медиана $D_i \geqslant 0$ — при альтернативе H_1: медиана $D_i < 0$.

Если H_1 справедлива, то можно сделать вывод о том, что значения x_i превосходят значения y_i, т. е. значения измерений изучаемого свойства существенно выше у объекта совокупности X, чем у соответствующего ему объекта совокупности Y.

Для $n \leqslant 20$ используется та же таблица B, что и для двустороннего критерия. При данном значении n H_0 отклоняется на уровне значимости α, указанном для одностороннего критерия, если наблюдаемое значение $T < W_\alpha$. В таблице даются критические значения T — W_α для $\alpha = 0{,}005$; $\alpha = 0{,}01$; $\alpha = 0{,}025$; $\alpha = 0{,}05$.

Если y_i имеют тенденцию превышать по значению x_i, то проводится проверка следующей нулевой гипотезы H_0: медиана $D_i \leqslant 0$ — при альтернативе H_1: медиана $D_i > 0$.

При $n \leqslant 20$ используется та же таблица, что и для двустороннего критерия. Для данного значения n H_0 отклоняется на уровне значимости α, указанном для одностороннего критерия, если наблюдаемое $T > W_{1-\alpha}$.

Таблица критических значений статистики критерия Вилкоксона построена на основе биномиального распределения. Для достаточно больших n биномиальное распределение аппроксимируется нормальным.

а) В случае двустороннего критерия для $n > 20$ критические значения статистики T — W_α определяются по формуле

$$W_\alpha = \frac{n\,(n+1)}{4} + x_{\frac{\alpha}{2}} \sqrt{\frac{n\,(n+1)\,(2n+1)}{24}}, \qquad (4.3.1)$$

$$W_{1-\alpha} = \frac{n\,(n+1)}{2} - W_\alpha,$$

где $x_{\frac{\alpha}{2}}$ — квантиль нормального распределения, соответствующий вероятности $\frac{\alpha}{2}$. Для $\alpha=0{,}05$ $x_{\frac{\alpha}{2}}=-1{,}96$; для $\alpha=0{,}02$ $x_{\frac{\alpha}{2}}=-2{,}58$; для $\alpha=0{,}01$ $x_{\frac{\alpha}{2}}=-3{,}29$. H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T<W_\alpha$ или $T>W_{1-\alpha}$. Если верно неравенство $W_\alpha \leqslant T \leqslant W_{1-\alpha}$, то полученные наблюдения не дают достаточных оснований для отклонения нулевой гипотезы.

б) В случае одностороннего критерия для $n>20$ критические значения статистики $T-W_\alpha$ и $W_{1-\alpha}$ определяются по формулам

$$W_\alpha = \frac{n\ (n+1)}{4} + x_\alpha \sqrt{\frac{n\,(n+1)\ (2n+1)}{24}}, \qquad (4.3.2)$$

$$W_{1-\alpha} = \frac{n\ (n+1)}{2} - W_\alpha, \qquad (4.3.3)$$

где x_α — квантиль нормального распределения, определяемый для вероятности α. Для $\alpha=0{,}05$ $x_\alpha=-1{,}64$; для $\alpha=0{,}02$ $x_\alpha=-2{,}05$; для $\alpha=0{,}01$ $x_\alpha=-2{,}58$.

При проверке гипотезы H_0: медиана $D_i \leqslant 0$ — и альтернативы H_1: медиана $D_i>0$ H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T>W_{1-\alpha}$. При проверке гипотезы H_0: медиана $D_i \geqslant 0$ — и альтернативы H_1: медиана $D_i<0$ H_0 отклоняется на уровне значимости α, если наблюдаемое значение $T<W_\alpha$.

2. Случай одной выборки

Данные. Пусть имеется серия измерений переменной X, сделанных по интервальной шкале при рассмотрении одной случайной выборки из некоторой совокупности:

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_N.$$

Распределение переменной X в совокупности предполагается симметричным. В этих условиях критерий Вилкоксона может быть использован для проверки следующих гипотез:

а) двусторонний критерий

H_0: медиана переменной $X=c$ (c — постоянная величина, имеющая определенное числовое значение) — и альтернатива H_1: медиана переменной $X \neq c$;

б) односторонний критерий

H_0: медиана переменной $X \geqslant c$ — и альтернатива H_1: медиана переменной $X<c$,

в) односторонний критерий

H_0: медиана переменной $X \leqslant c$ — и альтернатива H_1: медиана переменной $X>c$.

Критерий Вилкоксона может быть использован и для проверки гипотез о среднем значении переменной X, так как распределение переменной X предполагается симметричным, и, следовательно, медиана переменной X равна среднему значению переменной X.

На основе постоянного числа c, выбранного в качестве предполагаемого значения медианы переменной X, и наблюдений $x_1, x_2, \ldots, x_i, \ldots, x_N$ составляются N пар вида: $(x_1, c), (x_2, c), \ldots, (x_i, c), \ldots, (x_N, c)$. К этим парам и применяется без каких-либо изменений критерий Вилкоксона, описанный в первой части данного раздела работы.

Примеры использования критерия Вилкоксона. Одним из допущений при использовании критерия Вилкоксона является симметричность распределения изучаемого свойства (т. е. переменной X — состояние изучаемого свойства до «эксперимента» и переменной Y — состояние данного свойства после «эксперимента»). Симметрия распределения[1] означает, что с равной вероятностью в совокупности рассматриваемых явлений появляются объекты как с высоким[2] (большие значения переменной X), так и с низким[3] (малые значения переменной X) уровнем изучаемого свойства.

Проведенные исследования, в том числе и педагогические, показали, что распределения некоторых свойств личности существенно отличны от симметричного распределения. При изучении некоторых психических свойств личности, например знаний, умений, навыков, способностей учащихся, форма распределения изучае-

---

[1] См. допущение 3 на с. 61.

[2] По сравнению с некоторым средним уровнем, характерным для изучаемой совокупности.

[3] По сравнению с некоторым средним уровнем, характерным для изучаемой совокупности.

5 Зак. 224

мого свойства зависит от характера вопросов или заданий, проверяющих состояние этого свойства.

Критерий Вилкоксона, как было указано выше, применяется в тех случаях, когда измерения выполнены хотя бы по интервальной шкале. В педагогических исследованиях измерениями состояния некоторого свойства по такой шкале могут служить: число верных (неверных) ответов или число ошибок, данных учащимся при выполнении нескольких вопросов или заданий.

При таком способе измерения форма распределения некоторого свойства может быть близка к симметричной, если большинство вопросов (заданий) не являются слишком трудными, а также слишком легкими. В противном случае распределение изучаемого свойства может существенно отличаться от симметричного, а потому применение критерия Вилкоксона будет необоснованным.

П р и м е р 1. Изучалось влияние просмотра некоторого диафильма на уровень усвоения определенного понятия. Проверка усвоения понятия проводилась с помощью 10 контрольных заданий, рассчитанных на среднего ученика. Работа выполнялась дважды одними и теми же 12 учащимися — до просмотра и после просмотра диафильма. Выполнение работы каждым учащимся оценивалось числом верных ответов, т. е. в данном эксперименте отметки учащихся могли иметь значение от 0 до 10.

В условиях данного эксперимента возможно применение критерия Вилкоксона для выявления значимости различия в знаниях учащихся до и после просмотра диафильма, так как выполнены все допущения данного критерия. Результаты двукратного выполнения работы 12 учащимися запишем в виде табл. 6, в которую затем занесем все вычисления, необходимые для определения статистики критерия.

Проверяется гипотеза H_0: медиана $D_i \leqslant 0$ — просмотр диафильма не улучшает состояния знаний учащихся. Содержание диафильма позволяет обоснованно предполагать, что состояние знаний учащихся повысится после его просмотра. Поэтому альтернативная гипотеза формулируется следующим образом: H_1 — просмотр диафильма повышает состояние знаний учащихся (H_1: медиана $D_i > 0$).

Таблица 6

Учащиеся, № п/п	1	2	3	4	5	6	7	8	9	10	11	12
Число верных ответов при первой проверке	3	8	5	4	2	1	6	7	1	3	4	2
Число верных ответов при второй проверке	4	8	6	2	6	3	6	10	6	10	7	8
Разность отметок	1	0	1	—2	4	2	0	3	5	7	3	6
Ранг абсолютного значения разности	1,5		1,5	3,5	7	3,5		5,5	8	10	5,5	9
R_i	1,5		1,5	—3,5	7	3,5		5,5	8	10	5,5	9

В этих условиях для проверки гипотезы применяется односторонний критерий Вилкоксона для $n \leqslant 20$ (см. с. 63, случай 2). По правилу, указанному на с. 62, подсчитывается значение статистики критерия T, равное сумме положительных значений R_i. В данном случае (см. табл. 6) $T = 1{,}5 + 1{,}5 + 7 + 3{,}5 + 5{,}5 + 8 + 10 + 5{,}5 + 9 = 51{,}5$. Из 12 пар 2 пары значений переменных имеют разность, равную нулю; значит, $n = 12 - 2 = 10$.

Для $n = 10$ и уровня значимости $\alpha = 0{,}05$ таблица B дает в случае одностороннего критерия критическое значение статистики $T - W_{1-\alpha} = 44$. Следовательно, выполняется неравенство $T_{\text{наблюд}} > T_{\text{критич}}$ $(51{,}5 > 44)$. Поэтому в соответствии с правилом принятия решения на с. 63 (случай 2) нулевая гипотеза отклоняется на уровне $\alpha = 0{,}05$ и принимается альтернативная. Таким образом, на основе результатов эксперимента с достоверностью $\theta = 1 - \alpha = 0{,}95$ справедлив вывод об улучшении состояния знаний после просмотра диафильма.

П р и м е р 2. В начале четверти учитель проверил грамотность учащихся с помощью диктанта. Убедившись в том, что учащиеся делают много ошибок, он разработал систему упражнений, способствующую устранению обнаруженных ошибок. После выполнения этих упражнений в домашних условиях в течение неко-

3* Зак. 224

торого периода времени учащиеся снова писали тот же диктант. Выполнение работы каждым учащимся оценивалось числом ошибок того типа, над устранением которых работал учитель. Из 120 учащихся, участвовавших в эксперименте, было отобрано методом случайного отбора 27. Результаты двукратного выполнения диктанта этими учащимися запишем в виде табл. 7,

Таблица 7

Ученики, № п/п	Число ошибок первый раз	Число ошибок второй раз	Разность числа ошибок	Ранг абсолютного значения разности	R_i
1	7	0	—7	25,0	—25,0
2	8	2	—6	23,0	—23,0
3	6	2	—4	16,5	—16,5
4	4	5	1	2,5	2,5
5	2	4	2	7,5	7,5
6	5	3	—2	7,5	— 7,5
7	3	1	—2	7,5	— 7,5
8	10	4	—6	23,0	—23,0
9	8	5	—3	12,5	—12,5
10	7	2	—5	20,0	—20,0
11	6	4	—2	7,5	— 7,5
12	10	5	—5	20,0	—20,0
13	3	3	0		
14	1	4	3	12,5	12,5
15	5	1	—4	16,5	—16,5
16	8	0	—8	26,0	—26,0
17	7	4	—3	12,5	—12,5
18	6	4	—2	7,5	— 7,5
19	4	5	1	2,5	2,5
20	10	6	—4	16,5	—16,5
21	9	7	—2	7,5	— 7,5
22	3	0	3	12,5	12,5
23	5	6	1	2,5	2,5
24	7	2	—5	20,0	—20,0
25	8	2	—6	23,0	—23,0
26	3	4	1	2,5	2,5
27	5	9	4	16,5	16,5

которая будет использована для получения данных, необходимых для подсчета статистики критерия.

В условиях данного эксперимента возможно использовать критерий Вилкоксона для сравнения различия в выполнении диктанта учащимися до и после самостоятельной работы над системой упражнений, так как

выполнены все допущения данного критерия. Знакомство с содержанием системы упражнений позволяет обоснованно выдвинуть предположение о том, что работа с этими упражнениями будет способствовать повышению грамотности учащихся. Поэтому проверяется гипотеза H_0: медиана $D_i \geqslant 0$ — при альтернативе H_1: медиана $D_i < 0$ (повышение грамотности приводит к уменьшению ошибок).

В этих условиях для проверки гипотез применяется односторонний критерий Вилкоксона для $n > 20$ (см. с. 64, случай б). Подсчитаем значение статистики критерия T — сумму положительных значений R_i. Согласно табл. 7, $T = 2,5+7,5+12,5+2,5+12,5+2,5+2,5+16,5 = 59,0$. Из 27 пар значений переменных только одна пара имеет разность, равную нулю, поэтому $n = 27-1 = 26$. Подсчитаем критическое значение статистики критерия W_α по формуле (4.3.2), указанной на с. 64. Для $\alpha = 0,05$, $x_\alpha = -1,64$ при $n = 26$

$$W_\alpha = \frac{26\ (26+1)}{4} - 1,64 \sqrt{\frac{26\ (26+1)\ (2 \cdot 26+1)}{24}} = 110,93\,.$$

Следовательно, выполняется неравенство $T_{\text{набл}} < W_\alpha$ ($59,0 < 110,93$), поэтому в соответствии с правилом принятия решения на с. 64 нулевая гипотеза отклоняется на уровне $\alpha = 0,05$ и принимается альтернативная гипотеза. Таким образом, приходим к выводу о том, что выполнение разработанной системы упражнений способствует повышению уровня знаний.

П р и м е р 3. Группа учащихся выполняла тестовую работу из 19 заданий. Выполнение работы каждым учащимся оценивалось числом верных ответов, поэтому оценки учащихся могли иметь значение от 0 до 19. Для выявления медианы оценок учащихся данной группы было отобрано методом случайного отбора 16 учащихся. Оценки этих учащихся и были использованы для определения значения медианы, т. е. для проверки гипотезы вида H_0: медиана $= c$ (c — заданное число).

В условиях данного эксперимента проверку такой гипотезы можно провести с помощью критерия Вилкоксона для случая одной выборки при $n \leqslant 20$.

1 Зак. 224

Результаты выполнения работы 16 учащимися запишем в виде табл. 8. Опираясь на эти результаты, сформулируем гипотезу H_0 следующим образом: медиана оценок учащихся=7, а альтернативную гипотезу H_1 так: медиана оценок≠7.

Таблица 8

Ученики, № п/п	Число верных ответов, x_i	Разность, $D_i = 7 - x_i$	Ранг $\|D_i\|$	R_i
1	2	5	10,5	10,5
2	3	4	9	9
3	4	3	6,5	6,5
4	6	1	2	2
5	13	−6	12	−12
6	10	−3	6,5	−6,5
7	7	0		
8	12	−5	10,5	−10,5
9	10	−3	6,5	−6,5
10	14	−7	13	−13
11	15	−8	14	−14
12	4	3	6,5	6,5
13	6	1	2	2
14	8	−1	2	−2
15	7	0		
16	5	2	4	4

Подсчитаем значение статистики критерия T — сумму положительных значений R_i. Согласно табл. 8, $T=10{,}5+9+6{,}5+2+6{,}5+2+4=40{,}5$. Из 16 значений переменной только для двух $D_i=7-x_i=0$; значит, $n=16-2=14$. В табл. *В* для $\alpha=0{,}05$ и $n=14$ для двустороннего критерия даются следующие критические значения статистики критерия $T: W_{\frac{\alpha}{2}}=22$ и $W_{1-\frac{\alpha}{2}}=83$. Следовательно, выполняется неравенство $W_{\frac{\alpha}{2}}<T_{\text{наблюд}}<W_{1-\frac{\alpha}{2}}$ $(22<40{,}5<83)$, поэтому в соответствии с правилом принятия решений на с. 62 у нас нет достаточных оснований для отклонения нулевой гипотезы на уровне значимости $\alpha=0{,}05$, т. е. считать, что медиана оценок учащихся, выполнявших работу, не будет равна 7.

5 СРАВНЕНИЕ РЕЗУЛЬТАТОВ ДВУХ НЕЗАВИСИМЫХ ВЫБОРОК

В данном разделе будут рассмотрены 4 метода проверки статистических гипотез, основанных на сравнении измерений некоторого свойства у членов двух независимых выборок: медианный критерий, критерий Вилкоксона — Манна — Уитни, критерий χ^2 (хи-квадрат), критерий Колмогорова — Смирнова.

5.1. Медианный критерий

Критерий предназначен для выявления различия в центральных тенденциях состояния некоторого свойства в двух совокупностях на основе изучения членов двух независимых выборок из этих совокупностей. В данном случае показателем центральной тенденции служит медиана измерений изучаемого свойства в каждой из выборок.

Данные. Пусть случайная переменная X характеризует состояние изучаемого свойства в одной из рассматриваемых совокупностей, а случайная переменная Y — состояние того же свойства во второй совокупности.

Имеется две серии наблюдений над случайными переменными X и Y:

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_{n_1};$$

$$y_1,\ y_2,\ \ldots,\ y_j,\ \ldots,\ y_{n_2},$$

полученные при рассмотрении двух независимых выборок объема n_1 и n_2; x_i $(i=1,\ 2,\ \ldots,\ n_1)$ — результаты измерения некоторого свойства у объектов первой выборки, составленной из членов одной совокупности; y_j $(j=1,\ 2,\ \ldots,\ n_2)$ — результаты измерения того же свойства у объектов второй выборки, составленной из членов другой совокупности.

6 Зак. 224

В педагогических исследованиях объектами изучения могут быть учителя или учащиеся различных, независимых друг от друга групп, классов, школ и т. п. При этом x_i можно получить, например, в виде балловых оценок за выполнение некоторой деятельности учащимися школ одного города, а y_j — в виде балловых оценок за выполнение той же деятельности учащимися школ другого города.

Обе серии наблюдений объединяют в одну выборку, объем которой равен n_1+n_2 и обозначается N. Определяется медиана этой выборки m, после чего члены каждой выборки объемов n_1 и n_2 распределяются на две категории: больше общей медианы ($>m$) и меньше или равны общей медиане ($\leqslant m$). На основе полученных результатов составляется таблица 2×2:

	Выборка № 1	Выборка № 2
$>m$	A число x_i, больших m	B число y_j, больших m
$\leqslant m$	C число x_i, меньших или равных m	D число y_j, меньших или равных m

Рассмотрим способы определения медианы при различных значениях N. Медианой N измерений является такое значение, больше которого 50% или $\left(\frac{N}{2}\right)$ этих измерений и меньше которого тоже 50% $\left(\text{или } \frac{N}{2}\right)$ этих измерений.

