УДК 378

К ВОПРОСУ ПОВЫШЕНИЯ ДОСТОВЕРНОСТИ РЕЗУЛЬТАТОВ ПЕДАГОГИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Е.Н. Солодовникова[1,2], Н.П. Пучков[2]

Кафедра «Прикладная математика, информатика, физика и методика преподавания», ФГБОУ ВПО «Борисоглебский государственный педагогический университет» (1); кафедра «Высшая математика», ФГБОУ ВПО «ТГТУ» (2); eln__@mail.ru

Ключевые слова и фразы: критерии достоверности педагогических данных; педагогические эксперименты; статистические гипотезы; статистические методы.

Аннотация: Поставлена и решена задача обеспечения достоверности результатов педагогических экспериментов с выходными количественными данными за счет комплексного использования методов математической статистики. Сформированы основные требования к организации педагогических экспериментов, являющиеся необходимым условием их достоверности. Рассмотрены практические примеры применения методов математической статистики в реальных педагогических экспериментах.

В последние годы при обосновании достоверности результатов педагогических экспериментов все более требуется использовать математические методы, и, в частности, методы математической статистики. Педагогам необходимо владеть математической статистикой на уровне, позволяющем разбираться в особенностях той или иной статистической процедуры, уметь подобрать подходящие статистические методы, грамотно и осмысленно интерпретировать полученные результаты, так как произвольное применение статистических методов может привести к некачественным и недостоверным результатам.

Не вызывает сомнения тот факт, что применение математики в педагогике позволяет по точности и достоверности приблизить эту науку к естественным. Педагоги убедились, что математика не мешает им ставить эксперименты и проводить исследования, что не следует волноваться от того, что математики раньше них (на своем языке) все рассчитали и наперед определили. Более того, все большее число исследователей в области гуманитарных наук осознает, что математические методы являются эффективным средством корректного представления и сжатия данных. Это не только средство уточнения понятий, но и инструмент систематизации психолого-педагогических знаний.

Проблема измерения в психолого-педагогических науках требует во много раз бóльших усилий для своего решения, чем в других науках. В частности, если в области физического измерения можно применять довольно строгие правила, чтобы значения, приписываемые параметрам, обладали определенными алгебраическими свойствами, то в педагогике и психологии дело обстоит не всегда так.

В данной работе предпринята попытка адаптировать идеи математической обработки данных физического эксперимента, наиболее отработанные на практике, к результатам педагогических экспериментов.

В исследовательской практике «измерение» и «эксперимент» часто используются как синонимы. Однако при проведении педагогического или психологического эксперимента исследователя интересуют причинные связи между переменными, а результатом измерения является всего лишь отнесение испытуемого, либо оцениваемого им объекта к тому или иному классу, точке шкалы или пространству признаков.

Признаки и переменные – это измеряемые явления (количество допущенных ошибок, самооценка личности, социометрический статус и др.). Иногда вместо них используются понятия показателя или уровня (например, показатель успеваемости, уровень адекватности самооценки, коэффициент корреляции процессов и др.). Все они являются случайными величинами, поскольку заранее неизвестно, какое именно значение они примут.

Математическая обработка экспериментальных данных – это оперирование со значением признака, полученного у испытуемых в педагогическом исследовании, определенным при помощи специальных шкал измерений.

При обработке статистических педагогических данных необходимо, чтобы эти данные удовлетворяли следующим критериям.

1. *Содержательный критерий* (критерий операционной валидности), то есть экспериментальный метод должен соответствовать проверяемой гипотезе, а выборка должна определяться предметом и гипотезой исследования. В этой ситуации экспериментатор создает модель идеального объекта экспериментального исследования для своего частного случая, дает ему описание, которое учитывается при формировании выборки.

2. *Критерий эквивалентности испытуемых* (критерий внутренней валидности). Результаты, полученные при исследовании экспериментальной выборки, должны распространяться на каждый ее элемент, то есть должны учитываться все значимые характеристики объекта исследования, различия и выраженности которых могут существенно повлиять на зависимую переменную. При необходимости при обработке данных следует использовать нормировку результатов на величину значимого параметра.

