

УЧ. 2
У 54

**МЕТОДИКА ОРГАНИЗАЦИИ
И ПРОВЕДЕНИЯ
ПЕДАГОГИЧЕСКОГО
ЭКСПЕРИМЕНТА**

Электронный архив библиотеки МГУ имени А.А. Курашова

74.2
2154

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ

Учреждение образования
“МОГИЛЕВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. А.А. КУЛЕШОВА”

МЕТОДИКА ОРГАНИЗАЦИИ И ПРОВЕДЕНИЯ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА

Методические рекомендации

Автор-составитель

В.М. Кротов

БІБЛІЯТЭКА
Мастаўскага
дзяржаўнага
універсітэта
імя А. А. Куляшова



Могилев 2008



000001 728484

УДК 372.853(075.8)
ББК 74.265.1я73
М54

*Печатается по решению редакционно-издательского совета
УО "МГУ им. А.А. Кулешова"*

Рецензент
кандидат педагогических наук
заведующая кафедрой педагогики и психологии
МГОИПК и ПРР и СО
Л.И. Лунева

**Методика организации и проведения педагогическо-
М54 го эксперимента: метод. рекомендации / авт.-сост. В.М. Кро-
тов. – Могилев: УО "МГУ им. А.А. Кулешова", 2008. – 92 с.: ил.**

Данное издание включает учебный материал для чтения лекций и проведения практических занятий по курсу "Методика организации и проведения педагогического эксперимента", подготовки и проведения педагогического эксперимента при выполнении студентами курсовых и дипломных проектов.

Рекомендуется студентам физико-математического факультета. Будет весьма полезным учителям физики средних общеобразовательных учреждений при разработке и апробации дидактических материалов, методики изучения отдельных тем учебных курсов.

**УДК 372.853(075.8)
ББК 74.265.1я73**

© Кротов В.М., составление, 2008

© Оформление. УО "МГУ им. А.А. Кулешова", 2008

ВВЕДЕНИЕ

В настоящее время в сфере образования проявляется большой интерес к педагогическому эксперименту. Проведением педагогического эксперимента занимаются не только профессиональные исследователи, но и учителя, директора школ, методисты, не обладающие в достаточном объеме исследовательскими навыками.

В ситуации реформирования образования это явление имеет объективные предпосылки и носит достаточно распространенный характер. Учитель, директор школы, заместитель директора, инспектор и методист управления образования выступают в качестве разработчиков и экспериментаторов учебно-методической и другой продукции, обеспечивающей деятельность субъектов образовательной сферы в новой ситуации.

Сегодня школьным специалистам как никогда ранее предоставлена свобода педагогического творчества. Учитель может разрабатывать дидактические материалы, методики изучения отдельных тем, учебных курсов; методист – технологизировать и транслировать уникальный педагогический опыт, приспособлять научные разработки к практике образования, создавать авторские технологии; управленец – разрабатывать модели управления учебным процессом, адекватные реализуемым педагогическим технологиям, создавать учебные заведения нового типа и т.п. Все эти инициативы нуждаются в экспериментальной проверке.

Практику необходим эксперимент для того, чтобы проверить состоятельность (или несостоятельность) того нового, что создано в процессе творческой работы, показать ошибочность или прогрессивность предложенной идеи, заложенной в эксперименте, и раскрыть ее влияние на конечный результат.

Наиболее актуальным при этом становится овладение работниками образования, занимающимися данной деятельностью, как основами проектирования и создания разработок, так и профессиональными средствами их апробации, проверки, испытания. Поэтому в учебные планы подготовки студентов по специальности 1-31.03. 01. 02 “Физика (научно-педагогическая деятельность)” введена дисциплина “Методика организации и проведения педагогического эксперимента”.

Цель дисциплины:

Подготовить будущих преподавателей физики к организации и проведению педагогического эксперимента, развить методическое мышление студентов.

Задачи дисциплины:

- понимание студентами общих оснований экспериментальной деятельности преподавателя (учителя);
- освоение студентами умения по подготовке и проведению педагогического эксперимента;
- формирование у студентов умения применять количественные измерения при проведении педагогического эксперимента.

В результате изучения курса “Методика организации и проведения педагогического эксперимента” студент:

Помнит:

- содержание понятия “педагогический эксперимент”;
- основные этапы экспериментального цикла;
- примерную структуру программы педагогического эксперимента;
- структуру личности учащихся;
- основные методы математической статистики.

Понимает:

- необходимость проведения преподавателем (учителем) педагогического эксперимента;
- содержание основных этапов экспериментального цикла;
- возможности применяемых в педагогике диагностических материалов;
- критерии выбора методов количественных измерений в педагогическом эксперименте.

Умеет:

- составлять программу педагогического эксперимента;
- подбирать необходимую экспериментальную базу;
- составлять или подбирать необходимый диагностический материал;
- выбирать методы количественных измерений в педагогическом эксперименте.

§ 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОДГОТОВКИ И ПРОВЕДЕНИЯ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА

1.1. Педагогический эксперимент как метод исследования в педагогике.

Педагогическая наука исследует процесс развития подрастающего поколения при освоении культуры общества во взаимодействии с организаторами воспитывающей среды как необходимого условия воспитания.

Составной частью любой науки является ее методология, т.е. учение о научном методе познания или совокупность методов, применяемых в какой-либо науке. В педагогике методологию определяют как учение о принципах, формах и формах познания и преобразования педагогической действительности. Принято выделять четыре уровня методологии: *философский, общенаучный, конкретно-научный и технологический*. Стержнем методологии частных наук является теория познания.

Метод (от греч. *metodos* – путь) – способ построения и обоснования научного знания, способ, посредством которого познается предмет науки. Разнообразные методы и приемы исследовательской деятельности в теории познания образуют следующие группы методов:

1. *Общелогические* (общие принципы научного мышления: анализ, синтез, индукция, дедукция, абстрагирование, умозаключение и т.д.).
2. Методы исследования, *используемые только в научном познании*:
 - методы построения эмпирического знания (наблюдение, эксперимент, измерение);
 - методы построения теоретического знания (идеализация, формализация, выдвижение гипотез, аналогия, моделирование, мысленный эксперимент и т.д.).
3. *Сугубо специальные* методы и приемы, процедуры экспериментального характера, непосредственно связанные с сущностью явления и применяемые в узкой области или науке.

К теоретическим методам исследования относят анализ педагогической литературы, архивных материалов, документации и продуктов познавательной деятельности, контент-анализ; практические методы (анализ процесса практической деятельности; хронометрия, профессиография, метод независимых характеристик); моделирование и др.

Теоретические методы исследования необходимы для определения проблем, формулирования гипотез и для оценки собранных фактов. Они связаны с изучением литературы: трудов классиков по вопросам человекознания; общих и специальных работ по педагогике; периодической психолого-педагогической печати; справочной педагогической литературы и методических пособий по педагогике и смежным наукам.

К эмпирическим методам исследования относят педагогический эксперимент, наблюдение, беседу, анкетирование, социометрию, тестирование, экспертирование, изучение, обобщение и распространение передового педагогического опыта и др.

Слово “эксперимент” означает “проба”, “опыт”, “испытание”. Существует множество определений понятия “педагогический эксперимент”.

Педагогический эксперимент – это метод познания, с помощью которого исследуются педагогические явления, факты, опыт (М.Н. Скаткин).

Педагогический эксперимент – это специальная организация педагогической деятельности учителей и учащихся с целью проверки и обоснования заранее разработанных теоретических предположений или гипотез (И.Ф. Харламов).

Педагогический эксперимент – это научно поставленный опыт преобразования педагогического процесса в точно учитываемых условиях. (И.П. Подласый).

Педагогический эксперимент – это активное вмешательство исследователя в изучаемое им педагогическое явление с целью открытия закономерностей и изменения существующей практики. (Ю.З. Кушнер) [4].

Все эти определения понятия “педагогический эксперимент” имеют право на существование, так как в них утверждается общая мысль о том, что *педагогический эксперимент – это научно обоснованная и хорошо продуманная система организации педагогического процесса, направленная на открытие нового педагогического знания, проверки и обоснования заранее разработанных научных предположений, гипотез.*

Выделяя существенные признаки наиболее употребительных формулировок, можно построить следующее определение педагогического эксперимента:

Педагогический эксперимент – это исследовательская деятельность, по проверке выдвинутой гипотезы, разворачиваемая в естественных или искусственно созданных контролируемых и управляемых условиях, результатом которой является новое знание, включающее в себя выделение существенных факторов, влияющих на позитивные изменения в состоянии учащегося.

Педагогический эксперимент следует рассматривать в контексте организации и проведения педагогического исследования, при котором применяется определенная совокупность научных методов, одним из которых является педагогическое наблюдение [4].

Педагогическое наблюдение – это непосредственное восприятие, познание педагогического процесса в естественных условиях (например, в процессе учебы, внеклассной работы и т.д.). Наблюдение требует от исследователя точной фиксации фактов, объективного педагогического анализа.

Наблюдение в педагогическом исследовании может быть направлено на достижение различных целей. Оно может быть использовано как источник информации для построения гипотез, служить для проверки данных, полученных другими методами, с его помощью можно извлечь дополнительные сведения об изучаемом объекте.

Трудности применения наблюдения делятся на субъективные и объективные. К субъективным трудностям наблюдения относится то, что исследователь понимает и истолковывает поведение и действия учащихся и преподавателей через призму собственного “Я”, через свою систему ценностных ориентаций.

К объективным трудностям наблюдения, прежде всего, следует отнести ограниченность времени наблюдения. Кроме того, далеко не все педагогические факты поддаются непосредственному наблюдению. Чтобы получить необходимую для целей исследования информацию, не пропустить каких-то важных фактов или значимых сведений изучаемого объекта, следует заранее разработать программу педагогического наблюдения.

Составление программы наблюдения ставит исследователя перед необходимостью:

- изучать объект исследования всесторонне, в различных условиях и ситуациях;
- строить систему классификации фактов, явлений, которые отвечают целям исследования;
- проверять жизненность гипотезы;

- устанавливать сроки проведения наблюдения и определять средства сбора информации.

Видов наблюдения существует много. По степени охвата наблюдение может быть сплошным (когда внимание исследователя обращено на весь класс (группу)) и выборочным (когда ведется наблюдение за отдельным учащимся и небольшой группой ребят).

По регулярности проведения можно различать наблюдение систематическое и случайное. Систематическое наблюдение характеризуется, прежде всего, регулярностью фиксации действий, ситуаций, процессов в течение определенного периода времени. Он позволяет выявить динамику процессов, значительно повысить достоверность экстраполяции их развития. К случайному наблюдению относится наблюдение заранее незапланированного явления, деятельности, педагогической ситуации.

Наблюдение бывает прямое и косвенное. Прямое наблюдение характеризуется тем, что исследователь изучает процесс непосредственно. Косвенное наблюдение характеризуется тем, что об особенностях изучаемого явления педагог узнает через других лиц.

В зависимости от степени участия исследователя в изучаемой ситуации различают включенное (участвующее) и невключенное (неучаствующее) наблюдение. При невключенном (внешнем) наблюдении исследователь находится вне изучаемого объекта, он как бы со стороны наблюдает происходящие процессы, не вмешиваясь в их ход, не задавая никаких вопросов, – он просто регистрирует ход происходящих событий.

Внешнее наблюдение дает возможность фиксировать факты так называемого “открытого поведения”, “открытых отношений”, “открытых поступков”. Но исследователь, как человек посторонний, не может точно знать, что скрывается за этими фактами, поэтому его интерпретация не всегда может быть объективной.

Повысить объективность наблюдения можно путем увеличения числа наблюдателей (исследователей). Недостатком данного вида наблюдений является то, что при нем всегда может иметь место некоторая нарочитость в поведении учащихся. Длительность и систематичность наблюдений снимают этот недостаток.

Включенным (участвующим) называется такой вид наблюдения, при котором исследователь в той или иной степени непосредственно включен в изучаемый процесс, находится в контакте с наблюдаемыми людьми и принимает участие в их деятельности.

Ценность включенного наблюдения состоит в том, что оно позволяет собрать информацию, недоступную для получения другими методами. Исследователю здесь открываются более значимые для коллективной деятельности процессы и явления. При длительном наблюдении члены изучаемого коллектива успевают привыкнуть к исследователю, они возвращаются к своим привычным действиям и поведению, к своим обычным правилам и нормам. Исследователь постепенно начинает постигать более глубинные процессы коллективной жизни, проникать во внутрь взаимоотношений, воспринимать закономерности и противоречия, которые внешне не видны.

В исследовании педагогических проблем может использоваться длительное наблюдение за одними и теми же учащимися на протяжении нескольких лет. Это наблюдение охватывает большое количество в разной мере взаимосвязанных явлений. При этом наблюдении предоставляется возможность проследить развитие личности учащегося или ученического коллектива (но мере перехода из одного класса в последующий) на протяжении многих лет.

Благодаря этому многолетнему наблюдению удастся раскрыть объективные и субъективные тенденции, изменения интеллектуальных, познавательных и духовных потребностей учащегося, определить причины изменения учащегося к учебе, труду, товариществу, коллективу. А на основе собранного материала наметить конкретные пути совершенствования учебно-воспитательного процесса.

Эффективность метода наблюдения во многом зависит от соблюдения следующих педагогических требований:

- успех метода наблюдения в конкретном педагогическом исследовании во многом определяется личностью наблюдателя (его мировоззрением, способностями, профессионализмом, общительностью, и другими качествами);
- наблюдение должно строго соответствовать сформулированным научным задачам исследования и ничем не вредить наблюдаемым учащимся;
- поведение исследователя не должно влиять на наблюдаемую ситуацию, или это влияние должно быть минимальным;
- наблюдение не должно быть субъективным, исследователь обязан фиксировать все факты, а не только те, которые его устраивают;
- результаты наблюдений фиксируются в письменной форме в определенной системе, соответствующей системе параметров объектов наблюдения;

• результаты наблюдений необходимо сопоставлять с результатами, полученными с помощью других методов (например, анкетирования, тестирования, рейтинга, анализа результатов деятельности и т.д.).

1.2. Основные виды и этапы проведения педагогического эксперимента

В зависимости от цели, которую преследует экспериментатор, различают:

1) *констатирующий* эксперимент (изучается состояние изучаемой проблемы на практике). Этот эксперимент проводится в начале исследования с целью выявления как положительных, так и отрицательных сторон изучаемой проблемы;

2) *поисково-обучающий (поисково-формирующий)* эксперимент (проверяется эффективность применения отдельных элементов теоретической модели учебно-воспитательного процесса в различном их сочетании);

3) *обучающий (созидательно-формирующий)* эксперимент (проверяется эффективность применения теоретической модели как методической системы учебно-воспитательного процесса). Если результаты оказываются эффективными, а гипотеза подтверждается, то полученные данные подвергаются дальнейшему научно-теоретическому анализу и делаются необходимые выводы;

4) *контрольный* эксперимент – это завершающий этап исследования определенной проблемы. Целью его проведения является проверка полученных выводов и разработанной методики в массовой педагогической практике, апробация методики в работе других учебных заведений и педагогов. Если контрольный эксперимент подтверждает сделанные выводы, исследователь обобщает полученные результаты, которые и становятся теоретическим и методическим достоянием педагогики.

Чаще всего выделенные виды эксперимента применяются комплексно, составляют целостную взаимосвязанную и последовательную систему педагогического исследования.

В методике педагогических исследований различают естественный и лабораторный эксперименты. *Естественный* эксперимент проводится в естественных условиях в форме обычных уроков, внеклассных занятий. Суть этого эксперимента состоит в том, что исследователь, анализируя те или иные педагогические явления, стремится создать педагогические ситуации таким образом, чтобы они не нарушали привыч-

ного хода деятельности учащихся и педагогов и в этом смысле носили естественный характер. Объектом естественного эксперимента чаще всего становятся планы и программы, учебники и учебные пособия, методы и формы обучения и воспитания.

Сущность *лабораторного* эксперимента состоит в том, что он предполагает создание искусственных условий для того, чтобы влияние множества неконтролируемых факторов, различных объективных и субъективных причин свести к минимуму. Примером лабораторного эксперимента может быть экспериментальное обучение одного учащегося или небольшой группы учащихся в соответствии со специально разработанной методикой.

В ходе лабораторного эксперимента отчетливее прослеживается исследуемый процесс, обеспечивается возможность более глубоких измерений, применения комплекса специальных технических средств и аппаратуры. Однако лабораторный эксперимент упрощает педагогическую реальность тем, что он осуществляется в искусственных условиях. Именно искусственность экспериментальной ситуации является недостатком лабораторного эксперимента.

Результаты педагогического эксперимента имеют три взаимодополняющих аспекта: *объективного, преобразующего и конкретизирующего*. *Объективный аспект* отражает обоснованность результата на общенаучном или общепедагогическом уровнях. *Преобразующий аспект* раскрывает изменения, происходящие с объектом исследования, указывает на дополнения, уточнения или другие преобразования, которые могут происходить с ним.

Конкретизирующий аспект уточняет различные условия, факторы и обстоятельства, в которых происходит изменение объекта:

- конкретизация места и времени, в границах которого проявляются свойства объектов;
- указание необходимых условий для развития учащихся;
- перечень применяемых в обучении педагогических систем и технологий, способов мониторинга качества образования;
- уточнение подходов решения той или иной педагогической задачи.

Педагогический эксперимент будет эффективным, если соблюдать последовательность основных познавательных действий:

- осознание проблемной ситуации и постановка проблемы;
- поиск возможных причин проблемной ситуации;
- выработка нормы-гипотезы (способов воздействия), направленной на разрешение проблемы;

- проверка выработанной нормы-гипотезы на практике, в педагогической действительности;
- оценка результатов.

Отдельные этапы эксперимента необходимо специальным образом проектировать, ориентируясь на логику экспериментального цикла по реализации диагностической, прогностической, организационной, исполнительской, аналитической и внедренческой функций. При этом основные этапы, соответственно, могут быть названы как *подготовительный, практический, обобщающий, внедренческий*.

1.3. Организация педагогического эксперимента

Подготовительный этап довольно сложен, он выполняет следующие функции: *диагностическую, прогностическую и организационную*.

Диагностическая функция предполагает построение картины исследуемого объекта. В педагогических экспериментах это может быть:

- анализ индивидуального опыта и выявление конкретных затруднений в деятельности педагога, заместителя директора, директора школы;
- анализ состояния педагогического процесса в школе; выявление образовательных потребностей педагогического коллектива школы;
- формулирование противоречий воспитательно-образовательного процесса в учебном заведении;
- постановка и обоснование проблемы, на поиск решения которой и будет направлен эксперимент.

Реализация диагностической функции позволяет оценить состояние практики, построить конкретное “знание о незнании”, выявить узкие места, противоречия, затруднения в функционировании педагогической системы.

Дальнейшие шаги в организации эксперимента связаны с осуществлением экспериментатором *прогностической функции*. Выявленные проблемы формируют потребность их преодоления. Под них рождается или привносится идея эксперимента.

Экспериментальная идея – это наиболее общее представление о предполагаемом направлении деятельности экспериментатора в создавшейся проблемной ситуации, содержащее представление о желаемом результате, позволяющем разрешить проблему. Идея содержит мысль,

предположение о целесообразном пути организации деятельности обучаемых, о способе сочетания в ней известного и нового, о своеобразии педагогической помощи. Конкретизируется идея эксперимента, обретающая процессуальные составляющие в замысле эксперимента.

Замысел эксперимента претворяет идею в конкретные формы, связывает идею с методами ее воплощения. Замысел может содержать логическую схему разворачивания содержаний, отбор материала, выделение центральных мыслей, ведущих положений, методы, организационные формы, учет конкретных условий, в которых возникла экспериментальная идея, учет параметров, характеризующих состояние части образовательной практики, ограниченной проблемной ситуацией. *Замысел* – это конкретизация идеи через определенные шаги процессуального характера.

Идея и замысел эксперимента сопряжены для экспериментатора с выбором объекта и предмета экспериментального исследования.

Объект экспериментирования – педагогическое пространство, область, в границах которой содержится то, что будет изучаться. Функция объекта экспериментирования заключается в *пространствовании границ* воздействия. Сам объект устроен довольно сложно. Объект может быть рассмотрен как сложная совокупность свойств, связей, отношений.

Для того чтобы конкретизировать, на что именно направлено воздействие (свойства, связи, отношения), нужно выделить предмет экспериментирования. Аспект исследования, о котором будет получено новое знание, находит отражение в предмете эксперимента.

Выбор предмета экспериментирования предполагает ответы на вопросы: *как рассматривается объект? какие свойства, отношения, функции выделяются в объекте? какая реальность, какая часть объекта будет раскрываться в данном экспериментальном исследовании?* Функция предмета эксперимента – *конкретизация, удержание границ воздействия.*

Выбрав предмет эксперимента, важно поставить цель его проведения. При формулировании *цели эксперимента* следует ответить на вопрос: *что Вы хотите создать и проверить в результате эксперимента?*

Цель эксперимента выступает критерием оценки его результата. При завершении экспериментального исследования (или его этапа) результат, полученный в практической деятельности, соотносится с ожидавшимся результатом, отраженном в программе эксперимента как его цель.