Для нахождения медианы измерения записывают в ряд по возрастанию значений. Если число измерений N нечетное, то медиана численно равна значению этого ряда, стоящему точно в середине, или на $\frac{N+1}{2}$ месте. Например, медиана пяти ($N=5$) измерений: 10, 17, 21, 24, 25 — равна 21 — значению, стоящему на третьем месте $\left(\frac{N+1}{2}=\frac{5+1}{2}=3\right)$. Если число измерений четное, то медиана численно равна среднему арифметическому значений ряда, стоящих в середине, или на $\frac{N}{2}$ и $\frac{N}{2}+1$ местах. Например, медиана восьми измерений: 5, 5, 6, 7, 8, 8, 9, 9 — равна 7,5 $\left(\frac{7+8}{2}=7{,}5\right)$ — среднему арифметическому значений ряда, стоящих на четвертом и пятом местах $\left(\frac{N}{2}=\frac{8}{2}=4;\ \frac{N}{2}+1=4+1=5\right)$.

Допущения. Для применения медианного критерия необходимо выполнение следующих требований:

1) обе рассматриваемые выборки являются случайными выборками из некоторых совокупностей;

2) выборки независимы, и члены каждой выборки независимы между собой;

3) шкала измерений не ниже порядковой, так как при измерениях по шкале наименований невозможно нахождение медианы;

4) число членов в обеих выборках должно быть в сумме больше 20, т. е. $n_1+n_2>20$.

Гипотезы. Предположим, что законы распределения случайных величин X и Y одинаковы. Тогда выполняется и такое равенство: медиана значений x_i = медиане значений y_j. Справедливость этого равенства и проверяется с помощью медианного критерия. Таким образом, нулевая гипотеза имеет вид H_0: обе совокупности, из которых взяты выборки, имеют одну и ту же медиану. В качестве альтернативной гипотезы выбирается гипотеза H_1: совокупности имеют разные медианы.

Если гипотеза H_1 справедлива, то отсюда следует, что законы распределения изучаемого свойства различны в совокупностях, из которых взяты выборки, т. е. состояние изучаемого свойства в рассматриваемых совокупностях существенно различно.

Статистика критерия. Для проверки гипотез с помощью медианного критерия на основе наблюдений, записанных в форме таблицы 2×2, подсчитывается по формуле (5.1.1) величина T — статистика критерия:

$$T=\frac{N\left(|AD-BC|-\frac{N}{2}\right)^2}{(A+B)\,(C+D)\,(A+C)\,(B+D)}. \qquad (5.1.1)$$

Правило принятия решения. Вместо точного распределения статистики T используется приближенное, а именно для достаточно больших значений n_1 и n_2 (n_1, n_2 — объемы выборок) распределение величины T аппроксимируется достаточно точно распределением χ^2 с одной степенью свободы ($\nu=1$). Критические значения T для разных уровней значимости α даются в табл. Γ. Гипотеза H_0 отклоняется на данном уровне значимости α, если наблюдаемое значение $T > x_\alpha$, где x_α — квантиль распределения χ^2, определяемый по табл. Γ для выбранного α.

Не рекомендуется использовать медианный критерий для проверки гипотез, если $20 < n_1 + n_2 < 40$ и хотя бы одно из значений A, B, C, D, записанных в таблице 2×2, меньше 5.

Не рекомендуется также применять медианный критерий, если около середины ряда, составленного из упорядоченных по возрастанию значений обеих выборок, будет расположено значительное число одинаковых значений, принадлежащих обеим выборкам.

Если одинаковые значения, принадлежащие обеим выборкам, группируются около середины ряда, но составляют небольшую часть суммарного объема обеих выборок, то медиана подсчитывается по правилу, указанному на с. 72, и применение критерия возможно.

Некоторые исследователи [19, 21] в случае группировки нескольких одинаковых значений, принадлежащих обеим выборкам, около середины ряда рекомендуют подсчитывать медиану (m) данного ряда по следующей формуле:

$$m = L + \frac{\frac{N}{2} - \Sigma f}{f} \cdot h, \qquad (5.1.2)$$

где N — сумма объемов двух выборок; L — нижняя граница интервала значений ряда, к которому принадлежит значение ряда, стоящее на $\frac{N^*}{2}$ месте; Σf — число значений ряда, меньших L; f — число значений ряда, принадлежащих интервалу, включающему значение ряда, стоящее на $\frac{N}{2}$ месте; h — величина интервала, выбранная при группировке значений ряда.

Использование формулы (5.1.2) целесообразно, если значения обеих выборок записаны в форме интервального ряда. При этом необходимо иметь в виду, что использование формулы возможно только в тех случаях, когда изучаемое свойство педагогического явления измерено по шкале не ниже интервальной. Следует отметить, что применение формулы в подобных случаях имеет смысл, когда результаты измерений выражены дробными числами. Однако при выполнении этих требований возможно использование критериев Вилкоксона — Манна — Уитни или Колмогорова — Смирнова, позволяющих получить более достоверные выводы по сравнению с медианным критерием.

* При N нечетном берется значение ряда, стоящее на $\frac{N+1}{2}$ месте.

Использование медианного критерия целесообразно лишь в случае измерений изучаемого свойства по шкале порядка с достаточно большим числом градаций.

П р и м е р 1. В двух районах города проводилась проверка усвоения учащимися VIII класса метода решения задач с помощью составления уравнения по условию задачи. Учащимся предлагалось решить 3 задачи, различные по содержанию. Решение каждой из них разбивалось на четыре этапа: обозначение неизвестных, составление уравнения, решение уравнения, ответ на вопрос задачи.

Ответы учащихся при выполнении каждого этапа разбивались на две категории: верный, неверный. За верный ответ выставлялся один балл, за неверный — нуль. Таким образом, за верное выполнение всех этапов решения задачи учащийся мог получить 4 балла, за верное решение трех задач — 12 баллов. Если учащийся на предыдущем этапе решения допускал ошибку, то оценка выполнения последующих этапов выставлялась с учетом допущения, что результат предыдущего этапа верный. Например, при неверно составленном уравнении выполнение второго этапа решения оценивалось нулем. Далее оценивалась правильность решения этого уравнения и ответа на вопрос задачи с учетом корней данного уравнения. В тех случаях, когда учащийся не приступал к решению задачи, ему за данную задачу присваивался нуль. За каждый из невыполненных этапов неоконченного решения учащемуся присваивался нуль.

Из всех учащихся, писавших работу, методом случайного отбора в первом районе было отобрано 28 учащихся, которые составили первую выборку, во втором районе — 32 ученика, которые составили вторую выборку. На основе результатов сравнения выполнения заданий этими выборками учащихся предполагалось проверить гипотезу об отсутствии различий в состоянии умения решать задачи методом составления уравнения для учащихся различных районов города. В качестве показателя состояния умения в выборке учащихся использовалось распределение учащихся первой выборки (переменная X) и второй выборки (переменная Y) по числу баллов, выставленных каждому ученику за выполнение трех задач.

В условиях проведенного эксперимента возможно применение медианного критерия для выявления различия центральных тенденций (медиан) в состоянии проверяемого умения в двух совокупностях учащихся (учащиеся первого и второго районов города).

Результаты выполнения контрольной работы запишем в форме табл. 9, удобной для нахождения общей

Таблица 9

Число баллов	Абсолютная частота в первой выборке, f_1*	Абсолютная частота во второй выборке, f_2**	$f = f_1 + f_2$	Накопленная частота, Σf
12	2	1	3	60
11	1	2	3	57
10	—	1	1	54
9	—	5	5	53
8	3	3	6	48
7	4	7	11	42
6	6	3	9	31
5	5	5	10	22
4	4	3	7	12
3	1	—	1	5
2	2	2	4	4
1	—	—	—	—
0	—	—	—	—
	$n_1 = 28$	$n_2 = 32$	$N = 60$	

* f_1 — абсолютная частота появления конкретного числа баллов в первой выборке (например, 8 баллов получили 3 ученика первой выборки);

** f_2 — абсолютная частота появления конкретного числа баллов во второй выборке.

медианы ряда распределения учащихся двух выборок по числу баллов, полученных за решение контрольных задач.

На основе данных табл. 9 найдем медиану ряда распределения учащихся обеих выборок по числу баллов, полученных ими за выполнение работы. В данном случае медианой является число баллов, больше которого получили 50%, или $\frac{N}{2}$, учащихся обеих выборок, т. е. 30 учащихся ($N=60$, 50% от $60=30$), и меньше которого получили 50% учащихся обеих выборок, т. е. тоже 30 учащихся.

Число учащихся в двух выборках N равно 60 — число четное. Значит, медиана равна среднему арифметическому значений, стоящих в упорядоченном ряду измерений на $\frac{N}{2}$ и $\frac{N}{2}+1$ местах, т. е. на 30-м и 31-м местах. Однако в данном ряду (см. с. 76, последняя колонка табл. 9), начиная с 23-го и кончая 31-м номером, расположены одинаковые значения, каждое из которых равно 6[1]. Значит, среднее арифметическое значений, стоящих на 30-м и 31-м местах, равно 6 $\left(\frac{6+6}{2}=6\right)$, т. е. медиана равна 6.

Используя данные 1-й и 2-й колонок табл. 9, распределим значения каждой из выборок учащихся на две категории: больше медианы (больше 6) и меньше или равны медиане (меньше или равны 6). Запишем полученные результаты в форме таблицы 2×2, необходимой для подсчета статистики медианного критерия:

	Выборка № 1	Выборка № 2	
Число учащихся, отметки которых больше 6 баллов (>6)	10	19	29
Число учащихся, отметки которых меньше или равны 6 баллам ($\leqslant 6$)	18	13	31
	28	32	60

[1] В соответствии с замечанием на с. 74 использование медианного критерия возможно, так как одинаковые значения, группирующиеся около середины ряда, составляют небольшую часть объема двух выборок.

Проверяется гипотеза H_0: $m_1 = m_2$ — медианы распределения учащихся по числу баллов, полученных за выполнение работы, одинаковы в совокупностях учащихся первого и второго районов города. Альтернативная гипотеза H_1: $m_1 \neq m_2$.

В условиях проведенного эксперимента проверка данной гипотезы может проводиться с помощью медианного критерия, так как выполнены все допущения (см. с. 73), необходимые для применения критерия.

На основе данных таблицы 2×2 находим значение статистики критерия T по формуле (5.1.1) на с. 73:

$$T = \frac{60 \cdot \left(| 10 \cdot 13 - 18 \cdot 19 | - \frac{60}{2} \right)^2}{28 \cdot 32 \cdot 29 \cdot 31} = 2{,}47.$$

Для уровня значимости $\alpha = 0{,}05$ и одной степени свободы по табл. Г находим $T_{\text{критич}} = 3{,}84$. Значит, верно неравенство $T_{\text{наблюд}} < T_{\text{критич}}$. Следовательно, в соответствии с правилом принятия решения на с. 74 на уровне значимости $\alpha = 0{,}05$ у нас нет достаточных оснований для отклонения нулевой гипотезы, т. е. нет достаточных оснований считать различными медианы распределений учащихся по числу баллов в совокупностях учащихся первого и второго районов.

В то же время данные табл. 9 (см. с. 76) показывают, что результаты учащихся второй выборки выше, чем результаты учащихся первой выборки. Одной из причин того, что применение медианного критерия не позволило отклонить гипотезу и тем самым не позволило выявить эту тенденцию, может быть сравнительно небольшой объем каждой из выборок.

Увеличим объем каждой выборки до 40 учащихся. Для этого из учащихся первого района, писавших работу, методом случайного отбора выберем недостающие 12 человек, а из учащихся второго района — 8 человек. Запишем результаты выполнения работы учащимися полученных выборок в форме табл. 10, удобной для определения медианы.

Число учащихся в двух выборках N равно 80 — число четное. Значит, медиана равна среднему арифметическому значений, стоящих на $\frac{N}{2}$ и $\frac{N}{2}+1$ местах, т. е. на 40-м и 41-м местах, в упорядоченном ряду измерений обеих выборок. В данном случае (см. табл. 10) на 40-м

Таблица 10

Число баллов	Абсолютная частота в первой выборке, f_1	Абсолютная частота во второй выборке, f_2	$f = f_1 + f_2$	Накопленная частота, Σf
12	2	2	4	80
11	1	2	3	76
10	—	2	2	73
9	1	7	8	71
8	3	3	6	63
7	7	10	17	57
6	8	4	12	40
5	9	5	14	28
4	4	3	7	14
3	3		3	7
2	2	2	4	4
1	—	—	—	—
0	—	—	—	—
	$n_1 = 40$	$n_2 = 40$	$N = 80$	

месте стоит значение, равное 6, а на 41-м месте стоит значение, равное 7. Значит, среднее арифметическое значений, стоящих на 40-м и 41-м местах, равно 6,5 $\left(\frac{6+7}{2} = 6{,}5\right)$, т. е. медиана ряда равна 6,5.

Используя данные 1-го и 2-го столбцов табл. 10, распределим значения обеих выборок на две категории: больше медианы ($>6{,}5$) и меньше или равны медиане ($\leqslant 6{,}5$). Запишем результаты в виде таблицы 2×2:

	Выборка № 1	Выборка № 2	
Больше 6,5 ($>6{,}5$)	14	26	40
Меньше или равны 6,5 ($\leqslant 6{,}5$)	26	14	40
	40	40	80

На основе данных таблицы 2×2 подсчитаем статистику медианного критерия T по формуле (5.1.1) на с. 73.

$$T = \frac{80 \cdot \left(| 14 \cdot 14 - 26 \cdot 26 | - \frac{80}{2} \right)^2}{40 \cdot 40 \cdot 40 \cdot 40} = 7{,}2.$$

Для $\alpha=0{,}05$ и одной степени свободы по табл. Г находим критическое значение статистики T: $T_{\text{критич}}=3{,}84$. Так как верно неравенство $T_{\text{наблюд}}>T_{\text{критич}}$ ($7{,}2>3{,}84$), в соответствии с правилом принятия решения на с. 74 нулевая гипотеза (см. с. 78) отклоняется и принимается альтернативная гипотеза: медианы распределений учащихся по числу баллов за выполнение работы различны в совокупностях учащихся первого и второго районов города. При этом результаты учащихся второго района имеют тенденцию быть выше результатов учащихся первого района, т. е. состояние умения решать задачи методом составления уравнения (которое проверялось контрольной работой) выше во втором районе.

Пример 2. Проводилась проверка усвоения первого закона Ньютона учащимися выпускных классов городских и сельских школ. При этом использовалась методика проверки знаний по физике, разработанная В. И. Травинским[1]. Согласно этой методике, объективная оценка знаний учащихся по некоторому разделу программы может быть получена, если усвоение программы проверялось на разных уровнях. Таких уровней выделено четыре: фактический (знание фактов), операционный (осуществление действий или логических операций по образцу), аналитико-синтетический (определение связей, нахождение аналогий отдельным понятиям, разделам темы, выделение основных идей) и творческий (перенос знаний в новые ситуации, создание оригинальных алгоритмов познавательных и практических действий).

Для проверки усвоения первого закона Ньютона автор предлагает четыре вопроса, из которых каждый проверяет знания на одном из этих уровней.

Порядковую шкалу оценки верного ответа на вопросы, проверяющие знания выпускников средней школы на разных уровнях, автор строит с учетом процентов верных ответов на такие вопросы учащихся данной совокупности. Оценка процентов верных ответов на вопросы каждого из четырех уровней проводилась на основе результатов проверки знаний 11 000 учащихся.

---

[1] См.: *Травинский В. И.* Уровни знаний и критерии их усвоения. Канд. дис. М., 1971.

Выяснилось, что процент верных ответов среди выпускников различных по географическому расположению школ страны на вопросы, проверяющие усвоение основных понятий и разделов курса физики 10-летней школы, на фактическом уровне устойчиво колеблется около 70, на операционном уровне — около 50, на аналитико-синтетическом уровне — около 40, на творческом уровне — около 33.

В соответствии с полученными результатами было решено верный ответ на вопрос первого уровня оценивать 1 баллом, второго уровня — 2 баллами, третьего уровня — 3 баллами, верный ответ на вопрос четвертого уровня — 4 баллами.

Контрольную работу, состоящую из четырех вопросов, проверяющих усвоение первого закона Ньютона на четырех уровнях, писали 42 ученика из четырех сельских школ и 40 учащихся четырех городских школ. Работы нескольких учащихся по разным причинам (состояние здоровья, непонятные записи ответов и др.) не учитывались при получении конечных результатов выборки сельских и городских учащихся. Поэтому объем выборки из сельских школ оказался равным 41, а объем выборки из городских школ — 36 учащимся.

При верных ответах на все четыре вопроса работы учащийся получал 10 баллов ($4+3+2+1=10$), при всех неверных ответах — нуль. Таким образом результаты выполнения работы учащимися, выраженные в числе баллов, распределялись от 0 до 10. Ряды распределения учащихся каждой выборки по числу баллов, полученных за выполнение работы, использовались в качестве показателей усвоения учащимися сельских и городских школ первого закона Ньютона.

На основе сравнения этих рядов предполагалось проверить предположение о том, что усвоение проверяемых знаний не имеет существенных различий у учащихся городских и сельских школ. В условиях проведенного эксперимента выполняются основные допущения медианного критерия, который и будет применяться для проверки данного предположения.

Запишем результаты учащихся обеих выборок в форме табл. 11, удобной для подсчета медианы.

Число учащихся в обеих выборках N равно 77 — число нечетное. Значит, медиана численно равна значе-

Таблица 11

Число баллов	Абсолютная частота в первой выборке, f_1	Абсолютная частота во второй выборке, f_2	$f = f_1 + f_2$	Накопленная частота, Σf
10	1	—	1	77
9	—	2	2	76
8	5	—	5	74
7	3	2	5	69
6	7	4	11	64
5	3	7	10	53
4	2	3	5	43
3	8	10	18	38
2	3	4	7	20
1	4	6	10	13
0	—	3	3	3
	$n_1 = 36$	$n_2 = 41$	$N = 77$	

нию, стоящему на $\frac{N+1}{2}$, или на 39-м, месте $\left(\frac{77+1}{2} = 39\right)$. В упорядоченном ряду измерений, составленном по результатам двух выборок, это значение равно 4

Используя данные 1-го и 2-го столбцов табл. 11, распределим значения обеих выборок на две категории: больше медианы (>4) и меньше или равны медиане ($\leqslant 4$). Полученные результаты запишем в форме таблицы 2×2:

	Выборка № 1	Выборка № 2	
Больше 4 (>4)	19	15	34
Меньше или равны 4 ($\leqslant 4$)	17	26	43
	36	41	77

Проверяется гипотеза H_0: медианы распределения учащихся по числу баллов, полученных за выполнение работы, одинаковы в совокупностях учащихся городских и сельских школ, т. е. $m_1 = m_2$. Альтернативная гипотеза H_1: $m_1 \neq m_2$.

