

МЕТОДЫ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ ПРИ АПРОБАЦИИ ПЕДАГОГИЧЕСКИХ ИННОВАЦИЙ

Кротов Виктор Михайлович

Аннотация. Обосновывается важность освоения педагогами, участвующими в инновационной деятельности в учреждениях общего среднего образования, знаний и умений по применению методов математической статистики для оценки эффективности внедрения результатов инновационных проектов. Приводится информация по теории измерений и измерительным шкалам в педагогике, о табличном и графическом представлении результатов измерений, статистических гипотезах в психолого-педагогических исследованиях, о выборе и условиях применения параметрических и непараметрических статистических критериев (методов) и методов коррекционного анализа для проверки статистических гипотез.

Ключевые слова: метод, математическая статистика, педагогические инновации, апробация инноваций, измерения в педагогическом процессе, измерительные шкалы.

Abstract. The importance of mastering by teachers participating in innovative activities in institutions of general secondary education, knowledge and skills in the use of mathematical statistics methods to assess the effectiveness of their implementation is justified. Information on measurement theory and measurement scales in pedagogy, on tabular and graphical presentation of measurement results, static hypotheses in psychological and pedagogical studies, on selection and conditions of application of parametric and non-parametric static criteria (methods) and correction analysis methods for testing statistical hypotheses is given.

Key words: method, mathematical statistics, pedagogical innovations, approbation of innovations, measurements in the pedagogical process, measuring scales.

В настоящее время в сфере образования отмечается большой интерес к педагогическим инновациям, внедрением которых занимаются не только профессиональные исследователи, но и специалисты образования, не обладающие в достаточном объеме исследовательскими навыками. В случае апробации инноваций такая ситуация имеет объективные предпосылки и носит достаточно распространенный характер. Учитель, директор школы, заместитель директора, методисты учебно-методических кабинетов выступают в качестве организаторов и исполнителей инновационных проектов.

Педагогическая инновация рассматривается как нововведение в области педагогики, представляющее собой научно обоснованную и хорошо продуманную систему организации педагогического процесса, направленного на улучшающие его эффективности. Она имеет общие черты с педагогическим экспериментом:

- потребность в обновлении того или иного аспекта работы или педагогического процесса в целом;
- точное описание цели и задач;
- теоретическая обоснованность;
- описание условий применения;
- предположение об эффективности применения (гипотеза);
- проведение измерений проявления определенных свойств учащихся;
- оценка и анализ эффективности применения.

Поэтому актуальным при этом становится овладение занимающимися данной деятельностью, знаниями и умениями по теории измерений в педагогике, применению методов математической статистики при проведении педагогических исследований.

В ходе внедрения педагогической инновации

проводится измерение проявления определенных свойств учащихся. Невозможно измерить некоторое свойство у всех испытуемых (*генеральной совокупности*). Поэтому ограничиваются лишь относительно небольшой группой представителей соответствующей совокупности учащихся (*выборке*).

Выборка должна быть такой, чтобы была обоснована генерализация выводов выборочного исследования – обобщение, распространение 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, который предполагает обеспечение таких условий, чтобы каждый элемент генеральной совокупности имел равные с другими шансы попасть в выборку;
- стратифицированный случайный отбор, или отбор по свойствам генеральной совокупности, который предполагает предварительное определение тех качеств, которые могут влиять на изменчивость изучаемого свойства. Так в экспериментальных и контрольных классах могут быть учащиеся с разным уровнем обучаемости и обученности, развитием познавательного интереса.

Определяется процентное соотношение численности различающихся по этим качествам групп (страт) в генеральной совокупности и обеспечивается

идентичное процентное соотношение соответствующих групп в выборке. В каждую подгруппу выборки испытуемые подбираются по принципу простого случайного отбора.

Статистическая достоверность, или статистическая значимость, результатов исследования определяется при помощи методов математической статистики. Они предъявляют определенные требования к численности, или *объему выборки*. Наибольший объем выборки необходим при экспериментальной проверке диагностической методики, например, тестовой методики, от 200 до 1000 человек.

Если необходимо сравнивать две выборки по сформированности у учащихся, например, функциональной естественнонаучной грамотности, их общая численность должна быть не менее 50 человек и приблизительно одинаковой в сравниваемых выборках.