Задачи эксперимента можно определить как более конкретизированные по отношению к цели эксперимента частные цели. Их можно

определить как шаги достижения цели. Задачи определяют вер пер проблем, которые нужно решить в ходе эксперимента. В задачах *сформулированы параметры* достижения промежуточных результатов. При их выделении экспериментатор должен ответить, на вопрос: *какие промежуточные результаты нужно получить, чтобы достичь общей цели и какие шаги нужно сделать при этом?*

Направляет поиск решения сформулированных задач эксперимента посредством дополнения субъективно недостающей информации о способе их решения гипотеза. *Гипотезу* можно определить как научно обоснованное логическое предположение относительно принципа реализации идеи и замысла эксперимента.

Гипотезы могут быть общие и частные, интуитивные и логически обоснованные, рабочие (временные) и научные (сформированные и научно обоснованные). Возможная формула гипотезы: “если ..., то ..., так как...”. Гипотеза выступает в функции построения системы мер по реализации задач эксперимента.

Важным этапом прогнозирования результатов эксперимента является разработка его *программы*. В ней должны содержаться *цель эксперимента, идея, замысел, гипотеза, прогноз* ожидаемых результатов и учет возможных негативных проявлений, *компенсационные механизмы, этапы* эксперимента, *критерии* оценки полученных результатов.

Критерии оценки ожидаемых результатов эксперимента – важнейший компонент, необходимый для осуществления экспериментальной деятельности. Без критериев эксперимента невозможно оценить истинность или ложность выдвинутой гипотезы, проверить эффективность и результативность разработок, достоверность результатов, валидность инструментария.

Исследователю очень важно уметь отслеживать процесс экспериментальной работы. Это может быть: проведение констатирующих (исходных), уточняющих, преобразующих срезов; фиксация текущих результатов в ходе осуществления гипотезы; проведение итоговых срезов; анализ положительных, а также отрицательных результатов, анализ непредвиденных и побочных результатов эксперимента.

При разработке критериев экспериментатор отвечает на вопрос: *с помощью каких признаков, параметров будет оцениваться эффективность экспериментальных материалов?*

После разработки программы экспериментатор должен определить необходимые условия для ее реализации, т.е. реализовать *организационную* функцию. К выполнению данной функции могут относиться воп-

росы, связанные с подготовкой материальной базы эксперимента, распределением управленческих функций, поиском научного руководителя или научного консультанта, организацией специальной подготовки кадров, методическим обеспечением эксперимента и др.

Подготовив условия для проведения эксперимента, можно переходить к *практическому этапу*, в ходе которого реализуется *исполнительская функция*. В зависимости от типа эксперимента (*констатирующий, поисково-обучающий, обучающий*) характеристики данного этапа могут существенно различаться. В частности, результатом констатирующего эксперимента может быть анализ состояния предмета исследования на начальном этапе или описание фактического положения дел.

Для проведения подобного типа эксперимента должны быть заранее определены параметры сбора объективной информации, количество испытуемых образцов, необходимое для получения достоверного результата; определена выборка, достаточная для контроля и проведения последующего анализа состояния предмета исследования. Основной характеристикой констатирующего эксперимента является регистрация выявленных фактов и установление противоречий в исследуемом *предмете*.

Выявленные и проанализированные факторы могут служить основанием для зарождения идеи и замысла эксперимента, которые затем превращаются в рабочую гипотезу. Если изучаемая исследователем область относительно неизвестна и реальная (стабильная) гипотеза отсутствует, то говорят о поисковом, или пилотажном, эксперименте. Рабочая гипотеза может выступать в функции обоснованного и временно-го предположения для систематизации имеющегося фактического материала и применяться с целью проверки каких-либо зависимостей.

Далее рабочая гипотеза должна уточняться, развиваться и превращаться в реальную гипотезу или систему гипотез – научно обоснованных предположений о способах разрешения проблемы. Исследователь выбирает *предполагаемую систему мер для достижения выбранной цели*. Выстроенная система гипотез проверяется в формирующем эксперименте, существенным признаком которого является активное воздействие экспериментатора на личность учащегося (испытуемого).

Результатом завершения эксперимента формирующего типа является проверенная система гипотез, раскрывающая факторы, закономерности, динамику, механизмы, тенденции развития исследуемого объекта.

Ведущей функцией на *обобщающем этапе* является аналитическая функция. Реализация ее на практике предусматривает получение двух продуктов деятельности исследователя:

- анализ результатов эксперимента и соотнесение их с заявленными целями, зафиксированными в программе эксперимента (включая соотнесенность этапов эксперимента, ожидаемых промежуточных результатов с данными контрольных срезов или иных средств диагностики промежуточных результатов);
- оформление и описание хода и результатов эксперимента.

Экспериментатор устанавливает *подтвержденные закономерности и эффективность применяемых дидактических средств*.

На *внедренческом этапе* выявляются условия трансляции результатов эксперимента в массовую практику. Самостоятельное значение имеет публикация и экспертиза текста с описанием хода эксперимента с его результатами. Такой текст является основным инструментом широкого внедрения.

Этапы эксперимента определяют некоторые его части и их последовательность. Части позволяют преемственным образом реализовать систему мер, заложенную в гипотезе, и фиксируют последовательность действий, осуществляемых в конкретные временные интервалы. Этапы выступают в функции фиксации промежуточных результатов, их оценивания и корректировки.

1.4. Структура программы эксперимента

Можно выделить несколько субъективных и объективных факторов, влияющих на уровень требований к программе эксперимента. К числу субъективных относится *внутреннее самоопределение экспериментатора* относительно статуса эксперимента, его типа, характера.

К числу объективных относятся социальные нормы, предъявляемые к экспериментальной деятельности. Так, уровень требований зависит от:

- *типа* эксперимента (констатирующий, поисково-обучающий, обучающий и т.д.);
- *статуса* эксперимента (экспериментальная площадка республиканского, областного уровня, городского и районного, школьного и внутришкольного, индивидуального и коллективного эксперимента);
- *характера* экспериментальной деятельности (опытно-экспериментальный, внедренческий, поисковый, собственно исследовательский);
- *значимости* эксперимента (индивидуально значимый или социально значимый);

- *сложности предмета* эксперимента (вносимые изменения затрагивают отдельные составляющие образовательного процесса – содержание образования, методику и технологию преподавания, управленческие модели, обеспечивающие базовый процесс, структуру учебного заведения нового типа или любую совокупность этих элементов);
- *масштаба* эксперимента (в исследование вовлечены отдельные учащиеся, группы учащихся, классы, школы).

При разработке и описании программы эксперимента важно учитывать все эти факторы. Для этого рассматривается, из каких частей состоит экспериментальный процесс, как устроен экспериментальный цикл.

Программа эксперимента наиболее просто может быть представлена в виде таблицы 1, содержащей нумерацию пунктов, их название и содержание, перечень вопросов требующих ответов [9]. Каждому пункту программы (за исключением некоторых) соответствует несколько вопросов. Наличие по каждому пункту нескольких вопросов не означает необходимость ответа на каждый их них.

Таблица 1

	Пункт программы	Содержание	Вопрос для ответа
1	Тема эксперимента	<i>Название эксперимента</i>	Как называется эксперимент?
2	Исполнитель эксперимента	<i>Фамилия, имя, отчество, должность, звание</i>	
3	Научный руководитель, консультант	<i>Фамилия, имя, отчество, должность, звание, место работы, телефон</i>	
4	Актуальность темы	<i>Затруднения, проблемы, противоречия практики, из которых вытекает необходимость эксперимента по данной теме</i>	Что не устраивает, в чем состоит проблемная ситуация? Что хотелось бы изменить? Почему данную проблему нужно в настоящее время изучать?
5	Идея эксперимента	<i>Наиболее общее представление о проблемной ситуации, направлении деятельности исследователя, образе желаемого изменения в ученике</i>	Какое обстоятельство вызывает у Вас потребность в действиях, направленных на изменения в ученике? Что хотите изменить в ученике, за счет чего и как?

	Пункт программы	Содержание	Вопрос для ответа
6	Замысел эксперимента	<i>Конкретизация идеи эксперимента через определенные формы, методы, логику разворачивания содержаний, учет конкретных условий</i>	Как Вам видится процесс воплощения идеи эксперимента? Как воплотить идею эксперимента на практике?
7	Объект экспериментирования	<i>Границы исследования и изменения практики</i>	Что исследуется? Каковы границы педагогического воздействия? Какова область изменения практики?
8	Предмет экспериментирования	<i>Свойства, отношения, функции, выделяемые в объекте; часть объекта, раскрываемая в данном экспериментальном исследовании</i>	О чем в объекте экспериментирования будет получено новое знание? На что в объекте экспериментирования будете воздействовать? Какие свойства части, функции выделяются в объекте?
9	Педагогическая цель	<i>Ожидаемый результат педагогической деятельности, который выражен в позитивных изменениях в ученике, появившихся благодаря экспериментальной разработке</i>	Что хотите изменить в ученике? Какие качества личности хотите воспитать в Вашем ученике благодаря экспериментальным действиям, какие способности хотите развить? Какие изменения в обученности ученика предполагаете получить за счет экспериментальных действий?
10	Цель эксперимента	<i>Ожидаемый результат, представляемый в форме методических рекомендаций, разработок уроков, планов, авторских программ, концептуальных положений, педагогических технологий</i>	Что хотите разработать и апробировать? Что хотите создать в результате эксперимента? Какой текстовой результат намерены получить в эксперименте?
11	Задачи эксперимента	<i>Действия по достижению промежуточных результатов, направленных на достижение цели</i>	Какие действия необходимо совершить для того, чтобы достичь цели эксперимента? Какие промежуточные результаты необходимы для достижения цели?

	Пункт программы	Содержание	Вопрос для ответа
12	Гипотеза	Научно обоснованное логическое предположение относительно способа реализации идеи и замысла эксперимента, совокупность воздействий, система мер по реализации задач эксперимента, максимально подробно изложенная модель нововведения, благодаря которому ожидается получить определенную эффективность учебно-воспитательного процесса	Что будете проверять? В чем состоит Ваше предположение о том, как возможно реализовать идею и замысел эксперимента? Какие управляющие воздействия предполагаете по реализации задач эксперимента?
13	Диагностический инструментарий	Средства оценивания результатов эксперимента: анкеты, интервью, тексты контрольных работ, тесты, экспериментальные дидактические материалы	С помощью чего будет осуществляться контроль за результатами?
14	Критерии оценки ожидаемых результатов	Признаки или параметры, на основании которых производится оценка эффективности экспериментальной разработки	С помощью каких признаков, параметров будет оцениваться эффективность экспериментальных материалов: программ, методик, принципов и др.?
15	Сроки эксперимента	Время начала и предполагаемого завершения эксперимента	Какова продолжительность эксперимента?
16	Этапы эксперимента	Части, определяющие промежуточные результаты эксперимента и последовательность их достижения	Какие промежуточные результаты и в какой последовательности предполагаются для достижения цели?
17	Прогноз возможных негативных последствий	Замедление темпов прохождения тем, отклонения от базового компонента содержания образования, изменение образовательных целей	Какие возможны отрицательные последствия?

	Пункт программы	Содержание	Вопрос для ответа
18	Способы коррекции, компенсации негативных последствий	<i>Управляющие воздействия со стороны экспериментатора, устраняющие негативные последствия</i>	Какие конкретные действия могут компенсировать отрицательные последствия эксперимента?
19	Состав участников эксперимента	<i>Педагогический коллектив, участвующий в эксперименте по данной теме</i>	Кто участвует в эксперименте?
20	Функциональные обязанности	<i>Распределение функциональных обязанностей всех лиц, участвующих в эксперименте</i>	Кто и за что отвечает в эксперименте?
21	База эксперимента	<i>Группа учащихся, класс, параллель классов, вся школа</i>	На каком контингенте проводится эксперимент?
22	Масштаб эксперимента	<i>Продолжительность эксперимента по времени и охват по объему материала: один урок, несколько уроков темы, четверть, год, несколько лет</i>	Каков объем учебного материала, вводимого в поле эксперимента, и какова его продолжительность?
23	Тип эксперимента	<i>Констатирующий, поисковый, формирующий</i>	Какой тип эксперимента Вы осуществляете?
24	Статус эксперимента	<i>Правовое положение эксперимента, его состояние: внутришкольный, индивидуальный или коллективный; районного, областного или республиканского уровня</i>	Каковы индивидуальные притязания экспериментатора на статус эксперимента?
25	Форма представления результатов эксперимента для массовой практики	<i>Статья, отчет, методические рекомендации, программа и пр.</i>	В какой форме будут описаны итоги эксперимента?
26	Научно-методическая обеспеченность эксперимента	<i>Перечень экспериментальных материалов для педагогов, управленцев, обучаемых</i>	Какова обеспеченность эксперимента научно-методическими разработками?

1.5. Мониторинг педагогического эксперимента

Мониторинг в теории социального управления рассматривается как одно из важнейших, относительно самостоятельных звеньев в управленческом цикле. *Образовательный мониторинг* представляет собой

систему организации сбора, хранения обработки и распространения информации о функционировании педагогической системы.

Педагогическая система представляет собой определенную совокупность взаимосвязанных средств, методов и процессов, необходимых для создания организованного, целенаправленного и преднамеренного педагогического влияния на формирование личности учащихся с заданными качествами. В рамках мониторинга проводится выявление и оценивание проведенных педагогических действий.

В педагогике мониторинг часто определяют как постоянное наблюдение за педагогическим процессом с целью выявления его соответствия желаемому результату.

Средствами, с помощью которых осуществляется диагностика сформированности взглядов, мнений и качеств личности испытуемых, являются анкеты и тесты.

Анкетой принято называть перечень вопросов по определенной теме, на которые должны дать ответы учащиеся. С помощью анкеты можно выявить нравственные представления и понятия учащихся, их склонности, связи, оценочные суждения к окружающим людям, труду, профессии, показатели общественного мнения и т.д. В последнее время анкетирование получило в педагогической практике очень широкое распространение, настолько широкое, что некоторые учителя стали смотреть на него как на универсальный метод получения информации.

Анкетирование должно быть связано с острыми, “болевыми” проблемами, связанными с совершенствованием учебно-воспитательной работы и мотивировано потребностями жизни коллектива. *Принцип педагогической целесообразности здесь особенно важен.* Анкетирование поэтому не должно являться неожиданностью для учащихся и тем более для учителя; его следует планировать заранее.

Слабой стороной анкет является их стандартный характер, отсутствие живого контакта исследователя с опрашиваемыми, что не всегда обеспечивает достаточно полные и откровенные ответы. Не надо и переоценивать значение анкетирования и его результатов. Неверно думать, что с его помощью можно узнать все. С помощью анкетирования можно получить лишь эмпирический материал, в который необходимо внести коррективы и делать выводы.

Структура и характер анкет определяются содержанием и формой вопросов, которые задаются опрашиваемым. Поэтому главной трудностью в построении анкеты является подбор и формулировка вопросов, проверка их уместности, надежности.

В зависимости от предлагаемых вопросов они могут быть разными (по содержанию, направленности личности, форме и т.д.). По содержанию анкета может касаться выявления познавательных интересов, перспектив и оценки деятельности коллектива, тех или иных событий, качеств людей, степени общественной активности и ее мотивации.

По направленности личности анкеты могут касаться отношения к науке, социальной ответственности, нравственности и т.д.

По форме анкеты могут быть: открытыми, позволяющими учащимся отвечать по своему усмотрению, убеждению, взгляду, ценностным ориентациям. Они позволяют узнать личное мнение воспитанника. Отвечающий сам выбирает объем ответа и содержание информации, которую он считает необходимым дать.

В *закрытых* вопросах возможности выбора ограничиваются заранее определенным числом вариантов ответов, предусмотренным составителем анкеты. Закрытые вопросы в отличие от открытых легче подвергать статистической обработке. Однако уместно напомнить, что в них не предусмотрены некоторые маловероятные варианты ответов.

Если учащийся не найдет в списке тот ответ, который он намеревается избрать, то он не сможет ответить на вопрос анкеты. В этом случае инструкция должна разрешать учащемуся записать свой ответ.

По организации анкеты могут быть именные (с указанием Ф.И.О. отвечающего на вопросы) и *анонимные* (безымянные), когда анкета не подписывается. Эффективность анкетирования во многом зависит от соблюдения ряда психолого-педагогических требований:

- перед тем как проводить анкетирование, следует провести разъяснительную работу, рассказать учащимся, для чего оно проводится, с какой целью, учащиеся должны осознавать его целесообразность;
- очень важно правильно формулировать вопросы; они должны быть сформулированы ясно, конкретно и доступно; это предполагает в свою очередь ясные и конкретные ответы;
- анкета не должна быть громоздкой;
- анкетирование требует уважения к мнению отвечающего, нельзя навязывать собственное мнение опрашиваемым, необходимо сохранять анонимность;
- эффективное использование этого метода требует специальных умений (дидактических, коммуникативных, организационных, диагностических и др.);
- по результатам анкетирования надо вносить необходимые коррективы в программу исследования, в действия наставников.

Под тестами чаще всего подразумевают набор вопросов и заданий, из ответов на которые получают информацию об уровне сформированности интеллектуальных умений учащихся, освоения ими некоторого учебного материала.

Тесты достижений как инструмент оценивания имеют значительные отличия от контрольных работ [5]:

1) тесты – значительно более качественный и объективный способ оценивания;

2) показатели тестов ориентированы на измерение степени, определение уровня усвоения знаний тем учебных предметов, умений, навыков, а не на констатацию наличия у учащихся определенной совокупности формально усвоенных знаний.

Используемая в тестах достижений стандартизированная форма оценки позволяет соотнести уровень достижений учащегося по предмету в целом и по его отдельным разделам со средним уровнем достижений учащихся в классе и уровнем достижений каждого из них.

В текстологии (науке о тестах) рассматриваются различные виды тестов. Остановимся более подробно на составлении тестовых заданий разных видов.

Задания открытого типа. К ним относятся задания двух видов:

- свободного изложения или конструирования. На ответы учащихся не накладываются ограничения. Формулировки заданий должны обеспечить наличие только одного правильного ответа;

- дополнения. Вместо многоточий учащиеся записывают слова, символ, знак и т. д.

Инструкция для заданий свободного изложения может быть такой: закончите предложение, допишите определение и т.д. К заданиям-дополнениям инструкция может выглядеть так: *вместо многоточия впишите нужное слово, нужный символ* и т.п.

Положительными сторонами хорошо составленных заданий-дополнений и свободного изложения являются:

1. Краткость и однозначность ответов.
2. Необходимость воспроизведения ответа по памяти.
3. Отсутствие необходимости искать несколько вариантов ответа.
4. Простота формулировки вопросов.
5. Простота проверки.

При составлении заданий открытого типа имеет смысл учесть следующие рекомендации:

- использовать не более трех пропусков подряд, лучше 1-2;

- дополнять нужно самое важное, то, запоминание чего нужно проверить;
- дополнения лучше ставить в конце предложения;
- вопрос должен быть сформулирован четко;
- ответ на поставленный вопрос должен быть однозначным;
- чаще следует пользоваться количественными характеристиками.

Различают следующие типы тестовых заданий закрытого типа:

1. *Задания с альтернативным выбором ответа.* Испытуемый выбирает один из двух ответов: *да* или *нет*. Инструкция для заданий с альтернативным выбором ответов может быть такой: *отметьте любым знаком правильный вариант ответа, отметьте знаком X номера только тех вопросов, на которые вы даёте утвердительные ответы.*

Использование заданий альтернативных ответов в виде отдельного вопроса приводит, как правило, к тривиальному тестированию и применяется достаточно редко. Наиболее эффективно использование заданий этого вида в серии, когда для одного элемента знаний (определений, графиков, диаграмм) задается несколько вопросов.

Большим недостатком этих заданий является высокая вероятность угадывания ответов. Преодолеть этот недостаток можно с помощью увеличения размеров теста или увеличения количества заданий в серии.

Для получения объективных результатов также следует соблюдать ряд правил:

- вопрос должен содержать одну законченную мысль;
- в вопросе следует избегать слов, дающих возможность учащимся догадаться о правильном ответе;
- избегать вводных фраз или предложений, мало связанных с основной мыслью;
- не следует прибегать к пространным выражениям, т.к. они могут явиться явной подсказкой к выбору ответа;
- следует избегать модализованных вопросов (типа *Не считаете ли Вы, что...*);
- число ответов “да” и “нет” должны быть приблизительно равным, что исключает тенденцию отвечать одинаково;
- необходимо избегать двусмысленных утверждений.

2. *Задания с множественным выбором ответа,* состоят из двух частей – формулировки задания и вариантов ответов. Испытуемый должен выбрать один из предложенных вариантов, среди которых чаще всего только один правильный ответ. Ответы подбираются и формиру-

ются так, чтобы из них не менее трех были правдоподобными (похожими на правильные).