Значение статистики медианного критерия находим по формуле (5.1.1) на с. 73, используя данные таблицы 2×2:

$$T = \frac{77 \cdot \left(| 19 \cdot 26 - 15 \cdot 17 | - \frac{77}{2} \right)^2}{36 \cdot 41 \cdot 34 \cdot 43} = 1{,}43.$$

Для уровня значимости $\alpha = 0{,}05$ и одной степени свободы по табл. Г находим $T_{критич} = 3{,}84$. Значит, верно неравенство $T_{критич} > T_{наблюд}$ ($3{,}84 > 1{,}43$). Следовательно, согласно правилу принятия решения на с. 74, на уровне значимости $\alpha = 0{,}05$ у нас нет достаточных оснований для отклонения нулевой гипотезы H_0, т. е. нет достаточных оснований считать различными медианы распределения учащихся по числу баллов за выполнение работы, проверяющей усвоение первого закона Ньютона, а значит, считать различными уровни усвоения первого закона Ньютона учащимися сельских и городских школ.

5.2. Критерий Вилкоксона — Манна — Уитни

Критерий предназначен для выявления различий в распределениях изучаемого свойства у объектов двух совокупностей на основе сравнения результатов изучения данного свойства у членов независимых выборок, сделанных из этих совокупностей. Применение критерия позволяет проверить предположение о различии центральных тенденций[1] состояния изучаемого свойства в рассматриваемых совокупностях. В качестве показателей центральной тенденции рассматриваются медиана и среднее значение. Таким образом, с помощью данного критерия проверяется одновременно предположение о различии медиан, а также средних значений в рассматриваемых совокупностях. Следует иметь в виду, что различие медиан или средних значений свидетельствует о наличии тенденции членов первой совокупности превышать (быть меньше) по состоянию изучаемого свойства члены второй совокупности. Применение критерия основано на использовании рангов, приписанных упорядоченным объектам обе-

---

[1] См. введение, с. 11.

их выборок. Упорядочение объектов возможно только в том случае, когда шкала измерений изучаемого свойства не ниже порядковой.

Критерий может найти применение в педагогических исследованиях, так как позволяет проверить одновременно несколько предположений о характере различия распределений некоторого свойства в двух совокупностях (различие медиан, средних значений распределений обеих совокупностей, а также наличие тенденции объектов первой совокупности быть в среднем больше или, наоборот, меньше членов другой совокупности) на основе порядковых измерений состояния этого свойства. Таким образом, критерий более тонко улавливает различие распределений некоторого свойства в двух совокупностях по сравнению с медианным критерием. Кроме того, в педагогических исследованиях пока чрезвычайно редко удается измерить желаемое свойство личности по интервальной шкале, зато возможно во многих случаях разработать в соответствии с целями исследования порядковую шкалу для измерения этого свойства.

Данные. Будем считать, что случайная переменная X характеризует состояние изучаемого свойства в одной из рассматриваемых совокупностей, а случайная переменная Y характеризует состояние этого же свойства в другой совокупности.

Пусть имеется две серии наблюдений над случайными переменными X и Y:

$$x_1,\ x_2,\ \ldots,\ x_i,\ \ldots,\ x_{n_1};$$

$$y_1,\ y_2,\ \ldots,\ y_j,\ \ldots,\ y_{n_2},$$

полученные при рассмотрении двух независимых выборок объема n_1 и n_2; x_i $(i=1, 2, \ldots, n_1)$ — результаты измерения некоторого свойства у объектов первой выборки, y_j $(j=1, 2, \ldots, n_2)$ — результаты измерения того же свойства у объектов второй выборки.

Обе серии наблюдений объединяют в одну выборку, объем которой равен n_1+n_2 и обозначается N. Члены этой выборки записывают в ряд по возрастанию значений. Затем этот ряд ранжируют, т. е. каждому члену ряда приписывается некоторое число, которое называется рангом. Члену ряда, стоящему на 1-м от начала месте, приписывают число 1, стоящему на 2-м месте — число 2 и т. д. Таким образом, ранг члена ряда численно равен номеру места, которое занимает данный член в этом ряду. Ранг последнего члена ряда будет равен N $(N=n_1+n_2)$. Если несколько членов ряда, стоящих друг за другом, имеют одно и то же значение, то каждому из них приписывается один и тот же ранг, равный среднему арифметическому номеров мест, на которых стоят эти члены[1]. Далее находят сумму рангов, приписанных членам выборки меньшего объема.

Допущения. Применение критерия Вилкоксона — Манна — Уитни для проверки гипотез возможно при выполнении следующих требований:

1) обе выборки случайные;

2) выборки независимы, и члены каждой выборки также независимы между собой;

3) изучаемое свойство объектов распределено непрерывно[2] в обеих совокупностях, из которых сделаны выборки;

4) шкала измерений не ниже порядковой.

Гипотезы. Предположим, что законы распределения случайных величин X и Y в обеих совокупностях одинаковы. При этом условии будет верным равенство: $P(X<Y)=\frac{1}{2}$, т. е. с одинаковой вероятностью, равной 0,5, значения переменной X в первой совокупности будут больше или меньше значений переменной Y во второй совокупности. Справедливость данного равенства и проверяется с помощью критерия Вилкоксона — Манна — Уитни.

Двусторонний критерий позволяет проверить нулевую гипотезу H_0: $P(X<Y)=\frac{1}{2}$ — при альтернативной гипотезе H_1: $P(X<Y)\neq\frac{1}{2}$.

Если гипотеза H_1 истинна, то из этого следует, что законы распределения X и Y различны в обеих совокупностях, т. е. различны состояния изучаемого свойства в объектах первой и второй совокупностей.

Критерий также состоятелен для проверки гипотез относительно медиан изучаемых случайных величин X и

---

[1] См. пример приписывания рангов членам ряда на с. 91—92.
[2] См. введение, с. 8.

Y. Нулевая гипотеза имеет вид H_0: медиана (X) = медиане (Y). Если существуют[1] средние значения переменных X и Y, то критерий может применяться для проверки гипотез относительно средних значений изучаемых случайных величин X и Y. Нулевая гипотеза будет иметь вид H_0: $M(X)=M(Y)$, где M обозначает среднее значение.

Для проверки нулевой гипотезы при альтернативах более общего типа, чем указанные выше, критерий Вилкоксона — Манна — Уитни не является состоятельным[2].

Статистика критерия. Подсчитывается сумма рангов, приписанных членам одной из выборок. Обычно предпочтение отдается выборке с меньшим объемом. Обозначим сумму рангов, приписанных членам этой выборки, через S. Тогда

$$S=\sum_{i=1}^{n} R(x_i), \tag{5.2.1}$$

где $R(x_i)$ — ранг, приписанный i-му объекту данной выборки; n_1, n_2 — объемы выборок, а n равно минимальному из значений n_1 и n_2.

Для проверки гипотез с помощью критерия Вилкоксона — Манна — Уитни подсчитывается значение статистики критерия T по следующей формуле:

$$T=S-\frac{n(n+1)}{2}. \tag{5.2.2}$$

Правило принятия решения. Пусть объемы выборок равны n_1 и n_2 и α — принятый уровень значимости. Рассмотрим правила принятия решений при проверке различных гипотез.

1. Двусторонний критерий. Проводится проверка гипотезы H_0: $P(X<Y)=\frac{1}{2}$ — при альтернативе H_1: $P(X<Y)\neq\frac{1}{2}$.

Для $n_1\leqslant 20$ и $n_2\leqslant 20$ составлена табл. *Д* критических значений статистики T для различных значений α. В данной работе в табл. *Д* W_α — критические значения статистики T — даны для $\alpha=0{,}02$; $0{,}05$; $0{,}10$; $0{,}20$ при $5\leqslant n_1\leqslant 20$ и $5\leqslant n_2\leqslant 20$.

При данных значениях n_1 и n_2 H_0 отклоняется на уровне α, если для наблюдаемого значения T верно одно из неравенств $T<W_{\frac{\alpha}{2}}$ или $T>W_{1-\frac{\alpha}{2}}$; $W_{\frac{\alpha}{2}}$ определяется по табл. *Д*, а значение $W_{1-\frac{\alpha}{2}}$ можно найти по формуле

$$W_{1-\frac{\alpha}{2}}=n_1 n_2-W_{\frac{\alpha}{2}}. \tag{5.2.3}$$

Если хотя бы одно из значений n_1 и n_2 больше 20, то критические значения статистики T находят по формуле

$$W_{\frac{\alpha}{2}}=\frac{n_1 n_2}{2}+x_{\frac{\alpha}{2}}\sqrt{\frac{n_1 n_2(n_1+n_2+1)}{12}}, \tag{5.2.4}$$

где $x_{\frac{\alpha}{2}}$ — квантиль нормального распределения.

Укажем значения $x_{\frac{\alpha}{2}}$ для некоторых значений α: для $\alpha=0{,}01$ $x_{\frac{\alpha}{2}}=2{,}58$; для $\alpha=0{,}05$ $x_{\frac{\alpha}{2}}=1{,}96$; для $\alpha=0{,}10$ $x_{\frac{\alpha}{2}}=1{,}64$.

Нулевая гипотеза H_0 отклоняется на уровне α, если для наблюдаемого значения статистики критерия T верно одно из неравенств $T<W_{\frac{\alpha}{2}}$ или $T>n_1n_2-W_{\frac{\alpha}{2}}$.

2. Односторонний критерий. При наличии достаточных оснований для предположения о том, что значения переменной X имеют тенденцию в среднем превышать (быть меньше) значения переменной Y, для проверки гипотез вместо двустороннего критерия применяется односторонний.

а) Для случая, когда предполагается, что значения X в среднем больше значений Y, проводится проверка нулевой гипотезы H_0: $P(X<Y)\geqslant\frac{1}{2}$ — при альтернативе H_1: $P(X<Y)<\frac{1}{2}$.

Критические значения статистики критерия определяются по той же таблице *Д*, что и для двустороннего критерия. H_0 отклоняется на уровне значимости α, если

[1] Известны примеры случайных величин, для которых средние значения не существуют.

[2] См. работу [6].

наблюдаемое значение T больше $W_{1-\alpha}$ ($W_{1-\alpha}=n_1n_2-W_\alpha$, а W_α определяется по табл. *Д* для данных значений n_1, n_2, α). В таблице предлагаются значения W_α для $\alpha=0{,}01$; $0{,}025$; $0{,}05$; $0{,}10$.

б) В том случае, когда предполагается, что значения переменной X в среднем меньше значений переменной Y, проводится проверка гипотезы H_0: $P(X<Y)\leqslant\frac{1}{2}$ — при альтернативе H_1: $P(X<Y)>\frac{1}{2}$.

Нулевая гипотеза отклоняется на уровне значимости α, если наблюдаемое значение T меньше W_α (W_α определяется по табл. *Д* для данных значений n_1, n_2, α).

Если хотя бы одно из значений n_1 или n_2 больше 20, то критическое значение статистики критерия T определяется по формуле

$$W_\alpha=\frac{n_1\,n_2}{2}+x_\alpha\sqrt{\frac{n_1\,n_2\,(n_1+n_2+1)}{12}},\qquad(5.2.5)$$

где x_α — квантиль нормального распределения.

Укажем значения x_α для некоторых значений α: для $\alpha=0{,}01$ $x_\alpha=2{,}33$; для $\alpha=0{,}05$ $x^\alpha=1{,}64$; для $\alpha=0{,}10$ $x_\alpha=1{,}28$.

При применении критерия предполагается, что обе переменные X и Y имеют непрерывное распределение. В этих условиях при точных измерениях вероятность появления одинаковых значений переменных равна нулю. Однако на практике чаще всего не удается добиться высокой точности измерений, что и является причиной появления одинаковых значений. При ранжировании таким значениям приписывается ранг, равный среднему арифметическому номеров мест, которые они занимают среди членов двух выборок, упорядоченных по возрастанию значений.

Приписывание одинаковых рангов снижает точность критерия в том случае, когда одинаковые значения принадлежат обеим выборкам, и не оказывает никакого влияния на точность конечных выводов, если эти значения принадлежат только одной из выборок. Поэтому авторы критерия рекомендуют вводить коррекцию в формулу подсчета критического значения статистики критерия при наличии совпадающих значений в обеих

выборках. Формула для вычисления квантиля W_α имеет вид

$$W_\alpha=\frac{n_1\,n_2}{2}+x_\alpha\sqrt{\frac{n_1\,n_2\,(n_1+n_2+1)}{12}-\Sigma K},\qquad(5.2.6)$$

где

$$K=\frac{k^3-k}{12}\qquad(5.2.7)$$

и k — число членов ряда, имеющих одно и то же конкретное значение, ΣK — сумма значений K для всех цепочек совпадающих значений, принадлежащих обеим выборкам.

Используя приведенную выше формулу, следует иметь в виду, что эффект коррекции невелик даже в том случае, если около 90% измерений в обеих выборках составляют несколько цепочек совпадающих значений. Авторы рекомендуют использовать данную формулу, если имеются длинные цепочки одинаковых значений, принадлежащих обеим выборкам.

Пример. Проводилась проверка усвоения курса химии IX класса у учащихся экспериментальных и контрольных классов. Учащимся предлагалось четыре задания, каждое из которых состояло из нескольких вопросов, требующих иногда кратких, иногда развернутых ответов.

Для установления порядковой шкалы измерения знаний по химии проводилось взвешивание верных ответов на отдельные вопросы. Взвешивание проводилось на основе сравнения глубины и объема знаний, необходимых для получения верного ответа на отдельные вопросы. С этой целью по каждому вопросу выделялись элементы знаний, которые должны были быть использованы учащимся при ответе на данный вопрос. Эти элементы сравнили между собой и распределили на три категории трудности. За верное применение элемента знаний первой категории было решено присваивать учащемуся 1 балл, второй категории — 2 балла, третьей категории — 3 балла. Так, например, за верный ответ на все вопросы задания 1 учащемуся присваивалось 22 балла, так как ему приходилось применять 8 элементов знаний первой категории (8 баллов) и 7 элементов второй категории (14 баллов).

Приведем текст задания 1 и укажем некоторые элементы знаний и взвешенную оценку, которая выставляется учащемуся за их верное применение.

Задание 1. С какими из перечисленных веществ может вступать в реакцию гидроокись кальция: а) двуокись углерода; б) едкое кали; в) азотная кислота; г) окись железа; д) нитрат меди? Напишите полные и сокращенные уравнения возможных реакций.

Знание о возможности реакции гидроокиси кальция с веществами а), в), д) и невозможности его реакции с веществами б) и г) — пять элементов знаний первой категории (5 баллов).

Умение записать полные уравнения реакций с веществами а), в) и д) — три элемента знаний первой категории (3 балла).

Умение записать сокращенные уравнения трех реакций с веществами а), в) и д) — три элемента знаний второй категории (6 баллов).

При ответе на последующие вопросы задания 1 необходимо было применить еще четыре элемента знаний второй категории (8 баллов). Таким образом, за верный ответ на все вопросы задания 1 учащемуся присваивалось всего 22 балла $(5+3+6+8=22)$.

Приведем пример вопроса, при ответе на который учащийся должен был применить два элемента знаний третьей категории трудности: «Почему серная кислота сильнее фосфорной? Дайте ответ на основе сравнения зарядов ионов этих кислот».

Итак, на основе анализа вопросов было установлено, что за верные ответы на все вопросы 1-го задания учащийся будет получать 22 балла, 2-го задания — 6 баллов, 3-го задания — 10 баллов, 4-го задания — 4 балла. Таким образом, максимальная оценка учащегося могла быть равной 42 баллам $(22+6+10+4=42)$.

Контрольную работу по химии выполняли учащиеся IX класса двух районов города. Затем из учащихся первого района, писавших работу, была методом случайного отбора составлена выборка из 34 человек, а из учащихся второго района — выборка из 27 человек. Сравнение знаний учащихся двух районов по результатам выполнения работы учащимися этих выборок позволяло получить объективные выводы и значительно сэкономить время проверки всех контрольных работ.

Ответы учащихся обеих выборок объемом $n_1=34$ и $n_2=27$ были оценены следующими балловыми оценками, которые запишем по возрастанию значений отдельно по первой и по второй выборкам:

Выборка № 1 $n_1=34$	1, 5, 6, 6, 8, 9, 11, 12, 14, 16, 16, 21, 23, 25, 28, 29, 30, 32, 32, 33, 33, 34, 35, 35, 35, 35, 37, 37, 39, 40, 40, 40, 42, 42
Выборка № 2 $n_2=27$	6, 8, 11, 12, 14, 18, 20, 20, 23, 26, 28, 30, 30, 30, 30, 30, 31, 31, 31, 34, 34, 37, 37, 38, 40, 40, 40

На основе сравнения результатов выполнения работы учащимися этих выборок предполагалось проверить гипотезу об отсутствии различий в состоянии проверявшихся знаний по химии у учащихся двух районов города. В условиях проведенного эксперимента возможно применение критерия Вилкоксона — Манна — Уитни и медианного для проверки данной гипотезы. Безусловно, предпочтение следует отдать первому из критериев, так как он позволяет более точно исследовать изучаемое явление. Его использование позволяет выявить в данном случае различие медиан распределений балловых оценок учащихся двух районов и, если такое различие имеет место, установить, учащиеся какого района получают в среднем большие оценки.

Пусть случайная переменная X — число баллов, присвоенных учащемуся первой выборки, а случайная переменная Y — число баллов, присвоенных учащемуся второй выборки. Объем первой выборки равен 34; значит, будем иметь 34 значения x_i $(i=1, 2, ..., 34)$; объем второй выборки — 27; значит, будем иметь 27 значений y_j $(j=1, 2, ..., 27)$. Объединим все значения x и y в одну группу объемом $N=61$ $(34+27=61)$, запишем в ряд по возрастанию значений и ранжируем — припишем каждому значению x и y ранг R, численно равный месту, на котором оно (значение) стоит в этом ряду. Согласно правилу использования критерия, одинаковым значениям X и Y присваивается ранг R, равный среднему арифметическому номеров мест, которые эти значения занимают среди членов двух выборок, упорядоченных по возрастанию значений. Так, например, имеются два значения переменной X и одно значение переменной Y, равное 6. Эти значения стоят на 3-м, 4-м и 5-м местах. Следовательно, ранг каждого из них равен 4, так как среднее арифметическое номеров занимаемых ими мест равно $4\left(\frac{3+4+5}{3}=4\right)$.

Значения переменных X и Y и соответствующие им ранги запишем в форме табл. 12, удобной для подсчета значения статистики критерия Вилкоксона — Манна — Уитни.

