



управляем
предприятием



УПРАВЛЕНЧЕСКАЯ ТЕОРИЯ ИЗМЕРЕНИЙ

Часть 5

Проблемы качества измерений



Андрей Мицкевич

К. э. н., доцент Высшей школы финансов и менеджмента РАНХиГС и Научно-исследовательского университета ВШЭ, руководитель консультационного бюро Института экономических стратегий, преподаватель ведущих бизнес-школ Москвы. Консультант, автор четырех книг и более чем 80-и статей,

вышедших в последние годы, по контроллингу, мотивации, управленческому учету и сбалансированной системе показателей.

Незнание теории не освобождает от ответственности

Что такое качество измерений? Это способность передать эмпирическую систему отношений числами. Видим (пусть субъективно), Иванов работает лучше Петрова, а это значит, что и зарплата у него должна быть больше. Все понятно, но как же сложно... Например, в компаниях используется маркетинговая информация. Качество, то есть надежность и достоверность маркетинговой информации, определить весьма непросто. Еще большее значение качество играет при измерении показателей компании. Ведь плохое качество измерения показателя дискредитирует всю систему показателей компании. В этой части статьи мы расскажем об интересном примере вычисления качественного интегрального показателя на основе многих частных измерений, – о системе оценок в фигурном катании.



Качество измерений — это научная область, смежная с теорией измерений, теорией информации и математической статистикой объектов нечисловой природы. Теория предъявляет следующие общие требования к качеству измерений:

- информативность;
- достоверность;
- надежность.

Для практического применения эти общие требования дополняются требованиями конкретной предметной области. Ниже мы подробно рассмотрим эти общие требования к качеству измерений. Но сначала покажу, к каким проблемам может привести плохая система оценок и вычисления интегрального показателя.

Сага о системах оценок в фигурном катании

Приведу пример из фигурного катания, который наиболее ярко прояснит одну из главных проблем измерений. Почти каждому телезрителю примерно понятно, что именно оценивается в фигурном катании. Это дает возможность сосредоточиться на способах оценки и ошибках каждого из них.

Качественная система оценки интегрального показателя на основе многих частных измерений сложна, требует математических вычислений, и в далеко не просто разобраться, почему получились те или иные итоговые результаты.

Система оценок в фигурном катании очень наглядна: это вычисление качественного интегрального показателя на основе многих частных измерений, то есть именно то, что должна делать любая система показателей деятельности компании.

Шестибалльная система оценок в фигурном катании¹ оценивала мастерство фигуристов в исполнении программ с шагом 0,1 балла: 0,0 — «не выполнено», 6,0 — «безукоризненно». Причем техническая составляющая и артистичность оценивались отдельно. Отмечу, что оценка 6,0 за техническое исполнение ставилась крайне редко, а за артистичность — чаще. Первоначально, до 1980-х годов, две оценки (за произвольную и обязательную программы) просто складывались.

Долгое время система работала, но со сбоями — случались «казусы адекватности». Один такой пример приобрел всемирную известность. В начале 1980-х годов несколько раз подряд чемпионкой мира по фигурному катанию становилась австрийская фигуристка Беатрис Шуба. Она не была идеалом фигуристки. Чемпионкой она становилась именно благодаря сбою в системе измерения. В то время суммарная оценка складывалась из оценок за обязательную и произвольную программы. Обе оценки выставлялись по шкале из 6 баллов. Но дело в том, что критерии оценок были разными, оценки были выверены специалистами, но разными для разных упражнений. Лучшие фигуристки обычно получали за обязательную программу 5,1 — 5,2, а Шуба — 5,8. После этого ей достаточно было просто не упасть в произвольной программе.

«Формально правильно, а по существу издевательство», — как говорил Ленин. В идеале система оценок должна «сама расставлять по местам» всех фигуристов, даже субъек-

¹ Была введена в 1901 году Международным союзом конькобежцев и существовала с некоторыми изменениями до 2005 года, когда была заменена новой судейской системой.

тивные мнения не должны на нее слишком влиять. Идеала, конечно, достигнуть нельзя, но стремиться к нему надо. Если все или почти все «видят чемпиона», а система оценки позволяет регулярно ставить его ниже первого места, значит, что-то не так.

Результат выступления — сумма мест с весами. Когда сбой стал очевиден, пришлось обратиться к математикам, которые предложили систему оценки в соответствии с теорией измерений. Они спросили у специалистов по фигурному катанию: «Что вам надо от системы измерений?» Получили ответ: «Первого надо выбирать правильно». То есть фактически лица, принимающие решения, просили сделать акцент на надежности измерений. Тогда математики перевели баллы в места, так как порядковая шкала более надежная, чем балльная (об этом мы расскажем в следующей части статьи). Сумма мест с учетом разного веса обязательной и произвольной программ давала итоговую оценку. И Беатрис Шуба сразу окончила свою спортивную карьеру.

Субъективность системы оценок не может быть преодолена на 100%. Субъективность нельзя полностью исключить, но ее можно ограничить и постараться использовать во благо.

