# Практические вопросы подсчета корреляций результатов оценочных и конкурсных процедур при оценке объективности системы оценки качества образования образовательных организаций

Вопросы **объективности** сложившихся в образовательных организациях **внутренних систем оценки качества образования** невозможно рассматривать в отрыве от вопросов корреляций результатов участников с другими результатами этих же участников, с результатами других участников по этим же процедурам и с иными показателями качества образования.

1) В части 7 статьи 28 273-ФЗ устанавливает, что образовательная организация несет ответственность за:

* невыполнение или ненадлежащее выполнение функций, отнесенных к ее компетенции,
* реализацию не в полном объеме образовательных программ в соответствии с учебным планом,
* качество образования своих выпускников,
* жизнь и здоровье обучающихся, работников образовательной организации.

2) В пункте 6 Приказа Министерства образования и науки РФ от 14 июня 2013 г. №462 «Об утверждении Порядка проведения самообследования образовательной организацией» определяются направления самообследования:

* образовательной деятельности,
* системы управления организации,
* содержания и качества подготовки обучающихся,
* организации учебного процесса,
* востребованности выпускников,
* качества кадрового, учебно-методического, библиотечно-информационного
* обеспечения, материально-технической базы, функционирования внутренней
* системы оценки качества образования.

3) Пунктом 29 статьи 2 273-ФЗ установлено, что **качество образования** – комплексная характеристика образовательной деятельности и подготовки обучающегося, выражающая степень их соответствия федеральным государственным образовательным стандартам…. образования и (или) потребностям физического или юридического лица, в интересах которого осуществляется образовательная деятельность, в том числе степень достижения планируемых результатов образовательной программы.

3) На основании части 7 статьи 12 273-ФЗ организации, осуществляющие образовательную деятельность по имеющим государственную аккредитацию образовательным программам …., разрабатывают образовательные программы **в соответствии с федеральными государственными образовательными стандартами** (далее – ФГОС) и с учетом соответствующих примерных основных образовательных программ. Планируемые результаты подготовки обучающихся, указанные в требованиях ФГОС общего образования к структуре основной образовательной программы в разделе «**Оценка достижения учащимися** результатов освоения основной образовательной программы», должны являться одним из основных объектов оценки качества образования.

В соответствии с частью 3 статьи 11 273-ФЗ федеральные государственные образовательные стандарты (далее – ФГОС) включают в себя требования к: **структуре и содержанию** основных образовательных программ; **условиям** реализации основных образовательных программ, в том числе кадровым, финансовым, материально-техническим и иным условиям; **результатам освоения** основных образовательных программ.

3) Соответственно, **потенциальными источниками снижения качества образования** (рисками образовательной деятельности) являются: нарушение требований ФГОС общего образования, неудовлетворенность потребителей качеством образовательных услуг.

Результаты самообследований, констатирующие нарушения требований ФГОС крайне редки (нам найти не удалось, но наверняка имеются какие-то отдельные случаи), поскольку такие выводы повлекут за собой юридическую и дисциплинарную ответственность, а кроме того, вызовут явную неудовлетворённость потребителей услуг (допустим, родителей). В связи с этим выявленные недостатки и нарушения деятельности устраняются, условия оперативно адаптируются, структура и содержание основных образовательных программ меняется, но… Результаты учащихся, которые являются основным объектом оценки, оперативно изменить, безусловно, невозможно. Обучение и развитие компетенций – длительный процесс, не поддающийся мгновенной коррекции.

В этот момент появляется известный всем учителям и ученикам «диалог с совестью», в результате которого родилась известная фраза «три ставим, два в уме». Отметка «три», соответствующая освоению требуемого объёма материала, позволяет достичь сиюминутных целей формального соответствия требованиям к результатам освоения, но порождает целый комплекс стратегически важных проблем – от формирования устойчивого мнения о некачественной работы учителей и «плохих учебников», заниженного уровня требования к себе и своей работе, выгорания и других психологических проблем, до административных и уголовных дел за преднамеренное искажения результатов государственной итоговой аттестации, совершаемых должностным лицом по злому умыслу.

4) Любой директор рано или поздно задастся вопросом, каким образом выявить необъективность оценивания? Ведь процедура проходит единожды, и второго шанса провести её зачастую просто нет (учащиеся покинули заведение, другие условия проведения и т.п.).