При изучении взаимосвязи между какими-либо свойствами объектов, например, их познавательным интересом и обученностью, объем выборки должен быть не меньше 30–35 человек.

Необходимо различать зависимые и независимые выборки. Независимые выборки характеризуются тем, что вероятность отбора любого испытуемого одной выборки (например, из экспериментальных классов) не зависит от отбора любого из испытуемых другой выборки (например, контрольных классов). Зависимыми выборками называют выборки, которые формируются таким образом, что каждому испытуемому одной выборки поставлен в соответствие по определенному критерию испытуемый из другой выборки. Например, каждому испытуемому выборки из контрольных классов отбирается испытуемый из экспериментальных классов с одинаковым уровнем обучаемости.

К зависимым выборкам относят выборки, сформированные из одной и той же группы испытуемых, но в разные моменты времени. Например, до и после экспериментального обучения. Количество объектов в этих выборках всегда одинаковое [2].

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

Основное значение измерения состоит в том, что операции с числами, которые приписаны объектам, позволяют сравнивать между собой эти объекты по состоянию измеряемого свойства. Правило, по которому проводится измерение, порождает тип измерения — «измерительную шкалу».

Шкала — это средство фиксации результатов измерения свойств объектов путем упорядочивания их в определенную числовую систему, в которой отношение между отдельными результатами выражено в соответствующих числах. В процессе упорядочивания каждому элементу выборки ставится в соответствие определенный балл (шкальный индекс), устанавливающий положение наблюдаемого результата на шкале

Например, при исследовании познавательных интересов, учащихся устанавливают границы его проявления: очень высокий уровень развития — очень низкий уровень развития. Между этими границами определяется ряд ступеней. В результате складывается следующая шкала развития познавательных интересов учащихся: очень высокий (1); высокий (2); средний (3); низкий (4); очень низкий (5). Такую шкалу называют дискретной.

В ней множество возможных значений оцениваемой величины конечно.

В теории измерений принято выделять четыре основных вида шкал:

1. Шкала наименований (номинативная, или номинальная).
2. Ранговая шкала (порядковая, или ординарная).
3. Интервальная, или шкала равных интервалов.
4. Шкала отношений, или шкала равных отношений. [1; 3; 4].

Измерение в шкале наименований выражается в том, что определенному свойству или признаку учащихся присваивается соответствующее обозначение или символ (численный, буквенный и т. п.). В процессе измерения по данной шкале происходит процесс классификации или распределения объектов на непересекающиеся группы.

Примером измерения по шкале наименований может являться классификация учащихся по состоянию двух признаков: ведущий канал восприятия информации и успешность выполнения тестового задания.

По ведущему каналу восприятия информации учащихся можно разделить на 3 группы: визуалы, которые воспринимают большую часть информации с помощью зрения, аудиалы — те, кто в основном получает информацию через слуховой канал, кинестетики — люди, воспринимающие большую часть информации через эмоционально-тактильный анализатор (осознание, обоняние, вкус, восприятие движений собственного тела с подключением эмоций).

В состоянии первого признака выделяется в данном случае по три градации (*визуалы, аудиалы, кинестетики*), а второго — две градации (*верный или неверный ответ*). В результате по состоянию двух признаков класс разделяется на шесть групп: *визуалы*, успешно выполнившие тестовое задание; *визуалы*, не выполнившие тестовое задание; *аудиалы*, успешно выполнившие тестовое задание; *аудиалы*, не выполнившие тестовое задание; *кинестетики*, успешно выполнившие тестовое задание; *кинестетики*, не выполнившие тестовое задание. Всем объектам первого класса можно приписать, *например*, число 1, объектам второго класса — число 2, третьего класса — число 3, четвертого — число 4, пятого — число 5, шестого — число 6.

Ранговая шкала является способом оценки, при котором результаты измерений выстраиваются в порядке убывания или возрастания по степени выраженности измеряемого свойства. Поэтому второе название данной шкалы «порядковая шкала». Шкалы порядка

широко используются в педагогике. В частности, повсеместно распространенная шкала школьных отметок в баллах (пятибалльная, десятибалльная и т.д.) может быть отнесена к шкале порядка. При измерении в ранговой шкале из всех свойств чисел учитывается то, что они разные, и то, что одно число больше, чем другое.

