УДК 13+37.0+004

Жаркова Галина Алексеевна

Кандидат педагогических наук, доцент кафедры информационных технологий, zharkovaga@inbox.ru, Ульяновск

ИСПОЛЬЗОВАНИЕ ВЫЧИСЛИТЕЛЬНЫХ НОМОГРАММ ПРИ ОБРАБОТКЕ ДАННЫХ ПЕДАГОГИЧЕСКИХ ЭКСПЕРИМЕНТОВ

Аннотация. Статья посвящена особенностям подготовки и проведения педагогических экспериментов, математической обработке их результатов. Среди индикаторов выделяются те, которые допускают статистическую обработку, проверку различных статистических гипотез.

Ключевые слова: педагогический эксперимент, информационная культура, индикаторы педагогического эксперимента, статистические методы, номограммы специальных функций.

Zharkova Galina Alekseevna

Chair of Information Technologies, Ulyanovsk State University, zharkovaga@inbox.ru, Ulyanovsk

USING COMPUTATION NOMOGRAMS IN PROCESSING DATA OF PEDAGOGIC EXPERIMENTS

Abstract. This paper deals with preparation and conducting pedagogic experiments as well as processing their results. We select the indicators, which allow statistical processing and verification of different statistical hypotheses.

Keywords: Pedagogic experiment, informational culture, indicators of pedagogic experiment, statistical methods, nomograms of special functions.

Анализ любого педагогического факта или явления требует причинное объяснение его появления, установление связи с другими педагогическими явлениями. Как и во многих других науках, основным инструментом исследователя является эксперимент, в данном случае – учебно-воспитательный эксперимент. В ходе эксперимента исследователь получает новые опытные данные, характеризующие изучаемое педагогическое явление, ставит и решает задачи проверки гипотез об эффективности тех или иных методов, форм и средств обучения и воспитания.

Идеальным результатом эксперимента было бы выявление всех латентных (не наблюдаемых напрямую) величин, определяющих развитие педагогического явления, математическая формулировка зависимостей между входными, латентными и наблюдаемыми величинами, то есть полное описание динамики эксперимента [1]. После этого можно было бы ставить вопрос о синтезе педагогического явления с заранее заданными свойствами, о выборе оптимальной стратегии развития.

Однако ничего подобного современная педагогическая наука не предлагает, математический аппарат, используемый для изучения других управляемых систем: в технике, биологии, медицине, для этих целей не подходит, или, лучше сказать, мы его не умеем использовать.

Планируя педагогический эксперимент, педагог определяет объект исследования, педагогические методы воздействия на него, описывает средства измерения результатов эксперимента. В экспериментах объект это не абстрактное человеческое сообщество, а вполне конкретное множество людей, связанных одним учебным процессом, образованием, возрастом и т.п.

Среди бесконечного множества характеристик объекта – человека – нас будут интересовать только числовые переменные. Традиционно они подразделяются на индикаторы (их можно измерить непосредственно в ходе эксперимента) и латентные (скрытые, неявные) переменные, то есть их невозможно измерить непосредственно. Математической (иногда метрической) моделью эксперимента называется связь или выражение индикаторов через латентные переменные [2, С. 158]. Присвоение конкретных значений индикаторам производится в ходе специальной измерительной процедуры, в роли которой выступают: тестирование, экзамен, контрольная работа и т.п. При необходимости достичь высокого качества педагогических измерений к подобным процедурам необходимо предъявлять серьезные требования: объективность, полнота, воспроизводимость, непротиворечивость и многие другие. В данной работе мы не будем на этом останавливаться подробно, подразумевая выполнение всех необходимых требований.

Важно отметить, что числовое значение, присвоенное индикатору, представляет собой отметку на определенной шкале. Широко известны номинальная (наименований), порядковая (ранговая), интервальная (количественная) шкалы, шкала отношений (или пропорциональная шкала). В отдельных случаях приходится конструировать нелинейные шкалы, наиболее подходящие для данного эксперимента.

Опишем типичный «педагогический эксперимент», чуть ли не ежедневно проводимый в практической педагогике для оценивания успешности обучения школьника (студента, слушателя). Латентная переменная здесь – уровень знаний испытуемого по данной теме. Предположим, что n школьникам в ходе контрольной работы (теста, школьного экзамена, ЕГЭ), например, по информатике, предлагается m заданий по теме прошедшего цикла обучения. При проектировании задания автор (педагог) явно или неявно оценивает уровень трудности каждого задания, его валидность (способность отразить знания испытуемых), адекватность данной теме, творческую составляющую и т.п. Все эти оценки субъективны.