3. *Критерий репрезентативности* (критерий внешней валидности). Группа лиц, участвующих в эксперименте, должна представлять всю часть генеральной совокупности (популяции), по отношению к которой мы можем применить данные, полученные в эксперименте. При этом главная проблема состоит в том, чтобы ответить на вопрос: на какие интересующие нас группы можно распространить результаты проводимого исследования.

Самостоятельная проблема – численность экспериментальной выборки. Из статистических соображений рекомендуется, чтобы численность сравниваемых групп была не менее 30–35 человек (при таком числе испытуемых коэффициенты корреляции выше 0,35 и значимы на уровне 0,05).

Анализ задач педагогических экспериментов позволил обозначить те из них, где необходимо применять методы математической статистики:

1) можно ли считать отличными результаты (количественные) педагогического эксперимента при сопоставлении двух, трех и более эмпирических распределений одного и того же признака? Например, результаты оценки успеваемости студентов одной и той же учебной группы различными преподавателями, или оценки одним и тем же преподавателем успеваемости студентов различных групп;

2) при проведении экспериментов, имеющих количественные результаты, часто оперируют средними значениями признака и степенью «разброса» экспери-

ментальных данных относительно среднего значения (дисперсией). При этом возникает задача проверки гипотезы о достоверности разницы средних, разницы дисперсий. Например, средних значений успеваемости в двух группах, разброса оценок относительно этого среднего;

3) можно ли считать достоверным, что в результате действия каких-либо факторов произошли изменения в измеряемых параметрах. Например, изменение качества знаний в результате использования новой педагогической технологии;

4) можно ли считать достоверным, что результаты изменений обусловлены действием именно того фактора, на который нацелено педагогическое исследование? Например, насколько определяющим (доминирующим) в результатах повышения академической успеваемости явилось действие предлагаемой исследователем методики.

В некоторых, достаточно сложных педагогических экспериментах, все перечисленные задачи требуют одновременного разрешения.

В своей практической деятельности мы использовали (и рекомендуем использовать другим педагогам-исследователям) следующие четыре критерия статистической оценки различий: Пирсона, Стьюдента, Фишера и G-критерий знаков. Они, естественно, не решают все задачи статистической обработки данных педагогических экспериментов, но их знание необходимо для определенной гарантии достоверности результатов. Для тех, кто, имея соответствующую математическую подготовку, желает более глубоко осознать сущность этих (и других) критериев, можно рекомендовать такие литературные источники, как [1–3]; здесь же мы пытаемся изложить теорию (и практику) статистических методов доступно для педагогов, имеющих нематематическое базовое образование, но знакомых с основами статистики.

Представленные ниже критерии статистической оценки различий применяются для проверки статистических гипотез – утверждений относительно неизвестного параметра. Нулевая гипотеза H_0 – это гипотеза об отсутствии различий; ей альтернативная H_1 – о значимости различий, которая оценивается уровнем значимости критерия – вероятностью того, что различия считаются существенными, а на самом деле они случайные. Когда указывается, что различия достоверны на 5 % уровня значимости (или при $p \le 0{,}05$), то имеют в виду, что вероятность того, что они все-таки недостоверны, составляет 0,05 (низкий уровень статистической значимости). Для результатов педагогических исследований можно считать 0,01 достаточным уровнем статистической значимости.

При использовании представленных ниже критериев мы руководствовались следующими правилами: если эмпирическое значение критерия не меньше критического значения, соответствующего $p \le 0{,}05$, то H_0 отклоняется, но еще нельзя определенно принять H_1; если эмпирическое значение критерия не меньше критического значения, соответствующего $p \le 0{,}01$, то принимается H_1, если же эмпирическое значение критерия не достигает критического значения, соответствующего $p \le 0{,}05$, то принимается гипотеза H_0.

Основные задачи и используемые критерии

1. *Сравнение распределений одного и того же признака.* Такая задача возникает при сопоставлении двух, трех и более эмпирических распределений одного и того же признака или при их сопоставлении с известным теоретическим распределением, например нормальным, когда исследователь применяет методы математической статистики, справедливые для нормального закона распределения, и надо показать, что полученное эмпирическое распределение сходное нормальному.

