

УДК 311.2
51-7

Анализ номинальных данных педагогического эксперимента (на примере формирования конфликтологической компетенции)

Надежда В. Сиврикова¹, Елена В. Моисеева², Надежда А. Соколова³,
Татьяна Г. Пташко⁴, Елена Г. Черникова⁵, Наталья П. Артемьева⁶

¹ Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: Bobronv@cspu.ru

ORCID: <https://orcid.org/0000-0001-9757-8113>

² Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: mev64@inbox.ru

ORCID: <https://orcid.org/0000-0002-8119-8084>

³ Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: sokolovana-2011@mail.ru

ORCID: <https://orcid.org/0000-0003-2110-8320>

⁴ Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: ptashko75@mail.ru

ORCID: <https://orcid.org/0000-0002-0235-4190>

⁵ Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: Helen1268@mail.ru

ORCID: <https://orcid.org/0000-0002-8858-9790>

⁶ Южно-Уральский государственный гуманитарно-педагогический университет, Челябинск, Россия

E-mail: artemevanp@cspu.ru

ORCID: <https://orcid.org/0000-0001-9672-5484>

DOI: 10.26907/esd.17.1.12

Дата поступления: 21 января 2020; Дата принятия в печать: 30 ноября 2020

Аннотация

В статье рассматривается проблема использования методов математической статистики в педагогических исследованиях. Цель работы заключается в выявлении и описании возможностей применения методов анализа номинальных данных при обработке результатов педагогических экспериментов. Авторы провели анализ частоты использования статистических критериев в 100 публикациях в журналах, входящих в Web of Science и Российский индекс научного цитирования. Критериями отбора публикаций выступали: 1) год выпуска; 2) рецензирование рукописей; 3) наличие свободного доступа от издателя к публикации; 4) наличие количественных данных о результатах испытаний до и после педагогического воздействия;

5) наличие выводов об эффективности педагогического воздействия. Было установлено, что в 45 % публикаций не использовались методы математической статистики для проверки научных гипотез. Из применяемых авторами методов математической статистики чаще используются Т-критерий Стьюдента и χ^2 критерий Пирсона. На примере исследования, посвященного формированию конфликтологической компетенции у будущих педагогов, представлены возможности использования таких критериев, как: χ^2 Пирсона, точный критерий Фишера, критерий фи и критерий V Крамера, критерий D Сомерса.

Ключевые слова: номинальные данные, ординальные данные, статистические критерии, уровень значимости различий, величина эффекта, хи квадрат критерий Пирсона, точный критерий Фишера, критерий Крамера, критерий Сомерса.

Analysis of Nominal Pedagogical Experimental Data Using the Example of the Development of Conflict Competence

Nadezhda V. Sivrikova¹, Elena V. Moiseeva², Nadezhda A. Sokolova³,
Tatyana G. Ptashko⁴, Elena G. Chernikova⁵, Natalya P. Artemeva⁶

¹ South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: Bobronv@cspu.ru

ORCID: <https://orcid.org/0000-0001-9757-8113>

² South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: mev64@inbox.ru

ORCID: <https://orcid.org/0000-0002-8119-8084>

³ South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: sokolovana-2011@mail.ru

ORCID: <https://orcid.org/0000-0003-2110-8320>

⁴ South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: ptashko75@mail.ru

ORCID: <https://orcid.org/0000-0002-0235-4190>

⁵ South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: Helen1268@mail.ru

ORCID: <https://orcid.org/0000-0002-8858-9790>

⁶ South Ural State Humanitarian and Pedagogical University, Chelyabinsk, Russia

E-mail: artemevanp@cspu.ru

ORCID: <https://orcid.org/0000-0001-9672-5484>

DOI: 10.26907/esd.17.1.12

Submitted: 21 January 2020; Accepted: 30 November 2020

Abstract

The paper examines the problem of using mathematical statistics methods in pedagogical research. It identifies and describes the possibilities of using different methods of analysis of nominal data in processing the results of pedagogical experiments. The authors analyzed the frequency of use of statistical criteria in 100 publications in journals included in Web of Science and the Russian Index of Scientific Citation. The criteria for selecting publications were (1) year of issue; (2) reviews of manuscripts; (3) free access to the publication from the publisher; (4) availability of quantitative data on test results before and after pedagogical impact; (5) conclusions about the effectiveness

of pedagogical impact. It was found that 45% of publications did not use mathematical statistics methods to test scientific hypotheses. Of the mathematical statistics methods used by the authors, the T-Test and Chi-square Test are more often used. For the study example of developing conflict readiness in future teachers, it presents the possibilities of using: Chi-square Test, Fisher's Exact Test (Phi test), Cramer's V Test and Somers' d Test.

Keywords: nominal data, ordinal data, statistical criteria, difference significance level, effect magnitude, Pearson chi-square test, Fisher's exact test, Cramer's V test, Somers' d test.

Введение

Современная педагогика представляет собой сложную науку, которая имеет множественные взаимосвязи с разными дисциплинами. Другие науки обеспечивают ее необходимыми инструментами для экспериментального исследования педагогических явлений – к ним прежде всего относится высшая математика и математическая статистика.

В научной литературе отмечается, что важным условием востребованности результатов исследований международным сообществом (Grjibovski, Ivanov, & Gorbatova, 2016) и одним из признаков научной значимости работы (Rakhypbekov & Grjibovski, 2015) является грамотное использование методов математической обработки данных. Всесторонний, глубокий статистический анализ эмпирических данных является одним из критериев достоверности интерпретаций и обоснованности выводов научного исследования (Azevich & Alekseeva, 2016).

В научной литературе широко обсуждается необходимость применения методов математической статистики в научных исследованиях в целом (Rakhypbekov & Grjibovski, 2015; Unguryanu & Grjibovski, 2011) и в отдельных науках (Akanov et al., 2013; Grjibovski, Ivanov, & Gorbatova, 2016; Gumennikova, Kaydalova, & Ryabinova, 2015). Анализ качества педагогических работ, проводимый отечественными исследователями (Zagvyazinsky, 2009; Zakirova, 2015; Solnyshkov, 2009), не касается этой проблематики.

Практика использования методов математической статистики в педагогических экспериментах

Опыт знакомства с отечественными исследованиями позволил нам предположить, что статистическая обработка данных является слабым местом современных педагогических исследований. Поэтому мы провели анализ частоты использования в работах по педагогике различных статистических критериев.