Ранг 1 присваивается тому испытуемому, у которого наибольшая или наименьшая выраженность признака (свойства), первое место. Увеличение ранга проводится по уменьшению или увеличению уровня признака.

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 $\frac{N(N+1)}{2}$.

Интервальная (метрическая) шкала применяется при измерениях, для которых 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. Поэтому, применяя эту шкалу, можно судить, насколько больше или насколько меньше выражено свойство при сравнении объектов, но не можем судить о том, во сколько раз больше или меньше оно выражено.

Интервальная шкала, или шкала равных единиц, получается при использовании определенного критерия, с помощью которого удастся измерить интервал между объектами в состоянии изучаемого свойства. Для ее построения необходимо установить критерий, позволяющий определить интервалы между объектами педагогических измерений по состоянию измеряемых свойств, то есть определить на сколько единиц один объект отличается от другого.

Интервальные шкалы так же, как и ранговые, нашли применение в педагогических исследованиях. Так, результаты тестирования измеряются по интервальной шкале.

Шкалу отношений иногда называют шкалой равных отношений. Она является абсолютной метрической. Характерной особенностью данной шкалы является наличие фиксированного нуля, обозначающего полное отсутствие какого-либо свойства или признака. Шкалу отношений можно считать самой информа-

тивной шкалой, допускающей использование любых математических операций над результатами измерений по ней и разнообразных статистических методов.

Измерение в этой шкале отличается от измерения в интервальной только тем, что в ней устанавливается нулевая точка, соответствующая полному отсутствию выраженности измеряемого свойства.

В педагогических измерениях шкала отношений может быть использована при измерении времени выполнения того или иного задания (в секундах, минутах, часах и т.п.), количества ошибок или числа правильно решенных задач.

Анализ измерительных шкал показывает, что для обработки результатов педагогических исследований при определенных условиях могут использоваться все разновидности этих шкал. При этом выбор той или иной из них зависит от того, что и как измеряется.

Выбор исследователем измерительной шкалы во многом предопределяет выбор методики обработки полученных результатов с применением *параметрических* (в случае количественных измерений по интервальной шкале и шкале отношений) или *непараметрических* (в случае использования для этой цели шкалу наименований и порядка) критериев.

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, выступающему в качестве переменной величины, значения которой меняются от объекта к объекту.

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

Если результаты измерений принимают всего лишь несколько возможных значений (до 10–15), то в таблице распределения частот фиксируется частота встречаемости каждого значения признака (*абсолютная частота распределения*). Иногда целесообразно указывать долю результатов, приходящихся на то или иное значение признака (*относительная частота 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 разрядам или интервалам значений признака.

Предположим, в группе испытуемых численностью 40 человек измерено время решения тестовой задачи. Максимальное время составило 67 секунд, минимальное – 32 секунды. Построение таблицы (таблица 1) распределения частот (сгруппированных и накопленных) в этом случае производится поэтапно:

Интервал времени, с	Абсолютная частота (f_n)	Относительная частота (f_n)	Накопленная частота (f_{Σ})
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	

Таблица 1. – Распределение частот

1. Определение размаха: $67-32=35$.

2. Выбор желаемого числа разрядов и интервала разрядов (определяется произвольно). Обычное число разрядов – от 6 до 15. Учитывая, что начинать лучше с 30 или с 31 и заканчивать на 69 или 70, то число разрядов $(40/5=8)$.

3. Определение границ разрядов. Границы соседних разрядов не должны совпадать.

4. Расчет частот абсолютных и относительных частот значения признака для каждого интервала.

5. Расчет накопленных частот. Они показывают, как накапливаются частоты по мере возрастания значений признака.

Для более наглядного представления результатов измерения строится график распределения частот или график накопленных частот – гистограмма или сглаженная кривая распределения (рисунок 1).

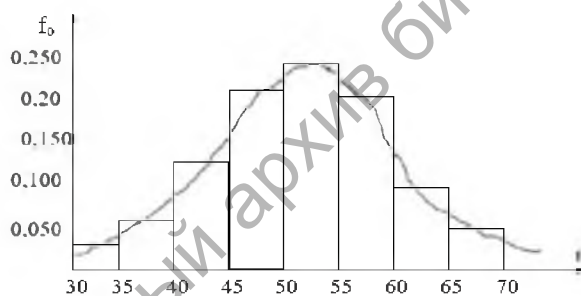


Рисунок 1. – Гистограмма относительных частот

Гистограмма распределения частот представляет собой столбиковую диаграмму, 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).