В этом случае используется критерий Пирсона (χ^2 – «хи» квадрат), который позволяет выяснить: с одинаковой ли частотой встречаются разные значения признака в двух и более эмпирических распределениях. Чем выше расхождения между двумя сопоставляемыми распределениями, тем больше эмпирическое значение χ^2.

Возможности этого критерия проявляются на больших выборках при $n \geq 30$; при малой выборке критерий дает лишь приближенное значение.

В процессе реализации метода рассматриваются гипотезы H_0: эмпирическое распределение некоторого признака не отличается от его теоретического распределения; альтернативная гипотеза H_1 состоит в том, что распределения отличаются.

Алгоритм вычисления χ^2 выражается формулой

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{\text{эмп}\,i} - f_{\text{теор}\,i})^2}{f_{\text{теор}\,i}},$$

где k – количество разрядов признака; i – порядковый номер разряда; $f_{\text{эмп}\,i}$, $f_{\text{теор}\,i}$ – эмпирическая и теоретическая частоты по i-му разряду признака соответственно.

Если сравниваются между собой эмпирические распределения, например X_1 и X_2 (табл. 1), то статистика критерия рассчитывается по формуле

$$\chi^2_{\text{эмп}} = \frac{1}{n_1 n_2} \sum_{i=1}^{k} \frac{(n_2 m_i - n_1 s_i)^2}{m_i + s_i}, \tag{1}$$

при этом число степеней свободы $k - 1$.

По статистическим таблицам (χ^2-распределения) для данного числа степеней свободы находится критическое значение $\chi^2_{\text{кр}}$. Если $\chi^2_{\text{эмп}} < \chi^2_{\text{кр}}$, то расхождения между распределениями статистически недостоверны, в противном случае они достоверны.

2. *Проверка гипотез о средних значениях.* В этом случае для проверки гипотез о достоверности разницы средних при анализе количественных данных в популяциях с нормальным распределением используется метод (критерий) Стьюдента.

Метод Стьюдента различен для независимых и зависимых выборок. Независимые выборки получаются при исследовании двух различных групп испытуемых (например, контрольной и опытной групп). К зависимым выборкам относятся,

Таблица 1

Распределение частот исследуемого признака

Разряды признака	Распределение X_1 (частоты)	Распределение X_2 (частоты)	$\sum$
α_1	m_1	s_1	$m_1 + s_1$
α_2	m_2	s_2	$m_1 + s_1$
…	…	…	…
α_k	m_k	s_k	$m_k + s_k$
$\sum$	n_1	n_2	$n_1 + n_2$

например, результаты одной и той же группы испытуемых до и после воздействия независимой переменной.

Гипотеза H_0 состоит в том, что разность между средними значениями двух выборок (статистически) равна нулю, альтернативная гипотеза состоит в том, что разность отлична от нуля.

В случае независимых выборок используют статистику

$$t_{\text{эмп}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \tag{2}$$

где $\bar{x}_1$, $\bar{x}_2$ – выборочные средние первой и второй выборок соответственно; σ_1^2, σ_2^2 – дисперсии первой и второй выборок соответственно; n_1, n_2 – объемы первой и второй выборок соответственно, $n_1, n_2 \geq 30$.

Число степеней свободы $\nu = n_1 - 1 + n_2 - 1$.

Для данного ν из статистических таблиц находят значения $t_{\text{кр}}$.

Если $|t_{\text{эмп}}| > t_{\text{кр}}$, то нулевая гипотеза отбрасывается и принимается альтернативная, если же $|t_{\text{эмп}}| < t_{\text{кр}}$, то разница средних недостоверна.

В случае зависимых выборок используют статистику

$$t_{\text{эмп}} = \frac{\sum d}{\sqrt{\frac{n\sum d^2 - \left(\sum d\right)^2}{n-1}}}, \tag{3}$$

где n – число пар данных (объемы выборок одинаковы); d – разность между результатами в каждой паре; $\sum d$ – сумма частных разностей; $\sum d^2$ – сумма их квадратов.

Число степеней свободы в случае зависимых выборок для определения $t_{\text{кр}}$ $\nu = n - 1$.

