

ВОПРОСЫ МЕТОДОЛОГИИ НАУКИ И ОБРАЗОВАНИЯ

УДК 519.6

Т. Е. Апанасенко

ПРОБЛЕМА КОРРЕКТНОГО ПРИМЕНЕНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ В НАУЧНЫХ РАБОТАХ ПО БАЛЕТНОЙ ПЕДАГОГИКЕ

«При создании этой установки мы рассчитывали получить следующие характеристики...». (Такие характеристики получились случайно, когда нам удалось наконец заставить установку начать работать).

«Был выбран сплав висмута со свинцом, поскольку именно для него ожидаемый эффект должен был проявиться наиболее отчетливо». (Другого сплава у нас вообще не было).

«Для детального исследования мы выбрали три образца». (Результаты, полученные на остальных двадцати образцах, не лезли ни в какие ворота).

«Типичные результаты приведены на...». (Приведены лучшие результаты).

«Согласие теоретической кривой с экспериментом:

– Блестящее... (Разумное.)

– Хорошее... (Плохое.)

– Удовлетворительное... (Сомнительное.)

– Разумное... (Вымышленное.)

– Удовлетворительное, если принять во внимание приближения, сделанные при анализе...» (Согласие вообще отсутствует.)

– Наиболее надежные результаты были получены Джонсом...» (Это мой дипломник)».

Н. Козлов раскрывает тайный смысл общеупотребительных выражений при описании экспериментальной методики [1, с. 154].

При написании научных работ педагогической направленности, особенно связанных с физическим развитием человека, исследователь сталкивается с необходимостью проведения эксперимента. Одна из задач, которая сразу же перед исследователем встает, — это задача перевода характеристик объекта исследования в какое-то измеримое состояние, в котором они могут быть подвергнуты формализации и дальнейшей обработке статистическими методами.

Сущностью эксперимента является изменение объекта в определенном направлении вследствие целенаправленного воздействия на него. Собственно, целью теоретического (теоретической части) исследования является описание этого воздействия, составляющих его элементов, его качеств, структуры, а также обзор литературы, содержащей заключения по поводу описываемого явления воздействия.

Однако чтобы данное теоретическое исследование имело какой-то смысл, надо доказать, что воздействие действительно изменило объект, да еще в желаемом направлении.

Чтобы доказать, что объект изменился, необходимо измерять его характеристики на протяжении какого-то временного отрезка. Но этого еще недостаточно. Надо доказать, что объект изменился не сам по себе, а именно в силу воздействия, описанного в теоретической части. Для этого требуется сопоставить динамику изменения объекта с динамикой изменения аналогичного (контрольного) объекта, не претерпевающего того же самого воздействия.

Как при измерении характеристик объекта, так и при сопоставлении их с характеристиками контрольного объекта исследователя подстерегают всевозможные ловушки. Ловушки эти связаны с двумя аспектами: аспектом злоупотреблений и аспектом добросовестного заблуждения исследователя. Злоупотребление в процессе создания или применения экспериментальной методики предполагает создание видимости доказанного того, что, на самом деле, доказанным не является. Добросовестное же заблуждение исследователя может быть связано с незнанием им некоторых тонкостей использования статистических методов.

Мы рассмотрим различные статистические ловушки, которые позволят нашим читателям в зависимости от целей, которые они ставят перед собой, как использовать статистику для подтасовки результатов, так и вовремя заподозрить себя в некорректном использовании определенных статистических методов.

Подбор рассматриваемых в настоящей статье статистических ловушек основан на произвольном предпочтении автора и обобщении его опыта при знакомстве с работами начинающих исследователей АРБ, которые, специализируясь на балетной педагогике, постоянно сталкиваются с проблемой измеримости функциональности тела и, соответственно, часто обращаются к статистике. Как правило, не обремененные знанием основ общей теории статистики, они в своей наивности не подозревают о коварстве этой науки и множестве соблазнов, подстерегающих исследователя, сталкивающегося с обычными бытовыми ограничениями своих возможностей при проведении эксперимента. Справедливости ради надо сказать, что обобщение такого опыта при знакомстве с защищенными диссертациями спортивной тематики ничем не хуже помогло бы достигнуть задач, ставящихся в настоящей статье, однако автор будет ориентироваться преимущественно на целевую аудиторию. Само собой разумеется, что о рассматриваемых ловушках свято упоминается в учебной литературе по статистике, иллюстрации которых неосознанно создали начинающие ученые в области балетной педагогики [см. 2].