Таблица 12

№ п/п	x_i	y_j	R	№ п/п	x_i	y_j	R
1	1		1	32		30	30,5
2	5		2	33		30	30,5
3	6		4	34		31	34
4	6		4	35		31	35
5		6	4	36		31	36
6	8		6,5	37	32		37
7		8	6,5	38	33		38
8	9		8	39	33		39
9	11		9,5	40	33		40
10		11	9,5	41	34		42
11	12		11,5	42		34	42
12		12	11,5	43		34	42
13	14		13,5	44	35		44
14		14	13,5	45	35		45
15	16		15	46	35		46
16	16		16	47	35		47
17		18	17	48	37		49,5
18		20	18	49	37		49,5
19		20	19	50		37	49,5
20	21		20	51		37	49,5
21		23	21,5	52		38	52
22	23		21,5	53	39		53
23	25		23	54	40		56,5
24		26	24	55	40		56,5
25	28		25,5	56	40		56,5
26		28	25,5	57		40	56,5
27	29		27	58		40	56,5
28	30		30,5	59		40	56,5
29		30	30,5	60	42		60
30		30	30,5	61	42		61
31		30	30,5				

В связи с тем что никаких предварительных данных о состоянии знаний, проверяемых контрольной работой,

в двух изучаемых районах города не имеется, применяется двусторонний критерий Вилкоксона — Манна — Уитни, т. е. проводится проверка гипотезы H_0: $P(X<Y)=\frac{1}{2}$ — при альтернативной гипотезе H_1: $P(X<Y)\neq\frac{1}{2}$. Гипотеза H_0 предполагает, что оценки учащихся первого района за выполнение контрольной работы (переменная X) с одинаковой вероятностью (равной 0,5) статистически больше или меньше оценок учащихся второго района, т. е. оценки учащихся первого района в среднем (статистически) не больше и не меньше оценок учащихся второго района.

На основе данных табл. 12 подсчитываем значение статистики критерия T по формуле (5.2.2) на с. 86. В соответствии с этой формулой сначала находим сумму рангов, приписанных членам выборки меньшего объема, состоящей из учащихся второго района[1]. В данном примере сумма рангов

$$S=\sum_{j=1}^{27} R\,(y_j)=832\,.$$

Минимальное из значений $n_1=34$ и $n_2=27$, обозначаемое n, равно 27. Тогда согласно формуле (5.2.2)

$$T=S-\frac{n\,(n+1)}{2}=832-\frac{27\,(27+1)}{2}=454.$$

Объемы выборок n_1 и n_2 больше 20, поэтому критическое значение статистики критерия T находим по формуле (5.2.4) на с. 87[2]. Для выбранного уровня значимости $\alpha=0{,}05$ $x_{\frac{\alpha}{2}}=1{,}96$ (см. с. 87). При $n_1=34$ и $n_2=27$ по формуле (5.2.4)

$$W_{\frac{\alpha}{2}}=\frac{34\cdot 27}{2}+1{,}96\sqrt{\frac{34\cdot 27\cdot\ (34+27+1)}{12}}=594.$$

[1] Слагаемые этой суммы подчеркнуты в табл. 12.

[2] В данном примере нет длинных цепочек одинаковых значений, принадлежащих обеим выборкам. Поэтому подсчет критического значения статистики критерия производится по формуле (5.2.4) без коррекции на приписывание одинаковых рангов, а не по формуле (5.2.6), включающей такую коррекцию.

Таким образом, оказывается верным неравенство $T_{\text{наблюд}} < W_{\frac{\alpha}{2}}$ (454<594). Согласно правилу принятия решений, при использовании двустороннего критерия (см. с. 87) нулевая гипотеза H_0 отклоняется на уровне $\alpha=0{,}05$ и принимается альтернативная гипотеза H_1 (с. 93). Принятие этой гипотезы означает, что анализ экспериментальных данных позволяет сделать вывод о различии законов распределения переменных X и Y или о различии в состоянии знаний, проверяемых работой, у учащихся первого и второго районов города.

Используем данные этого примера для иллюстрации подсчета критического значения статистики критерия по формуле (5.2.6), включающей коррекцию на приписывание одинаковых рангов совпадающим значениям переменных X и Y, принадлежащих обеим выборкам. Для этого необходимо найти значение величины ΣK, используя формулу (5.2.7) на с. 89.

В данном примере (см. табл. 12 на с. 92) имеется 11 групп совпадающих значений переменных, принадлежащих обеим выборкам. Таким образом, величина ΣK будет состоять из 11 слагаемых, значение каждого из которых (K) определяется по формуле (5.2.7) на основе числа членов (k), составляющих группу совпадающих значений.

На основе данных табл. 12 на с. 92 найдем значения k_i ($i=1, 2, ..., 11$):

1) $k_1=3$ (значение, равное 6, имеют три члена обеих выборок, поэтому $k_1=3$);
2) $k_2=2$ (значение, равное 8, имеют два члена обеих выборок, поэтому $k_2=2$);
3) $k_3=2$ (значение, равное 11, имеют два члена);
4) $k_4=2$ (значение, равное 12, имеют два члена);
5) $k_5=2$ (значение, равное 14, имеют два члена);
6) $k_6=2$ (значение, равное 23, имеют два члена);
7) $k_7=2$ (значение, равное 28, имеют два члена);
8) $k_8=6$ (значение, равное 30, имеют шесть членов);
9) $k_9=3$ (значение, равное 34, имеют три члена);
10) $k_{10}=4$ (значение, равное 37, имеют четыре члена);
11) $k_{11}=6$ (значение, равное 40, имеют шесть членов).

В соответствии с формулой (5.2.7) найдем значения 11 слагаемых суммы K:

$$K_1=\frac{k^3-k}{12}=\frac{3^3-3}{12}=2;\quad K_2=K_3=K_4=K_5=K_6=$$

$$=K_7=\frac{2^3-2}{12}=0{,}5;$$

$$K_8=\frac{6^3-6}{12}=17{,}5;\quad K_9=\frac{3^3-3}{12}=2;\quad K_{10}=\frac{4^3-4}{12}=5;$$

$$K_{11}=\frac{6^3-6}{12}=17{,}5.$$

Отсюда в данном примере $\Sigma K=\sum_{i=1}^{11}\frac{k_i^3-k_i}{12}=2+0{,}5+$ $+0{,}5+0{,}5+0{,}5+0{,}5+0{,}5+17{,}5+2+5+17{,}5=47$. По формуле (5.2.6), включающей коррекцию на приписывание одинаковых рангов совпадающим значениям переменных X и Y, для уровня значимости $\alpha=0{,}05$ при $n_1=34$ и $n_2=27$ найдем критическое значение статистики критерия T:

$$W_{\frac{\alpha}{2}}=\frac{n_1 n_2}{2}+x_{\frac{\alpha}{2}}\sqrt{\frac{n_1 n_2 (n_1+n_2+1)}{12}-\Sigma K}=$$

$$=\frac{34\cdot 27}{2}+1{,}96\sqrt{\frac{34\cdot 27\,(34+27+1)}{12}-47}=593{,}3.$$

Итак, значение $W_{\frac{\alpha}{2}}$, подсчитанное по формуле (5.2.6) с коррекцией, равно 593,3; подсчитанное по формуле (5.2.4) без коррекции—594 (см. с. 93). Таким образом, как этого и следовало ожидать в данном примере, внесение коррекции, учитывающей приписывание одинаковых рангов равным значениям переменных X и Y не оказало существенного влияния на значение $W_{\frac{\alpha}{2}}$. Этот факт объясняется тем, что в большинстве случаев одинаковые значения переменных X и Y не составляли длинных цепочек (только в двух случаях число совпадающих значений было сравнительно велико, $k_8=6$ и $k_{11}=6$).

5.3. Критерий χ^2 (хи-квадрат)

Критерий χ^2 (хи-квадрат) применяется для сравнения распределений объектов двух совокупностей по состоянию некоторого свойства на основе измерений по шкале наименований этого свойства в двух независимых выборках из рассматриваемых совокупностей.

Данные. Пусть имеются две выборки из двух совокупностей объемом n_1 и n_2 соответственно. Предположим, что состояние изучаемого свойства (например, выполнение определенного задания) измеряется у каждого объекта по шкале наименований, имеющей только две взаимоисключающие категории (например: выполнено верно — выполнено неверно). По результатам измерения состояния изучаемого свойства[1] у объектов двух выборок составляется четырехклеточная таблица 2×2.

	Категория 1	Категория 2	
Выборка № 1	O_{11}	O_{12}	$O_{11}+O_{12}=n_1$
Выборка № 2	O_{21}	O_{22}	$O_{21}+O_{22}=n_2$
	$O_{11}+O_{21}$	$O_{12}+O_{22}$	$n_1+n_2=N$

В этой таблице O_{11} — число объектов первой выборки, попавших в первую категорию по состоянию изучаемого свойства; O_{12} — число объектов первой выборки, попавших во вторую категорию; O_{21} — число объектов второй выборки, попавших в первую категорию; O_{22} — число объектов второй выборки, попавших во вторую категорию; N — общее число наблюдений, равное $O_{11}+O_{12}+O_{21}+O_{22}$ или n_1+n_2.

Допущения. Для применения критерия необходимо выполнение следующих требований:

1) обе выборки случайные;

2) выборки независимы, и члены каждой выборки также независимы между собой;

[1] Состояние изучаемого свойства в первой совокупности характеризуется случайной переменной X, во второй совокупности — переменной Y, причем каждая из переменных принимает одно из двух возможных значений.

3) шкала измерений может быть самой простой шкалой наименований с двумя категориями.

Гипотезы. Вероятность того, что случайно выбранный из первой совокупности объект будет принадлежать первой категории шкалы измерения проверяемого свойства, обозначим p_1. Вероятность того же события во второй совокупности обозначим p_2. Тогда на основе данных таблицы 2×2 (см. с. 96) можно проверить нулевую гипотезу о равенстве вероятностей попадания объектов первой и второй совокупностей в первую (вторую) категорию шкалы измерения проверяемого свойства, например гипотезу о равенстве вероятностей верного выполнения некоторого задания учащимися контрольных и экспериментальных классов.

Для двустороннего критерия нулевая гипотеза будет иметь вид H_0: $p_1=p_2$, а альтернативная гипотеза будет иметь вид H_1: $p_1\neq p_2$.

Для одностороннего критерия нулевая гипотеза принимает вид H_0: $p_1\leqslant p_2$, а альтернативная H_1: $p_1>p_2$.

При проверке нулевых гипотез не обязательно, чтобы значения вероятностей p_1 и p_2 были известны, так как гипотезы только устанавливают между ними некоторые соотношения (равенство, больше или меньше).

При использовании одностороннего критерия любая совокупность может быть обозначена первой, поэтому гипотеза вида H_0: $p_1\geqslant p_2$ специально не рассматривается, так как может быть заменена гипотезой H_0: $p_1\leqslant p_2$.

Статистика критерия. Для проверки рассмотренных выше нулевых гипотез по данным таблицы 2×2 (см. с. 96) подсчитывается значение статистики критерия T* по следующей общей формуле:

$$T=\Sigma\,\frac{(O-E)^2}{E}, \qquad (5.3.1)$$

где O — наблюдаемые частоты, а E — ожидаемые частоты попадания объектов выборки в первую или вторую категорию шкалы наименований, по которой измеряется изучаемое свойство, знак Σ означает сумму четырех слагаемых, соответствующих частотам, стоящим в четырех клетках таблицы 2×2. Наблюдаемые часто-

* В литературе по математической статистике для обозначения величины T употребляется также символ χ^2 (хи-квадрат).

ты обозначены O_{11}, O_{12}, O_{21}, O_{22} и расположены в клетках таблицы 2×2 (см. с. 96). Для каждой из этих наблюдаемых частот O_{ij} определяется ожидаемая частота по следующей формуле:

$$E_{ij} = \frac{(O_{1j} + O_{2j})\cdot n_i}{N}, \quad i=1,\ 2; \quad j=1,\ 2.$$

Так, например, для частоты O_{11} значение ожидаемой частоты E_{11} подсчитывается по формуле

$$E_{11} = \frac{(O_{11} + O_{21})\cdot n_1}{N}.$$

Для упрощения подсчетов значений статистики формулу (5.3.1) легко преобразовать к виду

$$T = \frac{N\cdot(O_{11}\cdot O_{22} - O_{12}\cdot O_{21})^2}{n_1\cdot n_2\ (O_{11}+O_{21})\ (O_{12}+O_{22})}, \qquad (5.3.2)$$

где n_1, n_2 — объемы выборок, $N = n_1 + n_2$ — общее число наблюдений.

Правило принятия решения. Из-за большого числа комбинаций возможных значений величин O_{11}, O_{12}, O_{21}, O_{22} трудно составить таблицу значений точного распределения статистики T в предположении справедливости нулевой гипотезы. Однако распределение статистики T с достаточной точностью можно аппроксимировать распределением χ^2 с одной степенью свободы ($\nu = 1$).

1) Двусторонний критерий. Проводится проверка гипотезы H_0: $p_1 = p_2$ — при альтернативе H_1: $p_1 \neq p_2$. Пусть α — принятый уровень значимости. Тогда значение статистики T, полученное на основе наблюдений, сравнивается с критическим значением статистики $T - x_{1-\alpha}$, которое определяется по «таблице χ^2 с одной степенью свободы»[1] с учетом выбранного значения α. Если наблюдаемое значение T больше критического, то нулевая гипотеза отклоняется на уровне α и принимается альтернативная гипотеза. В противном случае у нас нет достаточных оснований для отклонения нулевой гипотезы.

2) Односторонний критерий. При наличии достаточных оснований для предположения о том, что вероятность попадания объектов первой (второй) совокупности в первую (вторую) категорию шкалы измерения проверяемого свойства больше (меньше), чем в другой совокупности, вместо двустороннего критерия проверки гипотез применяется односторонний критерий.

[1] Значения $x_{1-\alpha}$ для некоторых значений α имеются в табл. Г.

Проводится проверка гипотезы H_0: $p_1 \leqslant p_2$ — при альтернативе H_1: $p_1 > p_2$. Пусть α — принятый уровень значимости. Тогда значение статистики T, полученное на основе экспериментальных данных, сравнивается с критическим значением статистики $x_{1-2\alpha}$, которое определяется по «таблице χ^2 с одной степенью свободы» с учетом выбранного значения α. Если верно неравенство $T > x_{1-2\alpha}$, то нулевая гипотеза отклоняется на уровне α и принимается альтернативная гипотеза. Если данное неравенство не выполняется, то у нас нет достаточных оснований для отклонения нулевой гипотезы.

В связи с тем что замена точного распределения статистики T распределением χ^2 с одной степенью свободы дает достаточно хорошее приближение только для больших выборок, применение критерия ограничено некоторыми условиями.

Критерий не рекомендуется использовать, если 1) сумма объемов двух выборок ($N = n_1 + n_2$) меньше 20; 2) хотя бы одна из абсолютных частот (O_{11}, O_{12}, O_{21}, O_{22}) в таблице 2×2, составленной на основе экспериментальных данных, меньше 5.

Некоторые исследователи второе условие заменяют следующим: хотя бы одна из ожидаемых частот

$$\frac{(O_{11}+O_{21})\ n_1}{N},\quad \frac{(O_{11}+O_{21})\ n_2}{N},\quad \frac{(O_{12}+O_{22})\ n_1}{N},\quad \frac{(O_{12}+O_{22})\ n_2}{N}$$

меньше 5. К сожалению, оба эти правила сформулированы на основе экспериментальных исследований и не имеют строгих доказательств.

Если хотя бы одна из абсолютных частот имеет значение, заключенное в пределах от 5 до 10, то применение критерия возможно при внесении некоторых изменений в формулу подсчета значения статистики критерия. Формула будет иметь вид

$$T = \frac{N\left(|O_{11}\cdot O_{22} - O_{12}\cdot O_{21}| - \dfrac{N}{2}\right)^2}{n_1\cdot n_2\cdot(O_{11}+O_{21})\cdot(O_{12}+O_{22})}. \qquad (5.3.3)$$

Применение критерия хи-квадрат возможно и в том случае, когда объекты двух выборок из двух совокупностей по состоянию изучаемого свойства распределяются более чем на две категории. Например, учащиеся экспериментальных и контрольных классов распределяются на четыре категории в соответствии с отметками (в баллах: 2, 3, 4, 5), полученными учащимися за выполнение некоторой контрольной работы.

Данные. Из двух совокупностей сделаны выборки объемом n_1 и n_2. Результаты измерения состояния изучаемого свойства у объектов каждой выборки распределяются на C категорий. На основе этих данных составляется таблица $2\times C$, в которой два ряда (по числу рассматриваемых совокупностей) и C колонок (по числу различных категорий состояния изучаемого свойства, принятых в исследовании).

	Категория 1	Категория 2	...	Категория i	...	Категория C	
Выборка № 1	O_{11}	O_{12}	...	O_{1i}	...	O_{1C}	n_1
Выборка № 2	O_{21}	O_{22}	...	O_{2i}	...	O_{2C}	n_2
	$O_{11}+O_{21}$	$O_{12}+O_{22}$		$O_{1i}+O_{2i}$		$O_{1C}+O_{2C}$	$N=n_1+n_2$

В этой таблице O_{1i} $(i=1, 2, ..., C)$ — число объектов первой выборки, попавших в i-ю категорию по состоянию изучаемого свойства; O_{2i} $(i=1, 2, ..., C)$ — число объектов второй выборки, попавших в i-ю категорию по состоянию изучаемого свойства.

Допущения. Для применения критерия необходимо выполнение следующих требований:

1) обе выборки случайные;

2) выборки независимы, и члены каждой выборки независимы между собой;

3) шкала измерений может быть не выше шкалы наименований с несколькими (C) категориями.

Гипотезы. Вероятность того, что случайно выбранный из первой совокупности объект будет принадлежать i-й категории шкалы измерения проверяемого свойства $(i=1, 2, ..., C)$, обозначим p_{1i} $(i=1, 2, ..., C)$. Вероятность того же события для объектов второй совокупности обозначим p_{2i} $(i=1, 2, ..., C)$. Тогда на основе данных таблицы $2\times C$ (с. 100) можно проверить нулевую гипотезу о равенстве вероятностей попадания объектов первой и второй совокупностей в каждую из i $(i=1, 2, ..., C)$ категорий, т. е. проверить выполнение всех следующих равенств: $p_{11}=p_{21}$, $p_{12}=p_{22}$, ..., $p_{1i}=p_{2i}$, ..., $p_{1C}=p_{2C}$. Возможна, например, проверка гипотезы о равенстве вероятностей получения отметок «5», «4», «3» и «2» за выполнение учащимися контрольных и экспериментальных классов некоторого задания.