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 (рисунок 1) [2].

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.

К первичным описательным статистикам обычно относят числовые характеристики распределения измеренного на выборке признака: мода, медиана и среднее арифметическое.

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. 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:

1) если данные содержат нечетное число значений, то медиана есть центральное значение. Пусть 3, 8, 10, 12, 14 – результаты измерений, то $Md=10$;

2) 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 арифметическое (M_x) 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 единственной подходящей мерой центральной тенденции является мода, или модальная категория — та градация номинативной переменной, которая встречается наиболее часто.

Наиболее часто используемой мерой центральной тенденции является среднее арифметическое значение.

Меры центральной тенденции чаще всего используются для сравнения групп по уровню выраженности признака и 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 меж индивидуальной вариации признака.

Для метрических данных используется дисперсия — величина, название которой в науке является синонимом изменчивости. Величина дисперсии получается при усреднении всех квадратов отклонений:

$$\bar{D}_x = \frac{\sum_{i=1}^N (X_i - M_x)^2}{N}$$

Важно 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. Теоретическая дисперсия (\bar{D}_x) не вычисляется, а для вычислений эмпирической дисперсии используется формула:

$$D_x = \frac{\sum_{i=1}^N (X_i - M_x)^2}{N-1}$$

Стандартное отклонение (σ_x , среднеквадратическое отклонение) — положительное значение квадратного корня из дисперсии: $\sigma_x = \sqrt{D_x}$. На практике чаще используется именно стандартное отклонение, а не дисперсия. [2; 4]

Любое психологическое или педагогическое явление, или процесс, рассматриваемое в ходе инновации, носит случайный характер и поэтому говоря о гипотезах, подразумевают статистическую гипотезу, которая является научной гипотезой, допускающей статистическую проверку. В самом общем смысле, *статистическая гипотеза* — это любое предположение о свойствах случайных величин или событий.

Чаще всего при внедрении педагогических инноваций проверяются следующие виды статистических гипотез:

- о свойствах тех или иных числовых характеристик (мода, медиана и др.), описывающих некоторое свойство изучаемого педагогического явления: не превосходит (не меньше) некоторого заданного значения или лежит в определенном диапазоне. Такие гипотезы проверяются на основе параметрических методов математической статистики. При этом требуются количественные измерения изучаемого явления;

- о стохастической (вероятностной) зависимости двух или более признаков, характеризующих некоторое свойство рассматриваемого явления: два или более свойств рассматриваемого педагогического явления стохастически зависимы. Для проверки таких гипотез используются методы корреляционного, но на основе данных количественных измерений;

- о равенстве или различии законов распределения случайных величин, характеризующих изучаемое свойство в двух или более совокупностях рассматриваемых явлений: состояние одного и того же свойства имеет одинаковое или различное распределение в каждой из двух (или более) совокупностей учащихся, отличающихся содержанием, методом или организацией обучения.

Проверка гипотез третьего вида проводится при помощи так называемых критериев значимости. Некоторую их часть можно применять только при количественных измерениях исследуемого свойства, остальные используются и при качественных измерениях.

При проверке гипотез двух последних видов применяются непараметрические методы математической статистики.

В процессе проверки статистических гипотез используют два понятия: нулевая гипотеза (обозначение H_0) и альтернативная гипотеза (обозначение H_1).

Нулевая гипотеза H_0 — это гипотеза о сходстве, а альтернативная H_1 — гипотеза о различии. Таким образом, принятие нулевой гипотезы H_0 свидетельствует об отсутствии различий, а гипотезы H_1 — о наличии различий.

В качестве нулевой H_0 можно рассматривать гипотезу, которая утверждает, что изучаемые выборки взяты из генеральных совокупностей с одинаковым законом распределения, а различие в результатах выборок объясняется чисто случайными причинами.

Примером нулевой гипотезы такого типа в педагогических новшествах является утверждение о том, что различие в результатах выполнения двумя группами учащихся одного и того же диагностического теста вызвано случайными причинами, а на самом деле уровень выполнения этой работы для той и другой группы учащихся одинаков.