Первая ловушка подстерегает исследователя уже при его попытке перевести характеристики объекта в измеримое состояние: это неадекватность шкалы измерений.

Д. А. Новиков и В. В. Новачадов выделили две основные самые распространенные проблемы, связанные с измеримостью экспериментальных данных — проблему адекватности и проблему агрегированных оценок. Мы рассмотрим примеры возникновения проблемы адекватности при ситуации начального выбора шкалы для систематизации исходных данных и при переходе от одной шкалы к другой. Рассматривая же проблему агрегированных оценок, особое внимание уделим ее частному случаю — агрегированию векторных оценок.

Неадекватность шкалы измерений связана с различием шкал по мощности.

Определенные критерии выступают в качестве оснований оценки свойств состояния объекта. Поскольку присутствуют оценки, должна иметь место и некая шкала их соизмерения. Шкала — это система соответствия отношений между какими-либо характеристиками объектов отношениям между знаками этих характеристик в определенной упорядоченной совокупности. Как правило, эта совокупность представляет собой множество чисел.

Шкалы различаются мощностью, т. е. разрешающей способностью.

Первое место по мощности занимает шкала отношений, или абсолютная шкала. Она характеризуется следующими особенностями: 1) наличествует естественная нулевая точка отсчета; 2) отсутствует естественная единица измерения; 3) имеет место условная единица измерения, или эталон, которым является один из объектов совокупности, отношения которых переводятся в свойства множества чисел шкалы. Шкала отношений показывает, во сколько раз измеряемый объект отличается от эталона.

Естественным нулем в шкале отношений может выступать время начала воздействия на объект или число правильных ответов на вопрос.

Более слабой разрешающей способностью по сравнению с шкалой отношений обладает шкала интервалов. Помимо отсутствия естественной единицы измерения в шкале интервалов отсутствует еще и естественная нулевая точка отсчета. Примером шкалы интервалов является шкала температур по Цельсию. Она была установлена следующим образом: точке замерзания воды и точке ее кипения в множестве чисел, соответствующим по своим отношениям температурным характеристикам, было присвоено значение нуля градусов и ста градусов соответственно, интервал температур между ними был поделен на 100 равных частей. Преобразование показателей в данной шкале ограничено тем, что длины отношений интервалов должны при данном преобразовании сохраняться: температура в 30 °C не является в три раза большей, чем температура 10 °C. Отношения между данными показателями можно выразить, сопоставляя длины отношений данных интервалов с длинами отношений других интервалов: температура в 30 °C отличается от температуры в 10 °C в 2 раза сильнее, чем температура в 30 °C отличается от температуры в 20 °C.

Порядковая (номинальная) шкала, или шкала рангов, отличается наименьшей мощностью, или разрешающей способностью, объекты в множестве чисел она

упорядочивает нестрогое, приписывая им определенное количество баллов. Пример порядковой шкалы: шкала землетрясений Рихтера, система балльных оценок в педагогическом процессе.

В медико-биологических исследованиях шкалы порядка встречаются сплошь и рядом и подчас весьма искусно замаскированы. Например, для анализа свертывания крови используется тромботест: 0 — отсутствие свертываемости в течение времени теста (а через минуту?), 1- слабые нити, 2 — желеподобный сгусток, 3 — сгусток, легко деформируемый, 4- плотный, упругий, 5 — плотный, занимающий весь объем и т. п. Понятно, что интервалы между этими плохо отличимыми и очень субъективными позициями произвольны.

Для анализа результатов измерений часто приходится преобразовывать данные эксперимента, переведенные в шкалу, с получением производных показателей. Каждая шкала допускает только определенные преобразования, детерминированные ее разрешающей способностью.

Порядковая шкала разрешает любое монотонное преобразование, когда сохраняется упорядоченность объектов, но не обязательно сохраняются их разности, отношения и интервалы между ними (умножение на одинаковое для всех членов шкалы положительное число, сложение с постоянным для всех членов множеств числом, возведение в квадрат и т. д.).

Для шкалы интервалов сохраняется возможность преобразования, оставляющего неизменным разности показателей, т. е. линейного преобразования (умножения на положительное число, сложение с числом, постоянным для всех членов множества).

Шкала отношений допускает лишь одно единственное действие преобразования, называемое преобразованием подобия, — умножение на положительное число,

Проблема, которая встает перед исследователем в первую очередь — как можно корректно перейти от одной шкалы к другой. При этом часто исходные данные оказываются в одной шкале, а производные — в другой. Эта проблема в теории измерений и получила название проблемы адекватности.