Инструкции для данного вида заданий могут быть сформулированы так:

– обведите кружком в бланке ответов букву, соответствующую варианту ответа;

– отметьте любым знаком номер варианта, который Вы считаете правильным;

– из предложенных вариантов выберите правильный и подчеркните его;

– выберите правильные ответы из предложенных вариантов и подчеркните их.

Для правильного составления заданий с множественным выбором важно соблюдать следующие требования:

- все варианты ответов должны быть грамматически согласованы с основной частью задания;
- вопросы должны содержать только одну мысль или утверждение;
- неправильные ответы должны быть разумны, умело подобраны, не должно быть явных неточностей;
- реже использовать в основной части отрицание;
- вопрос не должен содержать лишних деталей;
- место правильного ответа должно быть выбрано таким образом, чтобы оно не повторялось от вопроса к вопросу, без закономерностей в случайном порядке;
- правильные и неправильные ответы должны быть однозначны по содержанию и структуре;
- если ответы на вопрос носят количественный характер, то числа в ответах располагаются или в порядке возрастания, или в порядке убывания;
- лучше использовать длинный вопрос и короткий ответ, чем наоборот;
- среди ответов на вопрос может быть несколько правильных ответов.

3. *Задания на восстановление соответствия.* Это задания, где необходимо восстановить соответствия между элементами двух списков. Имеется достаточно много модификаций этих заданий.

Часто употребляемая форма установления соответствия между элементами двух списков – *рисование стрелочек*. Но эта форма обладает двумя существенными недостатками: сложность проверки, ученики трудно переключаются на другие формы.

Классической формой записи ответов является *запись сочетаний цифр и букв*, под которыми значатся элементы списков. В инструкции оговаривается форма установления соответствия.

При конструировании заданий на восстановление соответствия необходимо учитывать следующие рекомендации:

- число входных данных одного списка не должно превышать 10;
- если длина списков не совпадает, то об этом следует упомянуть в инструкции;
- все ответы по конструкции должны быть по возможности однородны.

Наряду с этими видами заданий в дидактических материалах встречаются и другие тестовые задания.

4. *Задания на преобразование.* Эти задания требуют от учащихся анализа имеющихся данных и изменения их последовательности в соответствии с поставленным условием.

5. *Задания на нахождение ошибок.* В материалах задания специально допущены ошибки. Учащиеся анализируют схемы, планы, высказывания и находят имеющиеся ошибки и неточности.

При построении моделей тестов необходимым является решение ряда педагогических задач, среди которых:

- определение содержания и параметров диагностики;
- выбор методов конструирования тестов;
- выделение способов определения измерительных качеств тестов;
- выбор измерительной шкалы и методов обработки полученных данных.

§ 2. МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПЕДАГОГИЧЕСКОМ ЭКСПЕРИМЕНТЕ [7]

2.1. Измерения и шкалы

В ходе педагогического эксперимента проводится измерение проявления исследуемых свойств у учащихся. Но невозможно измерить некоторое свойство у всех испытуемых (*генеральной совокупности*). Поэтому при проведении эксперимента ограничиваются лишь относительно небольшой группой представителей соответствующей совокупности учащихся (*выборке*).

Выборка должна быть такой, чтобы была обоснована генерализация выводов выборочного исследования – обобщение, распространение

ние их на генеральную совокупность. *Основные критерии обоснованности выводов эксперимента – это репрезентативность выборки и статистическая достоверность (эмпирических) результатов. Репрезентативность выборки – способность выборки представлять изучаемые явления достаточно полно с точки зрения их изменчивости в генеральной совокупности.*

Существуют приемы, позволяющие получить достаточную для исследователя репрезентативность выборки:

– *случайный отбор*. Он предполагает обеспечение таких условий, чтобы каждый элемент генеральной совокупности имел равные с другими шансы попасть в выборку. Случайный отбор обеспечивает возможность попадания в выборку самых разных представителей генеральной совокупности. При этом принимаются специальные меры, исключая появление какой-либо закономерности при отборе. И это позволяет надеяться на то, что в конечном счете в выборке изучаемое свойство будет представлено если и не во всем, то в максимально возможном его многообразии;

– *стратифицированный случайный отбор*, или отбор по свойствам генеральной совокупности. Он предполагает предварительное определение тех качеств, которые могут влиять на изменчивость изучаемого свойства. Затем определяется процентное соотношение численности различающихся по этим качествам групп (страт) в генеральной совокупности и обеспечивается идентичное процентное соотношение соответствующих групп в выборке. Далее в каждую подгруппу выборки испытуемые подбираются по принципу простого случайного отбора.

Статистическая достоверность, или статистическая значимость, результатов исследования определяется при помощи методов статистического вывода. Они предъявляют определенные требования к численности, или *объему выборки*. Наибольший объем выборки необходим при разработке диагностической методики – от 200 до 1000–2500 человек.

Если необходимо сравнивать 2 выборки, их общая численность должна быть не менее 50 человек; численность сравниваемых выборок должна быть приблизительно одинаковой.

Если изучается взаимосвязь между какими-либо свойствами, то объем выборки должен быть не меньше 30–35 человек. Чем больше *изменчивость* изучаемого свойства, тем больше должен быть объем выборки. Поэтому изменчивость можно уменьшить, увеличивая однородность выборки. При этом, естественно, уменьшаются возможности генерализации выводов.

Различают *зависимые* и *независимые* выборки. *Независимые* выборки характеризуются тем, что вероятность отбора любого испытуемого одной выборки не зависит от отбора любого из испытуемых другой выборки. Напротив, *зависимые* выборки характеризуются тем, что каждому испытуемому одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки.

В общем случае *зависимые* выборки предполагают попарный подбор испытуемых в сравниваемые выборки, а *независимые* выборки – *независимый* отбор испытуемых. Случаи “частично зависимых” (или “частично независимых”) выборок недопустимы: это непредсказуемым образом нарушает их репрезентативность.

Можно выделить две методологии в проведении педагогического (психологического) эксперимента:

- *R-методология* предполагает изучение изменчивости некоторого свойства испытуемого под влиянием некоторого воздействия, фактора либо другого свойства.
- *Q-методология* предполагает исследование изменчивости субъекта (единичного) под влиянием различных стимулов (условий, ситуаций и т.д.). Ей соответствует ситуация, когда *выборкой является множество стимулов*.

Педагогический эксперимент начинается с того, что исследователь фиксирует выраженность интересующего его свойства (или свойств) у объекта или объектов исследования, как правило, при помощи чисел. Таким образом, следует различать *объекты исследования* (в педагогике и психологии это чаще всего люди, испытуемые), их *свойства* (то, что интересует исследователя, составляет предмет изучения) и *признаки*, отражающие в числовой шкале выраженность свойств.

Под измерением понимают присвоение объекту числа по определенному правилу. Это правило устанавливает соответствие между измеряемым свойством объекта и результатом измерения – признаком. Если нет необходимости в измерении, то ограничиваются сравнительными суждениями.

Точность, с которой признак отражает измеряемое свойство, зависит от процедуры (операции) измерения. Всякое измерение имеет смысл при выборе измерительной шкалы. Выделяют следующие виды измерительных шкал:

- *Номинативная шкала (неметрическая)*, или шкала наименований (номинальное измерение). В ее основе лежит процедура, обычно не ассоциируемая с измерением. Пользуясь определенным пра-

вилом, объекты *группируются по различным классам так, чтобы внутри класса они были идентичны по измеряемому свойству*. Каждому классу дается наименование и обозначение, обычно числовое. Затем каждому объекту присваивается соответствующее обозначение.

- *Ранговая, или порядковая шкала (неметрическая)* (как результат ранжирования). Как следует из названия, измерение в этой шкале предполагает приписывание объектам чисел в зависимости от степени выраженности измеряемого свойства.

Существует множество способов получения измерения в порядковой шкале. Но суть остается общей: при сравнении испытуемых друг с другом мы можем сказать, больше или меньше выражено свойство, но не можем сказать, насколько больше или насколько меньше оно выражено. При измерении в ранговой шкале, таким образом, из всех свойств чисел учитывается то, что они разные, и то, что одно число больше, чем другое.

При ранжировании “вручную”, а не при помощи компьютера, следует иметь в виду два обстоятельства:

– Ранг 1 присваивается тому испытуемому, у которого наименьшая выраженность признака (свойства). Увеличение ранга проводится по увеличению уровня признака. Присвоение ранга можно проводить и в обратном порядке.

– В случае совпадения рангов у нескольких учащихся им присваивается ранг, равный среднему арифметическому их рангов. Сумма всех присваиваемых рангов для группы численностью N должна вне зависимости от наличия связей в рангах равняться $N(N - 1)/2$.

- *Интервальная шкала (метрическая)*. Применяется при измерении, при котором числа отражают не только различия между объектами в уровне выраженности свойства (характеристика порядковой шкалы), но и то, насколько больше или меньше выражено свойство. Равным разностям между числами в этой шкале соответствуют равные разности в уровне выраженности измеренного свойства. Измерение в этой шкале предполагает возможность применения *единицы измерения (метрики)*.

Объекту присваивается число единиц измерения, пропорциональное выраженности измеряемого свойства. Важная особенность интервальной шкалы – произвольность выбора нулевой точки: ноль вовсе не соответствует полному отсутствию измеряемого свойства. Произвольность выбора нулевой точки отсчета обозначает, что измерение в этой

шкале не соответствует *абсолютному* количеству измеряемого свойства. Следовательно, применяя эту шкалу, мы можем судить, насколько больше или насколько меньше выражено свойство при сравнении объектов, но не можем судить о том, во сколько раз больше или меньше выражено свойство.

- *Абсолютная шкала, или шкала отношений (метрическая)*. Измерение в этой шкале отличается от интервального только тем, что в ней устанавливается нулевая точка, соответствующая полному отсутствию выраженности измеряемого свойства.

Определение того, в какой шкале измерено явление (представлен признак), – ключевой момент анализа данных: любой последующий шаг, выбор любого метода зависит именно от этого.

2.2. Таблицы и графики

Результаты измерения для дальнейшего анализа чаще всего представляют в виде *таблицы исходных данных*. Каждая строка такой таблицы обычно соответствует одному *объекту*, а каждый столбец – одному измеренному *признаку*. Таким образом, исходной формой представления данных является таблица типа “объект – признак”. Каждый признак выступает в качестве переменной величины, или просто *переменной*, значения которой меняются от объекта к объекту.

Как правило, анализ данных начинается с изучения того, как часто встречаются те или иные значения интересующего исследователя признака (переменной) в имеющемся множестве наблюдений. Для этого строятся *таблицы и графики распределения частот*. Нередко они являются основой для получения ценных содержательных выводов исследования.

Если признак принимает всего лишь несколько возможных значений (до 10–15), то таблица распределения частот показывает частоту встречаемости каждого значения признака. Если указывается, сколько раз встречается каждое значение признака, то это – таблица *абсолютных частот* распределения, если указывается доля наблюдений, приходящихся на то или иное значение признака, то говорят об *относительных частотах* распределения.

Во многих случаях признак может принимать множество различных значений, например, если мы измеряем время решения тестовой задачи. В этом случае о распределении признака позволяет судить *таблица сгруппированных частот*, в которых частоты группируются по рядам или интервалам значений признака.

Пример. Предположим, в группе испытуемых численностью 40 человек измерено время решения тестовой задачи. Максимальное время составило 67 секунд, минимальное – 32 секунды. Построение таблицы (таблица 2) распределения частот (сгруппированных и накопленных) в этом случае производится поэтапно:

1. Определение размаха: $67-32=35$.
2. Выбор желаемого числа разрядов и интервала разрядов (определяется произвольно). Обычное число разрядов – от 6 до 15. Удобным интервалом разрядов в рассматриваемом случае может быть 5. 35 делим на 5, получаем число разрядов – 7. Учитывая, что начинать лучше с 30 или с 31 и заканчивать на 69 или 70, уточняем размах ($70-30=40$) и число разрядов ($40/5=8$).
3. Определение границ разрядов. Если начинать с 30, то первый разряд будет с 30 до 34, второй – с 35 до 39 и т. д., до восьмого – с 65 до 69. Границы соседних разрядов не должны совпадать.
4. Расчет частот *встречаемости значений* признака для каждого интервала.
5. Расчет *накопленных частот*. Они показывают, как накапливаются частоты по мере возрастания значений признака.

Таблица 2

Интервал времени, с	Абсолютная частота (f_n)	Относительная частота (f_n)	Накопленная частота (f_n)
30-34	1	0,025	0,025
35-39	2	0,050	0,075
40-44	5	0,125	0,200
45-49	8	0,200	0,400
50-54	10	0,250	0,650
55-59	8	0,200	0,850
60-64	4	0,100	0,950
65-69	2	0,050	1,000
Σ (сумма)	40	1,000	

Для более наглядного представления результатов измерения строится график распределения частот или график накопленных частот – гистограмма или сглаженная кривая распределения (рис. 1). *Гистограмма распределения частот* – это столбиковая диаграмма, каждый столбец которой опирается на конкретное значение признака или разрядный интервал (для сгруппированных частот). Высота столбика пропорциональна частоте встречаемости соответствующего значения.

Гистограмма накопленных частот отличается от гистограммы распределения тем, что высота каждого столбика пропорциональна частоте, накопленной к данному значению (интервалу).

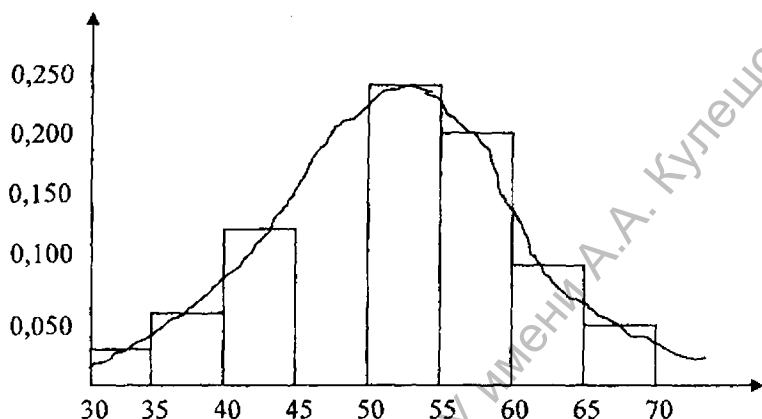


Рис. 1

Часто применяется построение *полигона распределения частот*. В гистограмме вершина каждого столбца, соответствующая частоте встречаемости данного значения (интервала) признака, – отрезок прямой. А для полигона отмечается точка, соответствующая середине этого отрезка. Далее все точки соединяются ломаной линией. Вместо гистограммы или полигона часто изображают *сглаженную кривую распределения частот* (рис. 1).

Таблицы и графики распределения частот дают важную предварительную информацию о *форме распределения признака*: о том, какие значения встречаются реже, а какие чаще, насколько выражена изменчивость признака. Обычно выделяют равномерные, симметричные, нормальные и ассиметричные типичные формы распределения.

Равномерное распределение – форма распределения, при которой все значения встречаются одинаково (или почти одинаково) часто. При *симметричном распределении* одинаково часто встречаются крайние значения результатов измерения.

Нормальное распределение – симметричное распределение, у которого крайние значения встречаются редко и частота постепенно повышается от крайних к серединным значениям признака. *Ассиметричные распределения* – *левосторонние* (с преобладанием частот малых значений), *правосторонние* (с преобладанием частот больших значений).

При проведении педагогического эксперимента применяют заполнение таблиц сопряженности номинативных признаков. *Таблицы сопряженности*, или кросстабуляции, – это таблицы совместного распределения частот двух и более номинативных признаков, измеренных на одной группе объектов. Эти таблицы позволяют сопоставить два или более распределения. Столбцы такой таблицы соответствуют категориям (градациям) одного номинативного признака, а строки – категориям (градациям) другого номинативного признака. Конечно, таблицы сопряженности могут включать номинативные признаки, имеющие и более двух градаций.

2.3. Первичные описательные статистики и меры изменчивости

К первичным описательным статистикам обычно относят числовые характеристики распределения измеренного на выборке признака. Каждая такая характеристика отражает *в одном числовом значении* свойство распределения *множества результатов измерения*: с точки зрения их *расположения* на числовой оси либо с точки зрения их *изменчивости*. Основное назначение каждой из первичных описательных статистик – замена множества значений признака, измеренного на выборке, одним числом (например, средним значением как мерой центральной тенденции). Компактное описание группы при помощи первичных статистик позволяет интерпретировать результаты измерений, в частности, путем сравнения первичных статистик разных групп.

Мера центральной тенденции – это число, характеризующее выборку по уровню выраженности измеренного признака. Существуют три способа определения “центральной тенденции”, каждому из которых соответствует своя мера: мода, медиана и выборочное среднее.

Мода – это такое значение из множества измерений, которое встречается наиболее часто. Моде, или *модальному интервалу* признака, соответствует наибольший подъем (вершина) графика распределения частот. Если график распределения частот имеет одну вершину, *то* такое распределение называется *унимодальным*.

Распределение может иметь и не одну моду. Когда все значения встречаются одинаково часто, принято считать, что такое распределение не имеет моды. *Бимодальное распределение* имеет на графике распределения две вершины, даже если частоты для двух вершин не строго равны. В последнем случае выделяют большую и меньшую моду. Во

всей группе может быть и несколько локальных вершин распределения частот. Тогда выделяют *наибольшую моду* и *локальные моды*.

Медиана (Md) – это такое значение признака, которое делит упорядоченное (ранжированное) множество данных пополам так, что одна половина всех значений оказывается меньше медианы, а другая – больше.

Первым шагом при определении медианы является упорядочивание (ранжирование) всех значений по возрастанию или убыванию. Далее медиана определяется следующим образом:

- если данные содержат нечетное число значений, то медиана есть центральное значение. Пусть 3, 8, 10, 12, 14 – результаты измерений, тогда $Md = 10$;
- если данные содержат четное число значений (5, 8, 9, 11), то медиана есть точка, лежащая посередине между двумя центральными значениями, т.е. $Md = (8+9)/2 = 8,5$.

Среднее (M_x , выборочное среднее, среднее арифметическое) – определяется как сумма всех значений измеренного признака, деленная на количество суммированных значений.

Каждая мера центральной тенденции обладает характеристиками, которые делают ее ценной в определенных условиях. Для *номинативных* данных, разумеется, единственной подходящей мерой центральной тенденции является мода, или *модальная категория*, – *та градация номинативной переменной, которая встречается наиболее часто*.

Наиболее очевидной и часто используемой мерой центральной тенденции является среднее значение. Но его использование ограничивается тем, что *на величину среднего влияет каждое отдельное значение*. Если какое-нибудь значение в группе увеличится на c , то среднее увеличится на c/N . Таким образом, среднее значение весьма чувствительно к “выбросам” – экстремально малым или большим значениям переменной.

На величину моды и медианы величина каждого отдельного значения не влияет. Например, если в группе из 20 измерений переменной наибольшее значение утроится по величине, то не изменится ни мода, ни медиана. Величина среднего при этом заметно изменится.

Меры центральной тенденции чаще всего используются для сравнения групп по уровню выраженности признака. Если исследователь при этом сомневается, какую меру использовать, то следует помнить, что выборочные средние можно сравнивать, если выполняются следующие условия:

- группы достаточно большие, чтобы судить о форме распределения;
- распределения симметричны;
- отсутствуют “выбросы”.

Если хотя бы одно из перечисленных условий не выполняется, то следует ограничиться модой и медианой. Альтернативой является «сквозное» ранжирование представителей сравниваемых групп и сравнение средних, вычисленных для рангов этих групп.

Помимо мер центральной тенденции в педагогике широко используются меры положения, которые называются *квантилями* распределения. *Квантиль* – это точка на числовой оси измеренного признака, которая делит всю совокупность упорядоченных измерений на две группы с известным соотношением их численности. Одним из квантилей является медиана. Это значение признака, которое делит всю совокупность измерений на две группы с равной численностью.

Кроме медианы часто используются процентиля и квартили. *Процентили* – это 99 точек-значений признака, которые делят упорядоченное (по возрастанию) множество наблюдений на 100 частей, равных по численности. Определение конкретного значения процентиля аналогично определению медианы. Например, при определении 10-го процентиля P_{10} сначала все значения признака упорядочиваются по возрастанию. Затем отсчитывается 10% испытуемых, имеющих наименьшую выраженность признака. P_{10} будет соответствовать тому значению признака, который отделяет эти 10% испытуемых от остальных 90%.

Квартили – это 3 точки – значения признака, которые делят упорядоченное (по возрастанию) множество наблюдений на 4 равные по численности части. Первый квартиль соответствует 25-му процентилю, второй – 50-му процентилю или медиане, третий к – 75-му процентилю.

Процентили и квартили используются для определения частоты встречаемости тех или иных значений (или интервалов) измеренного признака или для выделения подгрупп и отдельных испытуемых, наиболее типичных или нетипичных для данного множества наблюдений.

Меры центральной тенденции отражают уровень выраженности измеренного признака. Однако не менее важной характеристикой является выраженность индивидуальных различий испытуемых по измеренному признаку. *Меры изменчивости* применяются в педагогике и психологии для численного выражения величины межиндивидуальной вариации признака.

