



управляем
предприятием



УПРАВЛЕНЧЕСКАЯ ТЕОРИЯ ИЗМЕРЕНИЙ

Часть 6

Надежность и достоверность 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 не просто. Еще большее значение качество играет при измерении показателей компании. Ведь плохое качество измерения показателя дискредитирует всю систему показателей компании. В этой части статьи мы обсудим, от чего зависит надежность и достоверность измерений.



Надежность измерений

Надежность измерений — это широкое понятие. Описанные в теории измерений виды шкал не дают полного представления о свойствах полученных по ним оценок. Необходимо выявить присущие этим оценкам ошибки. Это и называется оценкой надежности измерения. Данная проблема решается путем выявления трех свойств измерения:

- **правильности;**
- **устойчивости;**
- **обоснованности.**

Правильность измерения — это приемлемость используемого способа измерения (шкалы или системы шкал). Оно связано с возможностью учета различного рода систематических ошибок, возникающих в результате измерения. Систематические ошибки измерения имеют некоторую стабильную природу возникновения: они либо постоянные, либо меняются по определенному закону. Возможно, что последующие этапы оценки надежности измерения окажутся излишними, если в начале выяснится полная неспособность инструмента измерения дифференцировать изучаемую характеристику на требуемом уровне. Самый простой возможный вариант — исследуемую характеристику объекта просто нельзя дифференцировать с нужной точностью. Систематические ошибки в результате измерения также будут, если окажется, что систематически не используется какая-то градация шкалы или вопроса. Например, исследования показывают, что при непосредственном присваивании коэффициентов люди пользуются ограниченным набором чисел [2].

Как добиться правильности измерений? Прежде всего, нужно ликвидировать или уменьшить такого рода недостатки шкалы и только затем использовать ее в прикладных целях. Например, изменение шкалы оценок, используемой в ВШЭ. Первоначально в ВШЭ использовалась 10-балльная шкала без пояснений к баллам. В результате анализа правильности измерений выяснилось, что есть градации шкалы, которые практически не применяются (оценки 1, 2 и 10). Тогда шкала оценок была заменена на новую 5-балльную — менее информативную, но более надежную (таблица 5).

Таблица 5. Шкала соответствия оценок ВШЭ.

Французская 20-балльная шкала	10-балльная шкала оценок ВШЭ	Новая 5-балльная шкала оценок ВШЭ
18 и выше	10	«Отлично» – 5
17	9	
15–16	8	
14	7	«Хорошо» – 4
12–13	6	
11	5	«Удовлетворительно» – 3
9–10	4	
Менее 9	1–3	«Неудовлетворительно»

Устойчивость измерения характеризует степень совпадения результатов при повторных применениях измерительной процедуры. Она описывается величиной случайной ошибки (чаще всего используется средняя квадратическая ошибка). Устойчивость характеризует постоянство подхода респондента к ответам на одинаковые или подобные вопросы. День на день не приходится, один и тот же человек в разных условиях может дать разные ответы на одинаковый вопрос. Эту неустойчивость оценок при использовании нефинансовых качественных показателей деятельности компании исключать или игнорировать никак нельзя.

В отличие от правильности и устойчивости, которые в идеале могут быть измерены достаточно строго, критерии обоснованности определяются либо экспертно, на основе логических рассуждений, либо на основе косвенных данных.

Как оценить устойчивость измерений? Существует несколько методов оценки устойчивости измерений:

- повторное тестирование;
- включение в анкету эквивалентных вопросов, то есть вопросов по той же проблеме, но сформулированных по-другому, и разделение выборки на две части (сравнение ответов на вопросы двух групп респондентов).

Устойчивость может быть измерена статистически достаточно строго и выражена в форме числового показателя. Правда, в системе показателей компании это практически не применяется, и устойчивость оценок остается «вещью в себе». Например, на Талнахской горно-обогатительной фабрике начальник может увеличить или уменьшить на 1 балл

оценку своему подчиненному при 4-балльной системе измерений. Попробуйте установить устойчивость такой оценки формально. Это не получится.

Тем не менее, этот прием (субъективное увеличение или уменьшение оценки на 1 балл) я оцениваю положительно. Дело в том, что абсолютно объективных оценок не бывает. Всегда надо как-то учесть субъективное мнение. Особенно если добавление (удаление) целого балла прописано и используется в системе как «дамоклов меч», то есть скорее как угроза, исключительная мера, а не система оценок «за голубые глазки» (именно так он используется, например, на Талнахской горно-обогатительной фабрике).

Обоснованность измерения — это доказательство соответствия между тем, что измерено, и тем, что должно было быть измерено. Это наиболее сложный вопрос надежности измерения. Трудно доказать, что измерено вполне определенное заданное свойство объекта, а не некоторое другое, более или менее на него похожее.