Недостатки системы «сумма мест с весами». Эта система оценки работала, однако, на Олимпийских играх 2002 года случился скандал. Борьбу за победу в соревнованиях спортивных пар вели две пары: российская и канадская². Они принадлежали к очень отличающимся друг от друга российской (советской) и канадской школам. Уровень мастерства и художественный уровень программ обеих пар можно оценивать по-разному. Судьи оказались в сложном положении, и, как часто бывало в истории фигурного катания, их мнения разошлись. В результате большинство судей

— пять из девяти — отдали первое место российской паре, которой и вручили золотые медали. Канадцы получили «серебро».

В тот же день после окончания соревнований член жюри, француженка Мари-Рейн Ле Гунь, встретила в гостинице буквально напавшую на нее председателя технического комитета Международного союза конькобежцев Сэлли-Энн Степлфорд. После этого с Ле Гунь случился нервный срыв: она призналась, что на неё оказал давление Д. Галагье, возглавлявший французскую Федерацию фигурного катания. По её словам, он просил Ле Гунь поставить российскую пару на первое место якобы «в обмен» на то, что российский член жюри поставит высшие баллы французской паре в спортивных танцах. На следующий день на судейском заседании после соревнований Ле Гунь повторила эти слова. Однако в письменном заявлении отказалась от этих утверждений, объяснив их давлением Степлфорд, а также подтвердила, что считает победу российской пары заслуженной.

Эта почти детективная история имела продолжение, но нас интересует другое.

- Что должен была сделать Международный союз конькобежцев, кого объявить победителем?
- Как надо изменить систему оценок, чтобы избежать подобных эксцессов в дальнейшем?

Сначала давайте поймем, что же делать в подобной ситуации? По сути, у руководителей Международного союза конькобежцев было 3 варианта решения:

1. оставить все как есть: «золото» — России, «серебро» — Канаде. Минус решения: игнорировать судейский скандал, то есть покрыть «взяточников». Плюс: сохранить тради-

² Елена Бережная — Антон Сихарулидзе (Россия) и Жами Сале — Давид Пеллетье (Канада).



цию: решения судей окончательны и обжалованию не подлежат. Но к системе судейства уже накопилось столько претензий, что плюс превращался в минус;

2. вручить «золото» — Канаде, а «серебро» — России. У этого решения одни минусы: с какой стати? Только из-за странного поведения одного члена жюри?
3. вручить два «золота» — и Канаде, и России.

Международный союз конькобежцев принял беспрецедентное решение: впервые за всю историю фигурного катания будут вручены две золотые медали. Российской паре сохранили золотые медали, поскольку её вины в Международном союзе конькобежцев и МОК не усмотрели, но «золото» вручили и канадцам. Судейский голос Ле Гунь был аннулирован. На мой взгляд, это решение «лучшее из трех зол».

Новая система оценок. Теперь давайте подумаем, как надо изменить систему оценок, чтобы избежать подобных эксцессов в дальнейшем? После скандала на Олимпийских играх 2002 года конгресс Международного союза конькобежцев принял решение заменить действовавшую многие годы систему из двух оценок на систему, в которой каждому из элементов ставится оценка и приписывается определенный вес, а итоговый результат выступления получается как сумма оценок за отдельные элементы в соответствующими весам, набранных фигуристами³. Суть новой системы измерения в том, что вместо выставления двух «общих» оценок (за сложность программы и за ее представление) судьи будут оценивать все элементы в программе по отдельности. Хотя разделение общей оценки на 2 категории, условно называемые «техника» и «артистичность», сохраняется.

«Техника». Каждому элементу фигурного катания приписывается определенная ценность (весовой коэффициент, или вес), которая является постоянной (по крайней мере, в течении одного сезона). Веса приведены в «Таблице ценности» (таблица 4). Во время соревнований трое судей, называемых техническим контролером, техническим специалистом и помощником технического специалиста, занимаются тем, что опознают исполняемые фигуристами элементы и вводят их в компьютер, в то время как судейская бригада оценивает исключительно качество исполнения каждого отдельно взятого элемента. Каждый из судей в бригаде оценивает качество исполнения элемента по шкале: +++, ++, +, 0, -, —. Каждому шагу оценки в таблице сложностей соответствует определенное числовое значение. Общая оценка каждого элемента получается путем сложения базовой сложности элемента (в таблице указана в графе 0) с оценкой качества его исполнения.

Таблица 4. Фрагмент таблицы ценности для прыжков.

Название	—	-	-	0	+	++	+++
Тулуп (3 оборота)	-3.0	-2.0	-1.0	4.0	1.0	2.0	3.0
Сальхов (3 оборота)	-3.0	-2.0	-1.0	4.5	1.0	2.0	3.0
...							
Флип (3 оборота)	-3.0	-2.0	-1.0	5.5	1.0	2.0	3.0
...							

³ Самая настоящая аддитивная функция полезности, сказал бы математик.