Наиболее простой и очевидной мерой изменчивости является размах, указывающий на диапазон изменчивости значений. *Размах* – это просто разность максимального и минимального значений. Ясно, что это очень неустойчивая мера изменчивости, на которую влияют любые возможные “выбросы”. Более устойчивыми являются разновидности размаха: *размах от 10-го до 90-го перцентиля* или *междуквартильный размах*. Последние две меры изменчивости находят свое применение для описания вариации в порядковых данных. А для метрических данных используется дисперсия – величина, название которой в науке является синонимом изменчивости.

Дисперсия – мера изменчивости для метрических данных, пропорциональная сумме квадратов отклонений измеренных значений от их арифметического среднего. Величина дисперсии получается при ус-

реднении всех квадратов отклонений:
$$\bar{D}_x = \frac{\sum_{i=1}^N (x_i - M_x)^2}{N}.$$

Следует отличать *теоретическую* (генеральную) дисперсию – меру изменчивости бесконечного числа измерений (в генеральной совокупности, популяции в целом) и *эмпирическую*, или *выборочную*, дисперсию (для реально измеренного множества значений признака). Выборочное значение в статистике используется для оценки дисперсии в генеральной совокупности. Выше указана формула для генеральной (теоретической) дисперсии (\bar{D}_x), которая, понятно, не вычисляется. Для вычислений используется формула выборочной (эмпирической) диспер-

сии, отличающаяся знаменателем:
$$D_x = \frac{\sum_{i=1}^N (x_i - M_x)^2}{N - 1}.$$

Стандартное отклонение (σ , среднеквадратическое отклонение) – положительное значение квадратного корня из дисперсии: $\sigma_x = \sqrt{D_x}$. На практике чаще используется именно стандартное отклонение, а не дисперсия. Это связано с тем, что сигма выражает изменчивость в исходных единицах измерения признака, а дисперсия – в квадратах исходных единиц.

Введение среднего арифметического и стандартного отклонения позволяет проводить стандартизацию результатов измерения. *Стандартизация*, или *z-преобразование* данных, – это перевод измерений в стандартную *z-шкалу* со средним $M_z = 0$ и D_z (или σ_z) = 1. Сначала для пере-

менной, измеренной на выборке, вычисляют среднее M_x и стандартное отклонение σ_x . Затем все значения переменной x_i пересчитываются по

$$\text{формуле } z_i = \frac{x_i - M_x}{\sigma_x}.$$

В результате преобразованные значения (z-значения) непосредственно выражаются в единицах стандартного отклонения от среднего. Если для одной выборки несколько признаков переведены в z-значения, появляется возможность сравнения уровня выраженности разных признаков у того или иного испытуемого. Для того чтобы избавиться от неизбежных отрицательных и дробных значений, можно перейти к любой другой известной шкале: IQ (среднее 100, сигма 15); Т-оценок (среднее 50, сигма 10); 10-балльной – стенов (среднее 5,5, сигма 2) и др. Перевод в новую шкалу осуществляется путем умножения каждого z-значения на заданную сигму и прибавления среднего: $S_i = \sigma_z z_i + M_z$.

2.4. Нормальный закон распределения и его применение

Начиная со второй половины XIX ст., измерительные и вычислительные методы в педагогике и психологии разрабатываются на основе следующего принципа: *если индивидуальная изменчивость некоторого свойства есть следствие действия множества причин, то распределение частот для всего многообразия проявлений этого свойства в генеральной совокупности соответствует кривой нормального распределения.* Это и есть закон нормального распределения.

Оно характеризуется своими параметрами: средним (M) и стандартным отклонением (σ). Только эти два значения отличают друг от друга бесконечное множество нормальных кривых, одинаковой формы, заданных уравнением

$$f(x_i) = \frac{1}{2\sigma\sqrt{\pi}} e^{-\frac{(x_i - M_x)^2}{2\sigma^2}}.$$

Среднее задает положение кривой на числовой оси и выступает как некоторая исходная, *нормативная величина измерения*. Стандартное отклонение задает ширину этой кривой, зависит от единиц измерения и выступает как *масштаб измерения* (рис. 2). 1-е распределение отличается от 2-го стандартным отклонением ($\sigma_1 < \sigma_2$), 2-е от 3-го – средним арифметическим ($M_1 < M_2$).

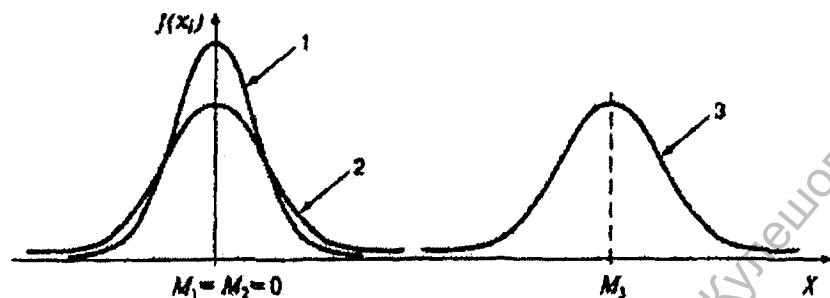


Рис. 2

Асимметрия – степень отклонения графика распределения частот от симметричного вида относительно среднего значения. Если исходные данные переведены в z-значения, показатель асимметрии вычисляется по формуле

$$As = \frac{\sum_{i=1}^N z_i^3}{N}$$

Для симметричного распределения асимметрия равна 0. Если чаще встречаются значения меньше среднего, то говорят о левосторонней, или *положительной*, асимметрии ($As > 0$). Если же чаще встречаются значения больше среднего, то асимметрия – правосторонняя, или *отрицательная* ($As < 0$). Чем больше отклонение от нуля, тем больше асимметрия.

Экссесс – мера плосковершинности или остроконечности графика распределения измеренного признака. Если исходные данные переведены в z-значения, показатель эксцесса определяется формулой:

$$Ex = \frac{\sum_{i=1}^N z_i^4}{N} - 3$$

Острове́ршинное распределение характеризуется положительным эксцессом ($Ex > 0$), а плосковершинное – отрицательным ($-3 < Ex < 0$). “Средневершинное” (нормальное) распределение имеет нулевой эксцесс ($Ex = 0$).

Все многообразие нормальных распределений может быть сведено к одной кривой, если применить z -преобразование по формуле

$$z_i = \frac{x_i - M_x}{\sigma_x}$$

ко всем возможным измерениям свойств. Тогда каждое

свойство будет иметь среднее 0 и стандартное отклонение 1. На рис. 3 построен график нормального распределения для $M=0$ и $\sigma=1$. Это и есть *нормальное единичное распределение*, которое используется как стандарт-эталон. Рассмотрим его важные свойства:

- единицей измерения нормального единичного распределения является стандартное отклонение;
- кривая приближается к оси Oz по краям асимптотически – никогда не касаясь ее;
- кривая симметрична относительно $M=0$. Ее асимметрия и эксцесс равны нулю;
- кривая имеет характерный изгиб: точка перегиба лежит точно на расстоянии в одну σ от M . Площадь между кривой и осью Oz равна 1.

Последнее свойство объясняет название нормальное *единичное* распределение и имеет исключительно важное значение. Благодаря этому свойству *площадь под кривой интерпретируется как вероятность, или относительная частота*. Действительно, вся площадь под кривой соответствует вероятности того, что признак примет любое значение из всего диапазона его изменчивости (от $-\infty$ до $+\infty$). Площадь под нормальной единичной кривой слева или справа от нулевой точки равна 0,5. Это соответствует тому, что половина генеральной совокупности имеет значение признака больше 0, а половина – меньше 0. Относительная частота встречаемости в генеральной совокупности значений признака в диапазоне от z_1 до z_2 равна площади под кривой, лежащей между соответствующими точками.

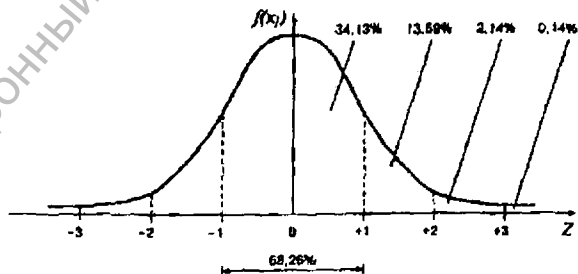


Рис. 3

Для любого нормального распределения существуют следующие соответствия между диапазонами значений и площадью под кривой:

$M \pm \sigma$ соответствует 68% (точно – 68,26%) площади;

$M \pm 2\sigma$ соответствует 95% (точно – 95,44%) площади;

$M \pm 3\sigma$ соответствует 100% (точно – 99,72%) площади.

Нормальное единичное распределение устанавливает четкую взаимосвязь стандартного отклонения и относительного количества случаев в генеральной совокупности для любого нормального распределения. Если распределение является нормальным, то:

90% всех случаев располагается в диапазоне значений $M \pm 1,64\sigma$;

95% всех случаев располагается в диапазоне значений $M \pm 1,96\sigma$;

99% всех случаев располагается в диапазоне значений $M \pm 2,58\sigma$.

Существует специальная таблица, позволяющая определять площадь под кривой справа от любого положительного z (приложение 1). Пользуясь ею, можно определить вероятность встречаемости значений признака из любого диапазона. Это широко используется при интерпретации данных тестирования.

Несмотря на исходный постулат, в соответствии с которым свойства в генеральной совокупности имеют нормальное распределение, реальные данные, полученные на выборке, нечасто распределены нормально. Более того, разработано множество методов, позволяющих анализировать данные безо всякого предположения о характере их распределения как в выборке, так и в генеральной совокупности.

2.5. Разработка тестовых шкал

Тестовые шкалы разрабатываются для того, чтобы оценить индивидуальный результат тестирования путем сопоставления его с тестовыми нормами, полученными на выборке стандартизации. *Выборка стандартизации* специально формируется для разработки тестовой шкалы. Она должна быть репрезентативна генеральной совокупности, для которой планируется применять данный тест. Впоследствии при тестировании предполагается, что и тестируемый, и выборка стандартизации принадлежат одной и той же генеральной совокупности.

Исходным принципом при разработке тестовой шкалы является предположение о том, что измеряемое свойство распределено в генеральной совокупности в соответствии с нормальным законом. Соответственно, измерение в тестовой шкале данного свойства на выборке стандартизации также должно обеспечивать нормальное распределение. Если это так, то тестовая шкала является метрической, точнее, равных

интервалов. Если это не так, то свойство удалось отразить в лучшем случае – в шкале порядка.

Большинство стандартных тестовых шкал являются метрическими, что позволяет более детально интерпретировать результаты тестирования с учетом свойств нормального распределения и корректно применять любые методы статистического анализа. Основная проблема стандартизации теста заключается в разработке такой шкалы, в которой распределение тестовых показателей на выборке стандартизации соответствовало бы нормальному распределению.

Исходные тестовые оценки – это количество ответов на те или иные вопросы теста, время или количество решенных задач и т.д. Они еще называются первичными оценками. Итогом стандартизации являются *тестовые нормы* – таблица пересчета первичных оценок в стандартные тестовые шкалы.

Существует множество стандартных тестовых шкал, основное назначение которых – представление индивидуальных результатов тестирования в удобном для интерпретации виде.

Общая последовательность стандартизации (разработки тестовых норм – таблицы пересчета первичных оценок в стандартные тестовые) состоит в следующем:

1) определяется генеральная совокупность, для которой разрабатывается методика и формируется репрезентативная выборка стандартизации;

2) по результатам применения первичного варианта теста строится распределение первичных оценок;

3) проверяют соответствие полученного распределения нормальному закону;

4) если распределение первичных оценок соответствует нормальному закону, производится *линейная стандартизация*;

5) если распределение первичных оценок не соответствует нормальному, то возможны два варианта:

- перед линейной стандартизацией производят эмпирическую нормализацию;
- проводят нелинейную нормализацию.

Линейная стандартизация заключается в том, что определяются границы интервалов первичных оценок, соответствующие стандартным тестовым показателям. Эти границы вычисляются путем прибавления к среднему первичных оценок (или вычитания из него) долей стандартных отклонений, соответствующих тестовой шкале. В общем случае границы интервалов определяются по формуле z преобразования:

$$z = \frac{x_i - M_x}{\sigma_x} = \frac{st_i - M_{st}}{\sigma_{st}} \rightarrow x_i = M_x + \frac{\sigma_x}{\sigma_{st}} (st_i - M_{st}),$$

где x_i – искомая граница интервала первичных оценок; st_i – граница интервала в стандартной тестовой шкале; M_x , σ_x , M_{st} , σ_{st} – средние и стандартные отклонения первичных оценок (x) и стандартной шкалы (st).

Эмпирическая нормализация применяется, когда распределение первичных баллов отличается от нормального. Она заключается в изменении содержания тестовых заданий. Например, если первичная оценка – это количество задач, решенных испытуемыми за отведенное время, и получено распределение с правосторонней асимметрией, то это значит, что слишком большая доля испытуемых решает больше половины заданий. В этом случае необходимо либо добавить более трудные задания, либо сократить время решения.

Нелинейная нормализация применяется, если эмпирическая нормализация невозможна или нежелательна, например, с точки зрения затрат времени и ресурсов. В этом случае перевод первичных оценок в стандартные производится через нахождение процентильных границ групп в исходном распределении, соответствующих процентильным границам групп в нормальном распределении стандартной шкалы. Каждому интервалу стандартной шкалы ставится в соответствие такой интервал шкалы первичных оценок, который содержит ту же процентную долю выборки стандартизации. Величины долей определяются по площади под единичной нормальной кривой, заключенной между соответствующими данному интервалу стандартной шкалы z -оценками.

Для проверки нормальности используются различные процедуры, позволяющие выяснить, отличается ли от нормального выборочное распределение измеренной переменной. *Если выборочное распределение не отличается от нормального, то это значит, что измеряемое свойство удалось отразить в метрической шкале* (обычно – интервальной).

Существует множество различных способов проверки нормальности, одним из которых является графический. Строят либо квантильные графики, либо графики накопленных частот. *Квантильные графики* строятся следующим образом. Сначала определяются эмпирические значения изучаемого признака, соответствующие 5, 10, ..., 95-процентиллю. Затем по таблице нормального распределения для каждого из этих

процентилей определяются z -значения (теоретические). Два полученных ряда чисел задают координаты точек на графике: эмпирические значения признака откладываются на оси абсцисс, а соответствующие им теоретические значения – на оси ординат. Для нормального распределения все точки будут лежать на одной прямой или рядом с ней. Чем больше расстояние от точек до прямой линии, тем меньше распределение соответствует нормальному.

Графики накопленных частот строятся подобным образом. На оси абсцисс через равные интервалы откладываются значения накопленных относительных частот, например 0,05; 0,1; ...; 0,95. Далее определяются эмпирические значения изучаемого признака, соответствующие каждому значению накопленной частоты, которые пересчитываются в z -значения. По таблице нормального распределения определяются теоретические накопленные частоты (площадь под кривой) для каждого из вычисленных z -значений, которые откладываются на оси ординат. Если распределение соответствует нормальному, то полученные на графике точки лежат на одной прямой.

Критерии асимметрии и эксцесса. Эти критерии определяют допустимую степень отклонения эмпирических значений асимметрии и эксцесса от нулевых значений, соответствующих нормальному распределению. Допустимая степень отклонения – та, которая позволяет считать, что эти статистики существенно не отличаются от нормальных параметров. Величина допустимых отклонений определяется так называемыми стандартными ошибками асимметрии и эксцесса. Для формулы асимметрии стандартная ошибка определяется по формуле

$$As = \sqrt{\frac{6(N-1)}{(N+1)(N+3)}}.$$

Для формулы эксцесса стандартная ошибка эксцесса:

$$Ex = \sqrt{\frac{24N(N-2)(N-3)}{(N+1)^2(N+3)(N+5)}},$$

где N – объем выборки.

Выборочные значения асимметрии и эксцесса не отличаются от нуля, если они не превышают (по абсолютной величине) значения своих стандартных ошибок. Это можно считать признаком соответствия выборочного распределения нормальному закону.

2.6. Коэффициенты корреляции

Кроме рассмотренных основных одномерных описательных статистик, меры центральной тенденции и изменчивости для описания одной переменной, применяются коэффициенты корреляции. *Коэффициент корреляции* – двумерная описательная статистика, количественная мера взаимосвязи (совместной изменчивости) двух переменных.

К настоящему времени разработано великое множество различных коэффициентов корреляции. Их *общей особенностью является то, что они отражают взаимосвязь двух признаков, измеренных в количественной шкале – ранговой или метрической.*

Коэффициент корреляции – это количественная мера силы и направления вероятностной взаимосвязи двух переменных; принимает значения в диапазоне от -1 до $+1$. Сила связи достигает максимума при условии взаимно однозначного соответствия: когда каждому значению одной переменной соответствует только одно значение другой переменной (и наоборот), эмпирическая взаимосвязь при этом совпадает с функциональной линейной связью. Показателем силы связи является *абсолютная* (без учета знака) величина коэффициента корреляции.

Направление связи определяется прямым или обратным соотношением значений двух переменных: если возрастанию значений одной переменной соответствует возрастание значений другой переменной, то взаимосвязь называется прямой (положительной); если возрастанию значений одной переменной соответствует убывание значений другой переменной, то взаимосвязь является обратной (отрицательной). Показателем направления связи является *знак* коэффициента корреляции.

Коэффициент корреляции r -Пирсона применяется для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке. Существует множество ситуаций, в которых уместно его применение. Влияет ли интеллект на успеваемость учащихся? Влияет ли настроение учащихся на успешность решения сложной физической задачи? Для ответа на подобные вопросы исследователь должен изменить два интересующих его показателя у каждого члена выборки.

Формула коэффициента корреляции Пирсона имеет вид:

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^N (x_i - M_x)^2 \times \sum_{i=1}^n (y_i - M_y)^2}}$$

Если значения той и другой переменной были преобразованы в z -значения по формулам

$$z_{x_i} = \frac{x_i - M_x}{\sigma_x}; z_{y_i} = \frac{y_i - M_y}{\sigma_y},$$

то формула для коэффициента корреляции r -Пирсона имеет вид:

$$r_{xy} = \frac{\sum_{i=1}^N z_{x_i} z_{y_i}}{N-1}.$$

На величину коэффициента корреляции не влияет то, в каких единицах измерения представлены признаки. Следовательно, любые линейные преобразования признаков (умножение на константу, прибавление константы: $y_i = x_i b + a$) не меняют значения коэффициента корреляции. Исключением является умножение одного из признаков на отрицательную константу: коэффициент корреляции меняет свой знак на противоположный.

Если обе переменные, между которыми устанавливается связь, представлены в порядковой шкале или одна из них в порядковой, а другая – в метрической, то применяются ранговые коэффициенты корреляции (r -Спирмена и τ -Кендалла). Тот и другой коэффициенты требуют для своего применения предварительного ранжирования обеих переменных.

Коэффициент корреляции r -Спирмена определяется по формуле

$$r_s = \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)},$$

где d_i – разность рангов для испытуемого с номером i . Коэффициент корреляции r -Спирмена равен коэффициенту корреляции r -Пирсона, вычисленному для двух предварительно ранжированных переменных.

Коэффициент корреляции τ -Кендалла рассчитывается по формуле

$$\tau = \frac{2(P - Q)}{N(N - 1)},$$

где P – число совпадений, Q – число инверсий, а $(P + Q) = N(N - 1)/2$.

Формулу для вычисления коэффициента корреляции τ -Кендалла может быть представлена и в виде

$$\tau = \frac{4P}{N(N-1)} - 1.$$

При подсчете τ -Кендалла “вручную” данные сначала упорядочиваются по переменной X . Затем для каждого испытуемого подсчитывается, сколько раз его ранг по Y оказывается меньше, чем ранг испытуемых, находящихся ниже. Результат записывается в столбец “Совпадения”. Сумма всех значений столбца “Совпадения” и есть P – общее число совпадений, подставляется в формулу для вычисления τ -Кендалла.

В измерениях часто встречаются одинаковые значения. При их ранжировании возникает проблема *связанных рангов*. В этом случае действует особое правило ранжирования: объектам с одинаковыми значениями приписывается один и тот же, средний ранг.

При наличии одинаковых (связанных) рангов формулы ранговой корреляции *Спирмена и Кендалла не подходят*. Хотя сумма рангов и не меняется, но изменчивость данных становится меньше. Соответственно, уменьшается возможность оценить степень связи между измеренными свойствами.

При использовании корреляции *Спирмена* в случае связанных рангов возможны два подхода:

- если связей немного (менее 10% для каждой переменной), то вычислить r -*Спирмена* приближенно по приведенной ранее формуле;
- при большем количестве связей применить к ранжированным данным классическую формулу r -*Пирсона* – это всегда позволит определить ранговую корреляцию независимо от наличия связей в рангах.