Пример альтернативной гипотезы — это утверждение о том, что уровни выполнения тестовых заданий учащихся контрольной и экспериментальной групп различаются и данные различия являются результатом воздействия неслучайных факторов, например,

различных методов или средств обучения.

Нулевая гипотеза H_0 обычно проверяется в процессе сравнения ее с альтернативной гипотезой H_1 .

При внедрении инноваций, основываясь на определенных правилах, необходимо сделать выбор: либо отвергнуть H_0 и тем самым принять H_1 , либо принять H_0 и тем самым отвергнуть H_1 .

Данную границу невозможно четко обозначить, так как в эксперименте всегда присутствуют случайные влияния.

Она основывается на понятии уровня значимости как вероятности ошибочного отклонения нулевой гипотезы. При обозначении данной вероятности чаще всего употребляют букву P .

Стандартными уровнями статистической значимости являются величины 0,05, 0,01 и 0,001.

Принято считать, что низшим уровнем статистической значимости является уровень $P = 0,05$, достаточным – уровень $P = 0,01$ и высшим – уровень $P = 0,001$. Поэтому в статистических таблицах, которые приводятся в приложениях к учебникам по статистике, обычно даются табличные значения для уровней $P = 0,05$, $P = 0,01$ и $P = 0,001$, иногда – для уровней $P = 0,025$ и $P = 0,005$.

При принятии статистического вывода можно пользоваться следующим правилом:

1. Подсчитать эмпирическое значение статистического критерия ($K_{\text{эмп}}$) по выбранному статистическому методу.

2. Сравнить эмпирическое значение статистического критерия ($K_{\text{эмп}}$) с двумя его критическими значениями, соответствующими уровням значимости $P = 0,05$ ($K_{\text{кр}1}$) и $P = 0,01$ ($K_{\text{кр}2}$) для выбранного статистического метода, найденными по таблицам критического значения критерия, приведенным в приложениях.

В разных статистических методах используются свои символические обозначения статистического критерия.

Процесс сравнения эмпирического значения критерия с двумя критическими значениями, которые были найдены в таблицах, можно оптимизировать, расположив все три числа ($K_{\text{эмп}}$, $K_{\text{кр}1}$, $K_{\text{кр}2}$) на оси значимости (рисунок 2).



Рисунок 2. – Ось значимости

На этой оси выделено три зоны: левая зона – зона незначимости, правая – зона значимости, промежуточная зона – зона неопределенности. Границами всех трех зон являются $K_{\text{кр}1}$ для $P = 0,05$ и $K_{\text{кр}2}$ для $P = 0,01$, как это показано на рисунке 2.

Выделяют четыре случая расположения $K_{\text{эмп}}$ на оси значимости.

1. $K_{\text{эмп}}$ находится в зоне незначимости. В этом случае принимается гипотеза H_0 об отсутствии разли-

чий.

2. $K_{\text{эмп}}$ принадлежит зоне значимости. В этом случае принимается гипотеза H_1 о наличии различий, гипотеза H_0 отклоняется.

3. $K_{\text{эмп}}$ находится в зоне неопределенности. В этом случае перед исследователем стоит дилемма. Так, в зависимости от важности решаемой задачи он может считать полученную статистическую оценку:

а) достоверной на уровне 5% ($P = 0,05$) и принять тем самым гипотезу H_1 ;

б) недостоверной на уровне 1% ($P = 0,01$) и принять тем самым гипотезу H_0 .

4. Значение $K_{\text{эмп}}$ точно совпадает либо с $K_{\text{кр}1}$, либо с $K_{\text{кр}2}$. Так, если $K_{\text{эмп}}$ точно совпадает с:

– $K_{\text{кр}1}$, то можно считать, что оценка достоверна точно на уровне в 5% и принять гипотезу H_1 или, напротив, гипотезу H_0 ;

– $K_{\text{кр}2}$, то, как правило, принимается альтернативная гипотеза H_1 о наличии различий, а гипотеза H_0 отклоняется.

При определении критического значения статистического критерия определяется число степеней свободы как количество возможных направлений изменчивости признака. Обычно число степеней свободы линейно зависит от объема выборки, от числа признаков: чем больше эти показатели, тем больше число степеней свободы. Важно, что каждая формула для расчета эмпирического значения критерия обязательно сопровождается правилом определения числа степеней свободы.

Можно выделить следующие этапы принятия статистического решения:

1. Формулирование нулевой и альтернативной гипотез.