Если $|t_{\text{эмп}}| > t_{\text{кр}}$, то принимаем альтернативную гипотезу, то есть считаем разницу средних достоверной. Если $|t_{\text{эмп}}| < t_{\text{кр}}$, то разница средних недостоверна.

3. *Проверка гипотез о дисперсиях.* Гипотеза H_0: разность между дисперсиями двух выборок равна нулю; альтернативная гипотеза H_1: разность между дисперсиями отлична от нуля.

При независимой выборке для проверки гипотез используется критерий отношения двух выборочных дисперсий Фишера: $F = \frac{\sigma_1^2}{\sigma_2^2}$, причем в качестве числителя обычно берут бóльшую из двух дисперсий. Количество степеней свободы определяется по формулам: $k_1 = n_1 - 1$, $k_2 = n_2 - 1$, где n_1, n_2 – объемы первой и второй выборок соответственно.

Из статистических таблиц F-распределения находят критические значения. Далее процедура проверки гипотез стандартная.

Если берутся две зависимые выборки из генеральной совокупности с нормальным распределением одинакового объема, то проверка гипотезы осуществляется по критерию Стьюдента

$$t = \frac{\sigma_1^2 - \sigma_2^2}{\sqrt{\frac{4\sigma_1^2\sigma_2^2}{n-2}(1-r_{12}^2)}}, \text{ где } r_{12} = \frac{\overline{x_1 x_2} - \bar{x}_1\bar{x}_2}{\sigma_1\sigma_2}, \quad (4)$$

где n – число пар наблюдений; r_{12} – коэффициент корреляции, найденный по парам наблюдений.

Количество степеней свободы определяется по формуле $\nu = n - 2$. Далее находят критические значения и делают выводы. Если $|t_{эмп}| > t_{кр}$, то принимают альтернативную гипотезу, то есть существует отличие дисперсий. Если же $|t_{эмп}| < t_{кр}$, то разница дисперсий недостоверна.

4. *G-критерий знаков* на наш взгляд недостаточно освещен в литературе, хотя отличается своей простотой и значимостью.

В педагогических исследованиях часто бывает необходимо доказать, что в результате действия каких-либо факторов произошли достоверные изменения («сдвиги») в измеряемых показателях.

Можно создать специальные экспериментальные условия, положительно влияющие на те или иные показатели, и сопоставить замеры, произведенные до и после экспериментального воздействия. Если сдвиги окажутся статистически достоверными, то экспериментальные воздействия можно считать существенными (эффективными).

Если нет контрольной группы, то сдвиг в экспериментальной группе может объясняться действиями самых различных причин. И хотя констатируется, что сдвиг произошел, его нельзя приписывать именно исследуемым факторам воздействия. Если же в экспериментальной группе сдвиги окажутся достоверными, а в контрольной – недостоверными, то это действительно может свидетельствовать об эффективности воздействия.

Если нет контрольной группы, но есть несколько экспериментальных групп, различающихся по условиям и способам воздействия на них, то имеется возможность уточнить специфические воздействия экспериментальных или естественно действующих факторов, хотя при этом воздействие неучтенных факторов может оказаться еще более мощным.

G-критерий знаков предназначен для установления общего направления сдвига исследуемого признака, а именно, в какую сторону в выборке в целом изменяются значения признака при переходе от первого измерения ко второму. Критерий знаков применим к сдвигам, которые можно определить качественно (например, изменение отрицательного отношения к учебе на положительное), и к сдвигам, которые можно измерить количественно (например, повышение успеваемости).

Если в процессе эксперимента у большинства испытуемых показатели во втором замере имеют какую-то определенную тенденцию, то необходимо доказать, что этот сдвиг является преобладающим. Сдвиги, которые кажутся доминирующими, называют типическими. А те, которые встречаются редко – нетипическими; если показатели не изменяются, то сдвиги называют нулевыми.

Суть критерия знаков состоит в том, чтобы определить «не слишком ли много наблюдается нетипических сдвигов», чтобы сдвиг в типическом направлении был преобладающим. Чем меньше нетипических сдвигов, тем более вероятно, что преобладание типического сдвига является доминирующим, то есть он статически достоверен.

Из-за ограничения таблиц критических значений для использования данного критерия количество наблюдений в обоих замерах должно быть не менее 5 и не более 300.