Для этого в реферативных базах данных публикаций в научных журналах мы выбрали 100 публикаций, описывающих результаты педагогического эксперимента: 50 публикаций из базы Web of Science (WoS) и 50 публикаций из Российского индекса научного цитирования (РИНЦ). В качестве ключевых слов для поиска использовались: педагогический и/или формирующий эксперимент. Критериями отбора публикаций стали: 1) год выпуска – не ранее 2015; 2) публикация в рецензированном издании; 3) наличие свободного доступа от издателя к публикации; 4) наличие в публикации количественных данных о результатах испытаний до и после педагогического воздействия; 5) наличие выводов об эффективности педагогического воздействия.

Для анализа частоты использования различных методов математической статистики использовался контент-анализ данных (Таблица 1). Следует отметить, что в 7 исследованиях использовалось 2 критерия, поэтому общее количество наблюдаемых случаев превышает значение 100, а сумма по столбцам с указанием процентов выше 100 %.

Таблица 1. Частота использования статистических критериев в научных публикациях по педагогике

Статистический критерий	WoS (50 публикаций)		РИНЦ (50 публикаций)		Всего (100 публикаций)
	Кол-во	%	Кол-во	%	Кол-во (%)
χ^2 Пирсона	7	14	9	18	16
Критерий ϕ	2	4	4	8	6
T-критерий Вилкоксона	6	12	2	4	8
U-критерий Манна-Уитни	2	4	2	4	4
T-критерий Стьюдента	9	18	6	12	15
Дисперсионный анализ (ANOVA)	4	8	0	0	4
G-критерий знаков	1	2	0	0	1
Иное	3	6	2	4	5
Отсутствует	17	34	28	56	45
Заявлен, но не указан	0	0	3	6	3

В 45 % публикаций не использовались методы математической статистики для подтверждения или опровержения гипотезы на основе эмпирических данных.

В 3 % работ авторы не указали методы математической статистики, которые они использовали. В разделе «методы исследования» в этом случае они просто писали «методы математической статистики». Что примечательно, раздел статьи с описанием результатов при этом содержал выражения типа « $p < 0,05$ ». Подобное явление было обнаружено и в публикациях, посвященных описанию медицинских исследований (Rakhymbekov & Grjibovski, 2015). В двух работах были обнаружены грубые ошибки в названиях используемых методов. Например, в одной из работ было указано на использование критерия χ^2 по Стьюденту (Lapshin, 2015), которого не существует.

Хотелось бы отметить, что в публикациях, входящих в РИНЦ, часто встречаются описания методов математической статистики с использованием большого количества терминов, определений, обозначений и формул. Эти описания призваны создать впечатление научности, значимости работы. Вместе с тем они мало связаны с самой работой и ее результатами и не позволяют читателю составить полное представление о структуре исследования и используемых математических методах. Используемые авторами описания не позволяют другим исследователям в случае необходимости воспроизвести описанный эксперимент.

Из применяемых авторами статистических критериев чаще всего в рассмотренных публикациях встречались критерии Стьюдента и χ^2 Пирсона. Реже остальных применялись критерий знаков и дисперсионный анализ.

Следует отметить, что среди публикаций, входящих в WoS, отмечается большее разнообразие в методах математической обработки данных, а также большая точность в их описании и интерпретации данных. Только среди этих публикаций нами были отмечены случаи, когда методы математической статистики приводятся в аннотации статьи. Среди публикаций в РИНЦ подобных случаев обнаружить не удалось.

Нами было замечено, что исследователи практически полностью игнорируют непараметрические критерии, предназначенные для анализа таблиц сопряженности. При этом в подавляющем большинстве исследований результаты представлены в виде частот распределения участников исследования по группам. Подобную картину можно объяснить тем, что эти критерии не предназначены для оценки из-

менений в исследованиях со схемой «до и после». Но, по мнению Д. А. Новикова, структура педагогического эксперимента позволяет получить косвенное подтверждение динамики изучаемых признаков без анализа изменений, произошедших у каждого испытуемого (Novikov, 2004).

Выявленные в ходе анализа недостатки педагогических исследований требуют рассмотрения возможностей использования методов анализа номинальных данных в педагогических экспериментах.

Структура педагогического эксперимента

Допускаемые исследователями ошибки, с нашей точки зрения, продиктованы непониманием роли статистических методов в педагогическом исследовании. Поэтому важно обозначить структуру педагогического эксперимента и место методов математической статистики в ней.

Педагогический эксперимент нацелен на обоснование факта, что состояние некоторого объекта (учащихся) изменилось в требуемую сторону благодаря специально организованному воздействию. Это можно сделать благодаря выполнению двух условий: 1) доказать, что произошли изменения объекта в ожидаемом направлении за время проведения эксперимента; 2) доказать, что подобные изменения без организованного воздействия за аналогичный период не появляются. Чтобы выполнить второе условие, достаточно найти схожий объект и посмотреть, какие изменения произойдут с ним за аналогичный период времени, но без воздействия. Поэтому традиционно в педагогических экспериментах объект исследования делится на 2 группы. В одной группе проводится воздействие (например, обучение по инновационной методике) – она называется экспериментальной. Во второй группе воздействие не проводится (обучение проходит по традиционной методике или не проводится вовсе) – она называется контрольной. Следует отметить, что в педагогике встречаются исследования с несколькими экспериментальными группами, в каждой из которых реализуются отдельные элементы (условия) инновационной методики обучения, и одной контрольной группой.

Д. А. Новиков (Novikov, 2004) предлагает графическое изображение структуры эксперимента (Рисунок 1). Стрелками отмечены процедуры сравнения результатов замеров исследуемых характеристик. При этом подразумевается, что методы сбора данных одинаковы при всех замерах, проводимых в эксперименте.

Констатации динамики исследуемого признака от начального (до воздействия) к конечному (после воздействия) состоянию экспериментальной группы (сравнение III) недостаточно для доказательства того факта, что эта динамика является следствием экспериментального воздействия. Поэтому педагогический эксперимент обязательно должен включать в себя анализ результатов сравнения IV.

