

УДК 004.62:519.2

М.С. МАЗОРЧУК, В.С. ДОБРЯК, К.А. БАЗИЛЕВИЧ, М.Ю. ХАЗАЙ

*Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина***ОЦЕНКА ЭФФЕКТИВНОСТИ ОДНОМЕРНЫХ ШКАЛ ПРИ АНАЛИЗЕ ДАННЫХ В СОЦИОЛОГИЧЕСКИХ ИССЛЕДОВАНИЯХ**

В работе рассмотрена оценка надежности измерения при анализе эффективности одномерных шкал. Раскрыты основные понятия данной области изучения. Описаны виды шкал, используемых в социологических измерениях, и их применение. Рассмотрены методы проверки первичного измерения на надежность: проверка обоснованности, устойчивости, точности и правильности измерения. Некоторые из описанных методов применены для решения реальной задачи. Приведены расчеты некоторых критериев надежности одномерной шкалы, показаны промежуточные вычисления. Проанализированы полученные результаты.

Ключевые слова: анализ качественных данных, измерение, эффективность, шкала, надежность, тестирование.

Введение

Задача измерения свойств объектов различной природы всегда являлась одной из наиболее сложных задач, т.к. успех конкретного исследования и качество получаемой информации всегда зависит от качества измерения. Необходимость измерения социальных характеристик объясняется как теоретическими, так и практическими задачами, которые возникают в социально-экономических системах. Измерение качественных свойств с помощью количественных параметров является необходимым условием любого социологического исследования [1].

В Украине многие организации в своей деятельности часто опираются на выводы социологических исследований. Если социологические данные достоверны, удовлетворяют научным критериям обоснованности и надежности, получены наиболее экономичным путем с помощью апробированных методик, они могут способствовать принятию оптимальных решений. Но иногда полученная информация искажается из-за ошибок, допущенных на этапе сбора данных (неправильная формулировка вопроса в анкетах, выбор неподходящей шкалы оценивания, нарушение методики проведения исследования и т.д.). Также недостоверные результаты могут возникнуть и в процессе анализа данных из-за неправильной обработки и интерпретации.

Проведение социологических исследований всегда является сложным и дорогостоящим процессом, как по времени и материальным затратам, так и по прилагаемым усилиям. Поэтому ошибки, допущенные на начальных этапах при составлении анкеты и выборе средств измерений, могут существенно

повлиять на конечный результат. Значит, проблема повышения качества и эффективности социологических исследований является актуальной проблемой.

Целью данной работы является поиск новых форм и методов для оценки эффективности надежности измерения качественных свойств объектов в социально-экономических системах.

Объектом исследования являются одномерные шкалы оценки качественных свойств объектов.

Предметом исследования являются методы оценки эффективности одномерных шкал.

Основные понятия

Прежде чем приступить к рассмотрению методов, которые используются для оценки эффективности, следует привести некоторые основные понятия.

Измерение – это процедура, с помощью которой измеряемый объект сравнивается с некоторым эталоном и получает числовое выражение в определенном масштабе или шкале.

Шкала – это знаковая система, для которой задано некоторое однозначное отображение, ставящее в соответствие реальным объектам тот или иной элемент шкалы. Формально шкалой называют кортеж $\langle X, \varphi, Y \rangle$, где X – реальный объект, φ – отображение, Y – знаковая система.

Процесс шкалирования заключается в конструировании шкалы по определенным правилам и включает в себя два этапа. На первом этапе осуществляется сбор данных, на втором – их обработка и анализ, по результатам которого выстраивается оценочная шкала (по возможности следует перейти от порядковой к интервальной шкале).

Существуют шкалы отношений, интервальные, порядковые (ранговые) и номинальные (наименований). Порядковая и шкала наименований – основные шкалы качественных признаков. Поэтому во многих конкретных областях результаты качественного анализа можно рассматривать как измерения по этим шкалам. Также выделяют псевдометрические шкалы, которые используются для измерения качественных свойств объектов с помощью принятых метрических систем оценивания.

Укажем основные виды шкал измерения и соответствующие группы допустимых преобразований [2].

Шкала наименований (номинальная) применяется для идентификации элементов множества. На ней определены две операции - «равно» и «неравно». Это могут быть имена собственные, названия городов и т.д. Значения на номинальной шкале дают возможность отличить один объект от другого. Эти значения не могут быть упорядочены и рассматриваются изолированно друг от друга. Номера студентов в учебном журнале – номинальная шкала.