Расчет критерия знаков G следующий:

1) подсчитывается число n, равное сумме положительных и отрицательных сдвигов; нулевые реакции из рассмотрения исключаются;

2) сдвиги в преобладающем направлении измерений считаются «типичными». Если типичный сдвиг установить невозможно (положительных и отрицательных сдвигов поровну), то данный критерий не используется;

3) количество нетипичных сдвигов считается эмпирическим значением G;

4) по таблицам критических значений критерия знаков G для данного n находят критические значения G;

5) если $G_{\text{эмп}} \leq G_{\text{кр}}$, то сдвиг в типичную сторону может считаться достоверным.

Комплексное применение рассмотренных методов проиллюстрируем методикой обработки данных двух реальных педагогических экспериментов, проведенных на базе ФГБОУ ВПО «Борисоглебский государственный педагогический институт».

Эксперимент № 1

1. В 2008 г. было организовано тестирование 55 студентов-выпускников после прохождения ими педагогической практики по оценке уровня их методической компетентности. В 2009 г. аналогичное исследование было организовано для 105 выпускников. Результаты оценки по 100-балльной шкале представлены в табл. 2, где, в зависимости от баллов, отрезок от 30 до 100 баллов разбит на 7 разрядов, частоты соответствуют количеству студентов. Сравним эти баллы.

Сформулируем гипотезы. H_0: эмпирические распределения оценок выпускников 2008 и 2009 гг. не различаются между собой; H_1 альтернативная гипотеза.

Р е ш е н и е. Имеем две выборки объемом $n_1 = 55$ и $n_2 = 105$. Средние выборочные для групп студентов 2008 и 2009 гг. выпуска равны $\bar{x}_1 = 72$ и $\bar{x}_2 = 61$ соответственно, а дисперсии $\sigma_1^2 = 432$ ($\sigma_1 = 20{,}78$), $\sigma_2^2 = 411{,}6$ ($\sigma_2 = 20{,}3$)*. Возникает вопрос: насколько достоверны различия средних выборочных и выборочных дисперсий?

Так как выборки независимые, то критерий Стьюдента по формуле (2) $t = 3{,}24$.

Таблица 2

Результаты оценки методической компетентности

Разряды по баллам	Выпускники 2008 (частоты)	Выпускники 2009 (частоты)	$\sum$
91–100	12	12	24
81–90	12	13	25
71–80	7	14	21
61–70	9	12	21
51–60	6	15	21
41–50	5	19	24
31–40	4	20	24
$\sum$	55	105	160

* Все промежуточные данные округляются до сотых.

Количество степеней свободы $n_1 + n_2 - 2 = 158$.

По статистическим таблицам распределения Стьюдента для $\nu = 158$ критические значения $t_{кр} \geq 1{,}975$ для $p \leq 0{,}05$ и $t_{кр} = 2{,}6$ для $p \leq 0{,}01$. Эмпирическое значение $t_{кр} = 3{,}24 > 2{,}6$ лежит в зоне значимости, следовательно различия между средними достоверны.

Критерий отношения двух выборочных дисперсий Фишера $F = \dfrac{\sigma_1^2}{\sigma_2^2} = \dfrac{432}{411{,}6} = 1{,}05$. По таблице F-распределения для $k_1 = 104$ и $k_2 = 54$ $F_{кр} = 1{,}51$ для $p \leq 0{,}05$ и $F_{кр} = 1{,}85$ для $p \leq 0{,}01$.

Эмпирическое значение $F = 1{,}05 < 1{,}51$ лежит в зоне незначимости, поэтому различия между дисперсиями выборок недостоверны.

Проведенные расчеты говорят о том, что в среднем выпуск 2008 г. более компетентен в методике, чем выпуск 2009 г. при одинаковых уровнях разброса оценок относительно средней.

2. Чтобы сделать окончательный вывод о том, как коррелируются между собой данные 2008 и 2009 гг., применим критерий Пирсона («хи» квадрат).

По формуле (1) и данным табл. 2 находим, что $\chi^2_{эмп} = 10{,}85$.

