

Особенности статистической обработки результатов анкетирования в целях получения нужного результата

Кононенко А.В.
Анохина И.Ю., к.т.н., доц., доцент, Кононенко И.Н., ассистент
каф. ВМиП
Донецкий Национальный Технический университет

Inna.Anohina@pochta.ru

Рассматриваются вопросы социологического анкетирования, какие требования предъявляются к правильно составленной анкете и самому процессу анкетирования. Важным является вопрос обработки анкет, а также выявления возможных вариантов манипулирования результатами при их статистической обработке.

Питання соціологічної анкети розглядаються, що вимоги проводяться до правильно зробленої анкети і процесу анкети. Важливий - питання лікування анкет, а також виставляння можливих варіантів результатів маніпулювання, в їх статистичному лікуванні.

The questions of sociological questionnaire are examined, what requirements are produced to the correctly made questionnaire and process of questionnaire. Important is a question of treatment of questionnaires, and also exposures of possible variants of manipulation results, at their statistical treatment.

В качестве исходного материала использованы результаты свободного анкетирования жителей Донецкой области по вопросам удовлетворенности такими товарами, как обувь и одежда.

Нами выделены следующие этапы программы исследования:

1. **Формулировка проблемы.** Под проблемой понимаем существующую реальную ситуацию, носящую массовый характер и затрагивающую интересы больших социальных групп.
2. **Определение объекта и предмета исследования.**
3. **Логический анализ основных понятий.** Структурирование понятий, определяющих предмет исследования. Итогом подобной процедуры является теоретическая модель предмета исследования.
4. **Выдвижение гипотез.** Выражают общую направленность исследования.
5. **Определение выборочной совокупности.** Выборочная совокупность - это выборка из объекта исследования, причем это может быть случайная выборка, а может быть выборка по определенным правилам, отражающим по социально-демографическим признакам структуру изучаемого объекта.

6. **Составление инструментария.** Это характеристика методов и приемов сбора первичной информации (анкетного опроса, интервью, анализа документов, наблюдения); математические методы и компьютерные программы для обработки полученной информации.

7. **Полевое обследование.** Получение первичной информации: анкетирование, оценки экспертов, итоги наблюдений.

8. **Обработка и интерпретация полученных данных.** Специальные математические процедуры, в основе которых лежит теория вероятностей, определяющая технологию составления выборочной совокупности и электронной обработки данных, а также процедура эмпирического обобщения.

Определение выборочной совокупности

Приняты несколько моделей отбора выборочной совокупности опрашиваемых. Одна из них – это *вероятностная, или случайная, выборка*, в которой соблюдается *принцип равенства шансов попадания в выборку*, т.е. рассчитывается процент попадания в выборку каждой категории населения, категории делятся по половому, возрастному, социальному и другим признакам.

При построении генеральной совокупности за основу брались данные Статуправления г.Донецка. В табл. 1 приведена численность жителей г.Донецкой области при дифференциации на городское и сельское население. Здесь и далее в связи с большим количеством эмпирического материала мы будем демонстрировать только начальные и конечные строки таблицы.

Таблица 1. Дифференциация на городское и сельское население в Донецкой области за период 1990-2009г.г.

Годы	Количество населения			Количество постоянного населения			Итого
	тыс.чел.	в том числе:		тыс.чел.	в том числе:		
		городское	сельское		городское	сельское	
1990	5346.7	4836	510.7	5323.7	4809.9	513.8	5323.7
1991	5352.6	4843	509.6	5329.2	4816.7	512.5	5329.2
2008	4500,5	4072,1	428,4	4487,6	4060.4	427.2	4487,6
2009	4466,7	4042,8	423,9	4453,9	4031.8	422.1	4453,9
Средний процент		90.2%					

В таблице 2 приведены аналогичные цифры для отдельных городов и районов Донецкой области.

Таблица 2. Дифференциация на городское и сельское население в Донецкой области

	На 1 сентября 2009 року			Средняя численность		
	все население	городское	сельское	все население	городское	сельское
Область	4540051	4099189	440862	4553862	4111922	441940
Донецк	987549	985947	1602	991726	990115	1611

Авдеевка	36148	36148	-	36164	36164	-
Артемовск	108239	108239	-	108423	108423	-
Ясиноватинский	29167	8868	20299	29134	8863	20271

На основании анализа таблиц нами было принято решение при составлении генеральной совокупности опрашивать 90,2% городского населения и 9,8% сельского с учетом того.

