

УДК 378.4.51

Жаров В.К., Форманов Ш.К., Хонкулов У.Х.

¹ Российский государственный университет туризма и сервиса (г. Москва)

² Институт математики АН Узбекистана (г. Ташкент)

³ Гулистанский государственный университет (Узбекистан)

О СТАТИСТИЧЕСКОМ МЕТОДЕ В ПЕДАГОГИЧЕСКОМ ЭКСПЕРИМЕНТЕ В УСЛОВИЯХ СОВРЕМЕННОГО УЧЕБНОГО ПРОЦЕССА

V. Zharov, Sh. Formanov, U. Honkulov

¹ Russian State University of Tourism and Services, Moscow

² Institute of Mathematics and Information Technologies,
Academy of Sciences of Uzbekistan, Tashkent

³ Gulistan State University, Uzbekistan

ON A STATISTICAL METHOD IN PEDAGOGICAL EXPERIMENT AND MODERN EDUCATIONAL PROCESS

Аннотация. В настоящей статье рассматриваются методы экспериментальной педагогики, которые могут применяться в изменившихся условиях современной средней школы. Изменившиеся условия потребовали введения новых оценок педагогических экспериментов, поэтому отдельное внимание в работе уделено использованию теории проверки статистических гипотез к задаче обнаружения влияния новых педагогических воздействий на учебный процесс. Доказано, что предлагаемая выборочная послойная оценка дает более содержательные результаты и адаптируется к изменяющимся условиям образовательной среды.

Ключевые слова: педагогический эксперимент, статистика, послойная выборка и их оценки, информационно-педагогическая среда, надежность эксперимента.

Abstract. The article presents the methods of experimental pedagogics which can be applied in the changed circumstances of the modern high school. As the changed circumstances required the use of new estimations of educational experiments it is proposed to use the theory of statistical hypotheses verification with the purpose of discovering the influence of new pedagogical impacts on the process of studying. It is proved that the results of the offered system of stratified sampling are more substantial. Besides, the system is easily adaptive to the changing conditions of the educational environment.

Keywords: pedagogical experiment, statistics, stratified sampling and evaluation, information and educational environment, the reliability of the experiment.

Задачи обнаружения воздействия применения новых педагогических технологий (например, использование интерактивных досок, индивидуализация обучения с помощью компьютерных технологий, новые формы представления содержания (право-, левополушарные методы представления знания), формы и методы управления учебной информацией согласно требованиям конкретного учебного заведения и т. д.) на текущий учебный процесс являются типичными для экспериментальной педагогики¹. В свою очередь, результаты и методы экспериментальной педагогики могут быть обоснованы с помощью теории статистических выводов, составляющей основную часть современной математической статистики (см. [5; 4]).

© Жаров В.К., Форманов Ш.К., Хонкулов У.Х., 2012.

¹ Здесь следует напомнить, что экспериментальная педагогика, насчитывающая более двухсот лет, изначально была теоретической наукой, построенной на прогнозах влияния того или иного методического изменения в преподавании и воспитании [7].

Статистические исследования обнаружения эффективности, полезности, удобства и преимущества предложенных новых технологий обучения в отношении к существующим (применяемым, используемым) педагогическим средствам в основном проводятся по следующей схеме: из совокупности (объекта) педагогического эксперимента (например, из групп учащихся) выделяются *экспериментальная группа*, которая сравнивается с *контрольной группой*. Различие эффектов педагогических технологий будет обосновано, если две эти группы, первоначально совпадающие по своим характеристикам, различаются после реализации новых педагогических воздействий. Следовательно, требуется провести два сравнения и показать, что при первом сравнении (до начала педагогического эксперимента) характеристики экспериментальной и контрольной группы совпадают, а при втором (после окончания эксперимента) – различаются.