Количество степеней свободы $\nu = 6$; для этого значения по статистическим таблицам находим критические значения $\chi^2_{кр} = 12{,}6$ для $p \leq 0{,}05$, $\chi^2_{кр} = 16{,}8$ для $p \leq 0{,}01$. Эмпирическое значение $10{,}85 < 12{,}6$ лежит в зоне незначимости. Поэтому приходим к выводу, что эмпирические распределения оценок 2008 и 2009 гг. не различаются между собой, хотя различия средних выборочных достоверны.

Эксперимент № 2

Для эксперимента были отобраны две группы: контрольная и экспериментальная. В экспериментальной группе занятия проводились с использованием принципов инновационного обучения. В начале и в конце эксперимента было проведено тестирование знаний студентов (табл. 3).

Сформулируем гипотезы.

1. H_0: сдвиг в сторону повышения успеваемости в экспериментальной группе по окончанию обучения является случайным; H_1 – альтернативная гипотеза.

2. H_0: сдвиг в сторону повышения успеваемости в контрольной группе по окончанию обучения является случайным; H_1 – альтернативная гипотеза.

Прежде чем проверять гипотезы, необходимо убедиться в том, что до эксперимента студенты экспериментальной и контрольной групп имеют одинаковый уровень академической подготовки. Это можно сделать двумя способами: используя критерии Стьюдента и Фишера, показать, что недостоверны разницы средних значений и дисперсий или, используя критерий Пирсона, показать, что распределение баллов тестирования в обоих группах не различаются между собой.

I способ. В результате использования известных формул для подсчета средних значений и дисперсий находим, что $\bar{x}_1 = 65$, $\bar{x}_2 = 67$; $\sigma_1^2 = 276{,}7$; $\sigma_2^2 = 262{,}7$ (индекс 1 соответствует экспериментальной группе, 2 – контрольной). Выборки независимые, так как получены при исследовании двух различных групп.

Согласно критерия Стьюдента (2) $t = 0{,}47$.

Количество степеней свободы $n_1 + n_2 - 2 = 58$.

По статистическим таблицам для $\nu = 58$ критические значения $t_{кр} = 2{,}002$ для $p \leq 0{,}05$ и $t_{кр} = 2{,}66$ для $p \leq 0{,}01$. Эмпирическое значение $t_{эмп} = 0{,}47$ лежит в зоне незначимости, поэтому различия между средними недостоверные.

Таблица 3

Результаты тестирования знаний студентов

Номер студента в списке группы	Экспериментальная группа			Контрольная группа		
	до	после	сдвиг	до	после	сдвиг
1	45	80	+	45	50	+
2	50	50		85	50	–
3	50	50		75	75	
4	55	80	+	85	85	
5	65	75	+	75	75	
6	55	75	+	50	50	
7	45	60	+	50	95	+
8	75	75		80	70	–
9	70	75	+	100	100	
10	75	75		85	65	–
11	100	100		55	45	–
12	50	70	+	60	50	–
13	75	75		75	75	
14	65	75	+	60	45	–
15	50	50		45	50	+
16	45	60	+	75	75	
17	50	50		45	70	+
18	75	75		75	75	
19	65	75	+	50	45	–
20	60	45	–	65	60	–
21	45	55	+	50	70	+
22	85	90	+	45	70	+
23	75	75		60	50	–
24	100	100		75	100	+
25	85	90	+	75	75	
26	80	60	–	70	80	+
27	50	65	+	75	60	–
28	75	50	–	45	55	+
29	45	50	+	60	75	+
30	90	95	+	80	100	+

Согласно критерия Фишера:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{276,7}{262,7} = 1,053; \quad k_1 = k_2 = 29.$$

$$F_{\text{кр}} = 1,85 \text{ для } p \le 0,05 \text{ и } F_{\text{кр}} = 2,41 \text{ для } p \le 0,01.$$

Эмпирическое значение $F = 1,053$ лежит в зоне незначимости, поэтому принимается гипотеза о том, что различия между дисперсиями выборок недостоверны.