Шкала порядка, как и шкала наименований, – качественная, но позволяет не только именовать, но и ранжировать элементы множества. Порядковая шкала допускает только монотонные преобразования, т.е. такие, которые не нарушают порядок следования значений измеряемых величин. Отличительная особенность порядковой шкалы – все значения упорядочены.

Шкала интервалов в отличие от шкалы порядка позволяет не только ранжировать элементы множества, но и задавать известные интервалы между элементами. Интервальная шкала допускает линейные преобразования вида

$$y = \alpha x + \beta,$$

где α - положительное число;

β - положительное или отрицательное число.

Изменение α приводит к изменению масштаба шкалы, а изменение β вызывает сдвиг по шкале, т.е. положение нуля на интервальной шкале не определено. Интервальные шкалы применяют, например, для измерения температуры. С помощью линейных преобразований современная теория тестов позволяет преобразовать результаты тестирования в интервальную шкалу.

Шкала отношений допускает линейные преобразования $y = \alpha x$.

Шкала отношений в отличие от интервальной обладает точкой нулевого отсчета. Эти шкалы используют для измерения массы тела, его длины и т.д. Например, длина может измеряться в метрах,

футах, парсеках – это определяется масштабным множителем α . Почти все физические величины измеряют шкалами отношений, но ею нельзя пользоваться в социальных измерениях

Индикатор – это доступная наблюдению и измерению характеристика изучаемого социального объекта. Индикатор является внешне хорошо различимым показателем измеряемого признака.

Надежность измерения – это мера устойчивости результатов, которая влияет на точность измерения того или иного признака.

Понятие надежности включает в себя три составляющие: обоснованность, устойчивость, точность (правильность).

Способы проверки надежности шкалы измерения

Рассмотрим способы проверки первичного измерения на надежность. Наиболее существенными требованиями и способами контроля надежности является проверка обоснованности, устойчивости, точности и правильности измерения [3].

Обоснованность шкалы – это уверенность в том, что с помощью данной шкалы целенаправленно измеряют именно то свойство, которое требуется для анализа, а не другое.

Чтобы повысить обоснованность измерения, используют ряд технических приемов:

1. Логические рассуждения на основе опыта и здравого смысла. Например, задача опроса: определить среднюю активность посещения спортзала. В анкете есть вопрос «Сколько раз в неделю Вы посещаете спортзал?». Логично, что этот вопрос следует детализировать и спрашивать о количестве посещений в будние дни, в субботу и воскресенье. Таким образом, полученная информация будет более точна при расчете средней посещаемости.

2. Тест по «эталонной» группе. Смысл проверки заключается в сопоставлении данных, полученных путем измерения по шкале, с достоверными сведениями об объекте измерения.

3. Поиск независимого критерия измерения, который бы фиксировало или измерял то же свойство объекта с достаточно высокой степенью надежности (регистрация прибором, данные независимых источников информации и т.д.).

4. Использование мнений экспертов для отбора пунктов шкалы. Так, при построении шкал на отношение к занятию спортом мнения мужчин и женщин об одних и тех же занятиях будут существенно разными. Для компоновки экспертной группы используют метод квоты, т.е. устанавливают пропорции экспертов по набору существенных призна-

ков, включая, например, пол, возраст, образование, выражающих позиции разных групп респондентов в предполагаемой выборке.

5. Совмещение нескольких показателей для регистрации одного определенного свойства, или построение индекса. Типы индексов крайне разнообразны. Суть индексной обоснованности в том, что, согласно гипотезе, данному свойству находится множество его проявлений, для каждого из которых формируют отдельную шкалу. Затем измерения по частным шкалам либо суммируются, либо из них образуют логические конструкции.

Устойчивость – это способность шкалы давать однозначную информацию при любом повторном измерении. Устойчивость часто отождествляют с надежностью процедуры в целом, однако, это различные характеристики шкалы. Существуют следующие способы проверки устойчивости:

1. Повторное измерение. Один и тот же объект измеряется дважды с двух-трехнедельным временным интервалом, и с помощью одинаковой процедуры измерения. Шкала считается устойчивой, если совпадения между первыми и вторыми сериями измерений будут достаточно высокими.

Устойчивость по повторному измерению вычисляется по формуле:

$$W = \frac{n}{N}, \quad (1)$$

где n – количество полностью совпавших пар;

N – общая численность испытуемых. Данный показатель может быть относительным или измеряться в процентах.