В психологии при построении экспериментов часто используют три стратегии составления групп, а также три схемы сравнения. Первая стратегия *равновероятного участия* в формируемых группах любого из добровольных испытуемых используется при выборе экспериментальных групп. Иначе говоря, с помощью генератора случайных чисел происходит распределение участников по группам¹. «Основное условие применения случайной стратегии – это большое число испытуемых в каждой экспериментальной группе» [3, с.168]. Другая стратегия – стратегия подбора пар. Вторая схема эксперимента с межгрупповыми сравнениями называется *попарным распределением групп*. Весьма любопытно, на наш взгляд, следующее замечание: «различие между двумя названными схемами (в отношении внутренней валидности эксперимента) не так уж существенно по сравнению с их огромным превосходством над другой схемой – *использованием реальных групп*. Примером мог бы служить случай, когда

¹ На самом деле эта стратегия условно может быть названа случайной, поскольку генератор случайных чисел сам алгоритмически организован.

преподаватели испанского языка решили бы проводить эксперимент с разговорным методом в одной школе, с письменным – в другой. Основной характеристикой испытуемых был бы при этом лишь сам факт учебы в школе. Но мы никак не можем рассчитывать на то, что студенты двух школ будут одинаковы по способности к изучению испанского языка. Ведь они живут в разных условиях, воспитываются в разных семьях, имеют разный опыт обучения перед поступлением в школу и т. д. Да и преподаватели в разных школах разные. Но даже если бы в двух школах занятия вел бы один и тот же преподаватель, систематическое смешение независимой переменной с индивидуальными различиями испытуемых было бы совершенно неизбежным» [3, с. 170]. Третья стратегия *случайного распределением групп с выделением слоев* является синтезом двух первых. «Для достижения столь же высокой надежности эксперимента потребуется меньшее количество испытуемых» [3, с. 171].

В современных условиях проведения педагогических экспериментов классическая, изложенная выше схема видоизменяется в силу требования первоначального совпадения по своим характеристикам групп. Выясняется, что в условиях ЕГЭ результаты, оцениваемые одним и тем же числом баллов, в разных регионах страны могут не коррелировать вовсе. Однако требования к уровню знания студентов, отраженные паспортом специальности, общие, поэтому опора на характеристики (начальные параметры), например, в условиях факультета университетов одного профиля, может приводить к неопределенным и некорректным статистическим выводам. В силу сделанного замечания, мы считаем, что необходимо внести некоторые изменения в условия проведения педагогического эксперимента.

1. Выделять типические группы, следовательно, проводить отдельное исследование групп.
2. Для чистоты эксперимента контрольная группа должна находиться под руководством другого преподавателя-исследователя.
3. Быть может, наиболее важное: в условиях первого курса (первый семестр) необхо-

димо подготавливать группы к эксперименту. Как сказано в связи с применением статистических методов Н. Бейли, «важнее научиться учитывать специфические особенности исследуемой проблемы, нежели овладеть навыками правильного применения выбранного метода» [1, с. 11]. Остановимся на каждом пункте подробнее. Начнем с последнего.

Мы считаем, что постановка педагогического эксперимента в исследованиях требует должной подготовки, предваряющей сам эксперимент. В самом деле, предварительно в педагогическом эксперименте необходимо оценить группы (контрольную и экспериментальную), которые будут участвовать в нем. Нашим обоснованием этому утверждению является следующее соображение. Общепринято, что оценка у студента должна как-то коррелировать относительно критериев соответствия базовым условиям ЕГЭ (на современном этапе образования). Учебная практика показывает очень часто несвязанность результатов двух уровней оценок, полученных на ЕГЭ и, например, в первой сессии или в первом модуле по математике. Возможно, это объясняется следующими обстоятельствами: унифицированных требований к проведению ЕГЭ нет, а те, что есть, легко, согласно местным условиям, подправляются; подготовка учеников слаба настолько, что невольно происходит проверка владения учебным материалом самих учителей; неприспособленность методических материалов к условиям ЕГЭ; есть и другие причины. Для уверенности в корректности оценки (потребителю: учебному учреждению, в конечном счете, государству) необходим многокритериальный подход хотя бы по трем составляющим, а именно: владение основами техническими операциями (операционная составляющая), уровень владения базовыми понятиями, решение задач достаточного уровня¹. В частности, глубина «владения» ба-