Таким образом, нулевая гипотеза будет иметь вид H_0: $p_{1i}=p_{2i}$ для всех C категорий, а альтернативная гипотеза будет иметь вид H_1: $p_{1i}\neq p_{2i}$ хотя бы для одной из C категорий.

Для проверки данной гипотезы можно и не знать значения вероятностей p_{1i} и p_{2i} $(i=1, 2, ..., C)$, так как гипотеза только устанавливает между ними соотношение равенства или различия.

Статистика критерия. Для проверки рассмотренной выше нулевой гипотезы с помощью критерия χ^2 на основе данных таблицы $2\times C$ подсчитывается значение статистики критерия T по следующей формуле:

$$T = \frac{1}{n_1\cdot n_2}\sum_{i=1}^{C}\frac{(n_1 O_{2i} - n_2 O_{1i})^2}{O_{1i}+O_{2i}}, \qquad (5.3.4)$$

где n_1 и n_2 — объемы выборок. Формула (5.3.4) получена из общей формулы (5.3.1) и позволяет упростить вычисление значений статистики критерия. Из-за большого числа комбинаций возможных значений O_{1i} и O_{2i} $(i=1, 2, ..., C)$ трудно составить таблицу значений точного распределения статистики T. Однако распределение статистики T можно аппроксимировать распределением χ^2 с $(C-1)$ степенью свободы $(\nu=C-1)$.

Правило принятия решения. Пусть α — принятый уровень значимости. Тогда значение T, полученное на основе экспериментальных данных, сравнивается с критическим значением $x_{1-\alpha}$, которое определяется по «таблице χ^2 с $C-1$ степенью свободы»[1] с учетом выбранного значения α. При выполнении неравенства $T>x_{1-\alpha}$ ну-

[1] Значения $x_{1-\alpha}$ можно найти для некоторых значений α в $(C-1)$-й строке табл. $Г$.

левая гипотеза отклоняется на уровне α и принимается альтернативная гипотеза. Это означает, что распределение объектов на C категорий по состоянию изучаемого свойства различно в двух рассматриваемых совокупностях. Если выполняется неравенство $T \leqslant x_{1-\alpha}$, то у нас нет достаточных оснований для отклонения нулевой гипотезы, т. е. нет достаточных оснований считать состояние изучаемого свойства различным в обеих совокупностях.

Аппроксимация распределения статистики T распределением χ^2 становится недостаточно точной, если хотя бы некоторые значения O_{1i} или O_{2i} в клетках таблицы $2 \times C$, составленной на основе экспериментальных данных, меньше 5. Поэтому для применения критерия рекомендуется объединить несколько категорий состояния изучаемого свойства в одну так, чтобы все значения в таблице $2 \times C$ стали не меньше 5.

При использовании критерия χ^2 необходимо иметь в виду следующее. Если изучаемое свойство измерено по шкале наименований, имеющей более двух категорий, или по шкалам более высокого порядка, то для сравнения распределений этого свойства в различных совокупностях существует только двусторонний критерий хи-квадрат.

Примеры. Проводился эксперимент, направленный на выявление лучшего из учебников, написанных двумя авторскими коллективами в соответствии с целями обучения геометрии и содержанием программы IX класса. Для проведения эксперимента методом случайного отбора были выбраны два района РСФСР, большинство школ которых относились по расположению к сельским. Учащиеся первого района (20 классов) обучались по учебнику № 1, учащиеся второго района (15 классов) обучались по учебнику № 2.

Оценка результатов эксперимента проводилась по нескольким показателям: уровень выполнения учащимися обоих районов текущих, четвертных, итоговых контрольных работ; ответы учащихся и учителей на устные контрольные вопросы, на анкеты; беседы с учителями и учащимися и др.

1. Рассмотрим методику сравнения ответов учителей экспериментальных школ двух районов РСФСР на один из вопросов анкеты: «Доступен ли учебник в целом для

самостоятельного чтения и помогает ли он усвоить учащимся материал, который учитель не объяснял в классе?» (Ответ: да — нет.)

Включение данного вопроса в анкету объясняется тем, что с целью выявления доступности каждого учебника ряд одинаковых разделов курса учащиеся обоих районов изучали самостоятельно по своим учебникам. На основе сравнения мнений учителей экспериментальных классов обоих районов предполагалось проверить гипотезу об отсутствии различия доступности двух учебников для самостоятельного чтения.

Отношение учителей к изучаемому свойству учебников измерено по шкале наименований, имеющей две категории: да, нет. Обе выборки учителей случайные и независимые. Таким образом, в условиях данного эксперимента выполняются все условия применения критерия χ^2 (хи-квадрат) для проверки сформулированной гипотезы[1]. В связи с тем что шкала измерений имеет только две категории, используется вариант двустороннего критерия, приспособленный для тех случаев, когда результаты измерений сведены в четырехклеточную таблицу 2×2.

Ответы 20 учителей первого района и 15 учителей второго района распределим на две категории (да, нет) и запишем в форме таблицы 2×2 (табл. 13).

Таблица 13

	Да	Нет	
Выборка № 1	$O_{11}=15$	$O_{12}=5$	$n_1=O_{11}+O_{12}=20$
Выборка № 2	$O_{21}=7$	$O_{22}=8$	$n_2=O_{21}+O_{22}=15$
	$O_{11}+O_{21}==22$	$O_{12}+O_{22}==13$	$N=n_1+n_2=35$

Обозначим p_1 вероятность того, что учитель школы первого района будет считать учебник № 1 доступным, а p_2 — вероятность того, что учитель второго района будет считать учебник № 2 доступным. На основе данных табл. 13 можно проверить нулевую гипотезу H_0: $p_1=p_2$—

[1] Малый объем обеих выборок не позволяет применить для проверки данной гипотезы параметрический u-критерий, рассмотренный в работе [8].

при альтернативной гипотезе H_1: $p_1 \neq p_2$. Все значения абсолютных частот в табл. 13 не меньше 5, но три из них меньше 10. Поэтому в соответствии с условиями использования критерия (см. с. 99) подсчет статистики критерия χ^2 производится по корректированной формуле (5.3.3) на с. 99.

На основе данных табл. 13.

$$T = \frac{N\left(|O_{11}\cdot O_{22} - O_{12}\cdot O_{21}| - \frac{N}{2}\right)^2}{n_1\cdot n_2\cdot(O_{11}+O_{21})\cdot(O_{12}+O_{22})} =$$

$$= \frac{35\left(|15\cdot 8 - 7\cdot 5| - \frac{35}{2}\right)^2}{20\cdot 15\cdot(15+7)\cdot(5+8)} = 1{,}86.$$

В соответствии с условиями применения двустороннего критерия хи-квадрат по таблице Г для одной степени свободы ($\nu=1$) и уровня значимости $\alpha=0{,}05$ найдем $x_{1-\alpha} = T_{\text{критич}} = 3{,}84$. Отсюда верно неравенство $T_{\text{наблюд}} < T_{\text{критич}}$ ($1{,}86 < 3{,}84$). Согласно правилу принятия решений для критерия χ^2, полученный результат не дает достаточных оснований для отклонения нулевой гипотезы, т. е. результаты проведенного опроса учителей двух экспериментальных районов не дают достаточных оснований для отклонения предположения об одинаковой доступности учебников № 1 и 2 для самостоятельного чтения учащимися.

2. Рассмотрим методику сравнения результатов письменной работы, проверявшей усвоение одного из разделов курса учащимися первого и второго районов.

Методом случайного отбора из учащихся первого района, писавших работу, была составлена выборка объемом 50 человек ($n_1=50$), из учащихся второго района — выборка объемом 50 человек ($n_2=50$). В соответствии со специально разработанными критериями оценки выполнения работы каждый ученик мог попасть в одну из четырех категорий: плохо, посредственно, хорошо, отлично. Результаты выполнения работы двумя выборками учащихся используем для проверки гипотезы о том, что учебник № 1 способствует лучшему усвоению проверяемого раздела курса, т. е. учащиеся первого экспериментального района в среднем будут получать более высокие оценки, чем учащиеся второго района.

Выборки учащихся случайные и независимые, измеряемое свойство (усвоение раздела программы) имеет непрерывное распределение и измерено по шкале порядка, имеющей четыре категории: плохо, посредственно, хорошо, отлично. Таким образом в данном случае выполнены все допущения критерия Вилкоксона — Манна — Уитни, позволяющего проверить сформулированную гипотезу. Однако в связи с небольшим числом категорий шкалы измерения (четыре категории) значительная часть экспериментальных данных представляет цепочки одинаковых значений, что снижает точность выводов, полученных на основе применения этого критерия. Поэтому остается воспользоваться двусторонним критерием χ^2 (хи-квадрат), приспособленным для тех ситуаций, когда экспериментальные данные записаны в форме таблицы $2\times C$ (в нашем примере 2×4, так как $C=4$).

Результаты выполнения работы учащимися обеих выборок запишем в виде таблицы 2×4 (табл. 14).

Таблица 14

	Категория 1 (плохо)	Категория 2 (посред.)	Категория 3 (хор.)	Категория 4 (отл.)
Выборка № 1 $n_1=50$	$O_{11}=3$	$O_{12}=19$	$O_{13}=18$	$O_{14}=10$
Выборка № 2 $n_2=50$	$O_{21}=9$	$O_{22}=24$	$O_{23}=12$	$O_{24}=5$

В таблице 14 O_{1i} обозначает число учащихся первой выборки, получивших оценку i ($i=1, 2, 3, 4$); O_{2i}—число учащихся второй выборки, получивших оценку i ($i=1, 2, 3, 4$).

Обозначим p_{1i} ($i=1, 2, 3, 4$) вероятность выполнения работы учащимися первого района на оценку i; p_{2i} ($i=1, 2, 3, 4$) — вероятность выполнения работы учащимися второго района на оценку i. На основе данных табл. 14 можно проверить нулевую гипотезу H_0: $p_{1i}=p_{2i}$ для всех $C=4$ категорий (т. е. $p_{11}=p_{21}$, $p_{12}=p_{22}$, $p_{13}=p_{23}$, $p_{14}=p_{24}$) — при альтернативе $H_1: p_{1i}\neq p_{2i}$ хотя бы для одной из $C=4$ категорий.

5 Зак. 224

Для проверки данной гипотезы подсчет значения статистики критерия χ^2 будем производить по формуле (5.3.4) на с. 101, учитывая, что число категорий $C=4$.

$$T = \frac{1}{n_1 \cdot n_2} \cdot \sum_{i=1}^{C} \frac{(n_1 O_{2i} - n_2 O_{1i})^2}{O_{1i} + O_{2i}} =$$

$$= \frac{1}{n_1 \cdot n_2} \sum_{i=1}^{4} \frac{(n_1 O_{2i} - n_2 O_{1i})^2}{O_{1i} + O_{2i}} = \frac{1}{n_1 \cdot n_2} \left[\frac{(n_1 O_{21} - n_2 O_{11})^2}{O_{11} + O_{21}} + \right.$$

$$\left. + \frac{(n_1 O_{22} - n_2 O_{12})^2}{O_{12} + O_{22}} + \frac{(n_1 O_{23} - n_2 O_{13})^2}{O_{13} + O_{23}} + \frac{(n_1 O_{24} - n_2 O_{14})^2}{O_{14} + O_{24}} \right] =$$

$$= \frac{1}{50 \cdot 50} \cdot \left[\frac{(50 \cdot 9 - 50 \cdot 3)^2}{3+9} + \frac{(50 \cdot 24 - 50 \cdot 19)^2}{19+24} + \right.$$

$$\left. + \frac{(50 \cdot 12 - 50 \cdot 18)^2}{18+12} + \frac{(50 \cdot 5 - 50 \cdot 10)^2}{10+5} \right] = 6{,}45.$$

По табл. Г для $\alpha=0{,}05$ и числа степеней свободы $\nu=C-1=4-1=3$ находим критическое значение статистики критерия T: $x_{1-\alpha}=7{,}815$. Отсюда верно неравенство $T_{\text{наблюд}}<T_{\text{критич}}$ $(6{,}45<7{,}815)$, т. е. в соответствии с правилом принятия решения (см. с. 102) полученные результаты не дают достаточных оснований для отклонения нулевой гипотезы (с. 105). Иначе говоря, полученные результаты выполнения контрольных работ учащимися двух экспериментальных районов не дают достаточных оснований для отклонения предположения о том, что учащиеся, обучавшиеся по разным учебникам, одинаково усвоили проверяемый раздел программы. Однако качественный анализ экспериментальных данных в табл. 14 дает некоторые основания считать полученный вывод следствием малого объема выборок. В связи с этим можно порекомендовать в данном случае проверить нулевую гипотезу еще раз, увеличив объем выборок.

5.4. Критерий Колмогорова—Смирнова

Критерий предназначен для выявления различия двух совокупностей по состоянию некоторого свойства. Критерий чувствителен в улавливании любого различия функций распределения этого свойства в рассматриваемых совокупностях (средних значений, дисперсий, эксцессов и др.).

Данные. Будем считать, что случайная переменная X характеризует состояние изучаемого свойства в одной из рассматриваемых совокупностей, а случайная переменная Y характеризует состояние этого же свойства в другой совокупности.

Пусть имеется n_1 наблюдений над случайной переменной X и n_2 наблюдений над случайной переменной Y, полученных при рассмотрении объектов двух выборок объемов n_1 и n_2; x_i $(i=1,2,\ldots, n_1)$ — результаты измерения некоторого свойства у объектов первой выборки; $y_k (k=1, 2, \ldots, n_2)$ — результаты измерения того же свойства у объектов второй выборки; $F(x)$ и $G(x)$ — неизвестные нам функции распределения изучаемого свойства в первой и второй совокупностях соответственно.

Результаты измерения объектов первой выборки запишем в ряд по возрастанию значений: $x_1, x_2, \ldots, x_{n_1}$. Затем для каждого значения переменной X найдем значение эмпирической функции распределения изучаемого свойства:

$$S_1(x) = \frac{a}{n_1},$$

где a — число значений x_i, меньших или равных x $(x_i \leqslant x)$; n_1 — объем выборки.

Аналогично составим эмпирическую функцию распределения изучаемого свойства по результатам второй выборки:

$$S_2(x) = \frac{b}{n_2},$$

где b — число значений y_k, меньших или равных x $(y_k \leqslant x)$; n_2 — объем второй выборки.

Для нахождения значений функций $S_1(x)$ и $S_2(x)$ удобно записать значения обеих выборок в форме таблицы. Покажем способ составления такой таблицы на конкретном примере.

Пример. Из двух совокупностей сделаны две выборки объемом 6 и 8:

$$x_i = 17,\ 10,\ 7,\ 9,\ 5,\ 4;$$

$$y_k = 2,\ 6,\ 7,\ 5,\ 8,\ 15,\ 3,\ 11.$$

Для нахождения значений эмпирических функций распределения изучаемого свойства в каждой из выборок составим табл. 15.

5* Зак. 224

Таблица 15

x_i	y_k	$S_1(x)$	$S_2(x)$	$S_1(x)-S_2(x)$
—	2	0	1/8	−1/8
—	3	0	2/8	−2/8
4	—	1/6	2/8	−1/12
5	5	2/6	3/8	−1/24
—	6	2/6	4/8	1/6
7	7	3/6	5/8	1/8
—	8	3/6	6/8	−1/4
9	—	4/6	6/8	−1/12
10	—	5/6	6/8	+1/12
—	11	5/6	7/8	−1/24
—	15	5/6	8/8	−1/6
17	—	6/6	8/8	0

$n_1=6$ $n_2=8$

Покажем способ определения значений функции $S_1(x)$ на примере значений первой выборки. Запишем значения x_i в ряд по возрастанию значений: $x_1=4$, $x_2=5$, $x_3=7$, $x_4=9$, $x_5=10$, $x_6=17$.

При $x=4$ и $n_1=6$ $S_1(x)=\frac{a}{n_1}=S_1(4)=\frac{1}{6}$ (числитель равен 1, так как в данной выборке только одно значение x_i удовлетворяет неравенству $x_i\leqslant 4$, а именно $x_1=4$).

При $x=5$ $S_1(5)=\frac{2}{6}$ (числитель дроби равен 2, так как только два значения x_i удовлетворяют неравенству $x_i\leqslant 5$ ($x_1=4$ и $x_2=5$). При $x=6$ $S_1(6)=\frac{2}{6}$ (числитель дроби равен 2, так как только два значения $x_i\leqslant 6$, а именно $x_1=4$ и $x_2=5$). При $x=7$ $S_1(7)=\frac{3}{6}$ (числитель дроби равен 3, так как три значения $x_i\leqslant 7$, а именно $x_1=4$, $x_2=5$, $x_3=7$) и т. д.

Такой способ определения значений функций $S_1(x)$ и $S_2(x)$ рекомендуется при небольших объемах обеих выборок.

При больших объемах выборок для подсчета значений данных функций составляется таблица, в которой значения наблюдений обеих выборок записываются в

форме интервального ряда[1]. Для каждого интервала значений наблюдений подсчитывается абсолютная частота f — число наблюдений рассматриваемой выборки, попавших в данный интервал значений. По каждой выборке на основе абсолютных частот подсчитываются накопленные частоты Σf — число наблюдений, имеющих значения, меньшие или равные значениям из данного интервала. Проиллюстрируем данный способ определения значений эмпирических функций распределения изучаемого свойства на конкретном примере.

Пример. Пример на все действия с многозначными числами писали 40 экспериментальных классов и 50 контрольных классов. Для каждого класса подсчитывался процент верных ответов на это задание. Значения этой величины изменялись в данном эксперименте от 41 до 100%.

В связи с большими объемами выборок полученные результаты было целесообразно записать в форме интервальных рядов. При величине интервала, равной 5%, получим 12 интервалов (100%−40%=60%, 60% : 5% = 12 (интервалов)). Подсчитаем число наблюдений, попавших в каждый из этих интервалов, по первой и второй выборкам в отдельности и составим табл. 16, два последних столбца которой содержат значения функций $S_1(x)$ и $S_2(x)$.

Значения эмпирической функции распределения $S_1(x)$, полученные на основе результатов выполнения контрольной работы выборкой учащихся из совокупности экспериментальных классов, используются в качестве оценки значений неизвестной нам теоретической функции $F(x)$ распределения изучаемого свойства (умение выполнять действия с многозначными числами) в данной совокупности. Значения функции $S_2(x)$, полученные на основе результатов выборки из совокупности контрольных классов, служат оценкой значений теоретической функции $G(x)$ распределения изучаемого свойства в данной совокупности.

---

[1] Рекомендуется число интервалов выбирать от 8 до 15. При числе интервалов, меньшем 8, использование критерия Колмогорова—Смирнова приводит к менее достоверным выводам. При числе интервалов, большем 15, значительно возрастает объем вычислений.