При использовании корреляции τ -Кендалла в случае наличия связанных рангов в формулу вносятся поправки и получается общая формула для вычисления τ коэффициента корреляции τ -Кендалла независимо от наличия или отсутствия связей в рангах:

$$\tau_b = \frac{P - Q}{\sqrt{\frac{N(N-1)}{2} - K_x} \sqrt{\frac{N(N-1)}{2} - K_y}},$$

где $K_x = \frac{\sum_{i=1}^N f_i(f_i - 1)}{2}$, где i – количество групп связей по X , f_i – численность каждой группы;

$$K_y = \frac{\sum_{i=1}^N f_i(f_i - 1)}{2}$$

(i – количество групп связей по y , f_i – численность каждой группы).

При изучении связей между переменными наиболее предпочтительным является случай применения r -Пирсона непосредственно к исходным данным. Применяя r -Пирсона, необходимо убедиться, что:

- обе переменные не имеют выраженной асимметрии;
- отсутствуют выбросы;
- связь между переменными прямолинейная.

Если хотя бы одно из условий не выполняется, можно попытаться применить ранговые коэффициенты корреляции: r -Спирмена или τ -Кендалла. Но и ранговые корреляции имеют свои ограничения. Они применимы, если:

- обе переменные представлены в количественной шкале (метрической или ранговой);
- связь между переменными является монотонной (не меняет свой знак с изменением величины одной из переменных).

2.7. Гипотезы научные и статистические

Обычно исследование проводится для проверки гипотезы, которая является следствием теоретических представлений. Эта гипотеза содержит утверждение о связи абстрактных категорий, относящихся к свойствам более или менее широкой совокупности объектов – генеральной совокупности.

Предположение, которое проверяется с применением научного метода, называют *научной гипотезой*. Следует отметить, что не всякая гипотеза, а только та, которая допускает для своей проверки применение научного метода, может претендовать на научность. Кроме того, можно научно проверять гипотезы относительно любых мелких проблем, обладающих незначительной научной или практической значимостью. *Сам факт применения научного метода вовсе не гарантирует, что проверяемая гипотеза представляет научный интерес.*

Применение научного метода для проверки гипотезы предполагает определенную последовательность действий исследователя. Исследование начинается с *операционализации* абстрактных категорий – определения операций, при помощи которых соответствующие этим категориям явления могут быть описаны количественно. Затем исследователь

организует выборку и проводит соответствующие измерения. Результаты измерения преобразуют с использованием описательных статистик к виду, допускающему статистическую проверку научной гипотезы.

Любое исследование сводится к выявлению связи между переменными. Связь эта может выражаться в величине и направлении различий между сравниваемыми группами или в знаке и величине коэффициента корреляции. То есть связь характеризуется своей силой и направлением. Однако есть еще одна не менее важная характеристика связи – ее *надежность*, “истинность”.

Надежность связи непосредственно связана с репрезентативностью выборки, с тем, насколько уверенно статистики выборки позволяют судить о соответствующих параметрах генеральной совокупности. Ведь связь, обнаруженная в выборке, интересует исследователя лишь в той мере, в какой она позволяет судить о связи, которая существует в генеральной совокупности.

Надежность связи определяется тем, насколько вероятно, что обнаруженная в выборке связь будет вновь обнаружена (подтвердится) на другой аналогичной выборке, извлеченной из той же генеральной совокупности.

Очевидный способ проверки надежности, обнаруженной в исследовании связи, – многократное проведение аналогичного исследования на разных выборках. Однако это и трудоемко, и не всегда возможно. Но можно сформулировать вопрос по-другому. Если в генеральной совокупности связи нет, то какова вероятность случайного получения данного результата исследования? Иначе говоря, какова вероятность того, что полученный результат является случайным, а на самом деле связи в генеральной совокупности нет? Вопрос, сформулированный таким образом, позволяет получить ответ с использованием методов статистики. Соответствующее проверяемое утверждение об отсутствии связи называется *статистической гипотезой*.

Статистическая гипотеза – это утверждение относительно неизвестного параметра генеральной совокупности, которое формулируется для проверки надежности связи и которое можно проверить по известным выборочным статистикам – результатам исследования. Обычно выделяют основную (нулевую) и альтернативную статистические гипотезы. *Основная (нулевая) гипотеза (H_0)* содержит утверждение об отсутствии связи в генеральной совокупности и доступна проверке методами статистического вывода.

Альтернативная гипотеза (H) принимается при отклонении H_0 и содержит утверждение о наличии связи. При этом нулевая и альтерна-

тивная гипотезы представляют собой в терминах теории вероятности “полную группу несовместных событий”: если верна одна из них, то другая является ложной, и наоборот, отклонение одной из них неизбежно влечет принятие другой. Если по результатам проверки эту гипотезу можно отклонить, то принимается альтернативная гипотеза H .

Следует отметить, что выборочное распределение средних значений соответствует нормальному виду, если $N > 100$. Для выборок меньшего объема распределение средних начинает зависеть от объема выборок (точнее – от числа степеней свободы, df) и соответствует другому теоретическому распределению *t-Стьюдента*.

Статистическая значимость, или *p-уровень значимости*, – основной результат проверки статистической гипотезы. Иначе это вероятность получения данного результата выборочного исследования при условии, что на самом деле для генеральной совокупности верна нулевая статистическая гипотеза: между категориями связь отсутствует. Иначе говоря, это вероятность того, что обнаруженная связь носит случайный характер, а не является свойством совокупности. Именно статистическая значимость, *p-уровень значимости* является количественной оценкой надежности связи: чем меньше эта вероятность, тем надежнее связь.

Предположим, что при сравнении двух выборочных средних было получено значение уровня статистической значимости $p=0,05$. Это означает, что вероятность случайного появления обнаруженных различий составляет не более 5%, т.е. существует 5%-я вероятность того, что обнаруженные различия носят случайный характер, а не являются свойством совокупности.

В отношении научной гипотезы уровень статистической значимости – это количественный показатель степени недоверия к выводу о наличии связи, вычисленный по результатам выборочной, эмпирической проверки этой гипотезы. *Чем меньше значение p-уровня, тем выше статистическая значимость результата исследования, подтверждающего научную гипотезу,*

Уровень значимости при прочих равных условиях выше (значение *p-уровня* меньше), если:

- величина связи (различия) больше;
- изменчивость признака (признаков) меньше;
- объем выборки (выборок) больше.

Статистический критерий – это инструмент определения уровня статистической значимости.

Помимо формулы эмпирического значения критерий задает формулу определения числа степеней свободы. *Число степеней свободы* (обозначается как df) – это количество возможных направлений изменчивости признака. Как правило, число степеней свободы линейно зависит от объема выборки, от числа признаков или их градаций: чем больше эти показатели, тем больше число степеней свободы. *Каждая формула для расчета эмпирического значения критерия обязательно сопровождается правилом (формулой) для определения числа степеней свободы.*

Назначение критерия – проверка статистической гипотезы путем определения p -уровня значимости (вероятности того, что H_0 верна). *Выбор критерия* определяется проверяемой статистической гипотезой.

Критерий включает в себя:

- формулу расчета эмпирического значения критерия по выборочным статистикам;
- правило (формулу) определения числа степеней свободы;
- теоретическое распределение для данного числа степеней свободы;
- правило соотнесения эмпирического значения критерия с теоретическим распределением для определения вероятности того, что H_0 верна.

Для проверки статистических гипотез применяются различные критерии. При этом одному теоретическому распределению могут соответствовать разные формулы критериев – в зависимости от проверяемой статистической гипотезы. Но принцип проверки является общим для всего этого многообразия: вычисленное по формуле эмпирическое значение критерия сопоставляется с теоретическим распределением для заданного числа степеней свободы, что позволяет определить вероятность того, что H_0 верна.

Для проверки гипотезы исследователь совершает последовательность действий, включающую применение специальных таблиц критических значений критерия:

1. Выбор критерия в зависимости от вида исходных данных и статистической гипотезы: теоретического распределения, формул расчета эмпирического значения критерия и числа степеней свободы.

2. Расчет по исходным данным (или по имеющимся статистикам) эмпирического значения критерия и числа степеней свободы.

3. Применение “Таблицы критических значений критерия” позволяет определить значение p -уровня для данного числа степеней свободы.

Таблица критических значений содержит значения (квантили) теоретического распределения, соответствующие наиболее важным – кри-

тическим значениям p -уровня (0,1; 0,05; 0,01 и т. д.) для различных чисел степеней свободы. По вычисленному эмпирическому значению критерия p -уровень значимости при помощи таких таблиц определяется следующим образом:

1) для данного числа степеней свободы по таблице определяются ближайшие критические значения и p -уровни, соответствующие им;

2) значение p -уровня определяется в виде неравенства по правилу, которое демонстрируется на рис.4 (значимость возрастает слева направо, в соответствии с убыванием p -уровня):

- если эмпирическое значение критерия ($K_э$) находится между двумя критическими значениями, то p -уровень *меньше* того критического p , которое находится левее;
- если $K_э$ находится левее крайнего левого критического значения (обычно это соответствует критическому $p=0,1$, реже – $p=0,05$), то p -уровень *больше*, чем крайнее правое критическое значение p ;
- если $K_э$ находится правее крайнего правого критического значения, то p -уровень *меньше* крайнего правого критического p .

Например, если эмпирическое значение критерия ($K_э$) находится между $K_{0,05}$ и $K_{0,01}$, то $p < 0,05$. Если $K_э$ находится левее $K_{0,1}$, то $p > 0,1$. Если $K_э$ находится правее $K_{0,001}$, то $p < 0,001$.

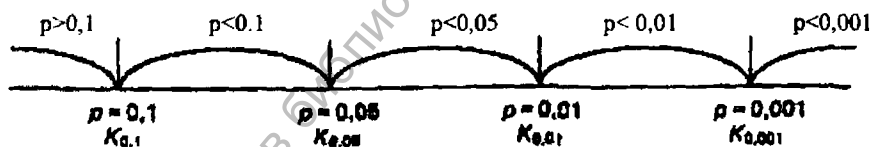


Рис. 4

В конечном счете проверка статистической гипотезы должна заканчиваться *принятием статистического решения* о том, какая же гипотеза верна: нулевая – об отсутствии связи или альтернативная – о ее наличии. Соответственно, от этого зависит и окончательный, содержательный вывод исследования: подтверждена или нет исходная научная гипотеза.

Вполне очевидно, что основанием для принятия исследователем решения о том, какая гипотеза верна, является p -уровень – вероятность того, что верна все-таки нулевая гипотеза. Чем меньше p -уровень, тем с большей уверенностью можно отклонить H_0 в пользу H_1 , тем самым подтвердив исходную содержательную гипотезу.

Не менее очевидно и то, что, принимая решение, исследователь всегда допускает вероятность его ошибочности, ведь исследование проведено на выборке, а вывод делается в отношении генеральной совокупности. При отклонении H_0 в пользу H исследователь рискует, что связи на самом деле в генеральной совокупности нет. И наоборот, решение в пользу H_0 вовсе не исключает наличие связи.

Основная проблема статистического вывода заключается в том, что заранее должно быть установлено оптимальное значение величины α (*допустимой вероятности ошибки*), удовлетворяющее двум противоречивым требованиям. Величина α должна быть достаточно мала, чтобы обеспечивать доверие к результатам исследования при отклонении H_0 . Величина α должна быть достаточно велика, чтобы отклонить H_0 при наличии связи (различий). Вопрос о том, какая же величина α является приемлемой, не имеет однозначного ответа.

Вопрос о величине α – вопрос о том, при каком же p -уровне исследователь может отклонить H_0 , решается преимущественно исходя из неформальных соглашений. Традиционная интерпретация различных уровней значимости исходит из $\alpha = 0,05$. В соответствии с ней приемлемым для отклонения H_0 признается уровень $p < 0,05$. Такая относительно высокая вероятность ошибки может быть рекомендована для небольших выборок. Если объемы выборок около 100 и более объектов, то порог отклонения H_0 целесообразно снизить до $p = 0,01$ и принимать решение о наличии связи (различий) при $p < 0,01$.

Основная (нулевая) статистическая гипотеза, как отмечалось, содержит утверждение о равенстве нулю (коэффициента корреляции) или о равенстве средних значений, дисперсий и т.д. Если по результатам статистической проверки основная гипотеза отклоняется, то принимается альтернативная гипотеза. Принимаемая альтернатива может быть как *направленной* (например, $H_1: r > 0$ или $H_1: M_1 > M_2$), так и *ненаправленной* (например, $H_1: r \neq 0$ или $H_1: M_1 \neq M_2$). Какая альтернатива должна быть принята по результатам проверки, зависит от применяемого для проверки метода и теоретического распределения. Обычно характер альтернативы явно указывается при описании метода, следовательно, в большинстве случаев исследователь избавлен от необходимости выбора.

Если процедура проверки гипотезы H_0 подразумевает ненаправленную альтернативу, то критические области, соответствующие ее отклонению (принятию альтернативы), поровну распределяются по обоим “хвостам” распределения. Чаще всего интервал принятия нулевой ги-

гипотезы ($1 - \alpha$) при этом охватывает диапазон теоретических значений, симметричный относительно нуля. Поэтому такие критерии часто называют *двусторонними*, имеющими “два хвоста” (для проверки ненаправленных гипотез). Заметим, что в этом случае, если принят уровень α для решения об отклонении H_0 , существует *два теоретических (критических) значения*: одно отсекает $\alpha/2$ справа, а другое, отрицательное, $-\alpha/2$ слева. Если проверяется направленная гипотеза, то процедура проверки допускает принятие *односторонней альтернативы*. В этом случае, если принят уровень α для решения об отклонении H_0 , существует одно теоретическое (критическое) значение (для $H_1: r > 0$ – положительное), и оно отсекает ровно α справа (или слева – в зависимости от направления альтернативы). Очевидно, что односторонняя альтернатива более “лояльная” к отклонению H_0 для одних и тех же выборочных результатов. *При двусторонней альтернативе, по сравнению с односторонней, нулевая гипотеза отвергается при больших значениях силы связи (корреляции, различий средних и пр.)*.

Важно отметить, что принятие по результатам проверки гипотезы ненаправленной альтернативы вовсе не означает ограничение выводов лишь “ненаправленными” суждениями типа “средние различаются”, “корреляция отличается от нуля”. Как следует из предыдущих рассуждений, проверка ненаправленной гипотезы является более “строгой” (при прочих равных условиях). *Принятие ненаправленной (двусторонней) альтернативы позволяет сделать вывод о направлении связи в генеральной совокупности в соответствии с выборочными данными*.

Для одного и того же эмпирического значения критерия p -уровень значимости для направленной альтернативы в 2 раза меньше p -уровня для ненаправленной альтернативы. Различие между направленной и ненаправленной альтернативами, кажется, еще более усложняет и без того непростую логику статистической проверки гипотез. Однако в большинстве случаев выбор альтернативы не является проблемой для исследователя: он определен самим методом (критерием) статистической проверки и исключает возможность произвола. При проверке гипотезы с помощью таблиц критических значений указывается, для какой альтернативы приведены критические значения.

Статистическое решение является основанием для содержательного вывода в отношении проверяемой гипотезы. Когда принимается H_0 , всегда остается вероятность того, что связь или различия все же есть. И мы ничего не можем сказать о том, насколько велика или мала эта вероятность. Принятие H_0 не означает, что различия отсутствуют или мера

связи равна нулю; из этого следует только то, что *статистически значимые результаты не обнаружены*.

Когда в результате исследования принимается H_0 , никакого содержательного вывода сделать нельзя. Поэтому выражение “Отрицательный результат исследования – тоже результат” имеет для исследователя исключительно психотерапевтическое значение: отрицательный результат исследования – это отсутствие какого бы то ни было результата.

В случае отклонения H_0 остается вероятность того, что H_0 все-таки верна и эта вероятность равна p -уровню значимости. Следовательно, нельзя утверждать, что результаты *доказывают* справедливость содержательной гипотезы. Корректным будет более осторожный вывод о том, что *получено свидетельство в пользу содержательной гипотезы*.

Следует отметить, что при оформлении исследовательского отчета (курсовой или дипломной работы, публикации) статистические гипотезы и статистические решения, как правило, не приводятся. Обычно при описании результатов указывают критерий, приводят необходимые описательные статистики (средние, сигмы, корреляции и т.д.), эмпирические значения критериев, степени свободы и обязательно – p -уровень значимости. Затем формулируют содержательный вывод в отношении проверяемой гипотезы с указанием (обычно в виде неравенства) достигнутого или не достигнутого уровня значимости.

Общая последовательность проверки содержательной гипотезы:

1. Формулировка содержательной гипотезы.
2. Планирование исследования (выборка, процедура, инструментарий), в том числе предварительная формулировка доступной проверке статистической гипотезы.
3. Проведение измерений и накопление исходных данных.
4. Окончательная формулировка статистической гипотезы, выбор статистического критерия, установление величины α – допустимой вероятности ошибки.
5. Определение p -уровня статистической значимости в результате применения статистического критерия.
6. Статистический вывод: статистическое решение о принятии или отклонении H_0 .
7. Формулировка содержательного вывода.

Приступая к операционализации содержательной гипотезы – к определению того, как будут измерены изучаемые явления, исследователь уже должен представлять себе, какому методу статистического вывода будут соответствовать получаемые в процессе исследования

исходные данные. В противном случае он рискует оказаться в ситуации, когда данные уже собраны, но невозможно определить метод их анализа.

Любая содержательная гипотеза научного исследования касается связи между явлениями (свойствами, событиями), независимо от того, содержит ли формулировка гипотезы указание на связь или на различия (между группами, условиями, событиями). Кроме того, независимо от своей формулировки, одна и та же содержательная гипотеза может быть проверена при помощи самых разных статистических методов. Ограничение на выбор статистического метода возникает только после определения того, как измерены (или будут измерены) явления, в отношении связи которых проверяется гипотеза.

2.8. Методы корреляционного анализа

Методы корреляционного анализа классифицируются по различным основаниям. *Первое основание* для классификации исследовательских ситуаций – это *типы шкал*, в которых измерены признаки X и Y , связь между которыми изучается. Признаки могут быть измерены либо в количественной шкале (порядковой, метрической), либо в качественной (номинативной) шкале.

Наиболее многочисленная группа методов относится к случаю, когда одна из переменных является количественной, а другая – качественной. Это широкий класс исследовательских ситуаций, когда задача сводится к сравнению групп (градаций номинативной переменной) по уровню выраженности признака (количественной переменной).

Для решения такой задачи (X – качественный, Y – количественный) применяются *методы сравнения*, которые можно классифицировать по трем основаниям (таблица 3):

а) *количество сравниваемых групп* (градаций номинативной переменной) – две или более двух;

б) *соотношение сравниваемых групп*: зависимые выборки или независимые выборки;

в) *шкала, в которой измерен количественный признак*: метрическая, ранговая.

Корреляция метрических переменных

Статистическая гипотеза о связи двух метрических переменных проверяется в отношении коэффициента корреляции r -Пирсона, который вычисляется по формуле

$$r_{xy} = \frac{\sum_{i=1}^N (x_i - M_x)(y_i - M_y)}{\sqrt{\sum_{i=1}^N (x_i - M_x)^2 \times \sum_{i=1}^N (y_i - M_y)^2}}$$

Таблица 3

Количество выборок (градаций X)		Две выборки		Больше двух выборок	
Зависимость выборок		Независимые	Зависимые	Независимые	Зависимые
Признак Y	Метрический	Параметрические методы сравнения			
		t-Стьюдента для независимых выборок	t-Стьюдента для зависимых выборок	ANOVA	ANOVA с повторными измерениями
	Ранговый	Непараметрические методы сравнения			
		U-Манна-Уитни, критерий серий	T-Вилкоксона, критерий знаков	H-Краскала-Уоллеса	χ^2 -Фридмана

Основной (нулевой) статистической гипотезой является равенство r -Пирсона нулю в генеральной совокупности ($H_0: r_{xy} = 0$). Определение p -уровня значимости осуществляется при помощи критерия t -Стьюдента:

$$t_s = \frac{r_{xy} \sqrt{N-2}}{\sqrt{1-r_{xy}^2}}, \quad df = N-2.$$

С целью упрощения проверки при обработке данных “вручную” обычно пользуются *таблицами критических значений* r_{xy} , которые составлены с помощью этого критерия (приложение 3).

Для статистического решения о принятии или отклонении H_0 обычно устанавливают $\alpha = 0,05$, а для выборок большого объема (около 100 и более) – $\alpha = 0,01$. Если $p < \alpha$, H_0 отклоняется и делается содержательный вывод о том, что обнаружена статистически достоверная (значимая) связь между изучаемыми переменными (положительная или отрицательная – в зависимости от знака корреляции). Когда $p > \alpha$, H_0 не отклоняется и содержательный вывод ограничен констатацией того, что связь (статистически достоверная) не обнаружена.

Пусть на выборке $N = 10$ (учащиеся 8-го класса) были измерены два показателя интеллекта: вербального (x) и невербального (y). Коэффициент корреляции составил $r_{xy} = 0,517$.

Проверку гипотезы о связи этих показателей можно провести двумя способами. Подставка величины $N_1 = 20$ и $r_{xy} = 0,517$ в формулу вычисления критерия t-Стьюдента позволяет получить $t_s = 2,562$; $df = 18$. Из таблицы критических значений t-Стьюдента (приложение 2) для $df = 18$ видно, что эмпирическое значение находится между критическими значениями для $p = 0,05$ и $p = 0,01$. Следовательно, для этого случая $p < 0,05$.