Банальная логика подсказывает решение – сравнение полученных результатов с другими результатами этих же учащихся. Например, с текущими оценками и олимпиадами по предмету. Возможно сравнение с другими предметами (того же цикла – например, естественнонаучного; с той же методикой – знаково-символическая, и т.п.).

С другой стороны, возможно сравнение результатов этих участников с результатами других участников в рамках одной процедуры или даже нескольких схожих. Например, можно сравнить учащихся 4-х классов школы А по русскому языку ВПР с учащимися 4-х классов школы Б по русскому языку ВПР; и сравнение результатов этих же учащихся в олимпиаде по русскому языку.

Логично ли будет предположить, что если средний балл среди учащихся школы А на ВПР был вдвое выше, чем у учащихся школы Б, то и на олимпиаде будет такое же соотношение? Ведь обе процедуры проверяют одно и то же. Конечно, возможна ситуация, когда учащихся школы Б хорошо подготовили к олимпиаде, в отличие от школы А. Но насколько объективны такие результаты?

Напомним, что результат не выражается в умении решать конкретные задачи. Результат в ФГОС сформулирован в виде компетенции, которая не зависит от конкретных формулировок задания и сиюминутного настроя участника.

Последний случай – связь между результатами участников с иными показателями качества образования. В пунктах 1 и 2 мы видим примерный перечень направлений внутреннего мониторинга. Исчерпывающего списка всех показателей качества, безусловно, нет. Но, мы хорошо понимаем, что такой показатель, как «наличие командировок учителей иностранного языка в страну-носитель» наверняка будет влиять на качество преподавания иностранного языка, а наличие лабораторий будет влиять на изучение химии, наличие футбольного и волейбольного полей, бассейна и лыжных трасс – на физическую подготовку. Отдельным вопросом является корреляция уровня педагогической компетенции учителя с результатами учащихся, вопрос неоднозначен и требует дополнительных дискуссий, выходящих за рамки данной статьи.

Вернёмся к тезису, с которого мы начали: вопросы **объективности** сложившихся в образовательных организациях **внутренних систем оценки качества образования** невозможно рассматривать в отрыве от вопросов корреляций результатов участников с другими результатами этих же участников, с результатами других участников по этим же процедурам и с иными показателями качества образования. Или проще: вопросы объективности внутренних систем оценки качества требуют рассмотрения корреляций результатов.

## Опыт расчёта показателя объективности на основе корреляций результатов

Опыт расчёта показателя объективности на основе корреляций результатов с есть у Федерального института оценки качества образования ([https://fioco.ru/Media/Default/Методики/Уровень объективности оценки образовательных результатов.pdf](https://fioco.ru/Media/Default/%D0%9C%D0%B5%D1%82%D0%BE%D0%B4%D0%B8%D0%BA%D0%B8/%D0%A3%D1%80%D0%BE%D0%B2%D0%B5%D0%BD%D1%8C%20%D0%BE%D0%B1%D1%8A%D0%B5%D0%BA%D1%82%D0%B8%D0%B2%D0%BD%D0%BE%D1%81%D1%82%D0%B8%20%D0%BE%D1%86%D0%B5%D0%BD%D0%BA%D0%B8%20%D0%BE%D0%B1%D1%80%D0%B0%D0%B7%D0%BE%D0%B2%D0%B0%D1%82%D0%B5%D0%BB%D1%8C%D0%BD%D1%8B%D1%85%20%D1%80%D0%B5%D0%B7%D1%83%D0%BB%D1%8C%D1%82%D0%B0%D1%82%D0%BE%D0%B2.pdf)). Их методика довольно проста и основана на следующем алгоритме:

Шаг 1. Применение математических методов для выявления результатов, сильно отклоняющихся от нормы.

Шаг 2. Маркировка образовательных организаций по признакам отклонения от нормы (признакам необъективности.

Шаг 3. Вычисление доли (процента) образовательных организаций, не промаркированных ни по одному признаку.

В указанном приложении приведён математический алгоритм маркировки образовательных организаций, основанный на нескольких этапах вычислений.