Таблица 16

Процент верных ответов на задание, x	Абсолютная частота в выборке № 1, f_1	Абсолютная частота в выборке № 2, f_2	Накопленная частота Σf_1	Накопленная частота Σf_2	$S_1(x) = \frac{\Sigma f_1}{n_1}$	$S_2(x) = \frac{\Sigma f_2}{n_2}$
100—96	15	18	40	50	1,00	1,00
95—91	10	8	25	32	0,63	0,64
90—86	4	2	15	24	0,38	0,48
85—81	3	5	11	22	0,28	0,44
80—76	2	6	8	17	0,20	0,34
75—71	1	2	6	11	0,15	0,22
70—66	1	2	5	9	0,13	0,18
65—61	1	4	4	7	0,10	0,14
60—56	—	1	3	3	0,08	0,06
55—51	1	1	3	2	0,08	0,04
50—46	1	—	2	1	0,05	0,02
45—41	1	1	1	1	0,03	0,02
	$n_1 = 40$	$n_2 = 50$				

Допущения. Для использования критерия необходимо выполнение следующих требований:

1) обе выборки случайные;

2) выборки независимы, и члены каждой выборки также независимы между собой;

3) изучаемое свойство имеет непрерывное распределение в обеих совокупностях, из которых сделаны выборки;

4) шкала измерений не ниже порядковой.

Гипотезы. Предположим, что законы распределения случайных величин X и Y одинаковы в обеих рассматриваемых совокупностях, т. е. функции распределения $F(x)$ и $G(x)$ величин X и Y равны между собой. Двусторонний критерий позволяет провести проверку этой гипотезы H_0: $F(x) = G(x)$ для всех возможных значений x — при альтернативе H_1: $F(x) \neq G(x)$ хотя бы для одного значения x. Если гипотеза H_1 истинна, то это означает, что законы распределения величин X и Y различны и, значит, существенно различны состояния изучаемого свойства в рассматриваемых совокупностях.

Односторонние критерии позволяют провести проверку следующих гипотез:

1) H_0: $F(x) \leqslant G(x)$ для всех возможных значений x — при альтернативе H_1: $F(x) > G(x)$. Если данная нулевая гипотеза истинна, то это означает, что стохастически[1] значения переменной X не меньше значений переменной Y. Если истинна альтернативная гипотеза, то это означает, что значения переменной X стохастически меньше значений переменной Y.

2) H_0: $F(x) \geqslant G(x)$ для всех возможных значений x — при альтернативе H_1: $F(x) < G(x)$. Если данная нулевая гипотеза истинна, то это означает, что значения переменной X стохастически не больше значений переменной Y. Если истинна альтернативная гипотеза, то это означает, что значения переменной X стохастически больше значений переменной Y.

Статистика критерия. Для использования критерия требуется определить значения статистик T_1, T_2, T_3 на основе значений эмпирических функций распределения $S_1(x)$ и $S_2(x)$. Вычисляем выражения вида

$$\left|S_1(x) - S_2(x)\right|, \quad \left(S_1(x) - S_2(x)\right), \quad \left(S_2(x) - S_1(x)\right).$$

Статистики критерия подсчитываются по формулам:

$$T_1 = \max\left|S_1(x) - S_2(x)\right|, \qquad (5.4.1)$$

т. е. значение T_1 равно максимальному значению модуля разности $|S_1(x) - S_2(x)|$,

$$T_2 = \max\left(S_1(x) - S_2(x)\right), \qquad (5.4.2)$$

т. е. значение T_2 равно максимальному значению разности $(S_1(x) - S_2(x))$,

$$T_3 = \max\left(S_2(x) - S_1(x)\right). \qquad (5.4.3)$$

Используем рассмотренный выше пример для иллюстрации способа определения статистик критерия. Из табл. 16 (с. 110) выпишем данные первого и двух последних столбцов и на их основе заполним последующие столбцы новой табл. 17.

[1] См. с. 8—9.

Таблица 17

Процент верных ответов на задание, x	$S_1(x)$	$S_2(x)$	$S_1(x)-S_2(x)$	$\lvert S_1(x)-S_2(x)\rvert$	$S_2(x)-S_1(x)$
100—96	1,00	1,00	0,00	0,01	0,00
95—91	0,63	0,64	—0,01	0,01	0,01
90—86	0,38	0,48	—0,10	0,10	0,10
85—81	0,28	0,44	—0,16	0,16	0,16
80—76	0,20	0,34	—0,14	0,14	0,14
75—71	0,15	0,22	—0,07	0,07	0,07
70—66	0,13	0,18	—0,05	0,05	0,05
65—61	0,10	0,14	—0,04	0,04	0,04
60—56	0,08	0,06	0,02	0,02	—0,02
55—51	0,08	0,04	0,04	0,04	—0,04
50—46	0,05	0,02	0,03	0,03	—0,03
45—41	0,03	0,02	0,01	0,01	—0,01

Согласно данным табл. 17, максимальное значение разности $|S_1(x)-S_2(x)|$ равно 0,16; значит, $T_1=0{,}16$. Максимальное значение разности $S_1(x)-S_2(x)$ равно 0,04; значит, $T_2=0{,}04$. Максимальное значение разности $S_2(x)-S_1(x)$ равно 0,16; значит, $T_3=0{,}16$. Отметим, что T_1 всегда равно наибольшему из значений T_2 и T_3.

Выборки одинакового объема. Подсчет значений статистик критерия значительно упрощается, если объемы обеих выборок одинаковы, т. е. $n_1=n_2=n$. В этих случаях значения статистик определяются по формулам:

$$T_1=\frac{1}{n}\cdot\max|\Sigma f_1-\Sigma f_2|, \tag{5.4.4}$$

$$T_2=\frac{1}{n}\cdot\max(\Sigma f_1-\Sigma f_2), \tag{5.4.5}$$

$$T_3=\frac{1}{n}\cdot\max(\Sigma f_2-\Sigma f_1), \tag{5.4.6}$$

где Σf_1 — сумма накопленных частот в выборке № 1, Σf_2 — сумма накопленных частот в выборке № 2. Следовательно, вместо нахождения значений функций $S_1=\dfrac{\Sigma f_1}{n_1}$ и $S_2=\dfrac{\Sigma f_2}{n_2}$, как это приходится делать при разных объемах выборок, достаточно подсчитать только накопленные частоты в каждой выборке.

Пример. Учащиеся двух школ писали диктант. Объемы обеих выборок одинаковы: $n_1=n_2=n=37$. Работы учащихся проверили и подсчитали для каждого ученика число допущенных ошибок, которое колебалось от 0 до 7. Полученные результаты запишем в форме табл. 18, удобной для определения значений статистик критерия.

Таблица 18

Число ошибок	Абс. частота, f_1	Абс. частота, f_2	Накопл. частота, Σf_1	Накопл. частота, Σf_2	$\Sigma f_1-\Sigma f_2$	$\lvert\Sigma f_1-\Sigma f_2\rvert$	$\Sigma f_2-\Sigma f_1$
7	3	5	37	37	0	0	0
6	4	6	34	32	2	2	—2
5	3	9	30	26	4	2	—4
4	8	9	27	17	10	10	—10
3	5	4	19	8	11	11	—11
2	8	2	14	4	10	10	—10
1	5	2	6	2	4	4	—4
0	1	0	1	0	1	1	—1
	37	37					

Согласно данным табл. 18 максимальное значение выражения $|\Sigma f_1-\Sigma f_2|$ равно 11; значит,

$$T_1=\frac{1}{n}\cdot\max|\Sigma f_1-\Sigma f_2|=\frac{1}{37}\cdot 11=0{,}297.$$

Максимальное значение выражения $(\Sigma f_1-\Sigma f_2)$ равно 11; значит,

$$T_2=\frac{1}{n}\cdot\max(\Sigma f_1-\Sigma f_2)=\frac{1}{37}\cdot 11=0{,}297.$$

Максимальное значение выражения $(\Sigma f_2-\Sigma f_1)$ равно 0; значит,

$$T_3=\frac{1}{n}\cdot\max(\Sigma f_2-\Sigma f_1)=\frac{1}{37}\cdot 0=0.$$

Правило принятия решения. Пусть α — принятый уровень значимости.

а) Проверяемая нулевая гипотеза H_0 отклоняется на уровне α, если наблюдаемое значение статистики критерия превышает критическое значение этой статистики $W_{1-\alpha}$. Если верно неравенство $T_1>W_{1-\alpha}$, то H_0: $F(x)=G(x)$ отклоняется и принимается альтернативная гипотеза H_1: $F(x) \neq G(x)$, что позволяет сделать вывод о различии распределений переменных X и Y, т. е. о различии состояний изучаемого свойства у объектов двух рассмотренных совокупностей.

б) Если верно неравенство $T_2 > W_{1-\alpha}$, то H_0: $F(x) \leqslant \leqslant G(x)$ отклоняется на уровне α и принимается гипотеза H_1: $F(x) > G(x)$, что приводит к выводу о различии распределений переменных X и Y в том смысле, что значения X стохастически меньше значений переменной Y, т. е. объекты совокупности X стохастически меньше объектов совокупности Y по состоянию изучаемого свойства.

в) При выполнении неравенства $T_3 > W_{1-\alpha}$ H_0: $F(x) \geqslant G(x)$ отклоняется на уровне α и принимается гипотеза H_1: $F(x) < G(x)$, в результате чего приходят к выводу о том, что значения переменной X стохастически больше значений переменной Y, т. е. объекты совокупности X стохастически больше объектов совокупности Y по состоянию изучаемого свойства.

Для $\alpha \leqslant 0{,}2$ критическое значение статистики $T_1 = W_{1-\alpha}$ равно квантилю $W_{1-\frac{\alpha}{2}}$ — критическому значению статистик T_2 и T_3. Это утверждение, в частности, справедливо и для уровня значимости $\alpha = 0{,}05$, принятого в данной работе.

Критические значения статистик критерия $(W_{1-\alpha})$ определяются с помощью специальных таблиц, составленных на основе законов распределения этих статистик. В том случае, когда изучаемые случайные переменные X и Y имеют непрерывные распределения (см. допущение 3 на с. 110), законы распределения статистики не зависят от неизвестных нам законов распределения этих переменных[1].

Определение критических значений критерия Колмогорова — Смирнова связано с большим объемом вычислений, поэтому таблицы этих значений составлены пока только для выборок небольшого объема, а именно при $n_1 = n_2 = n$ для $n \leqslant 40$, при $n_1 \neq n_2$ для $n_1 \leqslant 20$ и $n_2 \leqslant 20$ (см. табл. *Е*, *Ж*).

[1] В случае разрывных распределений переменных X и Y законы распределения статистик критерия зависят от законов распределения самих величин X и Y. Таким образом, незнание этих законов (что весьма характерно для величин, изучаемых в педагогических исследованиях) делает невозможным применение критерия.

Для выборок большого объема критические значения статистик определяются по приближенной формуле

$$W_{1-\alpha} \approx \lambda_\alpha \sqrt{\frac{n_1 + n_2}{n_1 \cdot n_2}}, \qquad (5.4.7)$$

где λ_α — квантиль функции Колмогорова $K(\lambda)$, отвечающий выбранному уровню значимости α. Величины λ_α находятся по таблице значений функции $K(\lambda)$ как решения уравнений вида $K(\lambda_\alpha) = 1 - \alpha$.

Приведем некоторые чаще других используемые значения λ_α: $\lambda_{0,2} = 1{,}07$, $\lambda_{0,1} = 1{,}22$; $\lambda_{0,05} = 1{,}36$; $\lambda_{0,02} = 1{,}52$; $\lambda_{0,01} = 1{,}63$.

Обоснование приближенной формулы (5.4.7) опирается на сходимость эмпирических функций распределения переменных X и Y — $S_1(x)$ и $S_2(x)$ — к теоретическим функциям распределения этих переменных — $F(x)$ и $G(x)$ (см. введение, с. 15, теорема Гливенко). Поэтому для корректного применения формулы (5.4.7) измерения наблюдаемых значений X и Y ($x_1, x_2, \ldots, x_{n_1}$; $y_1, y_2, \ldots, y_{n_2}$) должны быть выполнены по интервальной шкале или шкале отношений. Из теоремы Гливенко следует также, что в этом случае критерий является состоятельным для рассмотренных выше альтернатив.

Таким образом, в случае выборок большого объема[1] применение критерия возможно только при замене допущения 4 (см. с. 110) на более жесткое требование — значения случайных величин X и Y должны быть измерены по шкале не ниже интервальной.

Для выборок любого объема критические значения статистик критерия, полученные по приближенной формуле (5.4.7), всегда превосходят точные значения этих статистик. Поэтому использование данной формулы приводит к более осторожным выводам.

Примеры. 1. Проводился эксперимент с целью разработки методики обучения учащихся решению задач с помощью составления гипотез на основе условия задачи.

[1] В соответствии с существующими в настоящее время таблицами критических значений статистики критерия Колмогорова—Смирнова это утверждение справедливо при выборках равного объема для $n > 40$ и при выборках разного объема для $n_1 > 20$ или $n_2 > 20$.

В эксперименте участвовало несколько седьмых классов, 210 учащихся обучались по первому варианту методики, 240 учащихся — по второму варианту методики. В качестве одного из показателей эффективности проверявших-ся вариантов методики использовались результаты выполнения этими учащимися контрольной работы из 7 заданий, направленных на проверку усвоения формируемого приема. Выполнение работы каждым учащимся оценивалось числом верных ответов, которое варьировало от 0 до 7. В соответствии с выводом на с. 25 можно считать, что изучаемое свойство учащихся (усвоение приема решения задач с помощью составления гипотезы) измеряется по интервальной шкале.

В проведенном эксперименте выполняются все условия, необходимые для использования критерия Колмогорова — Смирнова для сравнения эффективности двух методик обучения учащихся на основе выполнения некоторой деятельности (семи контрольных заданий) двумя выборками учащихся достаточно большого объема. Проверке подлежала гипотеза H_0: $F(x)=G(x)$, или предположение об одинаковых функциях распределения числа верных ответов на контрольные задания среди учащихся, обучавшихся по разным вариантам методики. Альтернативная гипотеза H_1: $F(x)\neq G(x)$ предполагает, что функции распределения числа верных ответов различны в двух рассматриваемых совокупностях учащихся.

Из учащихся, писавших контрольную работу и обучавшихся по первому варианту методики, отобрали методом случайного отбора 200 человек; из учащихся, обучавшихся по второму варианту методики, тоже отобрали 200 человек. Составление выборок одинакового объема позволило упростить процесс вычислений, связанных с подсчетом статистики критерия, так как для проверки сформулированной гипотезы можно было использовать двусторонний критерий Колмогорова — Смирнова для выборок одинакового объема (см. с. 112—113).

Результаты выполнения работы учащимися двух выборок запишем в форме табл. 19, удобной для нахождения статистики критерия.

По формуле (5.4.4) найдем значение статистики двустороннего критерия для выборок одинакового объема $n=n_2=n_1=200$. Из таблицы экспериментальных данных (табл. 19) находим, что наибольшее значение выражения $|\Sigma f_1-\Sigma f_2|$ равно 37. Значит, в соответствии с формулой (5.4.4)

Таблица 19

Число верных ответов	Абсолютная частота в выборке № 1, f_1	Абсолютная частота в выборке № 2, f_2	Накопленная частота в выборке № 1, Σf_1	Накопленная частота в выборке № 2, Σf_2	$\lvert\Sigma f_1-\Sigma f_2\rvert$
7	56	42	200	200	0
6	80	60	142	158	16
5	25	24	62	98	36
4	9	25	37	74	37
3	10	20	28	49	21
2	7	16	18	29	11
1	7	8	11	13	2
0	4	5	4	5	1

$n_1=200 \quad n_2=200$

$$T_1=\frac{1}{n}\max|\Sigma f_1-\Sigma f_2|=\frac{1}{200}\cdot 37=0{,}185\,.$$

Критическое значение статистики критерия находим по формуле (5.4.7) (см. с. 115), составленной специально для выборок большого объема ($n>40$). Для $\alpha=0{,}05$ и соответственно $\lambda_\alpha=1{,}36$ (см. с. 115), согласно данной формуле,

$$W_{1-\alpha}=\lambda_\alpha\cdot\sqrt{\frac{n_1+n_2}{n_1\cdot n_2}}=1{,}36\cdot\sqrt{\frac{200+200}{200\cdot200}}=0{,}136\,.$$

Отсюда верно неравенство $T_{1\,\text{наблюд}}>W_{1-\alpha}$ ($0{,}185>$ $>0{,}136$). Поэтому в соответствии с правилом принятия решения (см. с. 113, случай а) нулевая гипотеза отклоняется и принимается альтернативная гипотеза H_1 (см. с. 116), что позволяет сделать вывод о различии распределений числа верных ответов на контрольные задания среди учащихся, обучавшихся по разным вариантам методики. Анализ экспериментальных данных (см. два первых столбца в таблице 19) позволяет нам уточнить полученный вывод, так как дает основания сказать, что учащиеся, обучавшиеся по первому варианту методики, дают стохастически больше верных ответов, т. е. первый вариант методики эффективнее в отношении подготовки учащихся к выполнению деятельности, проверяемой контрольной работой.

2. Проводилось сравнение математической подготовки учащихся двух районов города по состоянию некоторых знаний, умений и навыков. Остановимся на результатах проверки состояния знаний дат исторических событий. Учащимся предлагалось припомнить даты 8 событий. Удалось опросить 17 учащихся трех школ одного района и 20 учащихся четырех школ другого района. Результат устного опроса каждого ученика оценивался числом верных ответов, и оценки варьировали от 0 до 8. Учитывая характер вопросов (воспроизведение пройденного материала без всякого изменения), можно считать, что изучаемое свойство учащихся измеряется по шкале порядка. Таким образом, в условиях проведенного эксперимента выполнены все допущения критерия Колмогорова — Смирнова, пригодного для сравнения результатов двух выборок небольшого объема. В табл. *E* критические значения данного критерия даны для $n_1 \leqslant 16$ и $n_2 \leqslant 20$. Если объемы выборок превышают указанные значения, то критические значения статистик критерия подсчитываются по приближенным формулам, которые можно использовать только в случае измерений не ниже, чем по интервальной шкале. В данном примере $n_1 = 17$ и $n_2 = 20$, а измерения сделаны по шкале порядка, что не позволяет использовать приближенные формулы. Поэтому методом случайного отбора из 17 учащихся первой выборки составляется выборка объема $n_1 = 16$.