Свойства взаимосвязи шкал детерминируют допустимые для каждой шкалы преобразования: данные шкалы с высокой разрешающей способностью могут быть переведены в данные шкалы с более низкой разрешающей способностью, обратная операция невозможна.

В литературе, посвященной методам измерения данных, большое внимание уделяется проблеме адекватности при переводе данных из одной шкалы в другую. Однако еще больший потенциал для добросовестного заблуждения исследователя и умышленного придания видимости доказательства желаемого положения открывается при выборе шкалы для анализа исходных данных. Очень часто, желая формализовать какую-либо информацию об исследуемом объекте, переводя данные этой информации в систему оценок, исследователь выбирает неадекватную шкалу.

Шкала со слишком слабой разрешающей способностью обладает свойством способствовать утрате важной информации об исследуемом объекте. Это свойство

сглаживания вариации исследуемого признака маломощной шкалы является не-оценимым при необходимости подтасовки данных. Примером такого сглаживания является пятибалльная система оценок знаний обучающихся в какой-либо образовательной системе: «четверка» может быть ближе к «пятерке», а может быть ближе к «двойке», однако носители данного уровня оценки знаний попадают в один интервал. Но особенно благодатная почва для утаивания каких-либо данных ждет исследователя при измерении в маломощной шкале не качественных, а количественных данных. Здесь любому произвольному утверждению возможно придать видимость строгого статистического доказательства.

Для шкалы с высокой разрешающей способностью верны обратные утверждения. Здесь возможна не потеря информации, а очень большой произвол при формализации информации и распределении ее по интервалам. Особенно это верно при применении шкалы отношений к качественным показателям. Пример, когда маломощная шкала (шкала порядка), но с более высокой разрешающей способностью, чем в предыдущем случае, применяется к качественным показателям — показателям оценки знаний обучающихся в некой образовательной системе — с целью реализации произвола, можно привести противоположный предыдущему. Таким примером является оценка знаний не по пяти-балльной, а по сто-балльной системе. Ранжировать отвечающих на экзамене можно как угодно, дифференцировав ответы на «отлично» с разницей от 1 до 10 баллов.

Следующая проблема — проблема агрегирования оценок.

Индивидуальные оценки измерения показателей каких-либо объектов необходимо преобразовывать в коллективные (групповые, производные, или агрегированные) для сглаживания индивидуальных колебаний, получения обозримого числа характеристик с целью последующего анализа. Некорректное преобразование индивидуальных оценок в агрегированные широко распространено.

Самый распространенный и естественный перевод индивидуальных оценок в агрегированные — вычисление среднего показателя. Совершенно корректная процедура по отношению к абсолютным величинам, которые ранжировала шкала отношений, она также весьма распространена и по отношению к порядковым шкалам, где ее использование не является корректным. Часто встречающийся пример этого — взвешивание школьных оценок. Нередко различные авторы, предлагающие модернизацию современной системы оценивания учебных достижений в средней школе, вполне отдают себе отчет о недопустимости вычисления среднего арифметического в порядковой шкале, как и любых математических операций (за исключением операций сравнения), поскольку в этой шкале они лишены смысла. Эти авторы предлагают, тем не менее, ввести веса для оценок в зависимости от сложности задания [3]. Т. е., зная, что вычисление среднего арифметического недопустимо, они предлагают применять среднее арифметическое взвешенное.

Тривиальные утверждения о недопустимости математических операций, кроме операций сравнения, по отношению к членам ранговой шкалы расходятся с общепринятой практикой вычисления среднего балла (для усреднения в этом случае следует использовать медиану).

Частным случаем проблемы агрегированных оценок является проблема комплексной оценки вектора показателей. Вектор показателей — совокупность оценки показателей какого-либо одного явления или процесса. Нередко возникает потребность выражения этой совокупности оценок одной величиной. Так, на спортивных соревнованиях победителя могут определить, суммируя очки, набранные на отдельных этапах соревнования, в отдельных матчах, а бывает, и в отдельных видах спорта.

Возможность однозначной оценки данной совокупности представляет собой проблему.