### Математические модели

#### Математическая модель расчётов доверительных диапазонов(используется для одной процедуры, но разных групп участников)

1. Подготовка полного перечня результатов, где $X\_{i}$ – итоговый балл *i*-го участника, а *n* – количество участников.

2. Расчёт среднего арифметического $\overbar{X}=\frac{\sum\_{i=1}^{n}X\_{i}}{n}$. Простыми словами, нужно суммировать все баллы всех участников и полученную сумму поделить на их количество.

3. Считаем выборочное стандартное отклонение по региону или по школе по формуле:

$$S=\sqrt{\frac{\sum\_{i=1}^{n}\left(X\_{i}-\overbar{X}\right)^{2}}{n-1}}$$

Простым языком, теперь нужно из каждого результата вычесть средний результат, потом получившиеся значения возвести в квадрат и просуммировать. Полученную сумму поделить на количество результатов минус 1. Обращаем внимание на то, что для единственного результата посчитать стандартное отклонение невозможно, что нормально. Ибо как можно, чтобы единственный результат отклонился от среднего?

4. Примем уровень достоверности 95%, при котором коэффициент $t\_{a}=1.96$.

Тогда доверительный интервал для всего диапазона участников получается:

$\left[\overbar{X}-t\_{a}∙\frac{S}{\sqrt{n}};\overbar{X}+t\_{a}∙\frac{S}{\sqrt{n}};\right]$.

5. Производится расчёт для всех ОО (или групп учащихся).

6. Если левая (нижняя) граница доверительных интервалов образовательной организации находится правее, чем правая (верхняя) граница доверительного интервала среднего балла по региону, то образовательная организация помечается маркером «завышение». В случае, если правая (верхняя) граница доверительных интервалов образовательной организации находится левее, чем левая (нижняя) граница доверительного интервала среднего балла по региону, то образовательная организация помечается маркером «занижение».

7. Результат предоставляется по образовательным организациям, а для мунициального образования рассчитывается индекс объективности.

Посмотрим, как делать это в Microsoft Excel на примере диагностических работ по математике в 10-ых классах в октябре 2020 года (книга Excel «Электронное приложение.xlsx», лист «Доверительный диапазон»).



**Рис. 1.** Расчёт доверительного интервала по всему диапазону значений

В столбце А представлены первичные баллы по всей совокупности (всей Свердловской области). В столбце B – отметки, соответствующие этому первичному баллу[[1]](#footnote-1). Дальнейшие расчёты будем производить по первичным баллам.

Поясним формулы в следующей таблице

| **Ячейка** | **Формула** | **Пояснение** |
| --- | --- | --- |
| E1 | =СЧЁТЗ(A:A)-1 | Формула считает число непустых строк в столбце A; и вычитает 1 строку, в которой находятся заголовки таблицы |
| E2 | =СРЗНАЧ(A:A) | Формула считает среднее арифметическое по столбцу A |
| E4 | =СТАНДОТКЛОН(A:A) | Формула считает стандартное отклонение по столбцу A |
| E6 | =E2-1,96\*E4/КОРЕНЬ(E1)или=E2-ДОВЕРИТ.НОРМ(0,04;E4;E1) | Формула ссылается на ранее посчитанные ячейки и считает левую точку доверительного интервала |
| G6 | =E2+1,96\*E4/КОРЕНЬ(E1)или=E2+ДОВЕРИТ.НОРМ(0,04;E4;E1)или=E2+ДОВЕРИТ.СТЬЮДЕНТ(0,04;E4;E1) | Формула ссылается на ранее посчитанные ячейки и считает правую точку доверительного интервала.Отметим, что в Excel две функции, реализующие приведённую логику. |

Таким образом, для всего столбца A высчитан доверительный интервал от 13,63..13,81.

Если для школы, в которой 21 результат, среднее значение 12, а стандартное отклонение 3,27, то доверительный диапазон получится от 10,6 до 13,39. Получается, что правая граница школы ниже, чем левая граница региона.

Соответственно, школа помечается как «занижение».

В 5-балльной шкале доверительный интервал от 2,94 до 3,81.

#### Математическая модель расчётов соответствия результатов одной оценочной процедуры другой оценочной процедуре

Математическая модель сводится к вычислению коэффициента ранговой корреляции Спирмана между двумя процедурами. Для этого выполняются последовательно следующие шаги.