При решении каждого задания испытуемый пытается реализовать (отразить) достигнутый уровень знаний по данной теме, а также некоторые свои личностные характеристики (интеллект, способность мыслить нестандартно, творческие способности и т.п.). Результат выполнения задания (решение) рассматривается экспертом (преподавателем). Предположим, что его оценка представляет собой число x_{ij} (i-й школьник в списке, j-е задание в тесте, i=1, 2, …, n, j=1, 2, …, m, $m<n$).

В большинстве случаев в педагогике x_{ij} представляет собой ранговую (порядковую) оценку успешности выполнения данного задания. Часто это дихотомичная оценка: $x_{ij} \in \{1;0\}$ (то есть задание либо выполнено, либо нет). В лучшем случае $x_{ij} \in \{0, 1, \dots, x_{max}\}$, то есть экспертом устанавливаются критерии выполнения задания на некоторой числовой шкале. Для опытного педагога, знающего определенные приемы, достичь при проектировании теста величины x_{max}=5 или даже x_{max}=10 не представляется непреодолимой трудностью. Очевидно, что чем больше x_{max}, тем ближе мы приближаемся к количественной шкале оценивания выполнения задания, в которой только и можно достаточно надежно применять математические метода анализа данных. Увеличение числа заданий m и соответствующее кумулятивное вычисление итоговой оценки x_i еще более улучшают статистические характеристики этой оценки. Например, в ЕГЭ по информатике итоговая оценка (в первичных баллах) представляет собой число от 0 до 40.

Сама постановка вопроса, формулировка цели эксперимента приводит нас к пониманию того факта, что проверка гипотезы будет иметь статистический характер и будет проводиться статистическим методом. В педагогических экспериментах это – основной математический метод исследования. Его основная посылка – представление о бесконечной генеральной совокупности. В математической статистике это понятие является абстракцией, отражающей представление, что эксперимент можно повторять бесконечно много раз. При этом предполагается, что относительная частота интересующего нас события обладает свойством статистической устойчивости, более того, при неограниченном возрастании объема выборки эта частота достигает некоторого «истинного» значения [4, с. 165].

Однако эти предположения именно в педагогических экспериментах вызывают обоснованную критику. Причин здесь несколько: весьма ограниченные объемы выборок, слабая воспроизводимость результатов педагогических экспериментов, имеющая место из-за «человеческой» природы как субъекта измерения, так и «прибора» измерения, опасение, что если результат эксперимента зарегистрирован и измерен педагогом-исследователем, то имеются ли основания приписать изменение состояния обучаемого лишь экспериментальному воздействию, а не каким-либо другим факторам, вообще не связанным с экспериментальными воздействиями [2, с. 152].

Указанные обстоятельства иногда приводят исследователя к получению недостоверных результатов обработки данных, слабо обоснованных выводов, а то и к вульгаризации итогов исследования. Отсюда превалирование среди педагогов-исследователей качественных методов исследования, принципиальное неприятие вообще любых количественных методов.

На наш взгляд эти обстоятельства проистекают от плохого проектирования эксперимента. Под этим понимается ситуация, когда экспериментатор не отдает себе отчет, какие предпосылки (гипотезы) он выдвигает в эксперименте, не заботится о проверке их выполнения, а статистические выводы подгоняются под заранее известный ответ.

Статистический подход приводит к построению эмпирического распределения оценок, выделения интервалов, которым будет присвоена одинаковая итоговая оценка. При этом в исследованиях выдвигаются гипотезы о том или ином теоретическом распределении, которым якобы соответствуют полученные гистограммы, оцениваются параметры, проверяются критерии согласия и однородности, вычисляются доверительные вероятности и т.п. Все это, так или иначе, строится на предположении о статистической однородности результатов эксперимента и нормальности всех используемых распределений. Как уже отмечалось, эти предположения вызывают обоснованную критику и должны проверяться специальными процедурами.

Математические методы обработки данных и проверки статистических гипотез весьма развиты, разнообразны, однако содержат трудоемкие процедуры, включая табличные вычисления специальных функций. Здесь мы предложим алгоритмы, используемые в наиболее часто встречающихся в практике педагогических экспериментов статистических процедурах, но в которых вместо таблиц специальных функций используются вычислительные номограммы. При этом предварительные вычисления сведены к минимуму и могут быть осуществлены на обычном калькуляторе.