2. Привлечение нескольких лиц для измерения данного свойства. Шкала может быть неустойчивой, если ее пункты однозначно не определены самими исследователями. Чаще это характерно для шкал качественной классификации объектов. Для повышения устойчивости целесообразно привлечь одновременно сразу несколько лиц, которые будут проводить непосредственно процедуру измерения на одном объекте. Если полученные результаты будут согласовываться, то шкала устойчива. Если нет, то требуется искать другие индикаторы, однозначно отражающие свойства объекта.

3. «Расщепление шкалы». Шкала разбивается на две половины. Если эти половины коррелируют между собой, значит, они образуют единую, более надежную шкалу, чем каждая в отдельности.

Точность (правильность) – достаточная чувствительность шкалы как оптимальная дробность, дающая устойчивое измерение без наличия систематических погрешностей. Проверить надежность шкалы по данному критерию можно с помощью

устранения явных погрешностей: выявление работающей части шкалы, испытания нескольких шкал разной дробности с расчетом абсолютной ошибки каждой из них и сравнение относительных ошибок.

Достижение устойчивых данных при максимальной дробности метрики повышает точность измерения.

Правильность и точность измерения можно повысить путем расчета относительной ошибки измерения. Относительная ошибка позволяет сопоставлять правильность замеров по двум и более шкалам разной чувствительности и таким путем выбрать оптимальный инструмент [3]. Относительная ошибка вычисляется как некоторый нормированный показатель:

$$\bar{\Delta}_{\text{отн}} = \frac{W}{m}, \quad (2)$$

где W – оценка полной устойчивости шкалы;

m – количество зон реально работающей шкалы.

Минимизация относительной ошибки при высокой устойчивости повышает точность и правильность.

Расчет критерия надежности на основе тестового примера

Рассмотрим, как можно оценить надежность выбранных шкал для измерения качественных свойств объекта на реальном примере. В качестве тестового примера была взята анкета социологического исследования «Идеальный преподаватель глазами студента», состоящая из 23 вопросов. Каждый вопрос отражал измерение качественного признака – одно из свойств, которым должен обладать преподаватель. Измерение каждого свойства проводилось в 10-бальной порядковой шкале. Для простоты и наглядности расчетов результаты получены путем опроса группы студентов в количестве 11 человек.

Этап 1. Проверка обоснованности (экспертный метод). В данном опросе приняли участие 6 девушек и 5 парней. Как было сказано раньше, что для компоновки экспертной группы устанавливаются пропорции по существенному признаку такому, как пол, т.е. можно считать, что условие для метода экспертной оценки выполняется.

Этап 2. Проверка на устойчивость с помощью повторного измерения. Студенты одной и той же группы отвечали дважды на вопросы с интервалом в одну неделю.

По каждому студенту и вопросу был проведен подсчет совпавших пар ответов. Результаты приведены в таблице 1.

Таблица 1

Сравнение данных двух замеров: совпадение (1) и несовпадение (0) пар

Код студента	№ вопроса						Итог по строкам
	1	2	...	10	...	23	
GAM	0	0	...	1	...	1	16
ROA	1	1	...	1	...	1	15
HAO	1	1	...	1	...	0	17
KAU	1	0	...	0	...	0	16
BSA	1	1	...	1	...	0	16
ZEA	0	1	...	1	...	0	16
KTV	1	1	...	1	...	0	17
SNA	0	1	...	1	...	0	18
VAA	1	1	...	0	...	1	13
MUA	1	1	...	1	...	1	17
PAA	1	1	...	0	...	1	16
Количество совпадений	8	3		8		5	
W	0,72	0,81		0,72		0,45	

Этот вопрос касался личных характеристик преподавателя. Следовательно, его необходимо заменить или перефразировать, т.к. этот вопрос не устойчив. Аналогично проверяются все вопросы анкеты.

Основной критерий устойчивости информации — анализ данных по столбцам. Однако важным является и анализ данных по строкам. Если анализировать итоги по строкам, видно, что некоторые субъекты (VAA особенно) дали наибольший разброс, а некоторые (SNA) — почти не дали разброса. Те вопросы, в которых обнаружено несовпадение даже у весьма «устойчивых» субъектов, должны быть переформулированы.