1. Подготовка полного перечня результатов, где *n* – количество участников, $X\_{i}$ – результат участника по первой процедуре, $Y\_{i}$ – результат участника по второй процедуре.

2. Выставляем ранги в списках, где $d\_{xi}$ – ранг результата *i*-го участника в перечне результатов первой процедуры, а $d\_{yi}$ – ранг результата *i*-го участника в перечне результатов второй процедуры.

3. Высчитываем разницу рангов $d\_{i}=d\_{xi}-d\_{yi}$.

4. Считаем коэффициент Спирмана по формуле:

$$r\_{s}=1-\frac{6∙\sum\_{i=1}^{n}d\_{i}^{2}}{n^{3}-n}$$

а $d\_{i}$ – разность рангов по переменным (первичных баллов) *i*-го участника.

В качестве примера вычислим коэффициент корреляции баллов ЕГЭ по математике в 2020 году и баллов ОГЭ по математике в 2018 году для одних и тех же участников (книга Excel «Электронное приложение.xlsx», лист «Корреляция, коэфф. Спирмана»). При этом будем учитывать одну из школ области.



**Рис. 2.** Таблица для расчёта коэффициента корреляции Спирмана

Столбцы А и B заполняются значениями, которые необходимо сравнить между собой. В ячейке C2 вставляется формула =РАНГ(A2;A:A;1) для ранжирования. Впоследствии формула копируется на весь диапазон ранжирования.

Для расчёта квадрата разницы в ячейку E2 вставляется формула =(C2-D2)^2.

Теперь следует посчитать

* количество результатов в ячейке H4: =СЧЁТЗ(A:A)-1 ;
* сумму квадратов разницы в ячейке H5: =СУММ(E:E) ;
* коэффициент корреляции Спирмана в ячейке H6: =1-6\*H5/H4/(H4^2-1)

В результате проведённых расчётов хочется отметить два момента. В «стандартной» формуле Спирмана считается, что для нескольких одинаковых результатов должны быть выставлены одинаковые ранги. Так, для процента 34 по ЕГЭ в приведённом примере, должен быть РАНГ 1. При этом для 37 должен быть РАНГ 2. Однако, Excel при этом вычитывает РАНГ считает не так же, как это делает, например, DENSE\_RANK() или RANK() в Microsoft SQL Server, а указывает РАНГ 2. В связи с этим число, описывающая коэффициент Спирмана для ранговой корреляции отличается в информационных системах. Однако, суть остаётся такая же.

Например, для данного случая корреляция может быть представлена графиком, представленном на рисунке 3.



**Рис. 3.** Корреляция результатов ЕГЭ и ОГЭ для одной из школ

В приложении в Microsoft Excel Вы можете исследовать любые соотношения, заменяя данные в столбце A и столбце B на необходимые и копируя ячейки C3:E3 на весь диапазон.

Можно высчитать соответствие любых процедур. Важно научиться толковать коэффициент Спирмана.



**Рис. 4.** Трактовка коэффициента корреляции Спирмана

Абсолютное несоответствие получается при 100% «перевёрнутом» результате одной процедуры относительно другой. При полном же соответствии получается коэффициент +1.

Пример полного несоответствия – набор рангов 1, 2, 3 по первой процедуре и набор рангов 3, 2, 1 по второй процедуре. Набор значений от (-1;0] означает превалирующее несоответствие.

Школа, для которой результат первой процедуры не соответствует результату второй процедуры, помечаются как «Несоответствие результатов оценочных процедур».

Проведите опытную работу в Excel. Попробуйте посчитать процент соответствия оценок и какие результаты вызывают недоверие. Выделите их из общего списка. Каков процент этих результатов от общего количества? Сколько из них необходимо привести в соответствие, чтобы результаты перестали коррелировать с отрицательным коэффициентом.

#### Резкое изменение результатов одной параллели по одному предмету

1. Подготовка полного перечня результатов, где *n* – количество участников, $X\_{i}$ – результат участника по первой процедуре, $Y\_{i}$ – результат участника по второй процедуре; $X\_{max}$ – максимальный балл по первой процедуре, $Y\_{max}$ – максимальный балл по второй процедуре.