Однако само наличие контрольной группы требует доказать, что перед воздействием исследуемые группы (экспериментальная и контрольная) не имели различий по уровню измеряемых характеристик (т. е. были эквивалентны). Поэтому алгоритм действий при проведении педагогического эксперимента заключается в следующем:

- 1) на основании сравнения I (Рисунок 1) установить, что до воздействия между экспериментальной и контрольной группой отсутствуют различия;
- 2) воздействовать на экспериментальную группу;
- 3) сравнить экспериментальную и контрольную группы после эксперимента (измерение II на Рисунке 1), чтобы установить, что после воздействия между экспериментальной и контрольной группой наблюдаются различия.

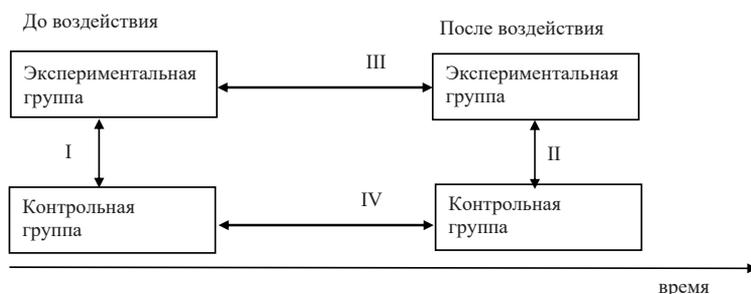


Рисунок 1. Структура педагогического эксперимента по Д. А. Новикову

Подобный алгоритм позволяет косвенным образом реализовать процедуру оценки динамики в экспериментальной группе (сравнение III на Рисунке 1), исключая влияние общих для экспериментальной и контрольной групп условий и факторов. Роль статистических методов в этом алгоритме заключается в корректном и достоверном обосновании отсутствия или наличия различий между экспериментальной и контрольной группами.

Мы предлагаем рассмотреть, как это делается, на примере исследования формирования конфликтологической компетенции. При выборе статистических критериев в данной статье мы опирались на наиболее распространенный в проанализированных статьях способ представления результатов исследования. Чаще всего в них представлены результаты срезов до и после проведения экспериментального воздействия в экспериментальной и контрольной группах без учета динамики результатов отдельных испытуемых.

Анализ таблиц сопряженности в педагогических исследованиях

Результаты исследования формирования конфликтологической компетенции у будущих педагогов

В качестве примера будут выступать данные исследования по формированию конфликтологической компетенции у будущих педагогов. Цель эксперимента состояла в оценке эффективности инновационной методики подготовки будущих педагогов к разрешению конфликтов в сфере образования.

Из числа студентов, принимавших участие в исследовании, были сформированы две группы: экспериментальная (ЭГ) и контрольная (КГ). В экспериментальной группе (60 человек) использовалась методика обучения, основанная на инкорпорировании обучающихся в информационное пространство конфликтологической подготовки и активизации конфликтологической деятельности обучающихся как субъектов. Контрольную группу составили 57 студентов, которые обучались по стандартной программе, предполагавшей лекционный курс конфликтологии. Они не подвергались воздействию, которое проводилось в экспериментальной группе.

С помощью специально разработанного комплекса методик до и после осуществления воздействия в обеих исследуемых группах определялся уровень конфликтологической компетенции. Под конфликтологической компетенцией в исследовании понимается определенный набор знаний и умений, необходимый специалисту для диагностики, профилактики и разрешения конфликтов. Поэтому в ходе эксперимента оценивались следующие элементы: уровень сформированности знаний о конфликте, уровень умений разрешать конфликты, уровень владения методами диагностики и прогнозирования конфликта. Полученные

оценки суммировались в общую оценку уровня сформированности конфликтологической компетенции. Полученные в ходе эксперимента результаты представлены в Таблице 2.

Таблица 2. Динамика уровня сформированности конфликтологической компетенции студентов

Сравниваемые показатели		ЭГ	КГ
		кол-во студентов	
до	низкий уровень	50	47
	средний уровень	10	10
после	низкий уровень	13	25
	средний уровень	29	24
	высокий уровень	18	8

Таблица составлена таким образом, чтобы результаты измерения «до» были представлены в виде четырехпольной таблицы, а результаты измерения «после» в виде многопольной таблицы. Это позволяет рассмотреть применение математических методов для разных видов таблиц.

Критерий χ^2 Пирсона

Наибольшей популярностью среди математических методов обработки результатов исследования, предназначенных для поиска различий в распределении частот, пользуются способы, в основе которых лежит сопоставление фактических частот, найденных в ходе исследования, с предполагаемыми частотами. Такая популярность связана с универсальностью, т. е. возможностью применения подобных методов для решения разного рода задач и сравнения данных, полученных на любых выборках. Одним из таких методов является критерий согласия χ^2 Пирсона. Он предназначен для определения уровня достоверности различий между двумя выборками.

Применение любого статистического критерия имеет свои ограничения. Для использования критерия χ^2 Пирсона необходимо соблюдать следующие условия (McHugh, 2013):

1) результаты исследования должны быть представлены в номинальной или порядковой шкале;

2) независимость наблюдений и независимость групп (метод не подходит для сравнения одной выборки в ситуации до и после воздействия, т. е. сравнение III и IV на Рисунке 1);

3) очень большое количество ячеек таблицы (более 20) может затруднить выполнение следующего условия и интерпретацию результатов, хотя данное условие не является строгим и критерий не предъявляет прямых ограничений к числу ячеек;

4) число наблюдений в каждой ячейке таблицы сопряженности должно быть равно или больше 5 (или 10) для таблиц, состоящих из 4-х ячеек. Для таблиц большей размерности важно, чтобы число ячеек со значениями меньше 5 (10) было меньше 20 % (Kroonenberg & Verbeek, 2018). По сути, это условие накладывает ограничение на объем выборки. Для использования критерия объем должен быть достаточно большим (когда общее количество $n \geq 20$ или $n \geq 40$), т. к. с увеличением числа наблюдений растет и точность критерия;

расчет эмпирического критерия производится с использованием только абсолютных значений частот, т. е. доли от целого или проценты использовать нельзя.

Обратимся к применению критерия χ^2 Пирсона в нашем исследовании (Таблица 2). Чтобы определить эквивалентность (сходство распределений исследуемого признака) экспериментальной и контрольной групп на начальном этапе эксперимента (до воздействия) с помощью данного критерия, необходимо установить число наблюдений для каждой ячейки таблицы, которое указывает на верность нулевой гипотезы (т. е. гипотезы об отсутствии различий). Это число принято называть ожидаемой или теоретической частотой.