В отличие от правильности и устойчивости, которые в идеале могут быть измерены достаточно строго и иногда даже могут быть выражены в форме числового показателя, критерии обоснованности определяются либо экспертно, на основе логических рассуждений, либо на основе косвенных данных. В социологии обычно применяется сравнение данных одной методики с данными других методик или исследований, однако в системе показателей компании это практически исключено. Остается экспертное профессиональное мнение. Однако наши менеджеры уверены, что их шкалы оценки той или иной характеристики верные. Уверяю вас, это очень далеко от истины! Наши руководители сами разрабатывают даже систему мотивации, тогда как на Западе принято отдавать это



на аутсорсинг фирмам-профессионалам. Что уж говорить о каких-то показателях. Неудивительно, что нефинансовые показатели используются пока редко.

Три составляющие измерения. Помимо перечисленного выше, при установлении надежности следует иметь в виду, что в процессе измерения участвуют три составляющие:

- **объект измерения;**
- **измеряющие средства, с помощью которых производится отображение свойств объекта на числовую систему;**
- **производящий измерение субъект (интервьюер, оценщик, разработчик ССП и т. д.).**

Достоверность системы показателей деятельности компании весьма важна. Она характеризует точность измерений по отношению к тому, что существует в реальности.

Предпосылки надежного измерения кроются в каждой отдельной составляющей. Может быть, что способ получения оценки не в состоянии дать максимально точных значений измеряемого свойства. Например, у респондента существует развернутая иерархия ценностей, а для получения информации используется шкала с вариациями ответов только «очень важно» и «совсем не важно». Как правило, если используется такая шкала, то все характеристики будут «очень важными», хотя реально у респондента имеется большее число уровней значимости.

Возможно, что субъект, производящий измерение, допускает грубые ошибки; нечетко составлены инструкции к анкете; интервьюер каждый раз по-разному формулирует один и тот же вопрос, используя различную терминологию. На практике это критически важно. Именно поэтому построение системы показателей деятельности сопровождается обучением, разработкой тезауруса, распространением информации, разбором опыта подобных измерений и оценок, приемами социологии и психологии.

Достоверность измерений

Достоверность характеризует другие аспекты измерений, чем надежность. Измерение может быть надежным, но недостоверным. Достоверность характеризует точность измерений по отношению к тому, что существует в реальности. Например, респонденту задали вопрос о целевой месячной выручке. Он ее оценивает в 250 000 долларов. Однако, не желая называть интервьюеру истинную цифру, респондент указал «более 100 000 долларов». При повторном тестировании он снова назвал данную цифру, демонстрируя высокий уровень надежности измерений. То есть измерение надежно, но недостоверно. Но ложь не является единственной причиной низкого уровня достоверности измерений, ее причиной может быть плохое знание респондентом действительности и т. п.

Достоверность — это доверие измерению и оценке, степень уверенности, что в измерении нет ошибки.

Достоверность системы показателей деятельности компании весьма важна. Например, какой вес присвоить одному из двух показателей, сделав его более важным? Точно посчитать вес невозможно, выбирать его придется субъективно из множества чисел

от 0,5 до 1. Но исследования показывают, что при присваивании коэффициентов люди пользуются ограниченным набором чисел, причем используют этот набор крайне неравномерно [2].

Отметим некоторые наиболее общие факторы, влияющие на достоверность измерений и не связанные с предметной областью.

1. **Достоверность измерений зависит от шкалы.** Наиболее достоверные измерения выполняются в шкале наименований, а наименее достоверные — в абсолютной шкале. Закономерность здесь такова: чем шире класс допустимых преобразований, тем легче оценивать и измерять, то есть давать оценки.

В социологии достоверность определяют как степень, в которой различия между оценками, полученными по шкале, отражают истинные различия между измеряемыми характеристиками объектов, а не обусловлены случайными и систематическими ошибками.

2. **Количество экспериментов или экспертов влияет на достоверность измерений.** Этот эффект наглядно показан во врезке «Четвертый объект лишний? Или третий эксперт?».

Как повысить достоверность измерений? Повышение достоверности измерений может быть связано с разными факторами. Отметим два наиболее общих способа.

1. **Укрупнение градаций шкалы** — это такая группировка значений измерения, которая сокращает число значимых оценок, делает измерение более грубым и одновременно более достоверным. Например, такие значения шкалы, как «полностью

согласен» и «скорее согласен» объединяются в одно значение «согласен». Достоверность измерений растет, так как если вы колеблетесь между «полностью согласен» и «скорее согласен», то уж точно в новой шкале с более грубыми градациями ответите «согласен». Но при этом страдает информативность, она падает, поскольку в шкале с более грубыми градациями исчезают оттенки.

