Использование критерия χ^2 (хи-квадрат) для проведения статистической обработки данных педагогического эксперимента
Набиулина Л. М.

Набиулина Луиза Махмудовна / Nabiulina Luiza Makhmudovna - кандидат педагогических наук,
кафедра «Информационные технологии»,
Ташкентский государственный педагогический университет им. Низами, г. Ташкент

Аннотация: *в статье рассматриваются основы отбора экспериментальных групп, выдвижение гипотезы, использование критерия χ^2 (хи-квадрат) для обработки и представления экспериментальных данных и формулирования выводов.*
Abstract: *the article deals with the basics of screening the experimental groups, the hypothesis was put forward, using the criterion χ2 (Chi-square) for the processing and presentation of experimental data and formulation of conclusions.*

Ключевые слова: *выборка, статистическая гипотеза, нулевая гипотеза, альтернативная гипотеза, критерий χ^2 (хи-квадрат).*
Keywords: *sampling, statistical hypothesis, null hypothesis, alternative hypothesis, criterion χ2 (Chi-square).*

При выполнении выпускных квалификационных работ, студенты педагогических вузов сталкиваются с проблемой оформления результатов педагогического эксперимента, в частности, неумением отбирать группы для проведения эксперимента, выдвигать гипотезы и использовать методы статистической обработки полученных данных на основе применения различных критериев. Одним из таких критериев является критерий хи-квадрат. Данный критерий применяется для сравнения распределений объектов двух совокупностей на основе измерений по шкале наименований в двух независимых выборках.

Прежде чем перейти к изучению критерия, необходимо остановиться на таких понятиях, как *выборка и гипотеза*.

Выборка или **выборочная совокупность** — часть *генеральной совокупности* элементов, которая охватывается наблюдением. Например, вы исследуете возможности применения метода проектов на уроках информатики в 5 классах общеобразовательной школы. Ученики 5 класса всех школ составляют генеральную совокупность, а отобранные для наблюдения ученики конкретной школы и конкретного 5 класса будут являться выборочной совокупностью. Обычно для наблюдения выбирают контрольную и экспериментальную группы. **Экспериментальная группа** – это группа, которая подвергается экспериментальному воздействию в процессе педагогического исследования, то есть группа, с которой непосредственно работает экспериментатор. Например, ученики класса 5А школы № 42. **Контрольная группа** (например, ученики класса 5Б школы № 42) помещается в те же условия, что и экспериментальная, но испытуемые в ней не подвергаются экспериментальному воздействию.

Статистика как метод исследования оперирует данными, которые могут быть искажены различными случайными факторами, поэтому статистические вычисления сопровождаются проверкой некоторых предположений или **гипотез** об источнике этих данных.

Статистическая гипотеза – это предположение о свойствах случайных величин или событий, которое мы хотим проверить по имеющимся данным. *Примеры статистических гипотез в педагогических исследованиях:*

Гипотеза 1. Усвоение возможностей программы Microsoft Excel не имеет существенных различий у учащихся, начавших их обучение в 5 классе или 6 классе.

Гипотеза 2. Проблемное обучение при обучении информатике эффективнее по сравнению с традиционной методикой обучения.

Нулевая гипотеза – это основное проверяемое предположение, которое обычно формулируется как отсутствие различий, равенство нулю значений выборочных характеристик и т. п. Примером нулевой гипотезы в педагогике является утверждение о том, что различие в результатах выполнения двумя группами учащихся одной и той же контрольной работы вызвано лишь случайными причинами.

Другое проверяемое предположение называется **альтернативной** гипотезой. Так, для упомянутого выше примера гипотезы H_0 в педагогике одна из возможных альтернатив H_1 будет определена так: уровни выполнения работы в двух группах учащихся различны, и это различие определяется влиянием, например, тех или других методов обучения.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость проверить ее. Данная проверка производится статистическими методами и поэтому называется статистической.

Альтернативная гипотеза принимаются тогда и только тогда, когда опровергается нулевая гипотеза. Это бывает в тех случаях, когда различия, например, в средних арифметических экспериментальной и контрольной групп настолько значимы, что риск ошибки отвергнуть нулевую гипотезу и принять альтернативную не превышает одного из трех принятых *уровней значимости* (α) статистического вывода:

1. первый уровень — 5 % ($\alpha = 0.05$); допускается риск ошибки в выводе в пяти случаях из ста теоретически возможных таких же экспериментов при строго случайном отборе испытуемых для каждого эксперимента;
2. второй уровень — 1 % ($\alpha = 0.01$), т. е. соответственно допускается риск ошибиться только в одном случае из ста;
3. третий уровень — 0,1 % ($\alpha = 0{,}001$), т. е. допускается риск ошибиться только в одном случае из тысячи.

Предположим, вы изучаете выполнение определенного задания, которое измеряется у каждого объекта по шкале наименований, имеющей две взаимоисключающие категории (например: выполнено верно — выполнено неверно). По результатам измерения у объектов двух выборок составляется таблица (см. табл. 1).