Для того чтобы вычислить ожидаемое количество наблюдений, необходимо сумму рядов умножить на сумму столбцов и разделить на общее число наблюдений во всех сравниваемых группах (Grjibovski, 2008; Kim, 2017). Чтобы наглядно продемонстрировать применение данного правила, составим таблицу с данными предварительного измерения, посчитав маргинальные итоги (суммы по строкам и столбцам) (Таблица 3).

Таблица 3. Таблица сопряженности данных на предварительном этапе эксперимента

Уровень сформированности конфликтологической компетенции	ЭГ	КГ	Всего
Низкий уровень	50 (А)	47 (В)	97
Средний уровень	10 (Б)	10 (Г)	20
Всего	60	57	117

Для определения ожидаемого числа людей с низким уровнем конфликтологической компетенции в экспериментальной группе (ячейка А) необходимо найти сумму по столбцу ЭГ (50+10=60) и умножить ее на сумму по строке «низкий уровень» (50+47=97). Полученное произведение $60 \cdot 97 = 5820$ необходимо разделить на общее число наблюдений, т. е. 117. В результате получается, что ожидаемая частота в ячейке А составляет $5820/117=49,7$.

Для ячейки, отражающей число студентов со средним уровнем конфликтологической компетенции в экспериментальной группе (ячейка Б) ожидаемое число наблюдений будет равно $20 \cdot 60/117 = 10,26$. Для ячейки, отражающей число людей с низким уровнем конфликтологической компетенции в контрольной группе (ячейка В), ожидаемая частота будет равна 47,26 (В), а для ячейки, отражающей число людей со средним уровнем конфликтологической компетенции в контрольной группе (ячейка Г), ожидаемая частота составит 9,74 (Г). Следует обратить внимание, что ожидаемые (в отличие от фактических) значения могут быть представлены дробными числами. В нашем примере все полученные ожидаемые значения оказались больше 5 (т. е. третье условие применения критерия соблюдено).

Для определения эмпирического значения критерия χ^2 Пирсона необходимо воспользоваться формулой (Grjibovski, 2008):

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^c \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

где i – номер ряда (строки, от 1 до r), j – номер столбца (от 1 до c), O_{ij} – фактическое количество наблюдений в ячейке ij , E_{ij} – ожидаемое число наблюдений в ячейке ij .

В нашем случае получится:

$$\chi^2 = \frac{(50 - 49,7)^2}{49,7} + \frac{(47 - 47,26)^2}{47,26} + \frac{(10 - 10,26)^2}{10,26} + \frac{(10 - 9,74)^2}{9,74} = 0,016$$

Чтобы сделать вывод о верности или неверности нулевой гипотезы, необходимо сравнить полученное эмпирическое значение критерия с критическим значением, которое рассчитывается исходя из числа степеней свободы. Для определения числа степеней свободы используется следующая формула (Grijbovski, 2008):

$$v = (r - 1) * (c - 1),$$

где r – число строк, c – число столбцов.

Для любой таблицы размерностью 2x2 ячейки число степеней свободы будет равно 1, т. к. такая таблица имеет 2 столбца и 2 строки, т. е. $(2 - 1) * (2 - 1) = 1$. Критические значения критерия χ^2 Пирсона для разных степеней свободы можно найти в математических справочниках. В нашем случае оно составит 3,841 при уровне значимости 0,05. При сравнении эмпирического (0,016) и критического (3,841) значений критерия делаем вывод, что первое меньше второго. Это позволяет отвергнуть альтернативную гипотезу и принять нулевую гипотезу, т. е. гипотезу об отсутствии различий между сравниваемыми группами на начальном этапе эксперимента.

Рассчитаем значение критерия χ^2 Пирсона для сравнения выборок после воздействия. Для этого составим таблицу, содержащую теоретические частоты (Таблица 4).

Таблица 4. Таблица сопряженности данных на завершающем этапе эксперимента

Уровень сформированности конфликтологической компетенции	Фактические частоты		Всего	Теоретические частоты	
	ЭГ	КГ		ЭГ	КГ
Низкий уровень	13	25	38	19,49	18,51
Средний уровень	29	24	53	27,18	25,82
Высокий уровень	18	8	26	13,33	12,67
Всего	60	57	117	60	57

Подставляем полученные значения в формулу:

$$\chi^2 = \frac{(13 - 19,49)^2}{19,49} + \frac{(25 - 18,51)^2}{18,51} + \frac{(29 - 27,18)^2}{27,18} + \frac{(24 - 25,82)^2}{25,82} + \frac{(18 - 13,33)^2}{13,33} + \frac{(8 - 12,67)^2}{12,67} = 8,03$$

В данном случае речь идет о таблице размерностью 3x2 ячейки. Следовательно, при расчете числа степеней свободы получаем $(3 - 1) * (2 - 1) = 2$. Согласно табличным значениям, при $v=2$ критическое значение критерия χ^2 Пирсона составит 5,991 (при $p \leq 0,05$). Эмпирическое значение критерия, полученное нами, составило 8,03. Оно оказалось больше критического, что дает основания для отвержения нулевой гипотезы и принятия альтернативной гипотезы.

Можно заметить, что от величины различий между эмпирическими и ожидаемыми частотами в отдельных ячейках зависит значение критерия χ^2 Пирсона. Чем больше фактические значения отличаются от теоретических, тем больше будет эмпирическое значение критерия, а уровень значимости (p) будет меньше.

Результаты применения критерия χ^2 Пирсона позволяют сделать вывод о том, что после воздействия экспериментальная группа стала отличаться от контрольной по частоте встречаемости в ней студентов с разным уровнем конфликтологической компетенции. Причем в экспериментальной группе наблюдается большее, чем в контрольной группе, число студентов с высоким и средним уровнем сформированности исследуемой компетенции.

Сопоставляя результаты исследования до и после воздействия, можно сделать вывод о том, что эти различия стали следствием педагогического воздействия, осуществляемого в экспериментальной группе.

Следует отметить, что в данном примере была установлена статистическая связь между распределением учащихся и фактором обучающего воздействия. Сама по себе она не дает основания для вывода о наличии причинно-следственных, или «достоверных», отношений (Grjibovski, 2008). Однако структура эксперимента (доказанность отсутствия различий между исследуемыми группами до воздействия) позволяет сделать вывод, что различия в конце эксперимента обусловлены обучением, которое проводилось в экспериментальной группе.