2. **Переход к измерениям в шкалах меньшей информативности.** Например, количественную шкалу заменить на порядковую. В этом случае достоверность вырастет за счет понижения информативности измерений.

Отмечу, что в социологии и маркетинге достоверность измерений существенно более сложное понятие, чем то, что описано выше. В социологии достоверность определяют как степень, в которой различия между оценками, полученными по шкале, отражают истинные различия между измеряемыми характеристиками объектов, а не обусловлены случайными и систематическими ошибками.

Полная достоверность — это отсутствие ошибок измерения.

Исследователи разделяют содержательную, критериальную и конструктивную достоверность.

Содержательная достоверность (content validity)¹ — субъективная систематическая оценка того, насколько хорошо содержание шкалы соответствует поставленной цели измерения. Например, шкала, разработанная для измерения имиджа магазина, считается неадекватной, если будет опущена одна из существенных характеристик объекта (например, качество и разнообразие (широта и глубина ассортимента) предлагаемой

¹ Иногда называемая лицевой достоверностью (face validity).



Четвертый объект лишний? Или третий эксперт?

Иногда при измерении по порядковым шкалам привлечение еще одного объекта экспертизы меняет отношения порядка при обработке с помощью суммарного балла порядковых измерений. Предположим, трем экспертам дали задание ранжировать 4 объекта. Результаты приведены в таблице:

Объекты измерения	A	B	C	D
Результаты измерений трех экспертов	1	2	3	4
	4	2	1	3
	2	3	4	1
Суммарная оценка	7	7	8	8

Вывод: объекты A и B равноценны и лучше, чем C и D, которые тоже равноценны. Теперь уберем объект D и попросим экспертов снова упорядочить между собой объекты A, B и C. Мы вправе ожидать, что предпочтения экспертов, то есть их оценки относительного порядка объектов, не изменятся. Тогда получим следующие результаты:

Объекты измерения	A	B	C	Нет данных
Результаты измерений трех экспертов	1	2	3	Нет данных
	3	2	1	Нет данных
	1	2	3	Нет данных
Суммарная оценка	5	6	7	Нет данных

Вывод: объект A лучше объекта B, который, в свою очередь, лучше, чем C. Согласитесь, что выводы различаются. И все из-за наличия «лишнего» объекта измерений. Значит, достоверность зависит от количества объектов измерения.

Вывод по результатам измерений в порядковой шкале меняется и в зависимости от количества экспертов. Уберем одного эксперта, а предпочтения оставшихся будем считать неизменными. В результате получим:

Объекты измерения	A	B	C	D
Результаты измерений трех экспертов	1	2	3	4
	4	2	1	3
	Нет данных	Нет данных	Нет данных	Нет данных
Суммарная оценка	5	4	4	7

Вывод опять изменился: объекты B и C равно хороши и лучше объекта A, который, в свою очередь, лучше, чем объект D. Значит, достоверность измерения зависит от количества экспертов. И не факт, что чем больше экспертов, тем выше достоверность выводов.

продукции). Оценка содержательной достоверности очевидно субъективна, но в любой системе показателей деятельности субъективизм неизбежен. Однако важно понимать: баллы выбраны и описаны адекватно измеряемому объекту или нет. Хотя выбор балльной шкалы — дело достаточно субъективное.

Критериальная достоверность (criterion validity) отражает, насколько используемая шкала соответствует цели измерений. В зависимости от временных рамок измерения критериальная достоверность может принимать две формы — текущую или прогнозную достоверность. **Текущая критериальная достоверность** определяется, когда данные, полученные на основе использования шкалы, и сведения о целях измерения (критериальных переменных) собираются одновременно. Для оценки **прогнозной критериальной достоверности** исследователь собирает измерения по шкале в один период времени, а по целям (критериальным переменным) — в другой. Например, отношение к различным маркам продуктов может использоваться для прогноза будущих объемов их продаж. У определенной выборки покупателей получают информацию об их отношении к продукту, а затем будущие закупки проверяются по данным исследования. Прогнозные и фактические данные по покупкам сравниваются для оценки прогнозной достоверности шкалы отношения к различным маркам продуктов.

При разработке и эксплуатации системы показателей деятельности компании критериальная достоверность вполне применима. Всем важно знать, как используются их оценки в системе мотивации и при принятии каких решений. Поэтому не лишне иметь представление и о проблеме критериальной достоверности.

Конструктивная достоверность (construct validity) связана с ответами на теоретические вопросы: почему должна использоваться эта шкала и какие выводы можно сделать из лежащей в ее основе теории. Конструктивную достоверность понять сложнее, останавливаться на ней здесь мы не будем.

В следующей части статьи речь пойдет об информативности измерений. Мы проанализируем приведенную в предыдущей части систему оценок в фигурном катании, а также обсудим использование теории измерений в системе показателей компании.