Таблица 1

	Категория 1	**Категория 2**	
Выборка №1	O_{11}	O_{12}	$O_{11}+O_{12}$
Выборка №2	O_{21}	O_{22}	$O_{21}+O_{22}$
	$O_{11}+O_{21}$	$O_{12}+O_{22}$	$n_1+n_2=N$

В этой таблице O_{ij} — число объектов в i-ой выборке, попавших в j-ую категорию по состоянию изучаемого свойства; $i=1,2$ – число выборок; $j=1,2$ – число категорий; $N=n_1+n_2$ — общее число наблюдений.

На основе данных таблицы 1 можно проверить нулевую гипотезу о равенстве вероятностей попадания объектов первой и второй совокупностей в первою (вторую) категорию шкалы измерения проверяемого свойства, например, гипотезу о равенстве вероятностей верного выполнения некоторого задания учащимися контрольных и экспериментальных классов.

Для проверки рассмотренных нулевых гипотез по данным таблицы 1 подсчитывается значение статистики критерия T по следующей общей формуле:

$$T = \frac{N \cdot (|O_{11} \cdot O_{22} - O_{12} \cdot O_{21}| - \frac{N}{2})^2}{n_1 \cdot n_2 \cdot (O_{11} + O_{21}) \cdot (O_{12} + O_{22})} \quad (1)$$

где n_1, n_2 — объемы выборок, $N = n_1 + n_2$ — общее число наблюдений.

Проводится проверка гипотезы H_0: $p_1=p_2$ — при альтернативе H_1: $p_1>p_2$. Пусть α — принятый уровень значимости. Тогда значение статистики T, полученное на основе экспериментальных данных, сравнивается с критическим значением статистики $T_{крит}$, которое определяется по таблице χ^2 с одной степенью свободы [2, С. 130] с учетом выбранного значения α. Если верно неравенство $T< T_{крит}$, то нулевая гипотеза принимается на уровне α.

В связи с тем, что замена точного распределения статистики T распределением χ^2 с одной степенью свободы дает достаточно хорошее приближение только для больших выборок, применение критерия ограничено некоторыми условиями. Критерий не рекомендуется использовать, если сумма объемов двух выборок меньше 20 и хотя бы одна из абсолютных частот в таблице 1, составленной на основе экспериментальных данных, меньше 5.

Пример 1. Проводился эксперимент, направленный на выявление лучшего из методов усвоения учащимися определенной темы, разработанных двумя авторами в соответствии с целями обучения информатике и содержанием программы. Для проведения эксперимента методом случайного отбора были выбраны два района. Учителя одного района (20 учителей) обучали по разработанной методике автора № 1, учителя второго района (15 учителей) обучали по другой предложенной методике.

Отношение учителей к предлагаемым методикам обучения измерено по шкале наименований, имеющей две категории (С=2): методика позволила достичь цели урока (да) – методика не позволила достичь цели урока (нет). Обе выборки учителей случайные и независимые.

Ответы 20 учителей первого района и 15 учителей второго района распределим на две категории и запишем в форме таблицы (табл. 2).

Таблица 2

	Да	Нет	
Выборка №1	O_{11}=14	O_{12}=6	n_1=O_{11}+ O_{12}=20
Выборка №2	O_{21}=8	O_{22}=7	n_2=O_{21}+ O_{22}=15
	O_{11}+ O_{21}=22	O_{12}+ O_{22}=13	N=n_1+n_2=35

Все значения в табл. 2 не меньше 5, поэтому в соответствии с условиями использования критерия χ^2 подсчет статистики критерия производится по формуле (1).

$$T = \frac{35 \cdot (|14 \cdot 7 - 6 \cdot 8| - \frac{35}{2})^2}{20 \cdot 15 \cdot (14+8) \cdot (6+7)} = 0{,}43$$

По таблице [2 , С. 130] для одной степени свободы (v=C-1=1) и уровня значимости α=0,05 найдем $T_{критич}$ = 3,84. Отсюда верно неравенство $T_{наблюд}$<$T_{критич}$ (0,43<3,84). Согласно правилу принятия решений для критерия χ^2, полученный результат не дает достаточных оснований для отклонения нулевой гипотезы, т. е. результаты проведенного опроса учителей двух экспериментальных районов не дают достаточных оснований для отклонения предположения об одинаковом достижении целей урока при применении предложенных методов обучения.

Применение критерия χ^2 возможно и в том случае, когда объекты двух выборок из двух совокупностей по состоянию изучаемого свойства распределяются более чем на две категории. Например, учащиеся экспериментальных и контрольных классов распределяются на четыре категории в соответствии с отметками (в баллах: 2, 3, 4, 5), полученными учащимися за выполнение некоторой контрольной работы.

Результаты измерения состояния изучаемого свойства у объектов каждой выборки распределяются на *С* категорий. На основе этих данных составляется таблица 2хС, в которой два ряда (по числу рассматриваемых совокупностей) и *С* колонок (по числу различных категорий состояния изучаемого свойства, принятых в исследовании).