Комплексные оценки чрезвычайно распространены и в науке, и в повседневной жизни. В повседневной жизни они носят конвенциональный характер: все заинтересованные лица признают их, поскольку они установлены каким-либо нормативным актом, как в приведенном выше спортивном примере. Относительно математически строгого использования комплексных оценок в научном исследовании, на первый взгляд, все детерминировано свойствами шкал измерения. Комплексная оценка допустима по отношению к абсолютным однородным величинам, измеренным в шкале отношений. Однако на практике суммирование баллов применяется в научных исследованиях, и не всегда в силу добросовестного заблуждения или необходимости в софизмах. Проблема агрегирования векторных оценок является открытым вопросом в высшей математике.

На том уровне, на котором на сегодняшний день исследована данная проблема, агрегирование векторных оценок допускается при соблюдении ряда условий (ограничении исследования определенными априорными допущениями, основанными на произвольных предпочтениях) и при посредстве сложных последовательных дополнительных процедур. При разработке данных процедур проблема агрегирования векторных оценок получила название проблемы принятия решений при многих критериях.

Первая процедура состоит в выявлении множества эффективных альтернатив¹.

Далее возможны варианты применения следующей процедуры, причем рациональное обоснование преимуществ того или иного выбора априори невозможно.

Первый самый легкий и оставляющий возможности для самого большого произвола вариант — упорядочить критерии сравнения альтернатив по важности в зависимости от вкусовых предпочтений исследователя.

Второй вариант — агрегировать оценки по исходным критериям в комплексный критерий, при этом в зависимости от целей выбора альтернативы будут получаться разные результаты этого агрегирования, и итоговый комплексный критерий будет диктовать разный выбор альтернативы.

Прочие процедуры варьируют или комбинируют две описанные [5, 6, 7, 8, 9, 10].

¹ Эффективная альтернатива, или недоминируемая по Парето альтернатива — альтернатива, не допускающая других допустимых альтернатив, которые были бы «не хуже» по всем критериям, а по одному из критериев — «строго лучше».

Следующей распространенной ошибкой является вычисление линейного коэффициента корреляции без проверки его надежности, т. е. без уточнения, что данная корреляция может считаться существенной при данном количестве наблюдений для определенного числа степеней свободы.

Наиболее распространенными и банальными являются следующие нарушения корректности применения статистических методов: неадекватность шкалы измерений, неадекватность преобразований при переходе в ту или иную шкалу измерений, применение агрегированных оценок к ранговой шкале, применение комплексной оценки к разнородным показателям, применение комплексной оценки к не абсолютным показателям, применение линейного коэффициента корреляции без оценки его надежности.

Пример некорректного преобразования индивидуальных оценок в агрегированные можно найти в статье С. А. Русиновой «Некоторые аспекты психолого-педагогической диагностики профессиональных компетенций будущих артистов балета», опубликованной в № 19 «Вестника...», где вычисляются среднеарифметические показатели эмпатии и макиавеллизма [4, с. 154].

Показатели эмпатии и макиавеллизма представляют собой балльные оценки, соответственно, оцениваемые с определенным уровнем эмпатии или макиавеллизма. Они располагаются по ранговой шкале, к членам которой недопустимо применять математические операции, помимо операций монотонного преобразования. В частности, вычисление среднего арифметического в ранговой шкале лишено смысла.

Более того, в рассматриваемом случае вычисления среднего арифметического не требовала даже логика исследования. С. А. Русинова пишет, что уровни эмпатии и макиавеллизма определялись ею с целью «проведения сопоставительного анализа и корреляционного анализа в дальнейшем [4, с. 154]. «Было проведено исследование студентов Академии Русского балета имени А. Я. Вагановой и Института телевидения и дизайна на эмпатийность и макиавеллизм. Для чистоты эксперимента опрашиваемые были разделены на группы по возрасту, специальности» [4, с. 154].

Однако для проведения сопоставительного анализа вполне достаточно было распределить по интервалам индивидуальные показатели уровня эмпатии и макиавеллизма и рассмотреть, сколько в интервалы попало студентов АРБ, сколько — Института телевидения и дизайна, аналогичную операцию провести относительно разных возрастных групп и специальностей, как это и было сделано, на сей раз корректно, ниже [4, с. 155–156].

Если же автору хотелось получить все же какой-то агрегированный показатель уровня эмпатии и макиавеллизма студентов определенного вуза, определенного факультета и определенной возрастной группы, следовало использовать медиану.

В третьем и четвертом столбце таблицы 1 приведены показатели среднего арифметического уровней эмпатии и макиавеллизма, которые не имеют смысла.