Первой таким алгоритмом является процедура проверки гипотезы о принадлежности экспериментальной выборки генеральной совокупности с нормальным распределением. Необходимость проверки этой гипотезы вызвана тем, что большинство статистических методов принятия решений основаны на предположении о нормальности распределений случайных величин. Причиной распространенности нормального распределения в прикладных науках, в том числе в педагогике, является общепринятое предположение о многочисленности и независимости факторов, влияющих на значение случайной величины – результата эксперимента, а также о том, что ни один фактор в отдельности не является определяющим.

Под выборкой будем понимать набор оценок, полученный в результате применения педагогического теста успешности обучения (или уровня формирования качества знаний, например, уровня информационной культуры личности [3]) к группе педагогических объектов (студентов, слушателей и т.п.). Предполагается, что с помощью применения различных методов шкалирования «оценка» допускает большее число фиксированных уровней, то есть представляет собой, по существу, интервальную оценку, к которой могут быть применены любые математические процедуры.

Итак, в результате эксперимента получена выборка $\{x_i\}, (i = 1, \dots, n)$. Требуется проверить гипотезу *H*, что эта выборка извлечена из некоторой нормально распределенной генеральной совокупности. Параметры этой совокупности обычно неизвестны и определяются по выборке: $a = \frac{1}{n}\sum x_i, \sigma = \sqrt{D}$, где $D = \frac{1}{n-1}\sum(x_i - a)^2$ – несмещенная выборочная оценка дисперсии. Вычислим нормированные данные выборки: $t_i = (x_i - a)/\sigma$. По известным свойствам нормального распределения, если гипотеза *H* верна, то $\{t_i\}$ будет выборкой из стандартной нормальной генеральной совокупности с параметрами $a = 0, \sigma = 1$.

Разобьем отрезок $[0; 1]$ на k $(k = 2, 3, \dots)$ одинаковых по длине промежутков и вычислим $z_j = \Phi^{-1}(j/k), (j = 1,2, \dots, k - 1)$,