Тот же результат получается без вычисления критерия t-Стьюдента. Из таблицы критических значений коэффициента корреляции r-Пирсона (приложение 3): для строки с $N_1 = 20$ видно, что эмпирическое значение корреляции находится между критическими значениями для $p = 0,05$ и $p = 0,01$. Следовательно, $p < 0,05$.

Статистическое решение: $H_0 (r_{xy} = 0)$ отклоняется для $\alpha = 0,05$.
Содержательный вывод: обнаружена статистически достоверная положительная связь вербального и невербального интеллекта для учащихся 8-го класса ($r_{xy} = 0,517$, $N_1 = 20$, $p < 0,05$).

Если связь статистически достоверна, то, прежде чем делать содержательный вывод, следует исключить возможность “ложной” корреляции.

1. Связь обусловлена выбросами: просмотреть график двухмерного рассеивания. При наличии выбросов перейти к ранговым корреляциям или исключить выбросы.

2. Связь обусловлена влиянием третьей переменной: просмотреть график двухмерного рассеивания на предмет наличия содержательно интерпретируемого деления выборки на группы, для которых согласованно меняются средние двух переменных. Если подобное явление возможно, необходимо вычислить корреляцию не только для всей выборки, но и для каждой группы в отдельности. Если “третья” переменная метрическая – вычислить частную корреляцию.

Частная корреляция

Если изучается связь между тремя метрическими переменными, то возможна проверка предположения о том, что связь между двумя переменными X и Y не зависит от влияния третьей переменной Z . Для этого можно вычислить коэффициент частной корреляции $r_{xy \cdot z}$:

$$r_{xy \cdot z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}}$$

Коэффициент r_{xy-z} тем больше по абсолютной величине (ближе к r_{xy}), чем меньше связь между X и Y обусловлена влиянием Z . Коэффициент r_{xy-z} близок к 0, если связь между X и Y близка к 0 при любом фиксированном значении Z , то есть связь между X и Y обусловлена влиянием Z .

Основной (нулевой) статистической гипотезой является равенство частной корреляции нулю в генеральной совокупности ($H_0: r_{xy-z} = 0$). Определение p -уровня значимости осуществляется при помощи критерия t -Стьюдента:

$$t_s = \frac{r_{xy-z} \sqrt{N-3}}{\sqrt{1-r_{xy-z}^2}}; \text{ df} = N - 3.$$

Если $p < \alpha$, то H_0 отклоняется и делается содержательный вывод о том, что обнаружена статистически достоверная связь X и Y при фиксированных значениях Z , то есть связь между X и Y не зависит от влияния Z . Когда $p > \alpha$, H_0 не отклоняется, и содержательный вывод ограничен констатацией того, что связь (статистически достоверная) между X и Y при фиксированных значениях Z обнаружена.

Проверка гипотез о различии корреляций

Задача сравнения корреляций имеет два варианта решения:

а) для независимых выборок – когда необходимо сравнить два коэффициента корреляции, полученных на разных выборках между одними и теми же переменными;

б) для зависимых выборок – когда необходимо сравнить корреляцию переменных X и Y с корреляцией переменных X и Z , при условии, что все три переменные измерены на одной и той же выборке.

Сравнение корреляций для независимых выборок. По результатам сравнения корреляций в данном случае можно делать вывод о различии корреляции признаков X и Y в двух сравниваемых совокупностях. Проверяемая H_0 содержит утверждение о равенстве корреляций в генеральной совокупности.

Задача статистической проверки решается при помощи z -преобразования Фишера коэффициентов корреляции и последующего применения z -критерия. z -преобразование Фишера – это пересчет коэффициентов корреляции r по формуле

$$z = \frac{1}{2} \ln \frac{1+r}{1-r}.$$

Эмпирическое значение z-критерия для определения p -уровня значимости различия корреляций вычисляется по формуле

$$z_3 = \frac{z_1 - z_2}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}},$$

где z_1 и z_2 – z-преобразованные значения сравниваемых корреляций, N_1 и N_2 – соответствующие объемы выборок. Уровень значимости определяется по формуле $p < 2S$, где S – площадь справа от z_3 под кривой нормального распределения.

Отметим, что одна и та же разность между корреляциями будет иметь более высокую статистическую значимость при больших значениях корреляции и меньшую – при более слабых корреляциях.

Сравнение корреляций для зависимых выборок. В данном случае предполагается сравнение корреляции X и Y с корреляции X и Z при условии, что все три признака измерены на одной и той же выборке. Проверяемая H_0 содержит утверждение о равенстве соответствующих корреляций. Для статистической проверки подобных гипотез применяется z-критерий, эмпирическое значение которого вычисляется по формуле

$$z_3 = \frac{(r_{xy} - r_{xz})\sqrt{N}}{\sqrt{(1 - r_{xy}^2)^2 + (1 - r_{xz}^2)^2 - 2r_{yz}^3 - (2r_{yz} - r_{xy}r_{xz})(1 - r_{xy}^2 - r_{xz}^2 - r_{yz}^2)}}.$$

Корреляция ранговых переменных

Если к количественным данным неприменим коэффициент корреляции г-Пирсона, то для проверки гипотезы о связи двух переменных после предварительного ранжирования могут быть применены корреляции г-Спирмена или τ -Кендалла.

Коэффициент корреляции r -Спирмена вычисляется либо путем применения формулы г-Пирсона к предварительно ранжированным двум переменным, либо, при отсутствии повторяющихся рангов, по упрощенной формуле

$$r_s = \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)}.$$

Поскольку этот коэффициент – аналог r -Пирсона, то и применение r -Спирмена для проверки гипотез аналогично применению r -Пирсона, изложенному ранее. *Преимущество r -Спирмена* по сравнению с r -Пирсона – в большей чувствительности к связи в случае:

- существенного отклонения распределения хотя бы одной переменной от нормального вида (асимметрия, выбросы);
- криволинейной (монотонной) связи.

Недостаток r -Спирмена по сравнению с r -Пирсона – в меньшей чувствительности к связи в случае несущественного отклонения распределения обеих переменных от нормального вида.

Частная корреляция и сравнение корреляций применимы и к r -Спирмена. τ -Кендалла применяется к предварительно ранжированным данным как альтернатива r -Спирмена. τ -Кендалла имеет более выгодную вероятностную интерпретацию. Общая формула для вычисления τ -Кендалла, вне зависимости от наличия или отсутствия повторяющихся рангов (связей):

$$t_b = \frac{P - Q}{\sqrt{\left[\frac{N(N-1)}{2} \right] - K_x} \sqrt{\left[\frac{N(N-1)}{2} \right] - K_y}}$$

где P – число совпадений, Q – число инверсий, K_x и K_y – поправки на связь в рангах.

Поскольку природа τ -Кендалла иная, чем у r -Спирмена и r -Пирсона, то p -уровень определяется по-другому: применяется z -критерий и единичное нормальное распределение. Эмпирическое значение вычисляется по формуле

$$z_s = \frac{|P - Q| - 1}{\sqrt{\frac{N(N-1)(2N+5)}{18}}}$$

При вычислениях “вручную” p -уровень определяется по следующему алгоритму:

- а) вычисляется эмпирическое значение z_s ;
- б) по таблице “Стандартные нормальные вероятности” (приложение 1) определяется теоретическое значение z , ближайшее меньшее к эмпирическому значению z_s ;
- в) определяется площадь S под кривой справа от z_s ;
- г) вычисляется p -уровень по формуле $p < 2S$.

Проверяемая статистическая гипотеза, порядок принятия статистического решения и формулировка содержательного вывода те же, что и для случая t -Пирсона или t -Спирмена.

Параметрические методы сравнения двух выборок

Сравнение двух выборок по признаку, измеренному в метрической шкале, обычно предполагает *сравнение средних значений с использованием параметрического критерия t -Стьюдента*. Следует различать три ситуации по соотношению выборок между собой: случай *независимых* и *зависимых* выборок (измерений признака) и дополнительно – случай сравнения одного среднего значения с заданной величиной (критерий t -Стьюдента для одной выборки).

К параметрическим методам относится и *сравнение дисперсий двух выборок по критерию F -Фишера*. Иногда этот метод приводит к ценным содержательным выводам, а в случае сравнения средних для независимых выборок сравнение дисперсий является *обязательной* процедурой.

При сравнении средних или дисперсии двух выборок проверяется *ненаправленная статистическая гипотеза* о равенстве средних (дисперсий) в генеральной совокупности. Соответственно, при ее отклонении допустимо принятие двусторонней альтернативы о конкретном направлении различий в соответствии с соотношением выборочных средних (дисперсий). Для принятия статистического решения в таких случаях применяются двусторонние критерии и, соответственно, критические значения для проверки ненаправленных альтернатив.

Сравнение дисперсий. Метод позволяет проверить гипотезу о том, что дисперсии двух генеральных совокупностей, из которых извлечены сравниваемые выборки, отличаются друг от друга. Проверяемая статистическая гипотеза $H_0: y = 0$. При ее отклонении принимается альтернативная гипотеза о том, что одна дисперсия больше другой.

Исходные предположения: две выборки извлекаются случайно из разных генеральных совокупностей с нормальным распределением изучаемого признака.

Структура исходных данных: изучаемый признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух сравниваемых выборок.

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального.

Формула для эмпирического значения критерия F -Фишера:

$$F_0 = \frac{\sigma_1^2}{\sigma_2^2}; df_{\text{числ}} = N_1 - 1; df_{\text{знам}} = N_2 - 1,$$

где σ_1^2 – большая дисперсия, а σ_2^2 – меньшая дисперсия.

Так как заранее не известно, какая дисперсия больше, то для определения p -уровня применяется *таблица критических значений для ненаправленных альтернатив* (приложение 5). Если $F_0 \geq F_{\text{кр}}$ для соответствующего числа степеней свободы, то $p = 0,05$ и статистическую гипотезу о равенстве дисперсий можно отклонить (для $\alpha = 0,05$).

Метод может применяться для проверки предположения о равенстве (гомогенности) дисперсий перед проверкой достоверности различия средних по критерию t -Стьюдента для независимых выборок разной численности. Однако содержательная интерпретация статистически достоверного различия дисперсий может иметь и самостоятельную ценность.

Критерий t -Стьюдента для одной выборки. Метод позволяет проверить гипотезу о том, что среднее значение изучаемого признака M отличается от некоторого известного значения A . Проверяемая статистическая гипотеза: $H_0: M_x = A$. При ее отклонении принимается альтернативная гипотеза о том, что M_x меньше (больше) A .

Исходное предположение: распределение признака в выборке приблизительно соответствует нормальному виду.

Структура исходных данных: значения изучаемого признака определены для каждого члена выборки, которая репрезентативна изучаемой генеральной совокупности.

Формула для эмпирического значения критерия t -Стьюдента:

$$t_0 = \frac{|M - A|}{\sigma / \sqrt{N}}; df = N - 1.$$

Критерий t -Стьюдента для независимых выборок. Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены сравниваемые независимые выборки, отличаются друг от друга. Допущение независимости предполагает, что представители двух выборок *не составляют пары коррелирующих значений признака*.

Проверяемая статистическая гипотеза: $H_0: M_1 = M_2$. При ее отклонении принимается альтернативная гипотеза о том, что M_1 больше (меньше) M_2 .

Исходные предположения для статистической проверки:

- одна выборка извлекается из одной генеральной совокупности, а другая выборка, *независимая* от первой, извлекается из другой генеральной совокупности;
- распределение изучаемого признака и в той, и в другой выборке приблизительно соответствует нормальному;
- дисперсии признака в двух выборках примерно одинаковы (гомогенны).

Структура исходных данных: изучаемый признак измерен у объектов (испытуемых), каждый из которых принадлежит к одной из двух сравниваемых независимых выборок.

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального; *в случае разной численности* сравниваемых выборок их дисперсии статистически достоверно не различаются (проверяется по критерию F-Фишера при вычислениях “вручную”).

Формулы для эмпирического значения критерия t-Стьюдента:

$$t_s = \frac{|M_1 - M_2|}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}} \quad \text{или} \quad t_s = \frac{|M_1 - M_2|}{\sqrt{\frac{(N_1 - 1)\sigma_1^2 + (N_2 - 1)\sigma_2^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2} \right)}}; \quad df = N_1 + N_2 - 2.$$

Первая формула применяется для приближенных расчетов, для близких по численности выборок, а вторая формула – для точных расчетов, когда выборки заметно различаются по численности.

Критерий t-Стьюдента для зависимых выборок. Метод позволяет проверить гипотезу о том, что средние значения двух генеральных совокупностей, из которых извлечены сравниваемые *зависимые* выборки, отличаются друг от друга. Допущение зависимости чаще всего значит, что признак измерен на одной и той же выборке дважды, например, до воздействия и после него. В общем же случае каждому представителю одной выборки поставлен в соответствие представитель из другой выборки (они попарно объединены) так, что два ряда данных положительно коррелируют друг с другом.

Проверяемая статистическая гипотеза, как и в предыдущем случае, $H_0: M_1 = M_2$. При ее отклонении принимается альтернативная гипотеза о том, что M_1 больше (меньше) M_2 .

Исходные предположения для статистической проверки:

- каждому представителю одной выборки (из одной генеральной совокупности) поставлен в соответствие представитель другой выборки (из другой генеральной совокупности);
- данные двух выборок положительно коррелируют;
- распределение изучаемого признака и в той, и в другой выборке соответствует нормальному закону.

Структура исходных данных: имеется по два значения изучаемого признака для каждого объекта (для каждой пары).

Ограничения: распределения признака и в той, и в другой выборке существенно не отличаются от нормального; данные двух измерений, соответствующих той и другой выборке, положительно коррелируют.

Альтернативы: критерий Т-Вилкоксона, если распределение хотя бы для одной выборки существенно отличается от нормального; критерий t-Стьюдента для независимых выборок – если данные для двух выборок не коррелируют положительно.

Формула для эмпирического значения критерия t-Стьюдента отражает тот факт, что единицей анализа различий является *разность (сдвиг)* значений признака для каждой пары наблюдений. Соответственно, для каждой из N пар значений признака сначала вычисляется разность $d_i = x_{1i} - x_{2i}$,

$$t_s = \frac{|M_d|}{\sigma_d / \sqrt{N}}; \text{ df} = N - 1,$$

где M_d – средняя разность значений; σ_d – стандартное отклонение разностей.

Непараметрические методы сравнения выборок

Непараметрические методы заметно проще в вычислительном отношении, чем их параметрические аналоги. Непараметрические аналоги параметрических методов сравнения выборок применяются в случаях, когда не выполняются основные предположения, лежащие в основе параметрических методов сравнения средних значений.

При решении вопроса о выборе параметрического или непараметрического метода сравнения необходимо иметь в виду, что параметрические методы обладают заведомо большей чувствительностью, чем их непараметрические аналоги. Поэтому исходной ситуацией является выбор параметрического метода. И решение о применении непараметрического метода становится определенным, если не выполняются исходные предположения, лежащие в основе применения параметрического метода.

Условия, когда применение непараметрических методов является оправданным:

- есть основания считать, что распределение значений признака в генеральной совокупности не соответствует нормальному закону;
- есть сомнения в нормальности распределения признака в генеральной совокупности, но выборка слишком мала, чтобы по выборочному распределению судить о распределении в генеральной совокупности;
- не выполняется требование гомогенности дисперсии при сравнении средних значений для независимых выборок.

На практике *преимущество непараметрических методов* наиболее заметно, когда в данных имеются выбросы (экстремально большие или малые значения).

Если размер выборки очень велик (больше 100), то непараметрические методы сравнения использовать нецелесообразно, даже если не выполняются некоторые исходные предположения применения параметрических методов. С другой стороны, если объемы сравниваемых выборок очень малы (10 и меньше), то результаты применения непараметрических методов можно рассматривать лишь как предварительные.

Структура исходных данных и интерпретация результатов применения для параметрических методов и их непараметрических аналогов являются идентичными.

При сравнении выборок с использованием непараметрических критериев, как и в случае параметрических критериев, обычно проверяются *ненаправленные статистические гипотезы*. Основная (нулевая) статистическая гипотеза при этом содержит утверждение об идентичности генеральных совокупностей (из которых извлечены выборки) по уровню выраженности изучаемого признака. Соответственно, при ее отклонении допустимо принятие двусторонней альтернативы о конкретном направлении различий в соответствии с выборочными данными. Для принятия статистического решения в таких случаях применяются двусторонние критерии и, соответственно, критические значения для проверки ненаправленных альтернатив.

Сравнение двух независимых выборок. Самым популярным и наиболее чувствительным (мощным) аналогом критерия t-Стьюдента для независимых выборок является *критерий U-Манна-Уитни*. Непараметрическим его аналогом является *критерий серий*, который еще проще в вычислительном отношении, но обладает заметно меньшей чувствительностью, чем критерий U.

Эмпирическое значение критерия (U-Манна-Уитни показывает, насколько совпадают (пересекаются) два ряда значений измеренного признака. Чем меньше совпадение, тем больше различаются эти два ряда. Основная идея критерия (основана на представлении всех значений двух выборок в виде одной общей последовательности упорядоченных (ранжированных) значений. Основной (нулевой) статистической гипотезе будет соответствовать ситуация, когда значения одной выборки будут равномерно распределены среди значений другой выборки, то есть когда два ряда значений пересекаются в наибольшей возможной степени. Напротив, отклонению этой гипотезы будет соответствовать ситуация, когда значения одной из выборок будут преобладать на одном из концов объединенного ряда – пересечение двух рядов тогда будет минимальным.

Формально, критерий U – это общее число тех случаев, в которых значения одной группы превосходят значения другой группы при попарном сравнении значений первой и второй групп. Соответственно, вычисляются два значения критерия: U_x и U_y .

Для вычислений используются следующие формулы:

$$U_x = mn - R_x + \frac{n(n+1)}{2}; U_y = mn - R_y + \frac{m(m+1)}{2},$$

где n – объем выборки X ; m – объем выборки Y , R_x и R_y – суммы рангов для X и Y в объединенном ряду. В качестве эмпирического значения критерия берется *наименьшее* из U_x и U_y . Чем больше различия, тем меньше эмпирическое значение U .

Поскольку критерий U отражает степень совпадения (перекрещивания) двух рядов значений, то значение p -уровня тем меньше, чем меньше значение U . При расчетах используют таблицы критических значений критерия U-Манна-Уитни (приложение 9).

Сравнение двух зависимых выборок. Самым чувствительным (мощным) аналогом критерия t-Стьюдента для зависимых выборок является критерий Т-Вилкоксона. Непараметрическим его аналогом является критерий знаков, который еще проще в вычислительном отношении, но обладает меньшей чувствительностью, чем критерий Т-Вилкоксона.

Критерий Т основан на упорядочивании величин *разностей* (сдвиг) значений признака в каждой паре его измерений (критерий знаков основан на учете только знака этой разности). Соответственно, критерий Т, будучи менее чувствительным аналогом t-Стьюдента, более чувствителен по сравнению с другими непараметрическими критериями для повторных измерений (зависимых выборок).

Критерий Т-Вилкоксона основан на ранжировании абсолютных разностей пар значений зависимых выборок. Далее подсчитывается сумма рангов для положительных разностей и сумма рангов для отрицательных разностей. Идея критерия заключается в подсчете вероятности получения минимальной из этих разностей при условии, что распределение положительных или отрицательных разностей равномерно и равно $1/2$.

Для расчетов “вручную” не требуется особых формул: достаточно подсчитать суммы рангов для положительных и отрицательных разностей. Затем меньшая из сумм принимается в качестве эмпирического значения критерия, значение которого сравнивается с табличным значением (приложение 10), рассчитанным для условия равной вероятности положительных и отрицательных разностей для данного объема выборки. Конечно, чем больше различия, тем меньше эмпирическое значение T , тем менее вероятно получение такого значения при условии равной вероятности встречаемости положительных и отрицательных разностей, следовательно, тем меньше значение p -уровня.

Сравнение более двух независимых выборок. Критерий Н-Краскала-Уоллеса является непараметрическим аналогом однофакторного дисперсионного анализа (ANOVA) для независимых выборок, поэтому другое его название – *однофакторный дисперсионный анализ Краскала-Уоллеса*. Он позволяет проверять гипотезы о различии более двух выборок по уровню выраженности изучаемого признака.

Н-Краскала-Уоллеса по идее сходен с критерием U-Манна-Уитни. Как и последний, он оценивает степень пересечения (совпадения) нескольких рядов значений измеренного признака. Чем меньше совпадений, тем больше различаются ряды, соответствующие сравниваемым выборкам. Основная идея критерия Н-Краскала-Уоллеса основана на представлении всех значений сравниваемых выборок в виде одной общей последовательности упорядоченных (ранжированных) значений с последующим вычислением среднего ранга для каждой из выборок. Если выполняется статистическая гипотеза об отсутствии различий, то можно ожидать, что все средние ранги примерно равны и близки к общему среднему рангу.