Таблица среднеарифметических показателей эмпатии и макиавеллизма²

		Эмпатия	Мак-шкала
Академия балета имени Вагановой (исполнительский факультет)	Молодые люди 17–18 лет (7 опр.)	60,4	73,1
	Девушки 17–18 лет (12 опр.)	73,75	68,6
	Молодые люди 19–20 лет (11 опр.)	71,7	74,9
	Девушки 19–20 лет (12 опр.)	73,4	73,5
ИТиД	Молодые люди 17–18 лет (7 опр.)	66,1	77
	Девушки 17–18 лет (33 опр.)	74,8	76,5
	Молодые люди 19–20 лет (12 опр.)	71	65,3
	Девушки 19–20 лет (19 опр.)	66,4	68,8
Академия балета имени Вагановой (педаг. Фак-т 1 курс)	Молодые люди (5 опр.)	74	78
	Девушки (22 опр.)	76,6	73,8

При корректном агрегировании данных таблица должна была бы выглядеть следующим образом:

Показатели эмпатии и макиавеллизма, значения медианы

		Показатели эмпатии, значение медианы	Показатели макиавеллизма, значения медианы
Академия балета имени Вагановой (исполнительский факультет)	Молодые люди 17–18 лет (7 опр.)		
	Девушки 17–18 лет (12 опр.)		
	Молодые люди 19–20 лет (11 опр.)		
	Девушки 19–20 лет (12 опр.)		
ИТиД	Молодые люди 17–18 лет (7 опр.)		
	Девушки 17–18 лет (33 опр.)		
	Молодые люди 19–20 лет (12 опр.)		
	Девушки 19–20 лет (19 опр.)		
Академия балета имени Вагановой (педаг. Фак-т 1 курс)	Молодые люди (5 опр.)		
	Девушки (22 опр.)		

Что касается корреляционного анализа, нам неизвестно, связь каких именно показателей собирался исследовать «в дальнейшем» автор статьи «Некоторые аспекты психолого-педагогической диагностики профессиональных компетенций

² Таблица воспроизводится по [4, с. 154]

будущих артистов балета». Если уровень эмпатии, либо макиавеллизма выступал в качестве результативного признака, то это требовало бы вычисления его средней величины, но в таком случае весь корреляционный анализ не имел бы смысла, поскольку средней величины в шкале рангов быть не может. Если же предполагалось, что уровень эмпатии, либо макиавеллизма сыграет роль признака-фактора, тогда потребовались бы никак не агрегированные, а индивидуальные данные с определениями показателей их вариации.

Неадекватность шкалы измерений, агрегирование оценок и векторных оценок, отсутствие проверки показателей тесноты связи на надежность — наиболее частые случаи некорректного применения статистических методов, к чему необходимо быть внимательным при проведении педагогического эксперимента.

ЛИТЕРАТУРА

1. Козлов Н. Философские сказки для обдумывания житъя, или Веселая книга о свободе и нравственности. 3-е изд., перераб. и доп. М.: Новая школа, 1997. — 428 с.
2. Новиков Д. А., Новочадов В. В. Статистические методы в медико-биологическом эксперименте (типовые случаи). Волгоград: Издательство ВолГМУ, 2005. — 84 с.
3. Воронов В. В. Современная система оценивания в средней школе. // Ярославский педагогический вестник. 2010. № 2. С. 71–74.
4. Русинова С. А. Некоторые аспекты психолого-педагогической диагностики профессиональных компетенций будущих артистов балета. // Вестник Академии Русского балета им. А. Я. Вагановой. 2008. № 1 (19). С. 144–159.
5. Новиков А. М., Новиков Д. А. Методология: основания методологии, методы научного исследования, методология практики, деятельности, введение в методологию художественной деятельности, методология учебной деятельности, введение в методологию игровой деятельности учебно-методическое пособие. М.: Синтег, 2007. 663 с.
6. Новиков Д. А. Теория управления организационными системами: учебно-методическое пособие. М.: Физматлит, 2007. 583 с.
7. Ногин В. Д. Принятие решений в многокритериальной сфере: количественный подход. М.: Физматлит, 2002. 175 с.
8. Орлов А. И. Устойчивость в социально-экономических моделях. М.: Наука, 1979. 296 с.
9. Подиновский В. В., Ногин В. Д. Парето-оптимальные решения многокритериальных задач. М.: Физматлит, 2007. 255 с.
10. Рыков А. С. Модели и методы системного анализа: принятие решений и оптимизация. М.: МИСИС, 2009. 607 с.