¹ Задачи достаточного уровня отличаются от предлагаемых ныне тем, что из пяти задач этого цикла три задачи адресованы тем, кто выбрал естественнонаучную или техническую специальность, для тех, кто выбрал экономическую или прикладную математику, – следующие две задачи, естественно включая и предыдущие три,

зовыми знаниями, которые в (современных) школьных испытаниях не проверяются. Кроме этого, возможно, составителям заданий по ЕГЭ необходимо включать средовые оценки базовых условий проведения испытаний на местах их проведения (опытный коэффициент, коэффициент поправки). Это ясно хотя бы из средовых изменений педагогического состава школ или востребованности педагогического труда государством, с одной стороны. Но, с другой – понятно, что среда информационная, микро-, макросреда индивидуума обладает свойством динамичности и поэтому требует фиксации изменений и, соответственно, влияния этих изменений на результаты обучения как непосредственно на экзаменах, так и в течение последнего этапа обучения в школе. Например, с помощью электронных дневников успеваемости, статус которых должно определить государство². В то же время тест-контроль успеваемости в процессе обучения, например, на младших курсах учебного заведения, особенно в первом семестре, должен соответствовать базовому уровню будущей специальности. Возникающие несоответствия должны быть отправными точками последующих оценок работ школ и выбора специальности индивидуумом. Также нельзя исключить возможность недооценки индивидуумом сложности будущей специальности. Но и в средней общеобразовательной школе на последнем этапе обучения можно сделать еще много, несмотря на наличие странного упорства государства игнорировать традиции российской школы.

Хорошо известно, что одной из задач статистических исследований является «проверка гипотезы о том, что мы делаем в предполо-

а для тех, кто займется профессионально математикой, – следующая задача, шестая. Причем понятно, что те, кто решает задачи из третьей группы, получают стимулирующие баллы. Могут быть, очевидно, и более гибкие шкалы с весовыми коэффициентами задач, коль скоро опыт проведения ЕГЭ в других странах в наших условиях не совсем применим (?!).

² Очевидно, если государству необходимо качество знаний будущих специалистов, то нарушение этого статуса приведет нарушителя к государственному преступлению.

жении о распределении случайной величины (например, о численных значениях одного или нескольких параметров функции распределения) и решаем, согласуются ли, в некотором смысле, эти значения параметров с полученными наблюдениями» [2, с. 77]. «Задача статистических выводов появляется именно тогда, когда нам необходимо дать наилучшие, в некотором смысле, ответы по ограниченному числу наблюдений» [2, с. 77]. Следовательно, проводя педагогический эксперимент даже в условиях типичности в каждой группе, мы на современном этапе развития математического образования в стране получаем явления неопределенности, поскольку нет информации о соответствии базовому уровню образования членов групп, участвующих в эксперименте (наличие некорректной оценки результатов ЕГЭ). Поэтому, «когда речь идет о задаче статистических выводов, то подразумевается, что мы в состоянии получить случайную выборку, т. е. совокупность, состоящую из реализаций независимых одинаково распределенных случайных величин» [2, с. 78].

Используя сочетание принципа выборки с принципом строгой группировки, можно обеспечить «значительно лучшую репрезентативность, чем та, которая получается при простой выборке» [8, с. 159]. В случае типичных выборок мы имеем возможность понять соответствие базовому уровню знаний студентов по многим параметрам¹, а с другой стороны, в целом оценить параметр, например глубины знаний по данной дисциплине или общей эрудиции принятых абитуриентов в университет или на конкретный факультет.