Эмпирическое значение критерия Н-Краскала-Уоллеса вычисляется после ранжирования всех значений сравниваемых выборок по формуле

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_i} - 3(N+1),$$

где N – суммарная численность всех выборок; k – количество сравниваемых выборок; R_i – сумма рангов для выборки; n_i – численность выборки i .

Чем сильнее различаются выборки, тем больше вычисленное значение H и тем меньше p -уровень значимости.

При расчетах “вручную” для определения p -уровня пользуются таблицами критических значений. Если объем каждой выборки больше 5 и количество выборок больше трех, то эмпирическое значение критерия сравнивается с χ^2 (приложение 7) для $df = k - 1$ (k – число выборок). Если сравниваются 3 выборки и объем каждой выборки меньше 5, то пользуются таблицей критических значений Н-Краскала-Уоллеса (приложение 11).

При отклонении нулевой статистической гипотезы об отсутствии различий принимается альтернативная гипотеза о статистически достоверных различиях выборок по изучаемому признаку – без конкретизации направления различий. Для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию U-Манна-Уитни.

Сравнение более двух зависимых выборок. Критерий χ^2 -Фридмана является непараметрическим аналогом однофакторного дисперсионного анализа (ANOVA) для повторных измерений. Он позволяет проверять гипотезы о различии более двух зависимых выборок (повторных измерений) по уровню выраженности изучаемого признака. Критерий χ^2 -Фридмана может быть более эффективен, чем его метрический аналог ANOVA в случаях повторных измерений изучаемого признака на небольших выборках.

Критерий χ^2 -Фридмана основан на ранжировании ряда повторных измерений для каждого объекта выборки. Затем вычисляется сумма рангов для каждого из условий (повторных измерений). Если выполняется статистическая гипотеза об отсутствии различий между повторными измерениями, то можно ожидать примерное равенство сумм рангов для этих условий. Чем больше различаются зависимые выборки по изучаемому признаку, тем больше эмпирическое значение χ^2 -Фридмана.

Эмпирическое значение χ^2 -Фридмана вычисляется после ранжирования ряда повторных измерений для каждого объекта по формуле

$$\chi^2 = \left[\frac{12}{Nk(k+1)} \sum_{i=1}^k R_i^2 \right] - 3N(k+1); df = k - 1,$$

где N – число объектов (испытуемых), k – количество условий (повторных измерений), R_i – сумма рангов для условия i .

При расчетах “вручную” для определения p -уровня пользуются таблицами критических значений. Если $k=3$, $N>9$ или $k>3$, $N>4$, то пользуются обычной таблицей для χ^2 , $df = k-1$ (приложение 7). Если $k=3$, $N < 10$ или $k=4$, $N < 5$, то пользуются дополнительными таблицами критических значений χ^2 -Фридмана (приложение 8).

При отклонении нулевой статистической гипотезы об отсутствии различий принимается альтернативная гипотеза о статистически достоверных различиях выборок по изучаемому признаку – без конкретизации направления различий. Для утверждений о том, что уровень выраженности признака в какой-то из сравниваемых выборок выше или ниже, необходимо парное соотнесение выборок по критерию Т-Вилкоксона.

ЛИТЕРАТУРА

1. **Загвязинский В.И.** Методология и методика дидактического исследования / В.И. Загвязинский. – М.: Педагогика, 1982. – 160 с.
2. **Кремень М.А.** Математические методы в научных исследованиях: Для педагогов и психологов / М.А. Кремень. – Мн.: НИО, 1998. – 92 с.
3. **Кухарев Н.В.** Диагностика познавательных интересов и умственной самостоятельности учащихся / Н.В. Кухарев. – М.: Педагогика, 1998. – 100 с.
4. **Кушнер Ю.З.** Методология и методы педагогического исследования: Учебно-методическое пособие / Ю.З. Кушнер. – Могилев: МГУ им. А.А. Кулешова, 2001. – 112 с.
5. **Майоров А.Н.** Тесты школьных достижений: конструирование и проведение / А.Н. Майоров. – СПб.: Образование и культура, 1997. – 304 с.
6. **Методы педагогических исследований / В.И. Журавлева, Г.П. Ников, М.Н. Скаткин и др.; Под ред. А.И. Пискунова, Г.В. Воробьева.** – М.: Педагогика, 1979. – 255 с.
7. **Наследов А.Д.** Математические методы психологического исследования. Анализ и интерпретация данных: Учебное пособие / А.Д. Наследов. – СПб.: Речь, 2006. – 392 с.
8. **Полонский В.М.** Оценка качества научно-педагогических исследований / В.М. Полонский. – М.: Педагогика, 1987. – 144 с.
9. **Сиденко А.С.** Нужен ли эксперимент практику? // Школьные технологии. – № 1. – 1997. – С. 71-79.
10. **Теория и практика педагогического эксперимента / Под ред. А.И. Пискунова, Г.В. Воробьева.** – М.: Педагогика, 1979. – 208 с.

Приложение 1

СТАНДАРТНЫЕ НОРМАЛЬНЫЕ ВЕРОЯТНОСТИ

В таблице указаны значения площади под кривой единичного нормального распределения, находящиеся справа от z .

Z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4444	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0019
2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
3,2	0,0007									
3,3	0,0005									
3,4	0,0003									

Приложение 2
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ t-СТЮДЕНТА
 (для проверки ненаправленных альтернатив –
 двусторонний критерий)

df	P				df	P			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
1	6,314	12,70	63,65	636,61	46	1,679	2,013	2,687	3,515
2	2,920	4,303	9,925	31,602	47	1,678	2,012	2,685	3,510
3	2,353	3,182	5,841	12,923	48	1,677	2,011	2,682	3,505
4	2,132	2,776	4,604	8,610	49	1,677	2,010	2,680	3,500
5	2,015	2,571	4,032	6,869	50	1,676	2,009	2,678	3,496
6	1,943	2,447	3,707	5,959	51	1,675	2,008	2,676	3,492
7	1,895	2,365	3,499	5,408	52	1,675	2,007	2,674	3,488
8	1,860	2,306	3,355	5,041	53	1,674	2,006	2,672	3,484
9	1,833	2,262	3,250	4,781	54	1,674	2,005	2,670	3,480
10	1,812	2,228	3,169	4,587	55	1,673	2,004	2,668	3,476
11	1,796	2,201	3,106	4,437	56	1,673	2,003	2,667	3,473
12	1,782	2,179	3,055	4,318	57	1,672	2,002	2,665	3,470
13	1,771	2,160	3,012	4,221	58	1,672	2,002	2,663	3,466
14	1,761	2,145	2,977	4,140	59	1,671	2,001	2,662	3,463
15	1,753	2,131	2,947	4,073	60	1,671	2,000	2,660	3,460
16	1,746	2,120	2,921	4,015	61	1,670	2,000	2,659	3,457
17	1,740	2,110	2,898	3,965	62	1,670	1,999	2,657	3,454
18	1,734	2,101	2,878	3,922	63	1,669	1,998	2,656	3,452
19	1,729	2,093	2,861	3,883	64	1,669	1,998	2,655	3,449
20	1,725	2,086	2,845	3,850	65	1,669	1,997	2,654	3,447
21	1,721	2,080	2,831	3,819	66	1,668	1,997	2,652	3,444
22	1,717	2,074	2,819	3,792	67	1,668	1,996	2,651	3,442
23	1,714	2,069	2,807	3,768	68	1,668	1,995	2,650	3,439
24	1,711	2,064	2,797	3,745	69	1,667	1,995	2,649	3,437
25	1,708	2,060	2,787	3,725	70	1,667	1,994	2,648	3,435
26	1,706	2,056	2,779	3,707	71	1,667	1,994	2,647	3,433
27	1,703	2,052	2,771	3,690	72	1,666	1,993	2,646	3,431
28	1,701	2,049	2,763	3,674	73	1,666	1,993	2,645	3,429
29	1,699	2,045	2,756	3,659	74	1,666	1,993	2,644	3,427
30	1,697	2,042	2,750	3,646	75	1,665	1,992	2,643	3,425
31	1,696	2,040	2,744	3,633	76	1,665	1,992	2,642	3,423
32	1,694	2,037	2,738	3,622	78	1,665	1,991	2,640	3,420
33	1,692	2,035	2,733	3,611	79	1,664	1,990	2,639	3,418

Окончание табл.

df	P				df	P			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
34	1,691	2,032	2,728	3,601	80	1,664	1,990	2,639	3,416
35	1,690	2,030	2,724	3,591	90	1,662	1,987	2,632	3,402
36	1,688	2,028	2,719	3,582	100	1,660	1,984	2,626	3,390
37	1,687	2,026	2,715	3,574	110	1,659	1,982	2,621	3,381
38	1,686	2,024	2,712	3,566	120	1,658	1,980	2,617	3,373
39	1,685	2,023	2,708	3,558	130	1,657	1,978	2,614	3,367
40	1,684	2,021	2,704	3,551	140	1,656	1,977	2,611	3,361
41	1,683	2,020	2,701	3,544	150	1,655	1,976	2,609	3,357
42	1,682	2,018	2,698	3,538	200	1,653	1,972	2,601	3,340
43	1,681	2,017	2,695	3,532	250	1,651	1,969	2,596	3,330
44	1,680	2,015	2,692	3,526	300	1,650	1,968	2,592	3,323
45	1,679	2,014	2,690	3,520	350	1,649	1,967	2,590	3,319

Приложение 3

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ Г-ПИРСОНА (Г-СПИРМЕНА)

(для проверки ненаправленных альтернатив, n – объем выборки)

n	p				n	p			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
5	0,805	0,878	0,959	0,991	46	0,246	0,291	0,376	0,469
6	0,729	0,811	0,917	0,974	47	0,243	0,288	0,372	0,465
7	0,669	0,754	0,875	0,951	48	0,240	0,285	0,368	0,460
8	0,621	0,707	0,834	0,925	49	0,238	0,282	0,365	0,456
9	0,582	0,666	0,798	0,898	50	0,235	0,279	0,361	0,451
10	0,549	0,632	0,765	0,872	51	0,233	0,276	0,358	0,447
11	0,521	0,602	0,735	0,847	52	0,231	0,273	0,354	0,443
12	0,497	0,576	0,708	0,823	53	0,228	0,271	0,351	0,439
13	0,476	0,553	0,684	0,801	54	0,226	0,268	0,348	0,435
14	0,458	0,532	0,661	0,780	55	0,224	0,266	0,345	0,432
15	0,441	0,514	0,641	0,760	56	0,222	0,263	0,341	0,428
16	0,426	0,497	0,623	0,742	57	0,220	0,261	0,339	0,424
17	0,412	0,482	0,606	0,725	58	0,218	0,259	0,336	0,421
18	0,400	0,468	0,590	0,708	59	0,216	0,256	0,333	0,418
19	0,389	0,456	0,575	0,693	60	0,214	0,254	0,330	0,414
20	0,378	0,444	0,561	0,679	61	0,213	0,252	0,327	0,411
21	0,369	0,433	0,549	0,665	62	0,211	0,250	0,325	0,408
22	0,360	0,423	0,537	0,652	63	0,209	0,248	0,322	0,405
23	0,352	0,413	0,526	0,640	64	0,207	0,246	0,320	0,402
24	0,344	0,404	0,515	0,629	65	0,206	0,244	0,317	0,399
25	0,337	0,396	0,505	0,618	66	0,204	0,242	0,315	0,396
26	0,330	0,388	0,496	0,607	67	0,203	0,240	0,313	0,393
27	0,323	0,381	0,487	0,597	68	0,201	0,239	0,310	0,390
28	0,317	0,374	0,479	0,588	69	0,200	0,237	0,308	0,388
29	0,311	0,367	0,471	0,579	70	0,198	0,235	0,306	0,385
30	0,306	0,361	0,463	0,570	80	0,185	0,220	0,286	0,361
31	0,301	0,355	0,456	0,562	90	0,174	0,207	0,270	0,341

Приложение 4
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ F-ФИШЕРА
 (для проверки направленных альтернатив)
 ($p=0,05, p=0,01$)

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	10,128	9,552	9,277	9,117	9,013	8,941	8,887	8,845	8,785	8,745	8,638	8,527
	5	6,608	5,786	5,409	5,192	5,050	4,950	4,876	4,818	4,735	4,678	4,527	4,366
	7	5,591	4,737	4,347	4,120	3,972	3,866	3,787	3,726	3,637	3,575	3,410	3,231
	10	4,965	4,103	3,708	3,478	3,326	3,217	3,135	3,072	2,978	2,913	2,737	2,539
	11	4,844	3,982	3,587	3,357	3,204	3,095	3,012	2,948	2,854	2,788	2,609	2,406
	12	4,747	3,885	3,490	3,259	3,106	2,996	2,913	2,849	2,753	2,687	2,505	2,297
	13	4,667	3,806	3,411	3,179	3,025	2,915	2,832	2,767	2,671	2,604	2,420	2,208
	14	4,600	3,739	3,344	3,112	2,958	2,848	2,764	2,699	2,602	2,534	2,349	2,132
	15	4,543	3,682	3,287	3,056	2,901	2,790	2,707	2,641	2,544	2,475	2,288	2,067
	16	4,494	3,634	3,239	3,007	2,852	2,741	2,657	2,591	2,494	2,425	2,235	2,011
	18	4,414	3,555	3,160	2,928	2,773	2,661	2,577	2,510	2,412	2,342	2,150	1,918
	20	4,351	3,493	3,098	2,866	2,711	2,599	2,514	2,447	2,348	2,278	2,082	1,844
	30	4,171	3,316	2,922	2,690	2,534	2,421	2,334	2,266	2,165	2,092	1,887	1,624
	40	4,085	3,232	2,839	2,606	2,449	2,336	2,249	2,180	2,077	2,003	1,793	1,511
	50	4,034	3,183	2,790	2,557	2,400	2,286	2,199	2,130	2,026	1,952	1,737	1,440
	70	3,978	3,128	2,736	2,503	2,346	2,231	2,143	2,074	1,969	1,893	1,674	1,355
100	3,936	3,087	2,696	2,463	2,305	2,191	2,103	2,032	1,927	1,850	1,627	1,286	
200	3,888	3,041	2,650	2,417	2,259	2,144	2,056	1,985	1,878	1,801	1,572	1,192	
∞	3,843	2,998	2,607	2,374	2,216	2,100	2,011	1,940	1,833	1,754	1,519		

$$p=0,01$$

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	34,116	30,816	29,457	28,710	28,237	27,911	27,671	27,489	27,228	27,052	26,597	26,126
	5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,051	9,888	9,466	9,022
	7	12,246	9,547	8,451	7,847	7,460	7,191	6,993	6,840	6,620	6,469	6,074	5,651
	10	10,044	7,559	6,552	5,994	5,636	5,386	5,200	5,057	4,849	4,706	4,327	3,910
	11	9,646	7,206	6,217	5,668	5,316	5,069	4,886	4,744	4,539	4,397	4,021	3,604
	12	9,330	6,927	5,953	5,412	5,064	4,821	4,640	4,499	4,296	4,155	3,780	3,362
	13	9,074	6,701	5,739	5,205	4,862	4,620	4,441	4,302	4,100	3,960	3,587	3,166
	14	8,862	6,515	5,564	5,035	4,695	4,456	4,278	4,140	3,939	3,800	3,427	3,005
	15	8,683	6,359	5,417	4,893	4,556	4,318	4,142	4,004	3,805	3,666	3,294	2,870
	16	8,531	6,226	5,292	4,773	4,437	4,202	4,026	3,890	3,691	3,553	3,181	2,754
	18	8,285	6,013	5,092	4,579	4,248	4,015	3,841	3,705	3,508	3,371	2,999	2,567
	20	8,096	5,849	4,938	4,431	4,103	3,871	3,699	3,564	3,368	3,231	2,859	2,422
	30	7,562	5,390	4,510	4,018	3,699	3,473	3,305	3,173	2,979	2,843	2,469	2,008
	40	7,314	5,178	4,313	3,828	3,514	3,291	3,124	2,993	2,801	2,665	2,288	1,806
	50	7,171	5,057	4,199	3,720	3,408	3,186	3,020	2,890	2,698	2,563	2,183	1,685
	70	7,011	4,922	4,074	3,600	3,291	3,071	2,906	2,777	2,585	2,450	2,067	1,542
	100	6,895	4,824	3,984	3,513	3,206	2,988	2,823	2,694	2,503	2,368	1,983	1,429
200	6,763	4,713	3,881	3,414	3,110	2,893	2,730	2,601	2,411	2,275	1,886	1,281	
∞	6,637	4,607	3,784	3,321	3,019	2,804	2,641	2,513	2,323	2,187	1,793		

Приложение 5
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ F-Фишера
 (для проверки ненаправленных альтернатив)
 ($p=0,05$)

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	?
Степени свободы для знаменателя	3	17,443	16,044	15,439	15,101	14,885	14,735	14,624	14,540	14,419	14,337	14,124	13,903
	5	10,007	8,434	7,764	7,388	7,146	6,978	6,853	6,757	6,619	6,525	6,278	6,017
	7	8,073	6,542	5,890	5,523	5,285	5,119	4,995	4,899	4,761	4,666	4,415	4,144
	10	6,937	5,456	4,826	4,468	4,236	4,072	3,950	3,855	3,717	3,621	3,365	3,081
	11	6,724	5,256	4,630	4,275	4,044	3,881	3,759	3,664	3,526	3,430	3,173	2,884
	12	6,554	5,096	4,474	4,121	3,891	3,728	3,607	3,512	3,374	3,277	3,019	2,726
	13	6,414	4,965	4,347	3,996	3,767	3,604	3,483	3,388	3,250	3,153	2,893	2,597
	14	6,298	4,857	4,242	3,892	3,663	3,501	3,380	3,285	3,147	3,050	2,789	2,489
	15	6,200	4,765	4,153	3,804	3,576	3,415	3,293	3,199	3,060	2,963	2,701	2,397
	16	6,115	4,687	4,077	3,729	3,502	3,341	3,219	3,125	2,986	2,889	2,625	2,318
	18	5,978	4,560	3,954	3,608	3,382	3,221	3,100	3,005	2,866	2,769	2,503	2,189
	20	5,871	4,461	3,859	3,515	3,289	3,128	3,007	2,913	2,774	2,676	2,408	2,087
	30	5,568	4,182	3,589	3,250	3,026	2,867	2,746	2,651	2,511	2,412	2,136	1,789
	40	5,424	4,051	3,463	3,126	2,904	2,744	2,624	2,529	2,388	2,288	2,007	1,639
	50	5,340	3,975	3,390	3,054	2,833	2,674	2,553	2,458	2,317	2,216	1,931	1,548
70	5,247	3,890	3,309	2,975	2,754	2,595	2,474	2,379	2,237	2,136	1,847	1,438	
100	5,179	3,828	3,250	2,917	2,696	2,537	2,417	2,321	2,179	2,077	1,784	1,351	
200	5,100	3,758	3,182	2,850	2,630	2,472	2,351	2,256	2,113	2,010	1,712	1,233	
?	5,027	3,692	3,119	2,788	2,569	2,411	2,290	2,194	2,051	1,947	1,643		

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ F-Фишера
 (для проверки ненаправленных альтернатив)
 ($p = 0,01$)

		Степени свободы для числителя											
		1	2	3	4	5	6	7	8	10	12	24	∞
Степени свободы для знаменателя	3	55,552	49,800	47,468	46,195	45,391	44,838	44,434	44,125	43,685	43,387	42,623	41,83
	5	22,785	18,314	16,530	15,556	14,939	14,513	14,200	13,961	13,618	13,385	12,780	12,14
	7	16,235	12,404	10,883	10,050	9,522	9,155	8,885	8,678	8,380	8,176	7,645	7,079
	10	12,827	9,427	8,081	7,343	6,872	6,545	6,303	6,116	5,847	5,661	5,173	4,641
	11	12,226	8,912	7,600	6,881	6,422	6,102	5,865	5,682	5,418	5,236	4,756	4,228
	12	11,754	8,510	7,226	6,521	6,071	5,757	5,524	5,345	5,085	4,906	4,431	3,907
	13	11,374	8,186	6,926	6,233	5,791	5,482	5,253	5,076	4,820	4,643	4,173	3,649
	14	11,060	7,922	6,680	5,998	5,562	5,257	5,031	4,857	4,603	4,428	3,961	3,439
	15	10,798	7,701	6,476	5,803	5,372	5,071	4,847	4,674	4,424	4,250	3,786	3,263
	16	10,576	7,514	6,303	5,638	5,212	4,913	4,692	4,521	4,272	4,099	3,638	3,114
	18	10,218	7,215	6,028	5,375	4,956	4,663	4,445	4,276	4,030	3,860	3,402	2,876
	20	9,944	6,987	5,818	5,174	4,762	4,472	4,257	4,090	3,847	3,678	3,222	2,693
	30	9,180	6,355	5,239	4,623	4,228	3,949	3,742	3,580	3,344	3,179	2,727	2,179
	40	8,828	6,066	4,976	4,374	3,986	3,713	3,509	3,350	3,117	2,953	2,502	1,935
	50	8,626	5,902	4,826	4,232	3,849	3,579	3,376	3,219	2,988	2,825	2,373	1,790
	70	8,403	5,720	4,661	4,076	3,698	3,431	3,232	3,076	2,846	2,684	2,231	1,622
100	8,241	5,589	4,542	3,963	3,589	3,325	3,127	2,972	2,744	2,583	2,128	1,490	
200	8,057	5,441	4,408	3,837	3,467	3,206	3,010	2,856	2,629	2,468	2,012	1,320	
?	7,886	5,304	4,284	3,720	3,355	3,096	2,901	2,749	2,523	2,363	1,903		