Этап 3. Точность (правильность) шкалы устанавливаются путем проверки чувствительности (дробности) и расчета относительной ошибки измерения.

При некотором навыке и достаточном исследовательском опыте можно интуитивно установить, какова должна быть дробность метрики, обеспечивающая устойчивую информацию. Однако, приступая к измерению новых или сложных социальных объектов или процессов, необходимо провести эксперименты для определения оптимальной дробности шкалы. Для проверки этого критерия необходи-

мо относительную ошибку измерения рассчитать по формуле (2).

Чтобы получить исходные данные для проверки дробности, было предложено оценить одни и те же вопросы анкеты в трех шкалах: 10-ти бальной, 5-ти бальной, 3-х бальной. Десятибалльная шкала имеет вид (табл.2).

Рассмотрим расчет относительной ошибки на примере вопроса: «Какой должен быть уровень эрудированности идеального преподавателя?». Полноса шкалы интерпретируются как «очень высокий» (10 баллов) и «очень низкий» (1 балл). Промежуточные пункты не имеют словесной интерпретации.

По таблице 2 видно, что оценки 5, 4, 3, 2, 1 вообще не используются.

Также, в случае, если оценки встречаются, но в сумме получается небольшой процент ответов от общего количества (приблизительно до 30%), то эту часть шкалы можно признать «плохой». Остается 5 пунктов работающей шкалы, т.е. $m = 10 - 5 = 5$. Проверив шкалу на устойчивость, как было описано выше, было получено $W = 0,73$, т.е. 73% полного совпадения ответов в двух последовательных пробах, что определенно недостаточно.

Относительная ошибка для десятибалльной шкалы будет равна

Таблица 2

Распределение ответов 11 опрошенных

Пункт шкалы в баллах	10	9	8	7	6	5	4	3	2	1
Частота использования балла	2	3	4	1	1	0	0	0	0	0

$$\bar{\Delta}_{\text{отн}} = \frac{W}{m} = \frac{0,73}{10-5} = 0,146.$$

Теперь проверим на устойчивость пятибалльную и трёхбалльную шкалы, задавая тот же вопрос той же группе испытуемых (группа из 11 человек). Результаты расчетов приведены в таблице 3.

Относительные ошибки равны

1) для пятибалльной шкалы:

$$\bar{\Delta}_{\text{отн}} = \frac{0,545}{5-4} = 0,545;$$

2) для трёхбалльной:

$$\bar{\Delta}_{\text{отн}} = \frac{0,909}{3-1} = 0,455.$$

На основе данных таблицы 3 определим надежность шкал.

Относительные ошибки трёхбалльной шкалы (0,455) и пятибалльной (0,545) практически одинаковы, а десятибалльной — существенно ниже (0,146). Сравнивая устойчивость шкал, можно сделать выводы, что по трёхбалльной и десятибалльной шкалам устойчивость сопоставима: 90,9% и 73%, т.е. градации шкал хорошо различимы респондентами. Однако устойчивость пятибалльной шкалы равна 54,5%, что является не высоким показателем.

Поэтому целесообразно выбирать из двух шкал десяти- и трёхбалльной. Если рассмотреть чувствительность этих шкал, то в отношении к трем градациям это составит $0,455:3=0,152$, а для десятибалльной: $0,146:10=0,014$ длины шкалы. Следовательно, десятибалльная шкала во много раз чувствительнее, а значит, правильнее и точнее.

Выводы

Проанализировав результаты опроса можно сделать следующие выводы:

1) в работе была рассмотрена надежность шкал измерений на основе трех критериев: обоснованности, устойчивости, точности (правильности);

2) как показал тестовый пример, наиболее простым методом для оценки обоснованности и устойчивости шкалы является метод повторного измерения на контрольной группе участников, которые выступают экспертами;

3) проверив критерий устойчивости, стало понятно, что некоторые вопросы необходимо перефразировать (в примере вопрос №23), потому что они дают большой разброс (45% несовпадений), а выбор более точной шкалы целесообразно выполнять путем сравнения величин относительной устойчивости измерения (например, десятибалльная шкала оказалось более чувствительная, чем пяти- и трехбалльные, несмотря на то, что много пунктов шкалы вообще не были выбраны респондентами);

4) результаты, полученные по тестовому примеру, не являются однозначными, поскольку эксперимент проводился на небольшой группе человек. Тем не менее, показано, что даже в этом случае возможно получить достоверные оценки надежности шкалы;

5) данный подход, который представлен в источнике [3], широко применяется для оценки надежности одномерных, а не многомерных шкал. Поэтому целесообразно работать над разработкой новых методов и моделей оценки надежности многомерных шкал.