Таблица 3

	Катег. 1	**Катег. 2**		**Катег. i**		**Катег. C**	
Выборка 1	O_{11}	O_{12}		O_{1i}		O_{1c}	n_1
Выборка 2	O_{21}	O_{22}		O_{2i}		O_{2c}	n_2
	O_{11} + O_{21}	O_{12} +O_{22}		O_{1i}+ O_{2i}		O_{1c} + O_{2c}	N = n_1+ n_2

На основе данных таблицы 3 можно проверить нулевую гипотезу о равенстве вероятностей попадания объектов первой и второй совокупностей в каждую из *i* (*i=l*, 2, ..., С) категорий, т. е. проверить выполнение всех следующих равенств: $p_{11}= p_{21}$, $p_{12} = p_{22}$, ..., $p_{1c} = p_{2c}$. Возможна, например, проверка гипотезы о равенстве вероятностей получения отметок «5», «4», «3» и «2» за выполнение учащимися контрольных и экспериментальных классов некоторого задания.

Для проверки нулевой гипотезы с помощью критерия χ^2 на основе данных таблицы 2хС подсчитывается значение статистики критерия *T* по следующей формуле:

$$\mathrm{T} = \frac{1}{n_1 \cdot n_2} \sum_{i=1}^{C} \frac{(n_1 O_{2i} - n_2 O_{1i})^2}{O_{1i} + O_{2i}} \quad (2)$$

где n_1 и n_2 — объемы выборок.

Значение *T,* полученное на основе экспериментальных данных, сравнивается с критическим значением $T_{критич}$, которое определяется по таблице χ^2 с v = С—1 степенью свободы с учетом выбранного уровня значимости α. При выполнении неравенства *T*>$T_{критич}$ нулевая гипотеза отклоняется на уровне α и принимается альтернативная гипотеза. Это означает, что распределение объектов на *С* категорий по состоянию изучаемого свойства различно в двух рассматриваемых совокупностях.

Пример 2. Рассмотрим методику сравнения результатов контрольной работы, проверявшей усвоение одного из разделов курса информатики учащимися первого и второго районов на основе выполнения определенного набора заданий.

Методом случайного отбора из учащихся первого района, писавших работу, была составлена выборка объемом 50 человек, из учащихся второго района — выборка объемом 50 человек. В соответствии со специально разработанными критериями оценки выполнения работы каждый ученик мог попасть в одну из четырех категорий: неудовлетворительно, удовлетворительно, хорошо, отлично. Результаты выполнения работы двумя выборками учащихся используем для проверки гипотезы о том, что набор заданий № 1 способствует лучшему усвоению проверяемого раздела курса, т. е. учащиеся первого экспериментального района в среднем будут получать более высокие оценки, чем учащиеся второго района.

Результаты выполнения работы учащимися обеих выборок запишем в виде таблицы 2х4 (табл. 4).

Таблица 4

	неуд.	**удовл.**	**хорошо**	**отлично**	
Выборка 1	$O_{11}=3$	$O_{12}=19$	$O_{13}=18$	$O_{14}=10$	$n_1=50$
Выборка 2	$O_{21}=9$	$O_{22}=24$	$O_{23}=12$	$O_{24}=5$	$n_2=50$
	12	43	30	15	N=100

В соответствии с условиями использования критерия χ^2 подсчет статистики критерия производится по корректированной формуле (2).

$$T = \frac{1}{n_1 \cdot n_2}\left(\frac{(n_1O_{21} - n_2O_{11})^2}{O_{11} + O_{21}} + \frac{(n_1O_{22} - n_2O_{12})^2}{O_{12} + O_{22}} + \frac{(n_1O_{23} - n_2O_{13})^2}{O_{23} + O_{13}} + \frac{(n_1O_{24} - n_2O_{14})^2}{O_{14} + O_{24}}\right)$$
$$= \frac{1}{16}(48 + 9{,}3 + 19{,}2 + 26{,}7) = 6{,}45$$

В соответствии с условиями применения двустороннего критерия хи-квадрат по таблице [2, С. 130] для одной степени свободы (v = 4-l = 3) и уровня значимости α = 0,05 найдем $T_{критич}$ = 7,815. Отсюда верно неравенство $T_{наблюд}<T_{критич}$ (6,45<7,815). Согласно правилу принятия решений для критерия χ^2, полученный результат не дает достаточных оснований для отклонения нулевой гипотезы, т. е. набор заданий № 1 способствует лучшему усвоению проверяемого раздела курса.

Литература

1. Введение в научное исследование по педагогике: Учеб. пособие для студентов пед. ин-тов. / Под ред. В. И. Журавлева. – М.: Просвещение, 1988. – 168 с.
2. *Грабарь М. И., Краснянская К. А.* Применение математической статистики в педагогических исследованиях. Непараметрические методы. М., «Педагогика», 1977. – 136 с.
3. *Набиулина Л. М., Тухташев У. Ф.* Актуальность изучения современных языков программирования в системе непрерывного образования Республики Узбекистан. // Проблемы современной науки и образования, 2014, № 9 (27), – С. 12-14.