Точный критерий Фишера

Рассмотренный выше метод дает возможность вычислить приблизительную (ее еще называют асимптотической) оценку вероятности принятия статистической гипотезы (Lin, Chang, & Pal, 2015). В математической статистике существует метод, который позволяет рассчитать точную вероятность для всевозможных четырехпольных таблиц. Поэтому он является более мощным по сравнению с критерием χ^2 Пирсона. Этот метод носит название «точный критерий Фишера» (Lin, Chang, & Pal, 2015).

Условия применения точного критерия Фишера схожи с условиями применения критерия χ^2 Пирсона (Warner, 2013):

- 1) сравниваемые переменные должны быть измерены в номинальной шкале;
- 2) сравниваемые наблюдения и группы должны быть независимыми.

Точный критерий Фишера имеет преимущества перед критерием χ^2 Пирсона. Он подходит для сравнения малых выборок ($n < 20$) и может использоваться в случае, если значения ожидаемого числа наблюдений в какой-либо ячейке таблицы менее 5.

Точный критерий Фишера определяется по формуле (Grjibovski, 2008; A. Hess & Hess, 2017; Nowacki, 2017):

$$P = \frac{(A + B)! (B + \Gamma)! (A + B)! (B + \Gamma)!}{N! A! B! B! \Gamma!}$$

где ! – факториал.

Для нашего примера данный вариант можно использовать только при сравнении результатов экспериментальной и контрольной групп до воздействия:

$$P = \frac{(50 + 47)! (10 + 10)! (50 + 10)! (47 + 10)!}{97! 50! 47! 10! 10!} = 0,54$$

Обращаем внимание, что в данном случае в результате вычисления сразу получаем информацию об уровне различий между сравниваемыми группами. Исследователь избавлен от вычисления эмпирического значения критерия и необходимости соотношения его с критическим значением. Однако при больших объемах выборки вычисление факториала становится сложной задачей. Поэтому мы реко-

мендуем использовать для расчетов специальные программы или онлайн калькуляторы.

Для расчета точной вероятности при использовании таблиц с размером, большим чем 2×2 , рекомендуется использовать расширение Фримена-Хэлтона для точного критерия Фишера. Однако расчет точного критерия Фишера в этом случае становится настолько трудоемким, что для некоторых таблиц его вычисление невозможно даже с использованием персонального компьютера. Поэтому мы воспользовались онлайн калькулятором, чтобы вычислить точное значение p для наших данных после осуществления воздействия ($p=0,018$).

Таким образом, на основании расчетов до экспериментального воздействия следует принять нулевую гипотезу (об отсутствии различий в распределении исследуемого признака в экспериментальной и контрольной группах), а после воздействия следует принять альтернативную гипотезу (о наличии различий в распределении исследуемого признака в экспериментальной и контрольной группах).

Критерий ϕ и критерий V Крамера

Методы математической статистики позволяют не только определить наличие различий между разными группами, но и оценить силу связи между исследуемыми переменными, т. е. величину эффекта. В нашем примере данная задача может заключаться в оценке силы связи между педагогическим воздействием и его результатом.

Для определения силы связи между номинальными переменными используются специальные критерии. Их значение варьируется в диапазоне от 0 до 1. В отличие от коэффициентов корреляции, эти критерии не могут иметь отрицательные значения, поскольку измерение признака в номинальной шкале не содержит информацию типа «больше-меньше». Для определения величины эффекта критерия χ^2 Пирсона чаще всего используются критерий ϕ (фи, phi) или критерий V Крамера (McHugh, 2013). Первый из этих критериев применяется только в случае анализа четырехпольных таблиц (2×2), т. к. для многопольных таблиц его значение может превышать 1. Критерий V Крамера предназначен для анализа многопольных таблиц. Поскольку в основе этих критериев лежит критерий χ^2 Пирсона, их применение имеет те же ограничения.

Эмпирическое значение критериев ϕ и V Крамера определяются по формулам:

$$\phi = \sqrt{\frac{\chi^2}{n}}$$

$$V = \sqrt{\frac{\chi^2}{n \times (r - 1) \times (c - 1)}}$$

При анализе таблиц, имеющих две строки и 2 столбца, значения этих критериев совпадут. Например, в представленном исследовании эмпирическое значение этих критериев при сравнении экспериментальной и контрольной групп до воздействия оказалось равным 0,012:

$$\phi = \sqrt{\frac{0,016}{117}} = 0,012$$

$$V = \sqrt{\frac{0,016}{117 \times 1}} = 0,012$$

При анализе многопольных таблиц предпочтительнее использовать только критерий V Крамера (0,185). В нашем случае для повторного замера критерий V Крамера составит:

$$V = \sqrt{\frac{0,8,03}{117 \times 2}} = 0,185$$

Чтобы интерпретировать результаты математических расчетов, рекомендуется использовать шкалу Л. М. Реа и Р. А. Паркера (Grjibovski, Ivanov, & Gorbatova, 2016), согласно которой значение критериев ϕ или V Крамера менее 0,1 свидетельствует о несущественной связи между переменными, от 0,1 до 0,2 – о слабой связи; от 0,2 до 0,4 – об относительно сильной связи; от 0,4 до 0,6 – о сильной связи; более 0,6 – об очень сильной связи.

Опираясь на эти ориентиры, можно сделать вывод, что в нашем эксперименте до воздействия между фактором принадлежности к группе и уровнем сформированности конфликтологической компетенции связь была несущественной, а после воздействия она стала слабой.

Следует отметить, что вероятность ошибки первого рода (отвержение верной нулевой гипотезы) зависит от объема выборки. В результате при небольшом числе наблюдений сильная статистическая связь будет обладать низким уровнем статистической значимости ($p > 0,05$), в то же время при большом числе наблюдений слабая связь может иметь $p \leq 0,05$.