2. Определение объема выборки N .

3. Выбор соответствующего уровня значимости или вероятности отклонения нулевой гипотезы.

4. Выбор статистического метода, непосредственно зависящий от того, какая психолого-педагогическая проблема решается при проведении исследования.

5. Вычисление необходимого эмпирического значения критерия $K_{\text{эмп}}$ по экспериментальным данным в соответствии с выбранным статистическим методом.

6. Нахождение в таблице критических значений, подходящих выбранному статистическому методу и соответствующих уровню значимости для $P = 0,05$ и для $P = 0,01$.

7. Сравнение $K_{\text{эмп}}$ с $K_{\text{кр}1}$ ($P = 0,05$) и $K_{\text{кр}2}$ ($P = 0,01$) с построением или без построения оси значимости.

8. Формулирование принятого статистического решения (выбор соответствующей гипотезы H_1 или H_0). [1]

Основными задачами использования методов математико-статистической обработки результатов педагогических исследований чаще всего являются:

– проверка статистических гипотез, достоверности различий между полученными результатами, например, насколько эффективно была подобрана методика обучения по тому или иному предмету;

– выявление меры связи между отдельными явлениями, объектами (корреляционный анализ).

Приступая к планированию измерения исследуемых свойств обучаемых важно определиться с выбором метода (критерия) статистической проверки гипотезы. Под статистическим критерием понимают правило позволяющее принимать или отвергать статистическую гипотезу. Одна и та же гипотеза может быть проверена при помощи самых разных статистических методов. При этом нужно определить, какой из методов годится в той или иной ситуации?

Кроме типов шкал, в которых измерены или представлены изучаемые признаки, на выбор метода статистической проверки гипотезы влияют такие параметры, как количество сравниваемых групп, зависимость или независимость сравниваемых выборок и пр.

Выделяют параметрические и непараметрические критерии математической статистики. Критерий различия, основанный на определенном типе распределения генеральной совокупности (как правило, нормальном) или использующий параметры данной совокупности (средние, дисперсии и т. д.), называют *параметрическим*.

Так, всесторонне используемые на практике критерии, основанные на t-распределении Стьюдента и F-распределении Фишера-Снедекора, являются примерами параметрических критериев. Необходимо отметить, что данные критерии исходят из предположения о *нормальном законе распределения* случайных величин в генеральных совокупностях: 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.

Оно характеризуется своими параметрами: средним (M) и стандартным отклонением (σ). Только эти два значения отличают друг от друга бесконечное множество нормальных кривых, одинаковой формы. 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* (рисунок 3). 1-е распределение отличается от 2-го стандартным отклонением ($\sigma_1 < \sigma_2$), 2-е от 3-го средним арифметическим ($M_1 < M_2$).

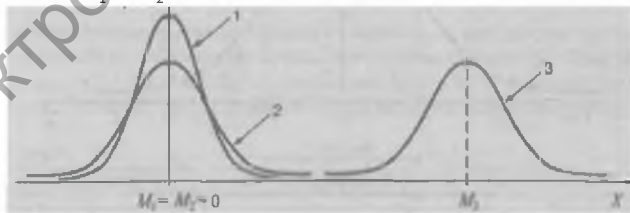


Рисунок 3. – График нормального распределения

Для любого нормального распределения существуют следующие соответствия между диапазонами значений и количеством результатов измерений (в процентах): $M \pm \sigma$ – 68,26%; $M \pm 2\sigma$ – 95,41%; $M \pm 3\sigma$ –

99,72%) (рисунок 4).

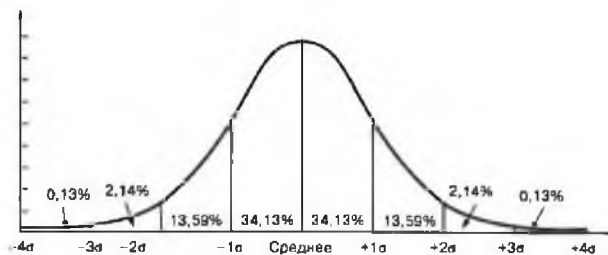


Рисунок 4. – Количество результатов измерений (в процентах) по диапазонам их значения при нормальном распределении

Достоинством методов параметрического анализа данных является тот факт, что они обладают достаточно высокой мощностью. Под *мощностью критерия* имеют в виду его способность избегать ошибок.