Таблицы ценности элементов — основополагающие документы современной системы судейства. В них указывается, сколько баллов спортсмен получит за исполнение каждого конкретного элемента, а также сколько баллов с него снимут или дадут дополнительно, если он исполнит этот элемент с ошибкой или, наоборот, очень хорошо. Например, если флип (3 оборота) выполнен с ошибками и качество его исполнения оценено как –, то оценкой элемента будет $5,5 - 2 = 3,5$. Если же он выполнен с хорошим заходом, отличным и нестандартным выходом, то судья может оценить качество его исполнения ++ и оценка будет совсем другой: $5,5 + 2 = 7,5$.

Через такой судейский фильтр пропускаются все элементы, включенные в программу. Сумма баллов за отдельные элементы в итоге превращается в общую оценку выступления за технику. Время от времени таблицы ценности пересматриваются (практически каждый год), иногда в них появляются новые разделы (например, подкрутки сначала различались только по числу оборотов; позднее для них появились уровни сложности, а потом еще и разделение на типы по варианту захода)⁴, но в целом это, пожалуй, самая консервативная часть системы.

«Артистичность». В новой системе это сумма пяти оценок за отдельные компоненты программы. Помимо оценок за технику, судейская бригада выставляет также пять оценок за компоненты программы, каждый из которых оценивается по десятибалльной шкале с шагом 0,25:

- базовый уровень катания;
- связующие элементы;
- хореография;
- прокат/выполнение;
- выразительность.

Приблизительное значение оценок таково: <1 — очень плохо, 1 — плохо, 2 — слабо; 3 — ровно, 4 — средне, 5 — выше среднего, 6 — хорошо, 7 — очень хорошо, 8 — прекрасно, 9—10 — великолепно. По окончании выступления все оценки складываются (иногда после

⁴ Обратите внимание, что это ежегодно изменяемая, улучшаемая система оценок. Какой и должна быть система показателей компании.

А судьи кто?

Этот популярный вопрос часто возникает в компаниях при построении системы показателей. И опыт Международного союза конькобежцев, на мой взгляд, полезен и консультантам, и заказчикам. В фигурном катании систему оценок разрабатывал административно-технический персонал. Первоначально мнения спортсменов и тренеров при этом абсолютно не учитывались. Да и потом: новые правила

сложны и требуют математических вычислений, порой сами фигуристы и тренеры не понимают, что и почему. А технические специалисты, как выяснилось, не имеют права на разъяснения. Но позвольте спросить, к чему такая закрытость и секретность? Судя по всему, при новой системе измерений проблемы субъективности оценок сохраняются, хоть и не в таком масштабе, как раньше.



умножения на определенные веса), и тот спортсмен, у кого итоговая сумма будет больше, объявляется победителем.

Общие оценки выступления. Общая оценка качества исполнения элемента получается путем вычисления усеченного среднего судейской бригады. Для этого отбрасывается равное количество самых высоких и самых низких оценок, а из оставшихся вычисляется арифметическое среднее (округляется до двух знаков после запятой).

Идеальной системы оценки качественных явлений нет и быть не может, но можно добиваться максимальной объективности.

Финальная оценка за каждый из элементов представления программы получается путем умножения усеченного среднего бригады на специальные коэффициенты и вычетов за разные нарушения. Например, в произвольной программе одиночников оценка для всех прыжков, исполненных во второй половине программы, умножается на 1.1, чтобы поощрить равномерное распределение наиболее сложных элементов. В случае нарушения общих правил применяются определенные штрафы.

Результат нескольких выступлений вычисляется путем сложения итоговых результатов короткой и произвольной программ. При совпадении общей суммы преимущество отдается тому спортсмену, чей результат в последней части соревнований был выше.

Выводы для системы показателей деятельности компании

Рассмотренный выше пример из фигурного катания весьма показателен и позволяет сделать некоторые важные выводы.

1. **Качественная система оценки интегрального показателя на основе многих частных измерений сложна**, требует математических вычислений, и в ней далеко не просто разобраться, почему получились те или иные итоговые результаты.
2. **Субъективность нельзя полностью исключить** (см. врезку «А судьи кто?»), но ее можно ограничить и даже постараться использовать во благо.
3. **Идеальной системы оценки качественных явлений нет и быть не может, но можно добиваться максимальной объективности.**

Приведенный пример также позволяет вывести правило:

Складывать значения разных показателей, даже если они выражены в одинаковых шкалах, не рекомендуется.

Для этого их надо сначала перевести в шкалу с меньшей информативностью, но с большей надежностью и достоверностью (об этом — дальше). А затем уже можно складывать, в том числе и с весами для принятия решений. Нарушение этого правила способно привести к абсурдным результатам, что мы зачастую и видим.

Пример системы оценок в фигурном катании наглядно показывает проблему адекватности (устойчивости) и надежности шкалы оценок. В следующей части статьи рассмотрим общие требования к качеству измерений: достоверности и надежности.