Критерий D Сомерса

Поскольку в педагогических исследованиях значения наблюдаемого признака часто представлены в порядковой шкале (низкий, средний, высокий уровень), то к ним можно применить критерии, которые позволяют оценить влияние одной переменной на другую. Поскольку порядковая шкала содержит данные о направленности различий (больше-меньше), то все критерии, входящие в данную группу, основываются на расчете числа нарушений порядка. В ходе расчетов сравнивается положение каждой ячейки таблицы с положением всех остальных ячеек попарно. Если номера столбца и строки первой сравниваемой ячейки ниже, чем у второй сравниваемой ячейки, то такая пара ячеек называется конкордантной парой. Если номер столбца первой сравниваемой ячейки больше, а номер ряда меньше, чем у второй сравниваемой ячейки, то такая пара ячеек называется дискордантной парой. Если у двух ячеек одинаковый номер ряда, то они называются связанными по рядам. Если у двух ячеек одинаковый столбец, то они называются связанными по столбцу (Grjibovski, Ivanov, & Gorbatova, 2016). Подробнее о подсчете конкордантных, дискордантных и связанных пар можно узнать из статьи А. М. Гржибовского и др. (Grjibovski, Ivanov, & Gorbatova, 2016).

Исходя из количества наблюдений, в каждой ячейке определяется число конкордантных, дискордантных, связанных по ряду и связанных по столбцу индивидуальных наблюдений. Оно представляет собой произведение числа наблюдений в каждой ячейке пары.

Примером критерия, относящегося к рассматриваемой группе, является предложенный в 1962 году Р. Х. Сомерсом критерий D Сомерса (Дельта Сомерса). Он является ассиметричным, т. е. его значение зависит от того, какая из переменных является зависимой, а какая – независимой. Обычно зависимая переменная записывается в столбцах, а независимая – в строках. Но если делать расчеты с помощью пакета статистических программ, например SPSS, то это условие является малоз-

начимым, т. к. программа рассчитывает два значения критерия: один – для случая, когда зависимая переменная обозначена в столбцах, а второй – для случая, когда зависимая переменная представлена в строках.

Критерий D Сомерса применяется для решения двух классов задач:

- 1) определение силы влияния (эффекта) независимой переменной на зависимую переменную;
- 2) определение индикатора эффективности прогноза предсказания значений зависимой переменной исходя из значений независимой переменной.

Применение критерия D Сомерса требует выполнения следующих условий (Grjibovski, 2008; Prorokowski, 2016):

- 1) должно быть определено, какая переменная является независимой, а какая – зависимой;
- 2) зависимая переменная должна быть представлена в порядковой шкале. Независимая переменная может быть как в шкале порядка, так и в номинальной шкале;
- 3) количество градаций зависимой переменной должно быть не больше 5. В противном случае возникают сложности с интерпретацией результатов;
- 4) объем меньшей из двух выборок должен быть более 7.

В рассматриваемом примере зависимой переменной является уровень сформированности конфликтологической компетенции, а независимой переменной – педагогическое воздействие, которое отличается в экспериментальной и контрольной группах.

Критерий D Сомерса рассчитывается по формуле (Grjibovski, Ivanov, & Gorbatoва, 2016):

$$D_{XY} = \frac{N_c - N_D}{N_c + N_D + N_t}$$

где N – количество пар; C – конкордантные пары, D – дискордантных пары, а t – количество пар, связанных по зависимой переменной.

В представленном выше примере расчет эмпирического значения критерия D Сомерса для определения влияния педагогического воздействия на уровень конфликтологической компетенции будет выглядеть следующим образом:

$$N_D = 13 \cdot 24 + 13 \cdot 8 + 29 \cdot 8 = 648$$

$$N_c = 25 \cdot 29 + 25 \cdot 18 + 24 \cdot 18 = 1607$$

$$N_t = 25 \cdot 13 + 24 \cdot 29 + 8 \cdot 18 = 1165$$

$$D_{XY} = \frac{1670 - 648}{1670 + 648 + 1165} = 0,28$$

Значение D Сомерса показывает разность между вероятностью того, что случайно выбранная пара наблюдений конкордантна, и вероятностью того, что эта пара дискордантна, при условии, что наблюдения не связаны по независимой переменной. Возможные значения D Сомерса варьируют от -1 до 1 (от полной прямо пропорциональной до полной обратной пропорциональной взаимосвязи), а 0 обозначает полную независимость переменных друг от друга (т. е. верность нулевой гипотезы) (Grjibovski, Ivanov, & Gorbatoва, 2016).

По результатам вычислений в нашем примере можно сделать вывод, что после воздействия между педагогическим воздействием и уровнем сформированности конфликтологической компетенции существует взаимосвязь средней силы. Значение $D_{xy} = 0,28$ также свидетельствует, что зависимая переменная (уровень конфликтологической подготовки студентов) может быть на 28 % лучше предсказана в

случае, если известно, в какой группе обучался студент (в экспериментальной или в контрольной).

Точное значение уровня значимости для принятия/отвержения нулевой гипотезы при использовании критерия D Сомерса является суммой гипергеометрических уровней значимости вариантов связанных пар, начиная с уровня $D_{xy} = 0,28$ или больше. На основе гипергеометрического распределения вероятности точная вероятность для гипотезы об увеличении признака в нашем случае составляет $p = 0,003$. Это позволяет отвергнуть нулевую гипотезу об отсутствии влияния независимой переменной на зависимую переменную и принять альтернативную гипотезу.

Подробный алгоритм расчета уровня значимости для критерия D Сомерса представлен в работе А. М. Либетрау (Liebetau, 1983). Поскольку для больших объемов выборок расчеты становятся очень трудоемкими, мы воспользовались для вычисления возможностями пакета статистических программ SPSS.

Хочется обратить внимание, что для педагогического эксперимента вывод о силе влияния педагогического воздействия на исследуемые характеристики объекта с помощью критерия D Сомерса можно сделать, только доказав, что до воздействия между контрольной и экспериментальной группами не было различий.

Обсуждение

Вопросы применения методов математической обработки данных в педагогических исследованиях широко обсуждаются в научной литературе. Признавая возможность использования стратегий качественного и количественного исследования, педагоги расширяют арсенал своих возможностей в решении научных задач. Вместе с тем, анализ научных работ, представленный в данном исследовании, показал, что даже при работе с накопленными статистическими данными педагоги не всегда используют методы математической обработки полученных данных. А. И. Азевич и С. И. Алексеева связывают это с тем, что выпускники педагогического вуза, особенно гуманитарных направлений обучения, недостаточно владеют математическими методами и способами использования статистических критериев в научных исследованиях (Azevich & Alekseeva, 2016). Авторы утверждают, что простое сравнение выборочных средних – без грамотной оценки величины доверительного интервала или ошибок средних арифметических – является недопустимым для уровня выпускных квалификационных работ. В. И. Загвязинский усматривает низкую математическую грамотность педагогов-исследователей в спекуляциях, основанных на математическом камуфляже (Zagvyazinsky, 2009). Он высказывает обоснованные опасения о том, что без должной проработки оснований и инструментария математический аппарат только тиражирует и усиливает ошибки, создавая видимость научной доказательности.