Приложение 6
ЗНАЧЕНИЯ Z-ПРЕОБРАЗОВАНИЯ ФИШЕРА
 (для коэффициентов корреляции)

r_s	0	1	2	3	4	5	6	7	8	9
0,0	0,0000	0,0100	0,0200	0,0300	0,0400	0,0501	0,0601	0,0701	0,0802	0,0902
1	0,1003	0,1105	0,1206	0,1308	0,1409	0,1511	0,1614	0,1717	0,1820	0,1923
2	0,2027	0,2132	0,2237	0,2342	0,2448	0,2554	0,2661	0,2769	0,2877	0,2986
3	0,3095	0,3206	0,3317	0,3428	0,3541	0,3654	0,3769	0,3884	0,4001	0,4118
4	0,4236	0,4356	0,4477	0,4599	0,4722	0,4847	0,4973	0,5101	0,5230	0,5361
5	0,5493	0,5627	0,5763	0,5901	0,6042	0,6184	0,6328	0,6475	0,6625	0,6777
6	0,6931	0,7089	0,7250	0,7414	0,7582	0,7753	0,7928	0,8107	0,8291	0,8480
7	0,8673	0,8872	0,9076	0,9287	0,9505	0,9730	0,9962	1,0203	1,0454	1,0714
8	1,0986	1,1270	1,1568	1,1881	1,2212	1,2562	1,2933	1,3331	1,3758	1,4219
9	1,4722	1,5275	1,5890	1,6584	1,7380	1,8318	1,9459	2,0923	2,2976	2,6467

Приложение 7
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2

ν	P				ν	P			
	0,10	0,05	0,01	0,001		0,10	0,05	0,01	0,001
1	2,706	3,842	6,635	10,829	46	58,641	62,841	71,221	81,431
2	4,605	5,992	9,211	13,817	47	59,774	64,013	72,463	82,752
3	6,251	7,815	11,346	16,269	48	60,907	65,183	73,703	84,069
4	7,779	9,488	13,278	18,470	49	62,038	66,351	74,940	85,384
5	9,236	11,071	15,088	20,519	50	63,167	67,518	76,175	86,694
6	10,645	12,593	16,814	22,462	51	64,295	68,683	77,408	88,003
7	12,017	14,068	18,478	24,327	52	65,422	69,846	78,638	89,308
8	13,362	15,509	20,093	26,130	53	66,548	71,008	79,866	90,609
9	14,684	16,921	21,669	27,883	54	67,673	72,168	81,092	91,909
10	15,987	18,309	23,213	29,594	55	68,796	73,326	82,316	93,205
11	17,275	19,677	24,729	31,271	56	69,919	74,484	83,538	94,499
12	18,549	21,028	26,221	32,917	57	71,040	75,639	84,758	95,790
13	19,812	22,365	27,693	34,536	58	72,160	76,794	85,976	97,078
14	21,064	23,688	29,146	36,132	59	73,279	77,947	87,192	98,365
15	22,307	24,999	30,583	37,706	60	74,397	79,099	88,406	99,649
16	23,542	26,299	32,006	39,262	61	75,514	80,232	89,591	100,887
17	24,769	27,591	33,415	40,801	62	76,630	81,381	90,802	102,165
18	25,989	28,873	34,812	42,323	63	77,745	82,529	92,010	103,442
19	27,204	30,147	36,198	43,832	64	78,860	83,675	93,217	104,717
20	28,412	31,415	37,574	45,327	65	79,973	84,821	94,422	105,988
21	29,615	32,675	38,940	46,810	66	81,085	85,965	95,626	107,257
22	30,813	33,929	40,298	48,281	67	82,197	87,108	96,828	108,525
23	32,007	35,177	41,647	49,742	68	83,308	88,250	98,028	109,793
24	33,196	36,420	42,989	51,194	69	84,418	89,391	99,227	111,055
25	34,382	37,658	44,324	52,635	70	85,527	90,531	100,425	112,317
26	35,563	38,891	45,652	54,068	71	86,635	91,670	101,621	113,577
27	36,741	40,119	46,973	55,493	72	87,743	92,808	102,816	114,834
28	37,916	41,343	48,289	56,910	73	88,850	93,945	104,010	116,092
29	39,087	42,564	49,599	58,320	74	89,956	95,081	105,202	117,347
30	40,256	43,780	50,904	59,722	75	91,061	96,217	106,393	118,599
31	41,422	44,993	52,203	61,118	76	92,166	97,351	107,582	119,850
32	42,585	46,202	53,498	62,508	78	94,374	99,617	109,958	122,347
33	43,745	47,408	54,789	63,891	79	95,476	100,749	111,144	123,595
34	44,903	48,610	56,074	65,269	80	96,578	101,879	112,329	124,839
35	46,059	49,810	57,356	66,641	90	107,565	113,145	124,116	137,208
36	47,212	51,007	58,634	68,008	100	118,498	124,342	135,807	149,449
37	48,363	52,201	59,907	69,370	110	129,385	135,480	147,414	161,582
38	49,513	53,393	61,177	70,728	120	140,233	146,567	158,950	173,618
39	50,660	54,582	62,444	72,080	130	151,045	157,610	170,423	185,573
40	51,805	55,768	63,707	73,428	140	161,827	168,613	181,841	197,450
41	52,949	56,953	64,967	74,772	150	172,581	179,581	193,207	209,265
42	54,090	58,135	66,224	76,111	200	226,021	233,994	249,445	267,539
43	55,230	59,314	67,477	77,447	250	279,050	287,882	304,939	324,831
44	56,369	60,492	68,728	78,779	300	331,788	341,395	359,906	381,424
45	57,505	61,668	69,976	80,107	350	384,306	394,626	414,474	437,487

Приложение 8

**КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2 -ФРИДМАНА
ДЛЯ ТРЕХ ВЫБОРОК ЧИСЛЕННОСТЬЮ $n < 10$
(для проверки ненаправленных альтернатив)**

$n=2$		$n=3$		$n=4$		$n=5$	
χ^2	P	χ^2	P	χ^2	P	χ^2	P
0	0,000	0,000	,000	0,0	1,000	0,0	1,000
1	0,833	0,667	0,944	0,5	0,931	0,4	0,954
3	0,509	2,000	0,528	1,5	0,653	1,2	0,691
4	0,167	2,667	0,361	2,0	0,431	1,6	0,522
		4,667	0,194	3,5	0,273	2,8	0,367
		6,000	0,028	4,5	0,125	3,6	0,182
				6,0	0,069	4,8	0,124
				6,5	0,042	5,2	0,093
				8,0	0,0046	6,4	0,039
						7,6	0,024
						8,4	0,0085

$n=6$		$n=7$		$n=8$		$n=9$	
χ^2	P	χ^2	P	χ^2	P	χ^2	P
0,00	1,000	0,000	1,000	0,00	1,000	0,000	1,000
0,33	0,956	0,286	0,964	0,25	0,967	0,222	0,971
1,00	0,740	0,857	0,768	0,75	0,794	0,667	0,814
1,33	0,570	1,143	0,620	1,00	0,654	0,889	0,865
2,33	0,430	2,000	0,486	1,75	0,531	1,556	0,569
3,00	0,252	2,571	0,305	2,25	0,355	2,000	0,398
4,00	0,184	3,429	0,237	3,00	0,285	2,667	0,328
4,33	0,142	3,714	0,192	3,25	0,236	2,889	0,278
5,33	0,072	4,571	0,112	4,00	0,149	3,556	0,187
6,33	0,052	5,429	0,085	4,75	0,120	4,222	0,154
7,00	0,029	6,000	0,052	5,25	0,079	4,667	0,107
8,33	0,012	7,143	0,027	6,25	0,047	5,556	0,069
9,33	0,0055	8,000	0,016	7,00	0,030	6,222	0,048
10,33	0,0007	8,857	0,0084	7,75	0,018	6,889	0,031
12,00	0,00013	10,286	0,0036	9,00	0,0099	8,000	0,019
		10,571	0,0027	9,25	0,0080	8,222	0,016
		11,143	0,0012	9,75	0,0048	8,667	0,010
		12,286	0,00032	10,75	0,0024	9,556	0,0060
		14,000	0,000021	12,00	0,0011	10,667	0,0035

Окончание табл.

$n=6$		$n=7$		$n=8$		$n=9$	
χ^2	P	χ^2	P	χ^2	P	χ^2	P
				12,25	0,00086	10,889	0,0029
				13,00	0,00026	11,556	0,0013
				14,25	0,000061	12,667	0,00066
				16,00	0,0000036	13,556	0,00035
						14,000	0,00020
						14,222	0,000097
						14,889	0,000054
						16,222	0,000011
						18,000	0,0000006

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ χ^2 -ФРИДМАНА
(для четырех выборок численностью $n < 5$)

$n=1$		$n=2$		$n=3$		$n=4$	
χ^2	P	χ^2	P	χ^2	P	χ^2	P
0,0	1,000	0,0	1,000	0,0	1,000	5,7	0,141
0,6	0,958	0,6	0,958	0,3	0,992	6,0	0,105
1,2	0,834	1,0	0,910	0,6	0,928	6,3	0,094
1,8	0,792	1,8	0,727	0,9	0,900	6,6	0,077
2,4	0,625	2,2	0,608	1,2	0,800	6,9	0,068
3,0	0,542	2,6	0,524	1,5	0,754	7,2	0,054
3,6	0,458	3,4	0,446	1,8	0,677	7,5	0,052
4,2	0,375	3,8	0,342	2,1	0,649	7,8	0,036
4,8	0,208	4,2	0,300	2,4	0,524	8,1	0,033
5,4	0,167	5,0	0,207	2,7	0,508	8,4	0,019
6,0	0,042	5,4	0,175	3,0	0,432	8,7	0,014
		5,8	0,148	3,3	0,389	9,3	0,012
		6,6	0,075	3,6	0,355	9,6	0,0069
		7,0	0,054	3,9	0,324	9,9	0,0062
		7,4	0,033	4,5	0,242	10,2	0,0027
		8,2	0,017	4,8	0,200	10,8	0,0016
		9,0	0,0017	5,1	0,190	11,1	0,00094
				5,4	0,158	12,0	0,000072

Приложение 9

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ U-МАННА-УИТНИ
(для проверки ненаправленных альтернатив $p=0,05$)

N_2	N_1													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	1	2	2	3	3	4	4	5	5	6	6	7	7	8
4	3	4	4	5	6	7	8	9	10	11	11	12	13	13
5	5	6	7	8	9	11	12	13	14	15	17	18	19	20
6	6	8	10	11	13	14	16	17	19	21	22	24	25	27
7	8	10	12	14	16	18	20	22	24	26	28	30	32	34
8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
9	12	15	17	20	23	26	28	31	34	37	39	42	45	48
10	14	17	20	23	26	29	33	36	39	42	45	48	52	55
11	16	19	23	26	30	33	37	40	44	47	51	55	58	62
12	18	22	26	29	33	37	41	45	49	53	57	61	65	69
13	20	24	28	33	37	41	45	50	54	59	63	67	72	76
14	22	26	31	36	40	45	50	55	59	64	67	74	78	83
15	24	29	34	39	44	49	54	59	64	70	75	80	85	90
16	26	31	37	42	47	53	59	64	70	75	81	86	92	98
17	28	34	39	45	51	57	63	67	75	81	87	93	99	105
18	30	36	42	48	55	61	67	74	80	86	93	99	106	112
19	32	38	45	52	58	65	72	78	85	92	99	106	113	119
20	34	41	48	55	62	69	76	83	90	98	105	112	119	127

$p=0,01$

N_2	N_1													
	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3			0	0	0	1	1	1	2	2	2	2	3	3
4	0	1	1	2	2	3	3	4	5	5	6	6	7	8
5	1	2	3	4	4	6	7	7	8	9	10	11	12	13
6	3	4	5	6	6	9	10	11	12	13	15	16	17	18
7	4	6	7	9	9	12	13	15	16	18	19	21	22	24
8	6	7	9	11	11	15	17	18	20	22	24	26	28	30
9	7	9	11	13	13	18	20	22	24	27	29	31	33	36
10	9	11	13	16	16	21	24	26	29	31	34	37	39	42
11	10	13	16	18	18	24	27	30	33	36	39	42	45	48
12	12	15	18	21	21	27	31	34	37	41	44	47	51	54
13	13	17	20	24	24	31	34	38	42	45	49	53	56	60
14	15	18	22	26	26	34	38	42	46	50	54	58	63	66
15	16	20	24	29	29	37	42	46	51	55	60	64	69	74
16	18	22	27	31	31	41	45	50	55	60	65	70	74	79
17	19	24	29	34	34	44	49	54	60	65	70	75	81	86
18	21	26	31	37	37	47	53	58	64	70	75	81	87	93
19	22	28	33	39	39	51	56	63	69	74	81	87	93	99
20	24	30	36	42	42	54	60	67	73	79	86	92	99	106

Приложение 10
КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ Т-ВИЛКОКСОНА
 (для проверки ненаправленных альтернатив)

n	Уровень значимости для одностороннего критерия				n	Уровень значимости для одностороннего критерия			
	0,05	0,025	0,01	0,005		0,05	0,025	0,01	0,005
	Уровень значимости для двустороннего критерия					Уровень значимости для двустороннего критерия			
	0,10	0,05	0,02	0,01		0,10	0,05	0,02	0,01
5	0				28	130	116	101	91
6	2	0	—	—	29	140	126	110	100
7	3	2	0	—	30	151	137	120	109
8	5	3	1	0	31	163	147	130	118
9	8	5	3	1	32	175	159	140	128
10	10	8	5	3	33	187	170	151	138
11	13	10	7	5	34	200	182	162	148
12	17	13	9	7	35	213	195	173	159
13	21	17	12	9	36	227	208	185	171
14	25	21	15	12	37	241	221	198	182
15	30	25	19	15	38	256	235	211	194
16	35	29	23	19	39	271	249	224	207
17	41	34	27	23	40	286	264	238	220
18	47	40	32	27	41	302	279	252	233
19	53	46	37	32	42	319	294	266	247
20	60	52	43	37	43	336	310	281	261
21	67	58	49	42	44	353	327	296	276
22	75	65	55	48	45	371	343	312	291
23	83	73	62	54	46	389	361	328	307
24	91	81	69	61	47	407	378	345	322
25	100	89	76	68	48	426	396	362	339
26	110	98	84	75	49	446	415	379	355
27	119	107	92	83	50	466	434	397	373

Приложение 11

КРИТИЧЕСКИЕ ЗНАЧЕНИЯ КРИТЕРИЯ Н-КРАСКАЛА-УОЛЛЕСА
 (для трех выборок численностью $n < 5$)
 (для проверки ненаправленных альтернатив)

Размеры выборки			H	P	Размеры выборки			H	P
n_1	n_2	n_3			n_1	n_2	n_3		
2	1	1	2,7000	0,500	4	3	2	6,4444	0,008
								6,3000	0,011
2	2	1	3,6000	0,200				5,4444	0,046
								5,4000	0,051
2	2	2	4,5714	0,067				4,5111	0,098
			3,7143	0,200				4,4444	0,102
3	1	1	3,2000	0,300	4	3	3	6,7455	0,010
								6,7091	0,013
3	2	1	4,2857	0,100				5,7909	0,046
			3,8571	0,133				5,7273	0,050
								4,7091	0,092
3	2	2	5,3572	0,029				4,7000	0,101
			4,7143	0,048					
			4,5000	0,067	4	4	1	6,6667	0,010
			4,4643	0,105				6,1667	0,022
								4,9667	0,048
3	3	1	5,1429	0,043				4,8667	0,054
			4,5714	0,100				4,1667	0,082
			4,0000	0,129				4,0667	0,102
3	3	2	6,2500	0,011	4	4	2	7,0364	0,006
			5,3611	0,032				6,8727	0,011
			5,1389	0,061				5,4545	0,046
			4,5556	0,100				5,2364	0,052
			4,2500	0,121				4,5545	0,098
								4,4455	0,103
3	3	3	7,2000	0,004					
			6,4889	0,011	4	4	3	7,1439	0,010
			5,6889	0,029				7,1364	0,011
			5,6000	0,050				5,5985	0,049
			5,0667	0,086				5,5758	0,051
			4,6222	0,100				4,5455	0,099
								4,4773	0,102
4	1	1	3,5714	0,200					
4	2	1	4,8214	0,057	4	4	4	7,6538	0,008
			4,5000	0,076				7,5385	0,011
			4,0179	0,114				5,6923	0,049

Размеры выборки			H	P	Размеры выборки			H	P
n_1	n_2	n_3			n_1	n_2	n_3		
							5,6538	0,054	
4	2	2	6,0000	0,014			4,6539	0,097	
			5,3333	0,033			4,5001	0,104	
			5,1250	0,052					
			4,4583	0,100	5	1	1	3,8571	0,143
			4,1667	0,105					
					5	2	1	5,2500	0,036
4	3	1	5,8333	0,021			5,0000	0,048	
			5,2083	0,050			4,4500	0,071	
			5,0000	0,057			4,2000	0,095	
			4,0556	0,093			4,0500	0,119	
			3,8889	0,129					
					5	4	4	7,7604	0,009
5	2	2	6,5333	0,008			7,7440	0,011	
			6,1333	0,013			5,6571	0,049	
			5,1600	0,034			5,6176	0,050	
			5,0400	0,056			4,6187	0,100	
			4,3733	0,090			4,5527	0,102	
			4,2933	0,122					
					5	5	1	7,3091	0,009
5	3	1	6,4000	0,012			6,8364	0,011	
			4,9600	0,048			5,1273	0,046	
			4,8711	0,052			4,9091	0,053	
			4,0178	0,095			4,1091	0,086	
			3,8400	0,123			4,0364	0,105	
5	3	2	6,9091	0,009	5	5	2	7,3385	0,010
			6,8218	0,010			7,2692	0,010	
			5,2509	0,049			5,3385	0,047	
			5,1055	0,052			5,2462	0,051	
			4,6509	0,091			4,6231	0,097	
			4,4945	0,101			4,5077	0,100	
5	3	3	7,0788	0,009	5	5	3	7,5780	0,010
			6,9818	0,011			7,5429	0,010	
			5,6485	0,049			5,7055	0,046	
			5,5152	0,051			5,6264	0,051	
			4,5333	0,097			4,5451	0,100	
			4,4121	0,109					
5	4	1	6,9545	0,008	5	5	4	7,8229	0,010
			6,8400	0,011			7,7914	0,010	
			4,9855	0,044			5,6657	0,049	

Окончание табл.

Размеры выборки	H	P	Размеры выборки	H	P
	4,8600	0,056		5,6429	0,050
	3,9873	0,098		4,5229	0,099
	3,9600	0,102		4,5200	0,101
5 4 2	7,2045	0,009	5 5 5	8,0000	0,009
	7,1182	0,010		7,9800	0,010
	5,2727	0,049		5,7800	0,049
	5,2682	0,050		5,6600	0,051
	4,5409	0,098		4,5600	0,100
	4,5182	0,101		4,5000	0,102
5 4 3	7,4449	0,010			
	7,3949	0,011			
	5,6564	0,049			
	5,6308	0,050			
	4,5487	0,099			
	4,5231	0,103			

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	3
§ 1. ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ПОДГОТОВКИ И ПРОВЕДЕНИЯ ПЕДАГОГИЧЕСКОГО ЭКСПЕРИМЕНТА	5
1.1. Педагогический эксперимент как метод исследования в педагогике	5
1.2. Основные виды и этапы проведения педагогического эксперимента	10
1.3. Организация педагогического эксперимента	12
1.4. Структура программы эксперимента	16
1.5. Мониторинг педагогического эксперимента	20
§ 2. МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ В ПЕДАГОГИЧЕСКОМ ЭКСПЕРИМЕНТЕ	26
2.1. Измерения и шкалы	26
2.2. Таблицы и графики	30
2.3. Первичные описательные статистики и меры изменчивости	33
2.4. Нормальный закон распределения и его применение	37
2.5. Разработка тестовых шкал	40
2.6. Коэффициенты корреляции	44
2.7. Гипотезы научные и статистические	47
2.8. Методы корреляционного анализа	55
ЛИТЕРАТУРА	70
Приложение 1	71
Приложение 2	73
Приложение 3	75
Приложение 4	76
Приложение 5	78
Приложение 6	80
Приложение 7	81
Приложение 8	82
Приложение 9	84
Приложение 10	85
Приложение 11	86