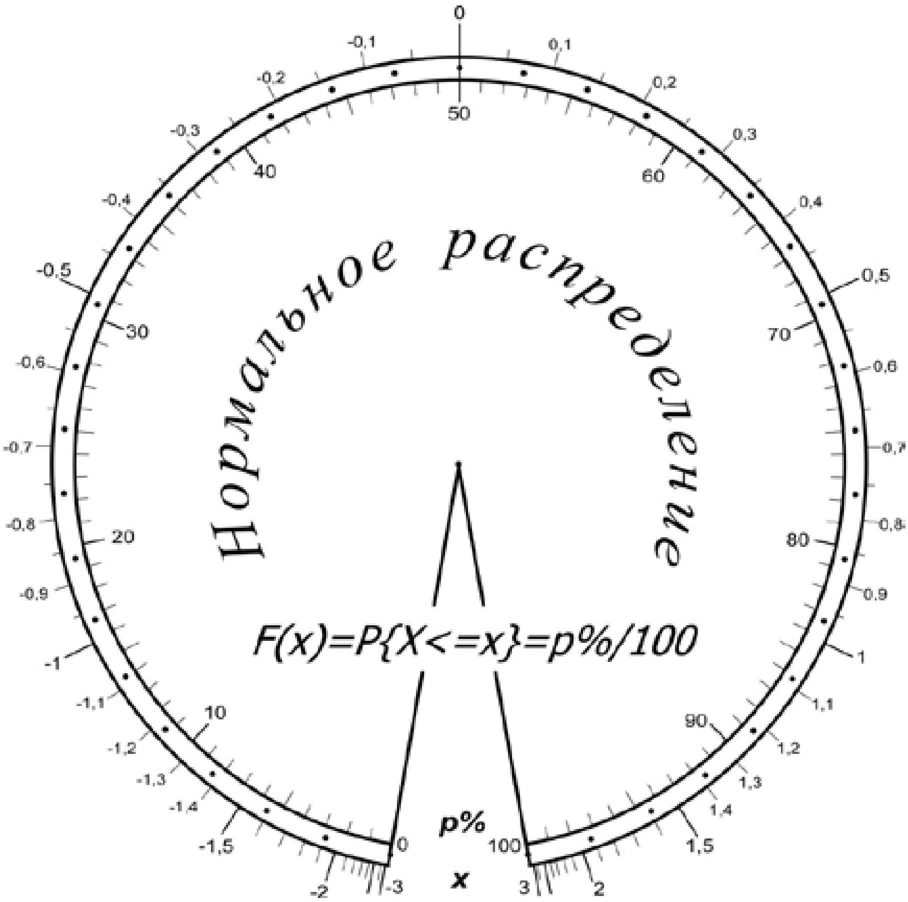

Рис 1. Круговая номограмма для вычисления прямой и обратной функций распределения стандартного нормального распределения

где $\Phi^{-1}(j/k)$ – обратная функция к функции стандартного нормального распределения.

Вычисление значений z_j упрощает номограмма, приведенная на рис. 1. Достаточно внутреннюю окружность на номограмме разбить на *k* одинаковых частей, что легко сделать по специально нанесенным точкам разбиения в том случае, если *k* является делителем числа 36 (то есть *k*=2,3,4,6,9,12,18,36). При этом на внешней окружности считываем значения z_j.

Тем самым мы получаем *k* непересекающихся промежутков:

$(-\infty; z_1], (z_1; z_2], \dots, (z_{k-1}; +\infty)$.

Подсчитаем количество элементов t_i выборки, попавших в каждый из этих интервалов, обозначим эти количества n_j $(j = 1,2, \dots, k)$.

Очевидно, $\sum_{j=1}^{k} n_j = n$. Подсчитаем число

$$\chi^2 = \frac{k}{n}\sum_{j=1}^{k} n_j^2 - n \qquad (1)$$

Это и есть величина χ^2 – мера отклонения экспериментального распределения выборки от стандартного нормального распределения. Поясним этот факт преобразованием величины χ^2 из [4, стр. 454]:

$$\chi^2 = \sum_{j=1}^{k} \frac{(n_j - np_j)^2}{np_j} = \sum \frac{n_j^2 - 2n_j p_j + n^2 p_j^2}{np_j} =$$

$$= \sum \frac{n_j^2}{np_j} - n = \frac{k}{n}\sum n_j^2 - n ,$$

где p_j – вероятность стандартной нормальной случайной величины попасть в *j*-й из указанных промежутков, причем $p_j = \frac{1}{k}$ для всех *j* и $\sum p_j = 1$.

Как известно [4, стр. 457], если гипотеза *H* верна, то величина χ^2 должна быть мала, точнее она отличается от 0 только в силу случайных причин, еще точнее: большие от-

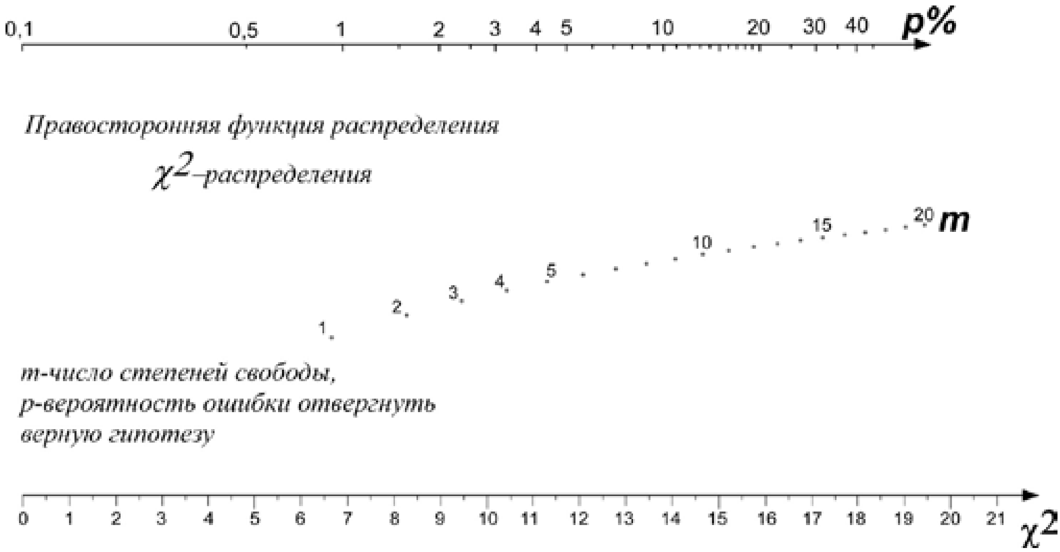

Рис 2. Номограмма из выравненных точек для вычисления правосторонней функции χ^2-распределения

клонения χ^2 от 0 маловероятны. Это означает, что если величина χ^2 велика (произошло редкое событие), то гипотезу *H* следует отвергать. Однако следует понимать, что при этом мы можем допустить ошибку: отвергнуть все же верную гипотезу *H*!