В данной работе представлен пример применения различных математических критериев для проверки статистических гипотез о наличии/отсутствии различий между исследуемыми группами. Он ограничен, во-первых, рамками сравнения использования критериев, предназначенных для анализа номинальных данных, а во-вторых, сравнением контрольной и экспериментальной групп до и после осуществления педагогического воздействия. Следует отметить, что представленный подход поддерживается рядом исследователей (Gumennikova, Kaydalova, & Ryabinova, 2015; Novikov, 2004). Но он является не единственным математическим методом оценки эффективности педагогического воздействия. В частности, педагог может оценить динамику исследуемого признака во время эксперимента. Для этого требуется иной способ представления данных и выбор других критериев (например, критерия Макнамары). Рассмотрение этого подхода анализа результатов

педагогического эксперимента может стать предметом следующей статьи. Также важным представляется рассмотрение возможностей использования современных статистических программ для оптимизации процесса обработки и анализа количественных данных в педагогическом эксперименте.

Заключение

Таким образом, можно сделать вывод, что, имея в качестве результатов данные, отражающие частоту распределения участников исследования с разным уровнем изучаемой компетенции из экспериментальной и контрольной группы на предварительном и завершающем этапах эксперимента, можно применить признанные в мировом научном сообществе методы математической статистики для подтверждения экспериментальной гипотезы.

При больших объемах выборки ($n \geq 30$) исследователь может доказать не только наличие/отсутствие различий в распределении исследуемого признака в разных группах (χ^2 Пирсона), но и оценить силу связи между воздействием и обучающим эффектом (критерии ϕ и V Крамера). Если значения наблюдаемого признака представлены в порядковой шкале (низкий, средний высокий уровень), то для анализа оценки силы эффекта можно использовать критерий D Сомерса. В случае небольших объемов выборки ($n < 30$) можно использовать точный критерий Фишера.

В заключение мы хотим отметить, что представленные в статье варианты и примеры являются не единственной стратегией статистического подтверждения экспериментальных гипотез в педагогических исследованиях. Например, важной является оценка достоверности сдвига значений исследуемого признака после педагогического воздействия.

Представленные в работе примеры призваны показать, как можно использовать методы математической статистики при анализе результатов формирующего эксперимента.

Исследование проводится при поддержке Мордовского государственного педагогического университета.

Грант № 28-05/2019 4.06.2019 г. «Формирование конфликтологической готовности студентов педагогических вузов».

Комментарий об открытом доступе к данным, этике, конфликте интересов

Доступ к данным может быть предоставлен при обращении по электронной почте к авторам данной работы. Авторы заявляют о гарантии прав лиц, вовлеченных в исследование, об отсутствии причинного вреда и соблюдении анонимности, а также об отсутствии конфликта интересов относительно публикуемой работы.

Список литературы

- Азевич, А. И., Алексеева, С. И. Педагогический эксперимент и средства описательной статистики // Вестник РУДН. Серия: Информатизация образования. – 2016. – № 2. – С. 57-65.
- Аканов, А. А., Турдалиева, Б. С., Изекенова, А. К., Рамазанова, М. А., Абдраимова, Э. Т., Гржибовский, А. М. Оценка использования статистических методов в научных статьях медицинских журналов Казахстана // Экология человека. – 2013. – № 5. – С. 61-64.
- Гржибовский, А. М. Анализ порядковых данных (независимые наблюдения) // Экология человека. – 2008. – № 8. – С. 56-62.
- Гржибовский, А. М., Иванов, С. В., Горбатова, М. А. Анализ номинальных и ранговых переменных данных с использованием программного обеспечения Statistica и SPSS // Наука и здравоохранение. – 2016. – № 6. – С. 5-39.

- Гуменникова, Ю. В., Кайдалова, Л. В., Рябинова, Е. Н. Применение методов математической статистики для обработки и анализа результатов педагогического эксперимента // Известия Самарского научного центра РАН. – 2015. – № 1-5. – С. 1032-1036.
- Загвязинский, В. И. Истоки и характер типичных ошибок в педагогических исследованиях // Образование и наука. – 2009. – № 11. – С. 5-12.
- Закирова, А. Ф. Методологический аппарат научного исследования в аспекте концептуализации педагогического знания // Образование и наука. – 2015. – Т. 129. – № 10. – С. 4-19.
- Лапшин, В. Е. Анализ результатов психолого-педагогического исследования аутодеструктивного поведения молодых людей // Вестник Владимирского юридического института. – 2015. – № 4 (37). – С. 95-98.
- Новиков, Д. А. Статистические методы в педагогических исследованиях (типовые случаи). – М.: МЗ-Пресс, 2004. – 67 с.
- Рахыпбеков, Т. К., Гржибовский, А. М. К вопросу о необходимости повышения качества казахстанских научных публикаций для успешной интеграции в международное научное сообщество // Наука и Здравоохранение. – 2015. – № 1. – С. 5-11.
- Солнышков, М. Е. Качество современных научно-педагогических исследований: анализ состояния проблемы // Ученые записки Российского государственного социального университета. – 2009. – № 7-1 (70). – С. 74-78.
- Унгурияну, Т. Н., Гржибовский, А. М. Краткие рекомендации по описанию, статистическому анализу и представлению данных в научных публикациях // Экология человека. – 2011. – № 5. – С. 55-60.
- Hess, A. S., Hess, J. R. Understanding tests of the association of categorical variables: The Pearson chi-square test and Fisher's exact test // Transfusion. – 2017. – Vol. 57. – No. 4. – P. 877-879. DOI: <https://doi.org/10.1111/trf.14057>
- Kim, H. Y. Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test // Restorative Dentistry & Endodontics. – 2017. – Vol. 42. – No. 2. – P. 152-155. DOI: <https://doi.org/10.5395/rde.2017.42.2.152>
- Kroonenberg, P. M., Verbeek, A. The Tale of Cochran's Rule: My Contingency Table has so Many Expected Values Smaller than 5, What Am I to Do? // The American Statistician. – 2018. – Vol. 72. – No. 2. – P. 175-183. DOI: <https://doi.org/10.1080/00031305.2017.1286260>
- Liebetrau, A. M. Measures of Association. – Newbery Park: Sage Publications, 1983. – 94 p.
- Lin, J-J., Chang, C.-H., Pal, N. A Revisit to Contingency Table and Tests of Independence: Bootstrap is Preferred to Chi-Square Approximations as Well as Fisher's Exact Test // Journal of Biopharmaceutical Statistics. – 2015. – Vol. 25. – No. 3. – P. 438-458. DOI: <https://doi.org/10.1080/10543406.2014.920851>
- McHugh, M. L. The Chi-square test of independence // Biochemia Medica. – 2013. – Vol. 23. – No. 2. – P. 143-149. DOI: <https://doi.org/10.11613/BM.2013.018>
- Nowacki, A. Chi-square and Fisher's exact tests // Cleveland clinic journal of medicine. – 2017. – Vol. 84. – No. 2. – P. e20-e25. DOI: <https://doi.org/10.3949/ccjm.84.s2.04>
- Prorokowski, L. Rank-Order Statistics for Discrimination Backtest of Credit Risk Models // Bank i Kredyt. – 2016. – Vol. 47. – No. 3. – P. 227-250.
- Warner, P. Testing association with Fisher's Exact test // Journal Family Planning and Reproductive Health Care. – 2013. – Vol. 39. – No. 4. – P. 281-284. DOI: <http://dx.doi.org/10.1136/jfprhc-2013-100747>