Таблица 4

Распределение частот признака для экспериментальной и контрольной групп

Оценка	Экспериментальная группа	Контрольная группа	$\sum$
5	5	5	10
4	11	12	23
3	14	13	27
$\sum$	30	30	60

II способ. С целью упрощения расчетов проведем группировку экспериментальных данных. Считаем, что получение от 81 до 100 баллов соответствует оценке «пять», от 61 до 80 – «четыре» и от 40 до 60 – «три». В этом случае имеет место табл. 4, аналогичная табл. 1.

Находим по формуле (1) соответствующее значение $\chi^2_{эмп} = 0{,}08$ при количестве степеней свободы, равным 2. По статистическим таблицам χ^2 для $\nu = 2$ критические значения $\chi^2_{кр} = 5{,}99$ для $p \le 0{,}05$ и $\chi^2_{кр} = 9{,}21$ для $p \le 0{,}01$. Экспериментальное значение $\chi^2_{кр} = 0{,}08 < 5{,}99$, поэтому принимается гипотеза H_0: распределение баллов в экспериментальной и контрольной группах не различаются между собой.

Таким образом, можно считать, что экспериментальная и контрольная группы до эксперимента не имеют различий по уровню знаний.

Если в эксперименте ставилась только задача определить наличие достоверных изменений (сдвигов) в измеряемом показателе, то для ее решения используется G-критерий знаков.

Подсчитываем количество положительных, отрицательных и нулевых сдвигов (в табл. 3 они отмечены символами «+», «–» и не отмечены) (табл. 5).

В экспериментальной группе $n = 19$, типический сдвиг – положительный, $G_{эмп} = 3$. Критические значения: $G_{кр} = 5$ при $p \le 0{,}05$; $G_{кр} = 4$ при $p \le 0{,}01$. Так как $G_{эмп} < G_{кр}$, то принимается гипотеза H_1. В контрольной группе $n = 21$, типический сдвиг – положительный, $G_{эмп} = 10$. Критические значения $G_{кр} = 6$ при $p \le 0{,}05$, $G_{кр} = 4$ при $p \le 0{,}01$. Поскольку $G_{эмп} > G_{кр}$, то принимается гипотеза H_0.

При сравнении результатов видно, что в экспериментальной группе сдвиги являются на самом деле достоверными. Это свидетельствует о том, что инновационное обучение является действенным.

Таблица 5

Подсчет количества сдвигов в группах

Количество сдвигов	Экспериментальная группа	Контрольная группа
Положительных	16	11
Отрицательных	3	10
Нулевых	11	9
n	19	21

Если в задачи эксперимента дополнительно входило желание обнаружить сравнительное изменение средних значений показателя академической успеваемости и доказать его достоверность, а также показать, что в результате эксперимента не произошло изменение разброса данных относительно среднего (что равносильно «сконцентрированному» уровню знаний), то следует опять же, как и на предварительном этапе обработки экспериментальных данных, воспользоваться критерием Стьюдента.

По результатам табл. 3 (экспериментальная группа) имеем средние значения $\bar{x}_1 = 65$, $\bar{x}_2 = 70$ и $\sigma_1^2 = 276{,}7$ и $\sigma_2^2 = 262{,}7$.

Выборки в данном случае зависимые и поэтому по формуле (3) находим, что $t_{\text{эмп}} = 2{,}24$.

Для $\nu = 58$ $t_{\text{кр}} = 2{,}002$ для $p \le 0{,}05$ и $t_{\text{кр}} = 2{,}663$ для $p \le 0{,}01$. Эмпирическое значение $t_{\text{эмп}} = 2{,}24$ лежит в зоне неопределенности $2{,}002 < 2{,}24 < 2{,}663$, поэтому различие средних значений можно считать имеющим место (гипотеза H_0 отвергается), но еще нельзя определенно принять гипотезу H_1: средние баллы отличаются. Насколько неустойчиво это состояние показывает следующее обстоятельство: если, например, в табл. 3 в 28-й строке для экспериментальной группы внести несущественные для основной задачи (наличие сдвига) изменения, заменив 75 на 65, а 50 на 60, то средние примут значения $\bar{x}_1 = 64{,}7$, $\bar{x}_2 = 69{,}5$ и соответствующее значение $t = 2{,}81$ окажется в зоне значимости $2{,}81 > 2{,}663$ и станет справедливой гипотеза H_1: отличие средних баллов достоверно.