Вероятность этой ошибки *p*%, то есть вероятность отвергнуть верную гипотезу о принадлежности рассматриваемой выборки некоторой нормальной генеральной совокупности, может быть подсчитана с помощью специальной номограммы, приведенной на рис. 2.

Номограмма предназначена для решения уравнения $P_m(\xi > \chi^2) = p\%/100$, где ξ – случайная величина с χ^2-распределением с m степенями свободы. При использовании номограммы потребуется указать параметр m – число степеней свободы. В нашем случае он подсчитывается так: $m = k - 3$, если параметры нормального распределения оценивались по выборке, $m = k - 1$, если параметры нормального распределения были заданы заранее.

Все «вычисления» на номограмме проводятся с помощью обыкновенной линейки. Достаточно известные два числа из трех: χ^2 и *m* или *p*% и *m*, отложить на соответствующих шкалах, полученные точки соединить по линейке прямой; тогда на третьей шкале можно прочитать значение недостающей переменной. Следует понимать, что номограмма дает лишь приближенно правильный ответ. Относительная погрешность по оси χ^2 не превосходит 1,5–2%.

Если полученная вероятность ошибки *p*% велика (например, более 20%), то гипотезу *H* можно принимать, то есть признать нормальное распределение рассматриваемой выборки. Более осторожным (но и более надежным) является совет повторить эксперимент, получить новую выборку и еще раз проверить гипотезу.

Если эта вероятность мала (например, 5% или меньше), то гипотезу о нормальности распределения выборки следует отвергать. В промежуточных случаях исследователь должен принимать решение, опираясь на свой опыт и уровень ответственности за то или иное принятое решение.

Можно следовать обычным рекомендациям справочников по статистической обработки данных: заранее задаться уровнем значимости *p*% (обычно *p*%=0,1%, 1% или 5%); зная число степеней свободы m, вычислить по номограмме значение $\chi^2 = \chi^2_{\text{крит}}$ (оно называется критическим *p*-процентным уровнем значимости). Теперь, если вычисленное значение $\chi^2 > \chi^2_{\text{крит}}$, то наша выборка обнаруживает значимое отклонение от гипотезы *H* (то есть экспериментальное распределение значимо отличается от нор-

Таблица 1

Успеваемость	Группа 1	Группа 2	Всего:	Доли
Уровень 1	g_1	h_1	$g_1 + h_1$	ω_1
Уровень 2	g_2	h_2	$g_2 + h_2$	ω_2
…	…	…	…	…
Уровень k	g_k	h_k	$g_k + h_k$	ω_k
Всего:	m	n	$m+n$	ω
			$\chi^2 =$	…

мального распределения), и мы должны гипотезу H отвергать на p%-м уровне значимости. Вероятность того, что такое положение возникнет в случае, когда гипотеза на самом деле справедлива и, следовательно, была отвергнута ошибочно, равна p%. Если вычисленное значение $\chi^2 \leq \chi^2_{\text{крит}}$, то это значение можно считать совместимым с гипотезой H.

Для проведения этапа планирования педагогического эксперимента следует отметить, что чем больше число степеней свободы (то есть число k – количество интервалов разбиения), тем надежнее будут результаты. Однако при этом должно быть выполнено условие, чтобы все $n_j > 10$ [4, с. 457].

В качестве примера математического моделирования нами было сгенерировано по алгоритму Марсалье [5, стр. 465] 1000 выборок по 1000 значений в каждой, выбранных из генеральной совокупности со стандартным нормальным распределением. Подсчитывая для каждой выборки величину χ^2 по описанной выше методике (для $k = 12$, $m = 11$), получили, что гипотеза о нормальности распределения выборки отвергается на 0,1%-м уровне значимости ($\chi^2 > \chi^2_{\text{крит}} = 31{,}3$) всего 1 раз, на 1%-м уровне значимости ($\chi^2 > \chi^2_{\text{крит}} = 24{,}7$) – 7 раз, на 5%-м уровне значимости ($\chi^2 > \chi^2_{\text{крит}} = 19{,}6$) – 46 раз. Эти данные укладываются в расчетные величины для χ^2-распределения, что подтверждает как справедливость предлагаемой методики, а так и хорошие качества использованного алгоритма генерации.