Проверка знаний учащихся двух районов, проводившаяся в предыдущие годы, показала, что учащиеся первого района обладают в среднем более высоким уровнем знаний. В связи с этим в данном эксперименте предполагалось проверить нулевую гипотезу H_0: $F(x) \geqslant G(x)$ — при альтернативе H_1: $F(x) < G(x)$. Нулевая гипотеза предполагает, что учащиеся первого района стохастически будут давать не больше верных ответов (столько же или меньше), чем учащиеся второго района. Альтернативная гипотеза предполагает, что учащиеся первого района будут давать стохастически больше верных ответов, чем учащиеся второго района. Проверка сформулированной нулевой гипотезы возможна с помощью одностороннего критерия Колмогорова — Смирнова, значение статистики которого подсчитывается по формуле (5.4.3).

Запишем результаты устного опроса двух выборок

учащихся в форме табл. 20, удобной для подсчета значения статистики критерия $T_3 = \max(S_2(x) - S_1(x))$ для выборок разного объема[1].

Таблица 20

Число верных ответов	Абс. частота, f_1	Абс. частота, f_2	Накопл. частота, Σf_1	Накопл. частота, Σf_2	$S_1 = \frac{\Sigma f_1}{n_1}$	$S_2 = \frac{\Sigma f_2}{n_2}$	$S_2 - S_1$
8	3	2	16	20	1,000	1,000	0,000
7	2	3	13	18	0,813	0,900	0,087
6	5	—	11	15	0,688	0,750	0,062
5	3	3	6	15	0,375	0,750	0,375
4	1	6	3	12	0,188	0,600	0,412
3	—	3	2	6	0,125	0,300	0,175
2	1	2	2	3	0,125	0,150	0,025
1	1	—	1	1	0,063	0.050	—0,013
0	—	1	0	1	0,000	0,050	+0,050

$n_1 = 16$ $n_2 = 20$

Из таблицы экспериментальных данных (табл. 20) находим, что максимальное значение выражения $(S_2 - S_1)$ равно 0,412. Отсюда значение статистики критерия по формуле (5.4.3)

$$T_3 = \max\ (S_2 - S_1) = 0{,}412\ .$$

По табл. *E* для $\alpha = 0{,}05$ при $n_1 = 16$ и $n_2 = 20$ критическое значение статистики одностороннего критерия $W_{1-\alpha} = \frac{31}{80} = 0{,}388$. Таким образом, верно неравенство $T_{\text{наблюд}} > T_{\text{критич}}$ $(0{,}412 > 0{,}388)$. Согласно правилу принятия решений при использовании одностороннего критерия (см. с. 114, случай в), нулевая гипотеза отклоняется и принимается альтернативная гипотеза H_1 (см. с. 118). Это означает, что учащиеся первого района дают стохастически больше верных ответов, чем учащиеся второго района, т. е. лучше усвоили проверяемый материал (даты исторических событий).

[1] Нежелание уравнять объемы выборок для облегчения вычислений объясняется тем, что выборки имеют небольшой объем и его уменьшение может повлиять на результаты сравнения деятельности учащихся двух районов города.

6 ВЫВОДЫ И РЕКОМЕНДАЦИИ

1. Сравнительный анализ критериев

Рассмотрим группу критериев, предназначенных для сравнения результатов двух зависимых выборок. Первый из них — критерий Макнамары допускает использование данных, полученных с помощью измерений по шкале наименований, и в этом смысле является уникальным критерием. Следует, однако, иметь в виду, что критерий Макнамары применим, если результаты эксперимента измерены по шкале наименований, имеющей только две категории (верно — неверно, согласен — не согласен, да — нет и т. п.). Сведение результатов эксперимента в таблицу 2×2, необходимое для подсчета значения статистики критерия, ограничивает возможности критерия в отношении использования информации, заключенной в экспериментальных данных. Это особенно проявляется в тех случаях, когда существует методика, позволяющая измерить изучаемое свойство по шкале наименований, имеющей более двух категорий, или по шкалам более высокого порядка.

Критерий знаков можно применять для сравнения данных, измеренных по шкале порядка. Критерий прост в употреблении, но использует далеко не всю информацию, заключенную в результатах выборок. Кроме того, критерий знаков состоятелен для проверки нулевых гипотез в том случае, когда альтернативная гипотеза утверждает, что плотности распределений сравниваемых случайных величин отличаются сдвигом, т. е. если $p(x)$ — плотность одной величины, то плотность другой величины имеет вид $p(x-a)$ (см. рис. 13, случай $a>0$).

Как показал Б. В. Гнеденко [6], критерий знаков не состоятелен для альтернативных гипотез более общего типа, утверждающих различие законов распределения рассматриваемых случайных величин без указания, в чем именно это различие проявляется[1]. Иначе говоря, возможны случаи, когда критерий знаков не позволяет отклонить нулевую гипотезу, хотя с достаточно большой вероятностью (вероятность ошибки второго рода) будет верна альтернативная гипотеза.

[1] Кроме сдвига распределения могут различаться, например, дисперсиями, асимметрией, эксцессом и др.

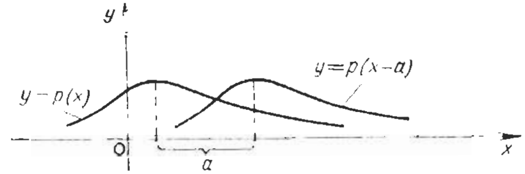


Рис. 13

Третий критерий — критерий Вилкоксона. Этот критерий весьма популярен среди исследователей, особенно психологов[1]. В связи с этим необходимо еще раз подчеркнуть довольно жесткие требования, которые обусловливают возможность применения критерия Вилкоксона. Во-первых, исходные данные должны быть измерены по интервальной шкале. Во-вторых, плотности распределения сравниваемых случайных величин должны быть симметричны, причем критерий состоятелен для альтернатив, утверждающих, что эти плотности распределения различаются только сдвигом. При проверке таких гипотез критерий Вилкоксона более чувствителен, чем критерий знаков, так как в некоторых случаях при одном и том же уровне значимости позволяет отклонить нулевую гипотезу, когда критерий знаков этого сделать не позволяет. Однако, подобно критерию знаков, критерий Вилкоксона не состоятелен при проверке нулевой гипотезы против альтернатив более общего вида. Таким образом, при использовании критерия Вилкоксона следует быть особенно осмотрительным, тем более что проверку свойства симметричности распределения в психолого-педагогических исследованиях осуществлять нелегко.

Резюмируя сказанное, приходим к следующим выводам в отношении возможностей использования трех рассмотренных критериев в педагогических исследованиях. Требования интервальных измерений и симметрии плотности распределения, необходимые для использования критерия Вилкоксона, ограничивают сферу его примене-

[1] См.: *Артемьева Е. Ю.* Сборник задач по теории вероятностей и математической статистике для психологов. М., 1969.

ния в педагогических исследованиях в настоящее время. Кроме того, критерий Вилкоксона и критерий знаков применимы только в тех случаях, когда альтернативная гипотеза предполагает, что распределения изучаемого свойства в двух рассматриваемых совокупностях различаются сдвигом. В то же время критерий Макнамары не имеет подобных ограничений для альтернативной гипотезы.

Таким образом, если предварительные исследования не позволяют сформулировать альтернативу о сдвиге (что характерно для большинства педагогических исследований), то для сравнения двух распределений изучаемого свойства следует использовать критерий Макнамары. В связи с этим приходится сделать вывод о том, что в настоящее время наиболее широкое применение в педагогических исследованиях может найти именно критерий Макнамары.

Рассмотрим теперь критерии для сравнения результатов независимых выборок.

Медианный критерий и критерий Вилкоксона — Манна — Уитни предназначены для улавливания различий в центральных тенденциях распределений рассматриваемых случайных величин, а именно различия медиан и средних значений. Поэтому указанные критерии, подобно критерию знаков и критерию Вилкоксона для зависимых выборок, более приспособлены для проверки нулевой гипотезы при альтернативах, утверждающих, что распределения случайных величин отличаются сдвигом. Критерии допускают использование данных, измеренных по шкале не ниже порядковой, но ни один из них (в отличие от критерия Вилкоксона для зависимых выборок) не требует, чтобы измерения были выполнены по шкалам более высокого порядка (интервальная шкала и шкала отношений). Следует также подчеркнуть, что использование этих критериев не опирается на предположение о симметричности распределений каждой из сравниваемых случайных величин.

Критерий Вилкоксона — Манна — Уитни более чувствителен, чем медианный, так как при одном и том же уровне значимости иногда позволяет отклонить нулевую гипотезу в тех случаях, когда медианный критерий этого сделать не позволяет. Поэтому среди указанных двух критериев некоторое предпочтение следует отдать критерию Вилкоксона — Манна — Уитни.

Другая группа критериев для независимых выборок, рассмотренная в данной работе, состоит из критерия χ^2 (хи-квадрат) и критерия Колмогорова — Смирнова. Эти критерии состоятельны для проверки нулевой гипотезы при самых общих альтернативных гипотезах. Наибольшую область применения имеет, бесспорно, критерий χ^2, так как он допускает использование данных, измеренных даже с помощью шкалы наименований и притом с любым числом категорий (сравните с критерием Макнамары, который применим только при двух категориях шкалы наименований). Критерий Колмогорова — Смирнова допускает использование данных, измеренных по шкале не ниже порядковой. Однако для порядковых статистик таблицы точных критических значений составлены для выборок небольшого объема, поэтому при выборках большого объема приходится пользоваться таблицами критических значений предельного распределения. Переход к предельному распределению возможен только в том случае, когда экспериментальные данные представляют измерения по интервальной шкале. Поэтому в настоящее время в педагогических исследованиях применение критерия Колмогорова — Смирнова возможно в случае оценки результатов педагогического эксперимента в двух независимых выборках учащихся с помощью времени изучаемых реакций или числа верных (неверных) ответов учащихся на контрольные вопросы (число вопросов должно быть не менее 7—8). Дело в том, что время реакции всегда является интервальным измерением, а число верных ответов учащихся на контрольные задания при выполнении некоторых условий (см. с. 25) можно также считать измерением, сделанным по интервальной шкале. В тех случаях, когда исходные данные допускают применение и критерия χ^2, и критерия Колмогорова — Смирнова, второй критерий несколько более чувствителен, позволяя в ряде случаев отклонять нулевую гипотезу при более низком уровне значимости, т. е. с меньшей вероятностью ошибки первого рода. В то же время статистика критерия χ^2 (хи-квадрат)

$$T = \Sigma \frac{(O-E)^2}{E},$$

где O — наблюдаемые частоты, а E — ожидаемые (согласно нулевой гипотезе) частоты попадания в интервалы группировки значений рассматриваемой случайной величины X (изучаемое свойство объектов), по-видимому, наилучшим образом улавливает различие вероятностей попадания объектов рассматриваемых совокупностей в одни и те же интервалы группировки значений изучаемого свойства [13].

При сопоставлении всех четырех критериев приходим к следующим выводам. Поскольку в педагогических исследованиях предварительный выбор специальных альтернативных гипотез к проверяемой нулевой гипотезе (например, альтернативы о сдвиге распределений) осуществить весьма трудно, в настоящее время можно рекомендовать сравнение результатов двух независимых выборок проводить с помощью критерия χ^2 и критерия Колмогорова — Смирнова. Причем критерий χ^2 следует применять в тех случаях, когда результаты педагогического эксперимента не могут быть измерены по интервальной шкале. Если интервальные измерения возможны и данные можно сгруппировать не менее чем в 8 интервалах, то лучше применять критерий Колмогорова — Смирнова. Если же число интервалов группировки меньше 8, то и в этом случае следует применять критерий χ^2.

2. Объективность статистических выводов, полученных на основе малых выборок

Рассмотренные в данной работе непараметрические критерии позволяют формулировать статистические выводы на основе данных, полученных при помощи обследования членов выборок сравнительно небольшого объема. Однако следует по-разному оценивать объективность этих выводов в случае отклонения или принятия (неотклонения) нулевой гипотезы. Дело в том, что, задавая заранее тот или другой уровень значимости α, можно гарантировать малую вероятность ошибки первого рода при отклонении нулевой гипотезы, если она верна, и, следовательно, вывод об отклонении нулевой гипотезы будет статистически достаточно объективным. Если же критерий не позволяет отклонить нулевую гипотезу, а на самом деле верна альтернативная гипотеза, то вероятность такой ошибки (ошибка второго рода) не обязательно будет малой. Даже состоятельные критерии гарантируют малую вероятность ошибки второго рода лишь в случае достаточно большого объема выборок. Поэтому при использовании результатов выборок малого объема с достаточным доверием следует относиться лишь к выводам об отклонении нулевой гипотезы. Особенно это относится к критериям, состоятельным для специальных, труднопроверяемых в педагогических исследованиях альтернатив (критерий знаков, критерий Вилкоксона и Вилкоксона — Манна — Уитни, медианный критерий). Если критерий не позволяет отклонить нулевую гипотезу, то ее следует перепроверить, увеличивая объемы выборок.

3. Объективность исходных данных для проверки статистических гипотез

Критерии, с помощью которых проверяются статистические гипотезы, позволяют формулировать достаточно обоснованные выводы только при условии, что исходные данные, используемые для вычисления статистики критерия, объективны.

Получение объективных исходных данных возможно только при соблюдении научных принципов планирования эксперимента. Допустим, например, что цель исследования состоит в обосновании эффективности какого-либо педагогического приема или метода и для этого сопоставляются результаты работы контрольной и экспериментальной групп учащихся. Если при планировании эксперимента не исключить влияния на качество проверяемых знаний учащихся других факторов, не связанных непосредственно с самим методом обучения, таких, как личность учителя, тип школы, предварительная подготовка учащихся и др., то даже в том случае, когда будет обнаружено статистически значимое различие в деятельности учащихся двух рассматриваемых групп, нельзя утверждать, что это различие определяется именно методом обучения.

Таким образом, научное планирование педагогических исследований является непременным условием как для самой постановки экспериментов, так и для проверки педагогических гипотез.

ЛИТЕРАТУРА

1. *Бернстайн А.* Справочник статистических решений. М., 1968.
2. *Большов Л. Н., Смирнов Н. В.* Таблицы математической статистики. М., 1965.
3. *Ван дер Варден Б. Л.* Математическая статистика. М., 1960.
4. *Вентцель Е. С.* Теория вероятностей. М., 1969.
5. *Гаек Я., Шидак З.* Теория ранговых критериев. М., 1971.
6. *Гнеденко Б. В., Беляев Ю. К., Соловьев А. Д.* Математические методы в теории надежности. М., 1965.
7. *Гнеденко Б. В., Хинчин А. Я.* Элементарное введение в теорию вероятностей. М., 1971.
8. *Грабарь М. И., Краснянская К. А.* Некоторые положения выборочного метода в связи с организацией изучения знаний учащихся. М., 1973.
9. *Дунин-Барковский И., Смирнов Н.* Теория вероятностей и математическая статистика в технике. М., 1955.
10. *Кендалл М., Стьюарт М.* Статистические выводы и связи. М., 1973.
11. *Леман З.* Проверка статистических гипотез. М., 1964.
12. Психологические измерения. Сб. переводов. М., 1967.
13. *Рао Р.* Линейные статистические методы и их применение. М., 1968.
14. *Стивенс С. С.* Экспериментальная психология (глава: Математика, измерения и психофизика). М., 1960.
15. *Урбах В.* Биометрические методы. М., 1964.
16. *Урбах В.* Математическая статистика для медиков и биологов. М., 1963.
17. *Фрес П., Пиаже Ж.* Экспериментальная психология. М., 1966.
18. *Bradley J. V.* Distribution — free Statistical Tests N. Y., 1968.
19. *Conover W. J.* Practical Nonparametric Statistics. N. Y., 1972.
20. *Gibbons J. D.* Nonparametric Statistical Inference. N. Y., 1971.
21. *Ferguson G. A.* Statistical Analysis in Psychology and Education, 2 ed. N. Y., 1966.
22. *Fraser D. A. S.* Nonparametric Methods in Statistis. N. Y., 1957.
23. *Noether G. E.* Elements of Nonparametric Statistics. N. Y., 1967.
24. *Siegel S.* Nonparametric Statistics for the Behavioral Sciences. N. Y., 1956.
25. *Hajek J.* Nonparametric Statistics. San Fransisco, 1969.

ТАБЛИЦЫ

Таблица *А**

Таблица вероятностей $P(T_2 \leqslant T_2$ наблюдаемое) для биномиального распределения при $p=q=0{,}5$**

n \ T_2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
5	031	188	500	812	969	+***										
6	016	109	344	656	891	984	+									
7	008	062	227	500	773	938	992	+								
8	004	035	145	363	637	855	965	996	+							
9	002	020	090	254	500	746	910	980	998	+						
10	001	011	055	172	377	623	828	945	989	998	+					
11		006	033	113	274	500	726	887	967	994	+	+				
12		003	019	073	194	387	613	806	927	981	997	+	+			
13		002	011	046	133	291	500	709	867	954	989	998	+	+		
14		001	006	029	090	212	395	605	788	910	971	994	999	+	+	
15			004	018	059	151	304	500	696	849	941	982	996	+	+	+
16			002	011	038	105	227	402	598	773	896	962	989	998	+	+
17			001	006	025	072	166	315	500	685	834	928	975	994	999	+
18			001	004	015	048	119	240	407	593	760	881	952	985	996	999
19				002	010	032	084	180	324	500	676	820	916	968	990	998
20				001	006	021	058	132	252	412	588	748	868	942	979	994
21				001	004	013	039	095	192	332	500	668	808	905	961	987
22					002	008	026	067	143	262	416	584	738	857	933	974
23					001	005	017	047	105	202	339	500	661	798	895	953
24					001	003	011	032	076	154	271	419	581	729	846	924
25						002	007	022	054	115	212	345	500	655	788	885

* Таблицы *А — Ж* заимствованы из книг [3, 19, 24].