Итак, на первом этапе подготовки к эксперименту были определены типические группы. Конечно, в эксперименте должны участвовать группы, близкие по послойной выборке. Критерий близости можно вводить разными способами, например, введя допустимую константу, которую не превосходит модуль разности средних показателей в дан-

ных выборках или разности существенных характеристик, взятых по модулю меньших определенного фиксированного числа. Если различия между показателями существенно больше какого-либо числа, то необходимо предварительно провести тренировочную работу со слабой группой. И наконец, в стадии активного эксперимента, скажем, воздействия традиционной и новаторской методик, можно оценивать по предлагаемой ниже схеме.

Возвращаясь к упомянутой проблеме соответствия базового уровня знаний начальному, предъявляемому паспортом специальности, технологически в рамках изменившихся условий, можно вполне корректно провести оценки характеристик соответствующих групп студентов по многим параметрам. Кроме того, оказывается, что эта проблема актуальна и, возможно, является главной или государственной проблемой повышения качества образования в стране.

Эту проблему в первом приближении можно разграничить на две составляющие: первая составляющая может соответствовать базовому уровню учебного заведения (рейтинг, социальный фон (мнение), традиционный уровень, тип учебного заведения и т. д.), второй составляющей будет являться базовый уровень конкретного факультета в учебном заведении. Оценка базового уровня будет различной, а следовательно, должны быть определены уровни, необходимые для выбранного абитуриентом учебного заведения (потенциальный) и соответствующие специальности (паспорту специальности). Понятно, что соответствие паспорту специальности и есть цель любого высшего учебного заведения. Потенциальный базовый уровень в известных учебных заведениях таков, что люди, выдержавшие конкурс, обладают некоторой типичностью, т. е. по некоторому нормальному распределению вероятности производят те или иные ошибки (неуспех²) или успехи в предлагаемых вопросах, заданиях, тестах, другими словами, в контрольно-

¹ Операционный арсенал первокурсников, освоенность конкретных тем в средней школе, направленность на творческие задания и многие другие параметры.

² Ясно, что неуспех достаточно точно определяется в рамках учебного процесса.

измерительном материале. В таком случае мы с некоторой долей вероятности можем ожидать успех в типичных условиях обучения. Иначе отношение успешных результатов к неудачам будет много больше единицы. Влияние информационно-педагогической среды тем выше, чем осознаннее был сделан выбор индивидуумом факультета, вуза, а в общем, – специальности. Ещё одним приложением из предлагаемого типа педагогических исследований является выделение общего значения какой-либо характеристики генеральной совокупности.

Положим, что $a = N^{-1} \sum_{j=1}^k \sum_{i=1}^{N_j} a_{ij}$ и $a_j = N_j^{-1} \sum_{i=1}^{N_j} a_{ij}$, a_{ij} – i -ый элемент j -го слоя. Но тогда в качестве $\bar{a} = N^{-1} \sum_{i=1}^k N_j a_j$, т. е. среднее значение для генеральной совокупности можно интерпретировать как средне взвешенное значение средних значений k слоёв, причем веса соответственно равны N_j/N , где $(j = 1, 2, \dots, k)$. Существует теорема о том, что полученная таким образом случайная величина \bar{a} абсолютно корректной оценкой характеристики генеральной совокупности, для которой

$$M(\bar{a}) = a, \quad D^2(\bar{a}) = \sum_{j=1}^k \left[\left(\frac{N_j}{N} \right)^2 (1 - f_j) \left(\frac{s_j^2}{n_j} \right) \right],$$

где

$$s_j^2 = \left[\frac{1}{N_j - 1} \right] \sum_{i=1}^{N_j} (a_{ij} - a_j)^2, \quad f_j = n_j / N_j$$

[7, с. 160].