На рис.1 дана графическая интерпретация выборок анкетлируемых с учетом процентного соотношения в приводимых выше таблицах.



Рис. 1.

На основании вышеизложенного определено процентное соотношение случайной выборки, табл.3.

Таблица 3. Выборка для проведения анкетирования

		Потребители 96%								Продавцы 4%
Район проживания	Городское				Сельское					
			90,2%				9,8			
Пол, профессия	Мужчины	Женщины								Не учитывались
	10,8%*	85,2%								
Возраст	Не учитывался	*	15-19 *	20-24	25-29	30-34	35-39	40-44	45-49 *	
		12.31%	11.14%	32.76%	25.55%	13.51%	4.57%	0.72%	0.03%	

Общая численность выборки определялась по формуле

$$\Delta = t \sqrt{\frac{\sigma^2}{n} \left(1 - \frac{n}{N}\right)}$$

где t-величина нормированного отклонения, определяемая по таблице нормального распределения вероятностей; σ^2 – квадрат среднего квадратического отклонения, n – необходимый объем выборки; N – численность генеральной совокупности.

Первоначально анкетирование производилось на основании методики, разработанной Ликертом. Шкала Ликерта состоит из ряда ответов на вопросы анкеты, выражающих степень согласия по определенному вопросу, например,

1. Полностью согласен (5), 2. Согласен (4), 3. Не могу решить (3), 4. Не согласен (2), 5. Решительно не согласен (1).

На первом этапе составлялся максимально возможный список вопросов, собранные ответы подвергались первичному отбору, причем дополнительно по каждому вопросу задавался вопрос о важности того или иного фактора, впоследствии были отсеяны вопросы, которые не считались важными для большинства респондентов.

Первичная обработка проводилась по методике Терстоуна. Идея методики заключается в определении процентов высказываний, относящихся к определенной градации, расчета кумулятивных процентов, медианы, верхнего и нижнего квартилей.

Медиана, или процентиль 50, в распределении накопленных частот - это такое значение на шкале <1> - <5>, относительно которого половина респондентов дала большие, а другая половина - меньшие оценки данного утверждения. Медиана, таким образом, делит пополам упорядоченное множество значений признака. Вычисление медианы производили по следующей формуле:

$$Md = \{ \text{Нижняя _ граница _ интервала} \} + \{ \text{ширина _ интервала} \} \cdot$$

$$\{ 50 - \text{кумулятивный _ процент _ для _ нижней _ границы _ интервала _ медианы} \} /$$

$$\{ \text{процент, соответствующий _ интервалу _ медианы} \}$$

(2).

Т.к. рассматривалась пятибалльная шкала, то ширина интервала между соседними значениями равна единице. Рассмотрим расчет медианы и квартилей на примере нескольких вопросов.

Фрагмент анкеты, используемой при проведении исследования.

Вопрос	Процент ответов по пятибалльной системе				
	1	2	3	4	5
Балы					
Укажите важность показателей при выборе и покупке обуви					
1. Наличие необходимого размера	0%	10%	5%	35%	50%
2. Соответствие товара по полноте	0%	5%	10%	35%	50%
3. Влияет ли на Ваш выбор эстетика и дизайн	15%	20%	20%	35%	10%

Таблица 4. Результаты расчетов числовых характеристик

Балы		1	2	3	4	5	Медиана и квантили	
Вопрос №1	Процент	0%	10%	5%	35%	50%	4	50%
	Кумулятивный процент	0%	10%	15%	50%	100%	3	25%
Вопрос	Процент	0%	5%	10%	35%	50%	4	50%

№2	Кумулятивный процент	0%	5%	15%	50%	100%	3.29	25%
							4.71	75%
Вопрос №3	Процент	15%	20%	20%	35%	10%	2.75	50%
	Кумулятивный процент	15%	35%	55%	90%	100%	2 4	25% 75%

Однако, не все ответы в равной мере пригодны для шкалирования, может не быть существенной разницы в ответах, соответствующей цифре 1 и 2, говоря обычным языком, не всегда присутствует разница между «очень плохо» и «плохо»; «совсем не важно» и «просто не важно».