** Значения вероятностей даны в десятичных дробях, например число 031 означает 0,031.

*** Знак «+» означает 1 или число, близкое к 1.

Таблица Б

Критические значения статистики критерия знаков

n	Уровень значимости для одностороннего критерия						n	Уровень значимости для одностороннего критерия					
	$\alpha=0{,}025$		$\alpha=0{,}01$		$\alpha=0{,}005$			$\alpha=0{,}025$		$\alpha=0{,}01$		$\alpha=0{,}005$	
	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$		t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$
5	0	5	0	5	0	5	8	1	7	1	7	1	7
6	1	5	0	6	0	6	9	2	7	1	8	1	8
7	1	6	1	6	0	7	10	2	8	1	9	1	9

n	Уровень значимости для одностороннего критерия						n	Уровень значимости для одностороннего критерия					
	$\alpha=0,025$		$\alpha=0,01$		$\alpha=0,005$			$\alpha=0,025$		$\alpha=0,01$		$\alpha=0,005$	
	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$		t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$
11	2	9	2	9	1	10	54	20	34	19	35	18	36
12	3	9	2	10	2	10	55	20	35	19	36	18	37
13	3	10	2	11	2	11	56	21	35	19	37	18	38
14	3	11	3	11	2	12	57	21	36	20	37	19	38
15	4	11	3	12	3	12	58	22	36	20	38	19	39
16	4	12	3	13	3	13	59	22	37	21	38	20	39
17	5	12	4	13	3	14	60	22	38	21	39	20	40
18	5	13	4	14	4	14	61	23	38	21	40	21	40
19	5	14	5	14	4	15	62	23	39	22	40	21	41
20	6	14	5	15	4	16	63	24	39	22	41	21	42
21	6	15	5	16	5	16	64	24	40	23	41	22	42
22	6	16	6	16	5	17	65	25	40	23	42	22	43
23	7	16	6	17	5	18	66	25	41	24	42	23	43
24	7	17	6	18	6	18	67	26	41	24	43	23	44
25	8	17	7	18	6	19	68	26	42	24	44	23	45
26	8	18	7	19	7	19	69	26	43	25	44	24	45
27	8	19	8	19	7	20	70	27	43	25	45	24	46
28	9	19	8	20	7	21	71	27	44	26	45	25	46
29	9	20	8	21	8	21	72	28	44	26	46	25	47
30	10	20	9	21	8	22	73	28	45	27	46	26	47
31	10	21	9	22	8	23	74	29	45	27	47	26	48
32	10	22	9	23	9	23	75	29	46	27	48	26	49
33	11	22	10	23	9	24	76	29	47	28	48	27	49
34	11	23	10	24	10	24	77	30	47	28	49	27	50
35	12	23	11	24	10	25	78	30	48	29	49	28	50
36	12	24	11	25	10	26	79	31	48	29	50	28	51
37	13	24	11	26	11	26	80	31	49	30	50	29	51
38	13	25	12	26	11	27	81	32	49	30	51	29	52
39	13	26	12	27	12	27	82	32	50	31	51	29	53
40	14	26	13	27	12	28	83	33	50	31	52	30	53
41	14	27	13	28	12	29	84	33	51	31	53	30	54
42	15	27	14	28	13	29	85	33	52	32	53	31	54
43	15	28	14	29	13	30	86	34	52	32	54	31	55
44	16	28	14	30	14	30	87	34	53	33	54	32	55
45	16	29	15	30	14	31	88	35	53	33	55	32	56
46	16	30	15	31	14	32	89	35	54	34	55	32	57
47	17	30	16	31	15	32	90	36	54	34	56	33	57
48	17	31	16	32	15	33	91	36	55	34	57	33	58
49	18	31	16	33	16	33	92	37	55	35	57	34	58
50	18	32	17	33	16	34	93	37	56	35	58	34	59
51	19	32	17	34	16	35	94	38	56	36	58	35	59
52	19	33	18	34	17	35	95	38	57	36	59	35	60
53	19	34	18	35	17	36	96	38	58	37	59	35	61

n	Уровень значимости для одностороннего критерия						n	Уровень значимости для одностороннего критерия					
	$\alpha=0,025$		$\alpha=0,01$		$\alpha=0,005$			$\alpha=0,025$		$\alpha=0,01$		$\alpha=0,005$	
	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$		t_α	$n-t_\alpha$	t_α	$n-t_\alpha$	t_α	$n-t_\alpha$
97	39	58	37	60	36	61	99	40	59	38	61	37	62
98	39	59	38	60	36	62	100	40	60	38	62	37	63
	Уровень значимости для двустороннего критерия							Уровень значимости для двустороннего критерия					
	$\alpha=0,05$		$\alpha=0,02$		$\alpha=0,01$			$\alpha=0,05$		$\alpha=0,02$		$\alpha=0,01$	
	$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$	$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$	$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$		$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$	$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$	$t_{\frac{\alpha}{2}}$	$n-t_{\frac{\alpha}{2}}$

Таблица *B*

Критические значения статистики критерия Вилкоксона

n	Уровень значимости для одностороннего критерия							
	$\alpha=0,05$		$\alpha=0,025$		$\alpha=0,01$		$\alpha=0,005$	
	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$	W_α	$W_{1-\alpha}$
6	3	17	1	20	0	21	0	21
7	4	24	3	25	1	27	0	21
8	6	30	4	32	2	34	1	35
9	9	36	6	39	4	41	2	43
10	11	44	9	46	6	49	4	51
11	14	52	11	55	8	58	6	60
12	18	60	14	64	10	68	8	70
13	22	69	18	73	13	78	10	81
14	26	79	22	83	16	89	13	92
15	31	89	26	94	20	100	16	104
16	36	100	30	106	24	112	20	116
17	42	111	35	118	28	125	24	129
18	48	123	41	130	33	138	28	143
19	54	136	47	143	38	152	33	157
20	61	149	53	157	44	166	38	172
	Уровень значимости для двустороннего критерия							
	$\alpha=0,10$		$\alpha=0,05$		$\alpha=0,02$		$\alpha=0,01$	
	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$	$W_{\frac{\alpha}{2}}$	$W_{1-\frac{\alpha}{2}}$

Таблица Г

Критические значения статистик, имеющих распределение χ^2 с числом степеней свободы ν, для уровней значимости α

Число степеней свободы, ν	$\alpha=0,10$ $1-\alpha=0,90$	$\alpha=0,05$ $1-\alpha=0,95$	$\alpha=0,025$ $1-\alpha=0,975$	$\alpha=0,01$ $1-\alpha=0,99$
1	2,706	3,841	5,024	6,635
2	4,605	5,991	7,378	9,210
3	6,251	7,815	9,348	11,34
4	7,779	9,488	11,14	13,28
5	9,236	11,07	12,83	15,09
6	10,64	12,59	14,45	16,81
7	12,02	14,07	16,01	18,48
8	13,36	15,51	17,53	20,09
9	14,68	16,92	19,02	21,67
10	15,99	18,31	20,48	23,21
11	17,28	19,68	21,92	24,73
12	18,55	21,03	23,34	26,22
13	19,81	22,36	24,74	27,69
14	21,06	23,68	26,12	29,14
15	22,31	25,00	27,49	30,58
16	23,54	26,30	28,85	32,00
17	24,77	27,59	30,19	33,41
18	25,99	28,87	31,53	34,81
19	27,20	30,14	32,85	36,19
20	28,41	31,41	34,17	37,57
21	29,62	32,67	35,48	38,93
22	30,81	33,92	36,78	40,29
23	32,01	35,17	38,08	41,64
24	33,20	36,42	39,37	42,98
25	34,38	37,65	40,65	44,31
26	35,56	38,89	41,92	45,64
27	36,74	40,11	43,19	46,96
28	37,92	41,34	44,46	48,28
29	39,09	42,56	45,72	49,59
30	40,26	43,77	46,98	50,89
40	51,81	55,76	59,34	63,69
50	63,17	67,50	71,42	76,15
60	74,40	79,08	83,30	88,38
70	85,53	90,53	95,02	100,4
80	96,58	101,9	106,6	112,3
90	107,6	113,1	118,1	124,1
100	118,5	124,3	129,6	135,8

Таблица Д

Критические значения статистики критерия Вилкоксона — Манна — Уитни

n_1	α	$n_2=$ 5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	0,01	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
	0,025	3	4	6	7	8	9	10	12	13	14	15	16	18	19	20	21
	0,05	5	6	7	9	10	12	13	14	16	17	19	20	21	23	24	26
	0,10	6	8	9	11	13	14	16	18	19	21	23	24	26	28	29	31
6	0,01	3	4	5	7	8	9	10	12	13	14	16	17	19	20	21	23
	0,025	4	6	7	9	11	12	14	15	17	18	20	22	23	25	26	28
	0,05	6	8	9	11	13	15	17	18	20	22	24	26	27	29	31	33
	0,10	8	10	12	14	16	18	20	22	24	26	28	30	32	35	37	39
7	0,01	4	5	7	8	10	12	13	15	17	18	20	22	24	25	27	29
	0,025	6	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35
	0,05	7	9	12	14	16	18	20	22	25	27	29	31	34	36	38	40
	0,10	9	12	14	17	19	22	24	27	29	32	34	37	39	42	44	47
8	0,01	5	7	8	10	12	14	16	18	21	23	25	27	29	31	33	35
	0,025	7	9	11	14	16	18	20	23	25	27	30	32	35	37	39	42
	0,05	9	11	14	16	19	21	24	27	29	32	34	37	40	42	45	48
	0,10	11	14	17	20	23	25	28	31	34	37	40	43	46	49	52	55
9	0,01	6	8	10	12	15	17	19	22	24	27	29	32	34	37	39	41
	0,025	8	11	13	16	18	21	24	27	29	32	35	38	40	43	46	49
	0,05	10	13	16	19	22	25	28	31	34	37	40	43	46	49	52	55
	0,10	13	16	19	23	26	29	32	36	39	42	46	49	53	56	59	63
10	0,01	7	9	12	14	17	20	23	25	28	31	34	37	39	42	45	48
	0,025	9	12	15	18	21	24	27	30	34	37	40	43	46	49	53	56
	0,05	12	15	18	21	25	28	32	35	38	42	45	49	52	56	59	63
	0,10	14	18	22	25	29	33	37	40	44	48	52	55	59	63	67	71
11	0,01	8	10	13	16	19	23	26	29	32	35	38	42	45	48	51	54
	0,025	10	14	17	20	24	27	31	34	38	41	45	48	52	56	59	63
	0,05	13	17	20	24	28	32	35	39	43	47	51	55	58	62	66	70
	0,10	16	20	24	28	32	37	41	45	49	53	58	62	66	70	74	79
12	0,01	9	12	15	18	22	25	29	32	36	39	43	47	50	54	57	61
	0,025	12	15	19	23	27	30	34	38	42	46	50	54	58	62	66	70
	0,05	14	18	22	27	31	35	39	43	48	52	56	61	65	69	73	78
	0,10	18	22	27	31	36	40	45	50	54	59	64	68	73	78	82	87

Продолжение

n_1	α	$n_2=5$	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
13	0,01	10	13	17	21	24	28	32	36	40	44	48	52	56	60	64	68
	0,025	13	17	21	25	29	34	38	42	46	51	55	60	64	68	73	77
	0,05	16	20	25	29	34	38	43	48	52	57	62	66	71	76	81	85
	0,10	19	24	29	34	39	44	49	54	59	64	69	75	80	85	90	95
14	0,01	11	14	18	23	27	31	35	39	44	48	52	57	61	66	70	74
	0,025	14	18	23	27	32	37	41	46	51	56	60	65	70	75	79	84
	0,05	17	22	27	32	37	42	47	52	57	62	67	72	78	83	88	93
	0,10	21	26	32	37	42	48	53	59	64	70	75	81	86	92	98	103
15	0,01	12	16	20	25	29	34	38	43	48	52	57	62	67	71	76	81
	0,025	15	20	25	30	35	40	45	50	55	60	65	71	76	81	86	91
	0,05	19	24	29	34	40	45	51	56	62	67	73	78	84	89	95	101
	0,10	23	28	34	40	46	52	58	64	69	75	81	87	93	99	105	111
16	0,01	13	17	22	27	32	37	42	47	52	57	62	67	72	77	83	88
	0,025	16	22	27	32	38	43	48	54	60	65	71	76	82	87	93	99
	0,05	20	26	31	37	43	49	55	61	66	72	78	84	90	96	102	108
	0,10	24	30	37	43	49	55	62	68	75	81	87	94	100	107	113	120
17	0,01	14	19	24	29	34	39	45	50	56	61	67	72	78	83	89	94
	0,025	18	23	29	35	40	46	52	58	64	70	76	82	88	94	100	106
	0,05	21	27	34	40	46	52	58	65	71	78	84	90	97	103	110	116
	0,10	26	32	39	46	53	59	66	73	80	86	93	100	107	114	121	128
18	0,01	15	20	25	31	37	42	48	54	60	66	71	77	83	89	95	101
	0,025	19	25	31	37	43	49	56	62	68	75	81	87	94	100	107	113
	0,05	23	29	36	42	49	56	62	69	76	83	89	96	103	110	117	124
	0,10	28	35	42	49	56	63	70	78	85	92	99	107	114	121	129	136
19	0,01	16	21	27	33	39	45	51	57	64	70	76	83	89	95	102	108
	0,025	20	26	33	39	46	53	59	66	73	79	86	93	100	107	114	120
	0,05	24	31	38	45	52	59	66	73	81	88	95	102	110	117	124	131
	0,10	29	37	44	52	59	67	74	82	90	98	105	113	121	129	136	144
20	0,01	17	23	29	35	41	48	54	61	68	74	81	88	94	101	108	115
	0,025	21	28	35	42	49	56	63	70	77	84	91	99	106	113	120	128
	0,05	26	33	40	48	55	63	70	78	85	93	101	108	116	124	131	139
	0,10	31	39	47	55	63	71	79	87	95	103	111	120	128	136	144	152

Таблица *E*

Критические значения статистики критерия Колмогорова — Смирнова для двух выборок различных объемов n_1 и n_2

Односторонний критерий:		1—α=0,90	0,95	0,975	0,99	0,995
Двусторонний критерий:		1—α=0,80	0,90	0,95	0,98	0,99
$n_1=5$	$n_2=6$	3/5	2/3	2/3	5/6	5/6
	7	4/7	23/25	5/7	29/35	6/7
	8	11/20	5/8	27/40	4/5	4/5
	9	5/9	3/5	31/45	7/9	4/5
	10	1/2	3/5	7/10	7/10	4/5
	15	8/15	3/5	2/3	11/15	11/15
	20	1/2	11/20	3/5	7/10	3/4
$n_1=6$	$n_2=7$	23/42	4/7	29/42	5/7	5/6
	8	1/2	7/12	2/3	3/4	3/4
	9	1/2	5/9	2/3	13/18	7/9
	10	1/2	17/30	19/30	7/10	11/15
	12	1/2	7/12	7/12	2/3	3/4
	18	4/9	5/9	11/18	2/3	13/18
	24	11/24	1/2	7/12	5/8	2/3
$n_1=7$	$n_2=8$	27/56	33/56	5/8	41/56	3/4
	9	31/63	5/9	40/63	5/7	47/63
	10	33/70	39/70	43/70	7/10	5/7
	14	3/7	1/2	4/7	9/14	5/7
	28	3/7	13/28	15/28	17/28	9/14

Продолжение

Односторонний критерий: / Двусторонний критерий:		1—α=0,90 / 1—α=0,80	0,95 / 0,90	0,975 / 0,95	0,99 / 0,98	0,995 / 0,99
$n_1 = 8$	$n_2 = 9$	4/9	13/24	5/8	2/3	3/4
	10	19/40	21/40	23/40	27/40	7/10
	12	11/24	1/2	7/12	5/8	2/3
	16	7/16	1/2	9/16	5/8	5/8
	32	13/32	7/16	1/2	9/16	19/32
$n_1 = 9$	$n_2 = 10$	7/15	1/2	26/45	2/3	31/45
	12	4/9	1/2	5/9	11/18	2/3
	15	19/45	22/45	8/15	3/5	29/45
	18	7/18	4/9	1/2	5/9	11/18
	36	13/36	5/12	17/36	19/36	5/9
$n_1 = 10$	$n_2 = 15$	2/5	7/15	1/2	17/30	19/30
	20	2/5	9/20	1/2	11/20	3/5
	40	7/20	2/5	9/20	1/2	
$n_1 = 12$	$n_2 = 15$	23/60	9/20	1/2	11/20	7/12
	16	7/8	7/16	23/48	13/24	7/12
	18	13/36	5/12	17/36	19/36	5/9
	20	11/30	5/12	7/15	31/60	17/30
$n_1 = 15$	$n_2 = 20$	7/20	2/5	13/30	29/60	31/60
$n_1 = 16$	$n_2 = 20$	27/80	31/80	17/40	19/40	41/80
Приближенные значения для больших выборок		$1,07\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1,22\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1,36\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1,52\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$	$1,63\sqrt{\frac{n_1+n_2}{n_1 \cdot n_2}}$

Т а б л и ц а *Ж*

Критические значения статистики критерия Колмогорова — Смирнова для двух выборок одинакового объема n

Односторонний критерий: 1—α=	0,90	0,95	0,975	0,99	0,995
Двусторонний критерий: 1—α=	**0,80**	**0,90**	**0,95**	**0,98**	**0,99**
$n = 3$	2/3	2/3			
4	3/4	3/4	3/4		
5	3/5	3/5	4/5	4/5	4/5
6	3/6	4/6	4/6	5/6	5/6
7	4/7	4/7	5/7	5/7	5/7
8	4/8	4/8	5/8	5/8	6/8
9	4/9	5/9	5/9	6/9	6/9
10	4/10	5/10	6/10	6/10	7/10
11	5/11	5/11	6/11	7/11	7/11
12	5/12	5/12	6/12	7/12	7/12
13	5/13	6/13	6/13	7/13	8/13
14	5/14	6/14	7/14	7/14	8/14
15	5/15	6/15	7/15	8/15	8/15
16	6/16	6/16	7/16	8/16	9/16
17	6/17	7/17	7/17	8/17	9/17
18	6/18	7/18	8/18	9/18	9/18
19	6/19	7/19	8/19	9/19	9/19

Односторонний критерий: 1—α=	0,90	0,95	0,975	0,99	0,995
Двусторонний критерий: 1—α=	**0,80**	**0,90**	**0,95**	**0,98**	**0,99**
$n = 20$	6/20	7/20	8/20	9/20	10/20
21	6/21	7/21	8/21	9/21	10/21
22	7/22	8/22	8/22	10/22	10/22
23	7/23	8/23	9/23	10/23	10/23
24	7/24	8/24	9/24	10/24	11/24
25	7/25	8/25	9/25	10/25	11/25
26	7/26	8/26	9/26	10/26	11/26
27	7/27	8/27	9/27	11/27	11/27
28	8/28	9/28	10/28	11/28	12/28
29	8/29	9/29	10/29	11/29	12/29
30	8/30	9/30	10/30	11/30	12/30
31	8/31	9/31	10/31	11/31	12/31
32	8/32	9/32	10/32	12/32	12/32
34	8/34	10/34	11/34	12/34	13/34
36	9/36	10/36	11/36	12/36	13/36
38	9/38	10/38	11/38	13/38	14/38
40	9/40	10/40	12/40	13/40	14/40
Приближенные значения для $n > 40$	$\frac{1,52}{\sqrt{n}}$	$\frac{1,73}{\sqrt{n}}$	$\frac{1,92}{\sqrt{n}}$	$\frac{2,15}{\sqrt{n}}$	$\frac{2,30}{\sqrt{n}}$

ОГЛАВЛЕНИЕ