2. Вычисляем средний процент выполнения работы по первой процедуре, средний процент выполнения работы по второй процедуре. Для этого считаем отношение первичного балла к максимальному баллу в соответствии со спецификацией процедуры:

$S\_{Xi}=\frac{X\_{i}}{X\_{max}}$*,* $S\_{Yi}=\frac{Y\_{i}}{Y\_{max}}$*,*

где $X\_{max}$ – максимальный балл по первой процедуре, а $Y\_{max}$ – максимальный балл по второй процедуре.

3. Считаем средний процент выполнения работы по процедурам:

$\overbar{S\_{X}}=\frac{\sum\_{i=1}^{n}S\_{Xi}}{n}$,$ \overbar{S\_{Y}}=\frac{\sum\_{i=1}^{n}S\_{Yi}}{n}$ .

4. Рассчитываем коэффициент отношения долей выполнения работы по процедурам для образовательной организации:

$$K\_{оо}=\frac{\overbar{S\_{X}}}{\overbar{S\_{Y}}}$$

5. После подсчёта коэффициентов для каждой организации, составляем ряд из коэффициентов по всему региону, и высчитываем для данного ряда 25-й и 75-й процентили, межквартальный размах.

Для подсчёта процентилей делается следующее:

5.1 Считается количество результатов.

5.2 Результаты сортируются по возрастанию.

5.3 Высчитывается указатель на 25-процентный результат (если всего 100 результатов, то указатель будет 100\*0,25 = 25). Выписывается результат, стоящий на указанном месте. Это и есть 1-й квартиль или 25-й процентиль.

5.4 Высчитывается указатель на 75-процентный результат (если всего 100 результатов, то указатель будет 100\*0,75 = 75). Выписывается результат, стоящий на указанном месте. Это и есть 3-й квартиль или 75-й процентиль.

5.5 Высчитывается межквартильный размах, являющийся разницей между 75-м и 25-м процентилем. (1 и 3 квартилем).

6. Считаем нижнее критическое значение, которое равно разнице значений 1-го квартиля и 3 межквартильных размахов.

7. Считаем верхнее критическое значение, которое равно сумме значений 3-го квартиля и 3 межквартильных размахов.

8. Маркируем образовательные организации, у которых коэффициент ниже нижнего критического предела как «Резкий спад», а организации, у которых выше критического предела как «Резкий подъём».

Рассмотрим практическую реализацию данной работы в Excel (книга Excel «Электронное приложение.xlsx», лист «Резкое изменение»). Возьмём данные из предыдущего примера – соотношение ОГЭ и ЕГЭ по математике. В обоих случаях максимальный балл – 32.



**Рис. 5.** Исходные данные о первичных баллах и их доли от максимального

Внесём вручную максимально возможные баллы по первой процедуре в G2, а по второй – в G3.



**Рис. 6.** Таблица с константными и расчётными величинами

Появляется возможность расчёта процента первичного балла по формуле для ячейки C2: =ОКРУГЛ(A2\*100/$G$2;0) или =ЦЕЛОЕ(A2\*100/$G$2).

Считаем средний процент

- по первой процедуре в ячейке G5: =СРЗНАЧ(C:C) ;

- по второй процедуре в ячейке G6: =СРЗНАЧ(D:D) .

Пользуясь формулой для подсчёта коэффициента, считаем отношение содержимого G5 к G6 (можно и наоборот).

Получается, что в ячейке G8 вводим формулу: =G5/G6.

Такое действие выполняется для всех школ. Получаем список из всех возможных отношений. В нашем примере заполним им столбец I.



**Рис. 7.** Таблица с константными и расчётными величинами

Продолжим наши расчёты с использованием специализированных функций Microsoft Excel:

=ПРОЦЕНТИЛЬ.ВКЛ(**I:I**;**0,25**) – считает 25% процентиль по столбцу I

=КВАРТИЛЬ.ВКЛ(**I:I**;**1**) – считает первый квартиль.