Для оценки внутренней согласованности отдельных ответов шкалы Терстоун применил меру разброса оценок - междуквартильный размах. Междуквартильный размах - это расстояние между нижним и верхним квартилем распределения. Те ответы, для которых разброс оценок, измеренный через междуквартильный размах, превышает разницу между квартилями, исключаются из шкалы. В результирующей шкале оставляют только ответы, попавшие в указанный диапазон. Например, при ответе на вопрос, насколько важно наличие необходимого размера, будет оставлено только три ответа: 3(не совсем важен, видимо, для случая, когда обувь покупается впрок) -25% ответов, 4- (важен) -35%, 5 (очень важен) -40%. Отметим, что отбрасываются ответы вида «совсем не важен», «вообще не играет роли», а полученные проценты присоединяются к ближайшему по бальности ответу.

Точно также и при ответах на второй и третий вопросы применяется трехбалльная маркировка. Произведя обработку всех полученных ответов, была введена уже не пятибалльная, а меняющая в зависимости от результатов анкетирования шкала.

Эта система снижает и уровень ошибки. Для оценки приемлемости наблюдаемых отклонений использован коэффициент воспроизводимости шкалы Re (от англ. Reproducibility - воспроизводимость):

$$Re = 1 - \frac{\text{возможное_число_ошибок}}{\text{Общее_число_ответов}} \quad (3)$$

Для нашего анкетирования коэффициент воспроизводимости Re составил 99,5%. На практике считается приемлемым $Re \geq 90\%$. Это позволило считать предложенную методику достаточно точной, при этом увеличение шкалы до максимальной приведет к увеличению ошибки в два раза.

В приводимой ниже таблицы приводятся номера вопросов (столбец 1), оставшиеся после расчета квартилей бальности ответов, например, для вопроса №1 оставлены ответы 2, 3 и 4, а для вопроса №2 – 3, 4, 5 (столбцы 2-4), переведенная в проценты частота получения положительных ответов на вопрос с учетом баллов, например, по первому вопросу поставили 3 балла- 15% респондентов, 4 балла- 35% и 5- 50% (столбцы 5-7, в восьмом столбце приведена медиана, полученная на основании расчетов по формуле (2).

Таблица 5. Ранжировка ответов по степени важности

Вопрос	Важность для	Важность для	Медиана
--------	--------------	--------------	---------

1	потребителя в баллах			потребителя в процентах, %			8
	2	3	4	5	6	7	
1	3	4	5	15	35	50	4
2	3	4	5	5	35	60	4.17
21	1	2	3	33	44	23	1.38
Суммарное значение медианы							56.85

Значение медианы принимается за шкальный балл ответа. Для определения удельного веса каждого вопроса определяли суммарное значение всех медиан, затем для перевода в относительные величины находили, какой процент составляет медиана одного вопроса относительно общей суммы баллов (столбец 3), на последнем этапе определяли удельный вес полученных путем анкетирования результатов в разработанной шкале с учетом веса фактора (столбец 3).

Таблица 6. Вопросы анкеты и вычисленные весовые коэффициенты

Вопрос	Значение медианы	Весовой коэффициент
1	2	3
1. Размер	4.17	7.34%
2. Качество материала	4	7.04%
3. Возможность стирки без ограничений	3.75	6.60%
4. Прочность ткани	3.58	6.30%
19. Вид материала	1.6	2.81%
20. Фактура материала	1.4	2.46%
21. Наличие украшений	1.38	2.43%

Как видно из таблицы, все вопросы ранжированы в порядке снижения их значимости для потребителя.

Выборки, расположенные в столбцах P и Q сравнивали с помощью t-тестов пакета “Statistica for Windows”. Как показали результаты тестирования, наличествует существенная разница между желательной степенью удовлетворенности потребителя и реальными результатами.

Таблица 7. Таблица результатов сравнения выборок с помощью t-тестов

Идеальное значение	Среднее	Среднее квадратическое отклонение	Коэффициент Стьюдента t	Уровень значимости р
	Реальное	16.19873		
	11.43667	5.542984	-6.34011	0.000003

1. Методы социологического исследования. Девятко И.Ф. <http://lib.socio.msu.ru/l/library>
2. Девятко И. Ф. Диагностическая процедура в социологии: очерк истории и теории. М.: Наука, 1993.
3. Метод суммируемых оценок Ликерта. <http://www.bakom.ru/>

4. А.Крыштановский. Анализ социологических данных (МВШСЭН) - http://www.rusinfomar.ru/education/article_3.htm
5. Гутман Л. Основные компоненты шкального анализа//Математические методы в современной буржуазной социологии. М., 1966;