Таблица 3

Выбор более точной шкалы путем сравнения величин относительной устойчивости измерения

Метрика шкал	Показатели распределения										W	$\bar{\Delta}_{\text{отн}}$	
	пункты шкалы	10	9	8	7	6	5	4	3	2			1
10 баллов	пункты шкалы	10	9	8	7	6	5	4	3	2	1	0,73	0,146
	частоты	2	3	4	1	1	0	0	0	0	0		
5 баллов	пункты шкалы	5		4		3		2		1		0,545	0,545
	частоты	2		8		0		1		1			
3 балла	пункты шкалы	3			2			1				0,909	0,455
	частоты	5			6			1					

Основным результатом работы является анализ методов, которые могут быть использованы для оценки надежности одномерных шкал с целью повышения эффективности проведения социологических исследований.

Литература

1. Клигер, С.А. Шкалирование при сборе и анализе социологической информации [Текст]: моногр. / С.А. Клигер, М.С. Косолапов, Ю.Н. Толстова. – М.: Наука, 1978. – 112 с.
2. Орлов, А.И. Эконометрика [Текст]: учеб. пособие / А.И. Орлов. – М.: Экзамен, 2002. – 576 с.
3. Ядов, В.А. Стратегия социологического исследования. Описание, объяснение, понимание социальной реальности [Текст]: Университетский учебник / В.А. Ядов. – 3-е изд., испр. – М.: Омега-Л, 2007. – 567 с.

Поступила в редакцию 8.02.2013, рассмотрена на редколлегии 13.02.2013

Рецензент: д-р техн. наук, проф., зав. каф. охраны труда, стандартизации и сертификации Р.М. Трищ, Украинская инженерно-педагогическая академия, Харьков.

ОЦІНКА ЕФЕКТИВНОСТІ ОДНОВИМІРНИХ ШКАЛ ПРИ АНАЛІЗІ ДАНИХ В СОЦІОЛОГІЧНИХ ДОСЛІДЖЕННЯХ

М.С. Мазорчук, В.С. Добряк, К.О. Базілевич, М.Ю. Хазай

В роботі розглянуто оцінку надійності вимірювання при аналізі ефективності одновимірних шкал. Розкрито основні поняття даної галузі вивчення. Описано види шкал, які використовуються при соціологічних вимірюваннях, та їх застосування. Розглянуто методи перевірки первинного вимірювання на надійність: перевірка обґрунтованості, стійкості, точності і правильності вимірювання. Деякі з описаних методів застосовано для вирішення реальної задачі. Наведено розрахунки деяких критеріїв надійності одновимірної шкали, показано проміжні обчислення. Проаналізовано отримані результати.

Ключові слова: аналіз якісних даних, вимірювання, ефективність, шкала, надійність, тестування.

ESTIMATION OF EFFICIENCY ONE-DIMENSIONAL SCALES IN DATA ANALYSIS IN SOCIOLOGICAL RESEARCH

M.S. Mazorchuk, V.S. Dobriak, K.A. Bazilevich, M.Yu. Khazay

The paper deals with measurement reliability evaluation in one-dimensional scales effectiveness analyzing. The basic concepts of this field of study are disclosed. The types of the sociological measurements scales and their applications are described. Methods of the primary measurement reliability verification: checking the validity, stability, accuracy and precision of measurement are considered. Some of these methods are used for real problems solving. The calculations of some of the one-dimensional scale reliability criteria and intermediate calculations are shown. The obtained results have been analyzed.

Keywords: analysis of qualitative data, measurement, efficiency, scale, reliability, testing.

Мазорчук Марія Сергеевна – канд. техн. наук, доцент, доцент каф. інформатики, Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: mazorchuk_mary@inbox.ru.

Добряк Виктория Сергеевна – аспирант каф. інформатики, Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: viktoriya--13@mail.ru

Базілевич Ксенія Алексеевна – студент каф. інформатики, Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: ksenia.bazilevich@gmail.com.

Хазай Маргарита Юрьевна – студент каф. інформатики, Национальный аэрокосмический университет им. Н.Е. Жуковского «Харьковский авиационный институт», Харьков, Украина, e-mail: margaritakhazay@gmail.com.