Ранее приведенную схему классического педагогического эксперимента можно формализовать следующим образом: пусть случайная величина (с.в.) X характеризует качество учебного процесса в контрольной группе (в более общем случае это может быть и вектор), проводимого по известным (традиционным) методическим схемам, а с.в. Y характеризует качество учебного процесса в экспериментальной группе (понятно,

что в общем случае координаты одноименны и векторы одной и той же размерности), где применялась некоторая новая педагогическая технология. Будем считать, что эти с.в. X и Y имеют нормальное распределение с параметрами (MX, σ_X^2) и (MY, σ_Y^2) соответственно, где MX, MY – математические ожидания (средние), σ_X^2 и σ_Y^2 – дисперсии с.в. X и Y . Предположения о нормально распределенных с.в. X и Y имеют довольно общий характер, поскольку эти величины формализуются как суммы достаточно большого числа независимых с.в., которые характеризуют какие-либо отдельно взятые показатели учебного процесса (согласно центральной предельной теореме теории вероятностей, суммы независимых с.в. имеют асимптотически нормальные распределения).

В работе Д.А. Новикова [9] приведен общий алгоритм использования статистических критериев в педагогических исследованиях (критерия Крамера-Уэлча, Вилкоксона-Манна-Уитни, хи-квадрат, Фишера).

В настоящей статье исследуется применение теории проверки статистических гипотез к задаче обнаружения влияния новых педагогических воздействий на учебный процесс. При этом используется статистическая модель, построенная на выборках, составленная из полученных данных с.в. X и Y . Пусть

$$X_1, X_2, \dots, X_{n_1},$$

$$Y_1, Y_2, \dots, Y_{n_2},$$

две независимые выборки объемами n_1 и n_2 из генеральных совокупностей X и Y . Рассмотрим случай, когда дисперсии σ_X^2 и σ_Y^2 генеральных совокупностей X и Y известны. Для того, чтобы обнаружить влияние примененной новой педагогической технологии на учебный процесс, необходимо проверить нулевую гипотезу H_0 , заключающуюся в том, что $MX = MY$, относительно альтернативной гипотезы H_1 , состоящей в том, что $|MX - MY| > 0$. Таким образом, целью исследуемого педагогического эксперимента является проверка гипотезы о совпадении центров рас-

пределений двух нормальных совокупностей при известных дисперсиях (гипотеза H_0). Так как о значениях математических ожиданий (средних) MX и MY нам ничего не известно, то для проверки гипотезы H_0 используются их наилучшие оценки

$$\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \quad \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i.$$

Как известно из математической статистики, средние \bar{X} и \bar{Y} имеют нормальное распределение с параметрами $[MX, \sigma_1^2/n_1]$ и $[MY, \sigma_2^2/n_2]$, где $\sigma_1^2 = \sigma_X^2$, $\sigma_2^2 = \sigma_Y^2$. Выборки независимы, поэтому \bar{X} и \bar{Y} также независимы и с.в. $\bar{X} - \bar{Y}$ имеет нормальное распределение, причем дисперсия

$$D(\bar{X} - \bar{Y}) = D\bar{X} + D\bar{Y} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Если гипотеза H_0 справедлива, то

$$M(\bar{X} - \bar{Y}) = M\bar{X} - M\bar{Y} = MX - MY = 0,$$

следовательно, нормированная разность

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

имеет стандартное нормальное распределение с параметром $(0,1)$.

Следуя общей теории проверки статистических гипотез, математические вопросы, которые разработаны в трудах известных математиков Ю. Нейман и Э. Пирсона (см. [9]), фиксируя вероятность $p = 1 - \alpha$, можно определить статистику Z_p , которая разделит множество значений Z на два непересекающихся подмножества: область допустимых значений Z ($|Z| \leq Z_p$) и множество значений, для которых $|Z| > Z_p$, образует критическую

область (вероятность α считается уровнем значимости, а вероятность p называется надежностью критерия).

Таким образом, для вероятности (надежности) $p = 1 - \alpha$ критическая область определяется неравенством

$$\frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} > Z_p$$

В частности, если $\sigma_1^2 = \sigma_2^2 = \sigma^2$, то критическая область определяется неравенством

$$\frac{\bar{X} - \bar{Y}}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} > Z_p$$

Вероятность α характеризует те события, которые при данных условиях исследования считаются практически невозможными. Чем меньше α , тем меньше вероятность отклонить проверяемую гипотезу, если она верна (совершить ошибку первого рода); с уменьшением уровня значимости α увеличивается область допустимых значений, следовательно, увеличивается риск принять проверяемую гипотезу, когда она неверна (увеличивается вероятность ошибки второго рода). Ошибки первого и второго родов в практику введены Ю. Нейманом (см. [6]).