Перечислим функции во всех ячейках (см. рисунок 8):

| **Ячейка** | **Формула** | **Пояснение** |
| --- | --- | --- |
| G10 | =ПРОЦЕНТИЛЬ.ВКЛ(I:I;0,25) | Вычисление процентиля |
| G11 | =ПРОЦЕНТИЛЬ.ВКЛ(I:I;0,75) |
| G12 | =КВАРТИЛЬ.ВКЛ(I:I;1) | Вычисление межквартильного размаха |
| G13 | =КВАРТИЛЬ.ВКЛ(I:I;3) |
| G14 | =G13-G12 |
| G16 | =G11+3\*G14 | Вычисление верхнего критического значения |
| G17 | =G10-3\*G14 | Вычисление нижнего критического значения |
| G19 | =СЧЁТЕСЛИ(I:I;">"&G16) | Подсчёт количества школ, выходящих за критические значения |
| G20 | =СЧЁТЕСЛИ(I:I;"<"&G17) |



**Рис. 8.** Таблица с константными и расчётными величинами

Иллюстрировать данный пример можно так, как показано на рисунке 9.



**Рис. 9.** Распределение школ по коэффициентам решаемости

Левая закрашенная сторона – это количество школ до 25 процентиля. Правая – количество школ, лежащих правее 75 процентиля. Мажквартильный размах – это разница между 75 и 25 процентилями.

В итоге данных вычислений получилось, что у 9 школ превышен верхний диапазон. То есть, произошло резкое (по сравнению с другими школами) повышение баллов ОГЭ в сравнении с баллами ЕГЭ (мы определяли отношения в G8, когда прописали отношение ОГЭ к ЕГЭ).

Школы, у которых превышен верхний диапазон, помечаются маркером «Резкое изменение».

Проведите опытную работу в Excel. Попробуйте посчитать процент соответствия оценок и какие результаты вызывают недоверие. Выделите их из общего списка. Каков процент этих результатов от общего количества? Сколько из них необходимо привести в соответствие, чтобы результаты перестали коррелировать с отрицательным коэффициентом.

## Результаты маркировки по результатам процедур

Результаты маркировки школ по разным процедурам сводятся в одной таблице, в которой указывается наличие или отсутствие маркера. Затем по каждой школе считается количество маркеров.

Допустим, что в муниципальном образовании 12 школ. Если у 4 из них есть хотя бы один маркер, то коэффициент объективности по муниципальному образованию (12-4)/12 = 8/12 = 77%.

## Как школа может использовать данные математические модели в своей работе

Эпизодические измерения коэффициентов не дают значимых результатов. Фактически, директор узнаёт уровень объективности процедуры постфактум, после того, как федеральный оператор или региональный посчитали доверительные «коридоры». Конечно, любой директор хотел бы иметь возможность предсказывания событий. Но, это не самое главное.

Нахождение в доверительном статистическом диапазоне – это основы диагностических карт Шухарта. Диагностические карты – это инструмент, которым управленец, контролёр качества, может замерить уровень (качество продукции) знаний учащихся на текущий момент, соответствие их оценок ожидаемым. Но самое важное – любое изменение относительно диагностической карты может быть отнесено к внешним или внутренним факторам.

 Подробнее смотрите в методических рекомендациях ИРО «Внутришкольный контроль: подходы и инструменты для реализации технологии всеобщего управления качеством (TQM)» (Алейникова С.В., Никитин С.В., ИРО, Екатеринбург, 2020), а также «Организация ВСОКО в ОО в условиях реализации ФГОС» (совместная разработка кафедры управления и РЦОИ).

Чтобы контролировать объективность оценивания, школа должна перейти от оценки объективности к мониторингу, для чего понадобятся несколько важных составляющих:

- наличие необходимых статистических данных для сравнения;

- наличие инструментальных средств расчёта показателей;

- наличие человека, способного анализировать изменения показателей, относить их к внешним или внутренним факторам;

- поддержка руководством стратегических целей повышения качества образования в ущерб сиюминутным выгодам от необъективных результатов.

Переход к долгосрочному планированию и внедрение управление качеством (и как часть его – управление объективностью) должен быть осознанным, кропотливым и поддерживаемым в первую очередь руководством организации.

1. Мы не будем приводить расчётов по 5-балльной шкале, однако, уважаемый читатель, Вы можете сделать это самостоятельно, просто изменив столбец подсчёта. [↑](#footnote-ref-1)