References

- Akanov, A. A., Turdalieva, B. S., Izenkova, A. K., Ramazanova, M. A., Abdraimova, E. T., & Grzhibovskiy, A. M. (2013). Assessment of use of statistical methods in scientific articles of the Kazakhstan's medical journals. *Ekologiya cheloveka – Human Ecology*, 5, 61-64.
- Azevich, A. I., & Alekseeva, S. I. (2016). Teaching experiment and means of descriptive statistics. *Vestnik RUDN. Seriya: Informatizatsiya obrazovaniya – RUDN Journal of Informatization in Education*, 2, 57-65.
- Grzhibovski, A. M. (2008). Analysis of nominal data (independent observations). *Ekologiya cheloveka – Human Ecology*, 8, 56-62.

- Grjibovski, A. M., Ivanov, S. V., & Gorbatova, M. A. (2016). Analysis of nominal and ordinal data using Statistica and SPSS software. *Nauka i zdravookhranenie – Science and Healthcare*, 6, 5-39.
- Gumennikova, Yu. V., Kaydalova, L. V., & Ryabinova, E. N. (2015). Application of methods of mathematical statistics for processing and analyzing the results of a pedagogical experiment. *Izvestiya Samarskogo nauchnogo tsentra RAN – Izvestia of Samara Scientific Center of the Russian Academy of Sciences*, 17(1-5), 1032-1036.
- Hess, A. S., & Hess, J. R. (2017). Understanding tests of the association of categorical variables: The Pearson chi-square test and Fisher's exact test. *Transfusion*, 57(4), 877-879. DOI: 10.1111/trf.14057
- Kim, H. Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. *Restorative Dentistry & Endodontics*, 42(2), 152-155. doi:10.5395/rde.2017.42.2.152
- Kroonenberg, P. M., & Verbeek, A. (2018). The Tale of Cochran's Rule: My contingency table has so many expected values smaller than 5, what am I to do? *The American Statistician*, 72(2), 175-183. doi:10.1080/00031305.2017.1286260
- Lapshin, V. E. (2015). Analysis of the results of psychological and pedagogical research self-destructive behavior of young people. *Vestnik Vladimirskogo yuridicheskogo instituta – Bulletin of the Vladimir Law Institute*, 37(4), 95-98.
- Liebetrau, A. M. (1983). *Measures of Association*. Newbery Park: Sage Publications.
- Lin, J-J., Chang, C.-H., & Pal, N. (2015). A revisit to contingency table and tests of independence: Bootstrap is preferred to Chi-square approximations as well as Fisher's exact test. *Journal of Biopharmaceutical Statistics*, 25(3), 438-458. doi:10.1080/10543406.2014.920851
- McHugh, M. L. (2013). The Chi-square test of independence. *Biochemia Medica*, 23(2), 143-149. doi:10.11613/BM.2013.018
- Novikov, D. A. (2004). *Statistical methods in pedagogical research (typical cases)*. Moscow: MZ-Press.
- Nowacki, A. (2017). Chi-square and Fisher's Exact tests. *Cleveland clinic journal of medicine*. 84(2), e20-e25. doi:10.3949/ccjm.84.s2.04
- Prorokowski, L. (2016). Rank-order statistics for discrimination backtest of credit risk models. *Bank i Kredyt*, 47(3), 227-250.
- Rakhyzbekov, T. K., & Grjibovski, A. M. (2015). The need for improvement of the quality of Kazakhstani publications for successful integration in the international research community. *Nauka i Zdravookhranenie – Science and Healthcare*, 1, 5-11.
- Solnyshkov, M. E. (2009). Quality of modern scientific and pedagogical research: analysis of the state of the problem. *Uchenye zapiski Rossiyskogo gosudarstvennogo sotsial'nogo universiteta – Scientific Notes of the Russian State Social University*, 70(7-1), 74-78.
- Unguryanu, T. N., & Grzhibovsky, A. M. (2011). Brief recommendations on description, analysis and presentation of data in scientific papers. *Ekologiya cheloveka – Human Ecology*, 5, 55-60.
- Warner, P. (2013). Testing association with Fisher's Exact test. *Journal Family Planning and Reproductive Health Care*, 39(4), 281-284. doi:10.1136/jfprhc-2013-100747
- Zagvyazinsky, V. I. (2009). The origins and nature of typical errors in pedagogical research. *Obrazovanie i nauka – Education and Science*, 11, 5-12.
- Zakirova, A. F. (2015). The methodological unit of scientific research in terms of conceptualisation of pedagogical knowledge. *Obrazovanie i nauka – Education and science journal*, 129(10), 4-19.