При проверке гипотезы о равенстве центров двух нормальных генеральных совокупностей (гипотезы H_0 , которая означает отсутствие какого-либо влияния применения новых педагогических воздействий) при заданном уровне значимости α контролируется лишь ошибка первого рода, но нельзя делать выводы о степени риска, связанного с принятием неверной гипотезы, т. е. с возможностью совершения ошибки второго рода.

Пример. Пусть количества учащихся в контрольной и экспериментальной группах соответственно равны $n_1 = 25$ и $n_2 = 50$. В результате проведения педагогического экс-

перимента, связанного с обнаружением влияния применения некоторой новой педагогической технологии получены выборки

$$X_1, X_2, \dots, X_{25} \text{ и } Y_1, Y_2, \dots, Y_{50},$$

и средние квадратические отклонения (дисперсии) в обеих сериях измерений $\sigma_1^2 = \sigma_2^2 = \sigma^2 = 0,30$. Предположим, что средние значения исследуемой величины в контрольной и экспериментальной группах соответственно равны

$$\bar{X} = \frac{X_1 + \dots + X_{25}}{25} = 9,79,$$

$$\bar{Y} = \frac{Y_1 + \dots + Y_{50}}{50} = 9,60.$$

Можно ли с надежностью $p = 0,99$ объяснить это расхождение случайными причинами (т. е. не было никакого влияния применения новой педагогической технологии)?

Решение. Вычислим нормированную разность

$$|Z| = \frac{\left| \bar{X} - \bar{Y} \right|}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{9,79 - 9,60}{0,30 \sqrt{\frac{1}{25} + \frac{1}{50}}} = 2,59$$

Путем использования числовых таблиц нормального распределения (или непосредственным вычислением на компьютере), получим критическое значение

$$Z_p = Z_{0,99} = 2,579.$$

Имеем $2,59 > 2,576$, поэтому с надежностью $0,99$ можно считать расхождение средних

неслучайным (значимым), т. е. применение новых педагогических воздействий имело влияние и изменило характеристику учебного процесса, поскольку попадание в область значений $|Z| > 2,576$ при гипотезе H_0 практически невозможно. Заметим, однако, что для значений $|Z| \leq 2,576$ еще нельзя утверждать, что гипотеза H_0 подтвердилась. Следовательно, с помощью проверки статистических гипотез можно лишь отвергнуть проверяемую гипотезу об отсутствии влияния новой педагогической технологии на учебный процесс, но никогда нельзя доказать ее справедливость.

Таким образом, многослойная выборка, представление статистических данных, по сравнению с «классическими» оценками педагогических экспериментов, дают более глубокие результаты и адаптируются к изменяющимся условиям в образовательных средах.

ЛИТЕРАТУРА:

1. Бейли Н. Статистический метод в биологии. – М., 1963. – 327 с.
2. Браунли К.А. Статистическая теория и методология в науке и технике. – М., 1977. – 408 с.
3. Готтсданкер Р. Основы психологического эксперимента. – М., 1982. – 464 с.
4. Грабарь М.И., Краснянская К.А. Применение математической статистики в педагогических исследованиях. Непараметрические методы. – М., 1973. – 136 с.
5. Кендалл М, Стьюарт М. Статистические выводы и связи. – М., 1973. – 789 с.
6. Леман З. Проверка статистических гипотез. – М., 1964. – 403 с.
7. Мейман Эрнст, Экспериментальная педагогика. Т. 3. – М., 1913. – 269 с.
8. Михок Г., Урсяну В. Выборочный метод и статистическое оценивание. – М., 1982. – 245 с.
9. Новиков Д.А. Статистические методы в педагогических исследованиях (типовые случаи). – М., 2004. – 67 с.