Высокая мощность параметрических критериев, обусловлена тем, что данные методы требуют, чтобы имеющиеся данные были описаны в *метрической шкале* (интервальной шкале или шкале отношений).

Нормальное распределение с большей вероятностью получается при выборках более 100 испытуемых (может получиться и при меньшем количестве, а может не получиться и при большем). При использовании параметрических критериев необходима проверка нормальности распределения.

Параметрические статистические критерии позволяют:

- прямо оценить различия средних, полученных в двух выборках (t-критерий Стьюдента);
- прямо оценить различие в дисперсиях (критерий Фишера) [2].

Непараметрическим называют критерий различия, не основанный на предположении о типе распределения генеральной совокупности и не использующий параметры данной совокупности.

Достоинства непараметрических статистических критериев в том, что 1 – они не требуют длительных и сложных расчетов; 2 – не требуют знания характера распределения; 3 – могут быть использованы при любом числе наблюдений; 4 – применимы для признаков, имеющих количественное и полуколичественное выражение.

Применение непараметрических методов является оправданным при условиях, если есть основания считать, что распределение значений признака в генеральной совокупности не соответствует нормальному закону; если есть сомнения в нормальности распределения признака в генеральной совокупности, но выборка слишком мала, чтобы по выборочному распределению судить о распределении в генеральной совокупности; если 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. Если же 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 [1; 3; 4].

Кроме одномерных описательных статистик применяются коэффициенты корреляции. Коэффициент корреляции – двумерная описательная статистика, количественная мера взаимосвязи (совместной изменчивости) двух переменных и 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.

На величину коэффициента корреляции не влияет то, в каких единицах измерения представлены признаки.

К настоящему времени разработано много различных коэффициентов корреляции. Их общей особенностью является то, что они отражают взаимосвязь двух признаков, измеренных в разных шкалах.

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 *t*-Кенделла. Тот и другой коэффициенты требуют для своего применения предварительного ранжирования обеих переменных. Для изучения взаимосвязи двух метрических переменных, измеренных на одной и той же выборке, применяют *коэффициент корреляции r*-Пирсона.

В тех случаях, когда признаки X и Y измеряются в дихотомической шкале наименований и могут принимать значения 0 или 1 в качестве меры связи применяется дихотомический коэффициент корреляции [1; 3; 4].

Приведем примеры наиболее применяемых параметрических и непараметрических статистических критериев для метрических и не метрических шкал, зависимых и независимых выборок (таблица 2).

	Шкалы измерения				
	Метрическая		Ранговая		Наименований
Выборки	зависимые	независимые	зависимые	независимые	зависимые
Параметрические критерии	t-Стьюдента	t-Стьюдента, F- Фишера			
Непараметрические критерии	T-критерий Вилкоксона	U-критерий Манна-Уитни	T-критерий Вилкоксона, G-критерий знаков	U-критерий Манна-Уитни Медианный критерий	Критерий Макнамара, χ^2 -Пирсона
Методы корреляционного анализа	r_{xy} – Пирсона		r_s -Спирмена, t -Кенделла		Дихотомический коэффициент корреляции

Таблица 2. – Статистические критерии

Список литературы

1. Ахметжанова, Г.В. Применение методов математической статистики в психолого-педагогических ис-

следованиях: электронное учебное пособие / Г.В. Ахметжанова, И.В. Антонова. – Тольятти: Изд-во ТГУ, 2016. – 147 с.

2. Методика организации и проведения педаго-

гического эксперимента: метод, рекомендации / авт. сост. В.М. Кротов. — Могилев: МГУ им. А.А. Кулешова, 2008. — 92 с.

3. Новиков, Д.А. Статистические методы в педагогических исследованиях (типовые случаи) / Д.А. Новиков. — М.: МЗ-Пресс, 2004. — 67 с.

4. Петров, П.К. Математико-статистическая об-

работка и графическое представление результатов педагогических исследований с использованием информационных технологий: учеб. пособие / П.К. Петров. — 2-е изд., исправ., и доп. — Ижевск: Издательский центр «Удмуртский университет», 2016. — 176 с.

Дата поступления в редакцию: 27.12.2023

Электронный архив библиотеки МГУ имени А.А. Кулешова