Эти результаты позволяют сделать вывод о том, что нельзя бездоказательно (без использования методов математической статистики) увеличение среднего значения, например успеваемости или уменьшение дисперсии (разброса данных), ставить в заслугу использованной методике (инновации).

Можно убедиться, что аналогичные расчеты для контрольной группы показывают справедливость гипотезы H_0: различие средних значений недостоверно, то есть рост средней успеваемости – случайное явление.

Критерий Стьюдента для зависимых выборок определяется по формуле (4), где дополнительно определено, что $\overline{x_1 x_2} = 4451{,}7$; $r_{12} = 0{,}377$ и $t = 0{,}32$. Эмпирическое значение $t_{\text{кр}}$ лежит в зоне значимости ($t_{\text{кр}} = 1{,}85$ для $p \le 0{,}05$), следовательно, различие дисперсий недостоверно, эксперимент не изменил разброс данных относительно их среднего значения.

Заключение

Достоверность результатов педагогических экспериментов непременно нуждается во всестороннем, комплексном применении методов математической статистики, так как результаты педагогических измерений образуют пространство случайных чисел, где действуют отличные от детерминированных процессов законы.

Список литературы

1. Новиков, Д.А. Статистические методы в педагогических исследованиях (типовые случаи) / Д.А. Новиков. – М. : М3 Пресс, 2004. – 67 с.

2. Глас, Д. Статистические методы в педагогике и психологии / Д. Глас, Д. Стенли. – М. : Прогресс, 1976. – 495 с.

3. Крамер, Г. Математические методы статистики / Г. Крамер. – М. : Мир, 1975. – 648 с.

The Problem of Improving the Accuracy of Pedagogical Experiment Results

E.N. Solodovnikova[1,2], N.P. Puchkov[2]

Department "Applied Mathematics, Computer Science, Physics and Teaching Methods", Borisoglebsk State Training University (1);
Department "Higher Mathematics", TSTU (2);
eln__@mail.ru

Key words and phrases: pedagogical experiments; statistical hypotheses; statistical methods; validation criteria of educational data.

Abstract: The problem of ensuring the reliability of the results of pedagogical experiments with quantitative data output by the integrated use of mathematical statistics is formulated and solved. The basic requirements for organizing pedagogical experiments, which are a prerequisite for their authenticity, are formed. Practical examples of application of mathematical statistics in real pedagogical experiments are considered.

Zur Frage der Erhöhung der Richtigkeit der Ergebnisse der pädagogischen Experimente

Zusammenfassung: Es ist die Aufgabe der Gewährleistung der Richtigkeit der Ergebnisse der pädagogischen Experimente mit den quantitativen Ausgabedaten auf Rechnung der Komplexbenutzung der Methoden der matematischen Statistik gestellt und gelöst. Es sind die Hauptanforderungen zur Organisation der pädagogischen Experimente formuliert. Es sind die praktischen Beispiele der Anwendung der Methoden der matematischen Statistik in den realen pädagogischen Experimenten betrachtet.

Sur le problème de l'augmentation de l'authenticité des résultats des expériments pédagogiques

Résumé: Est posé et résolu le problème de l'assurance de l'authenticité des résultats des expériments pédagogiques avec les données de sorie quantitatives compte tenu de l'utilisation complexe de la statistique mathématique. Sont formulées les essentielles exigences pour l'organisation des expériments pédagogiques présentant une condition nécessaire de leur authenticité. Sont examinés les exemples pratiques de l'application des méthodes de la statistique mathématique dans les expériments pédagogiques réelles.

Авторы: ***Солодовникова Елена Николаевна*** – аспирант кафедры «Высшая математика», ФГБОУ ВПО «ТГТУ»; старший преподаватель кафедры «Прикладная математика, информатика, физика и методика преподавания», ФГБОУ ВПО «Борисоглебский государственный педагогический университет»; ***Пучков Николай Петрович*** – доктор педагогических наук, профессор, заведующий кафедрой «Высшая математика», ФГБОУ ВПО «ТГТУ».

Рецензент: ***Молотков Николай Яковлевич*** – доктор педагогических наук, профессор кафедры «Физика», ФГБОУ ВПО «ТГТУ».