Отметим, что формула (1) подсчета величины χ^2 остается справедливой и для любого другого теоретического распределения генеральной совокупности. В подсчетах меняется только набор промежутков $(z_{j-1}; z_j]$. Например, для равномерного распределения на отрезке [a; b] достаточно этот отрезок разделить на k одинаковых частей. Разумеется, номограммой 1 при этом пользоваться не имеет смысла. Расчеты по номограмме 2 сохраняют свой смысл, только надо иметь в виду, что величина χ^2 будет иметь необходимое распределение только в асимптотическом (при $n \to \infty$) смысле.

В педагогических экспериментах часто ставится задача статистически проверить влияние или эффективность того или иного педагогического воздействия (новой методики, формы, средства, приема и т.п.) [2, с. 150]. В этом случае формируются две группы испытуемых: группа 1 из испытуемых, подвергнутых изучаемому воздействию и так называемая «контрольная группа 2», оставшаяся без указанного воздействия. В обеих группах отслеживается одинаковый признак, например, достигнутый уровень информационной культуры личности [3] (или просто успеваемость по рассматриваемому предмету). Признак квантуется на k уровней, которые могут быть выражены как в порядковой, так и в номинальной шкале.

По окончании эксперимента составляется таблица («таблица сопряженности») количества обучающихся в каждой группе, находящихся на каждом из уровней успеваемости (см. табл 1).

Здесь: $\omega_i = g_i/(g_i + h_i)$, $\omega = m/(m+n)$. Величину χ^2 теперь можно вычислить по формуле, удобной для практических расчетов [4, стр. 485]

$$\chi^2 = \frac{1}{\omega(1-\omega)}\left(\sum_{i=1}^{k} g_i\omega_i - m\omega\right) \quad (2)$$

Вычислив в (2) по данным проведенного эксперимента величину χ^2 и сравнивая ее с $\chi^2_{\text{крит}}$, вычисленным по номограмме 2 при $m = k - 1$ степенях свободы, принимаем решение о значимости изучаемого педагогического воздействия: если $\chi^2 > \chi^2_{\text{крит}}$, то гипотезу об однородности групп следует отвергать, и при этом вероятность ошибиться в случае, если гипотеза все же верна, будет не больше *p%*. В противном случае гипотезу об однородности групп по данному признаку можно принимать, что трактуется как отсутствие влияния педагогического воздействия. Наоборот, неоднородность распределений трактуется как значимость исследуемого педагогического воздействия. Величину *p%* заранее берут малым числом (обычно 1%, 5% или 10%).

Статистическое обоснование данной процедуры следующее: число χ^2 является случайной величиной (так как зависит от случайной выборки), причем асимптотически распределение этой величины известно. Если верна гипотеза об однородности двух выборок (то есть они извлечены из одной и той же генеральной совокупности), то значение χ^2 должно быть близко к нулю, и можно считать, что отклонение χ^2 от 0 вызвано случайными, незначащими причинами, причем вероятность этого отклонения может быть подсчитано.

В статье обсуждаются принципы и алгоритмы статистической обработки результатов педагогических экспериментов. Предложены новые процедуры проверки статистических гипотез, часто встречающихся в педагогических исследованиях. Получены новые формулы вычисления статистических характеристик, более простые и удобные. Использование статистических таблиц специальных функций полностью исключено и заменено расчетами на вычислительных номограммах, разработанных нами.

Библиографический список

1. *Аванесов В. С.* Педагогическое измерение латентных качеств // Педагогическая диагностика. – 2003. – №4. – С. 69–78.

2. *Михеев В. И.* Моделирование и методы теории измерений в педагогике. Изд. 2-е. – М.:Едиториал УРСС, 2004. – 200 с.

3. *Жаркова Г. А.* Метрическая модель информационной культуры учащихся // Сибирский педагогический журнал. – 2011. – №3. – С. 84ж93.

4. *Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 648 с.

5. *Каханер Д., Моулер К., Нэш С.* Численные методы и математическое обеспечение: пер. с англ. – М.: Мир, 1998. – 575 с.