

В.Н. КАРНАВСКИЙ, О.А. МЕККА

**Методы контроля и ремонта  
выборки в ходе социального  
исследования**

УДК 316.013

Муромский институт  
(филиал) ФГБОУ ВПО  
«Владимирский  
государственный  
университет имени  
А.Г. и Н.Г. Столетовых»

*Цель статьи – показать важность математических методов в изучении социально-гуманитарных процессов. Основное внимание уделено особенностям применения известных формул из теории вероятностей для расчета репрезентативной выборки, а также для оценки процесса исследования в целом.*

*The purpose of this work is to show the importance of mathematical method while calculating representativeness in researches of social process. Main attention is paid to the peculiarities of applying the most known formulas from the theory of probability to calculate a representative selection as well as to estimate the process of research by itself.*

*Работа выполнена по материалам социологического исследования «Восприятие феноменов социальной инноватики населением провинциальных городов Владимирской области». Грант РГНФ № 11-13-33602е/Ц - 2011 г. Руководитель О.А. Мекка.*

Ошибки выборки могут иметь как систематический, так и случайный характер, т.е. могут быть обусловлены как неправильным построением выборки, так и особенностями распределения признака внутри изучаемой совокупности. Например, это может быть различие мнений студентов внебюджетной и бюджетной форм обуче-

ния, работников государственных и частных предприятий, работающих и неработающих пенсионеров [1].

Основой контроля предельно допустимой ошибкой выборки в ходе исследования может служить формула обеспечения репрезентативности бесповторной выборки вида:

$$\mu = \frac{\delta}{\sqrt{n^*}},$$

где

$\Delta \approx \mu$  – предельная ошибка выборки;

$\delta$  – стандартное отклонение от среднего, полученное в реальных условиях эмпирического) исследования;

$n^*$  - объем выборки (бесповторной), рассчитанной как  $n^* = \frac{nN}{n + N}$ ,

в которой

$n$  – рассчитанное значение повторной выборки, формула которой имеет вид [2]:

$$n = \frac{t^2 \cdot pq}{\Delta^2},$$

где  $\Delta \approx \mu$  – предельная ошибка;

$p$  – доля данного качественного признака в выборке,

$q$  – отсутствия признака (то есть  $1 - p$ ),

$t$  – параметр, определяемый заданной доверительной вероятностью  $P_{\text{доп.}}$  и определяемый по таблицам для значения интеграла

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-x}^{+x} e^{-\frac{t^2}{2}} dt \quad [2].$$

От нее мы переходим к определению объема бесповторной выборки, учитывая, что при одинаковых предельных ошибках выборки и доверительной вероятности необходимый объем бесповторной выборки  $n^*$  всегда меньше необходимого объема повторной выборки  $n$  [2]:

$$n^* = \frac{nN}{n + N},$$

рассчитывая при этом значение  $n^*$  для каждой конечной реальной генеральной совокупности  $N$ .

В условиях той же доверительной вероятности в отношении каждой (конечной) генеральной совокупности мы стремимся сократить объем выборки из соображений оптимизации процесса исследования.

Это проявляется в стремлении снизить количество респондентов каждой категории, соответственно – снизить количество анкетного материала, участников и организаторов опроса, поездок и т.д. С одной стороны, инструментарий исследования (анкета) используется бесповторно относительно определенной генеральной совокупности (социальной группы). С другой стороны, анкета, фиксирующая набор необходимых признаков, по отношению к респонденту, как правило, «проявляет себя» статистически повторно, так как многие вопросы исследования имеют не только внешнюю и открытую корреляцию, но и связь индивидуально-психологическую через респондента. Например, если респондент отрицательно реагирует на первый вопрос анкеты, то велика вероятность, что и последующие будут восприниматься аналогично. Иными словами, социально ориентированный набор вопросов анкеты относительно респондента «проявляет себя» статистически повторно в силу взаимосвязи комплекса самих социальных вопросов.

Что касается ошибки выборки, то в любом случае при получении эмпирических результатов в ходе исследований мы можем контролировать ее по значению дисперсии в каждой социальной группе. Мы считаем, что достаточно высокую планку достижения достоверности необходимо ставить каждый раз при проектировании объема выборки с учетом затрат на исследование, осознавая при этом, что реальная репрезентативность может быть получена только эмпирически. Ее коррекция осуществляется в ходе исследования в зависимости от полученных результатов достоверности (через оценку дисперсии и математического ожидания), а также и с учетом стоимости исследования.

Во-вторых, реальное эмпирическое исследование не в полной мере может быть отражено в модели приведенных формул, рассчитанных с позиций теории вероятностей и математической статистики для выборки, когда респондент фиксирует мнение в пределах дихотомии: «годен-негоден», «плюс–минус», «черный – белый», «согласен – не согласен», «оцениваю положительно» – «оцениваю отрицательно» и т.п.

В социальных исследованиях эти сильные модальные характеристики дополняются, как правило, слабыми: «поддерживаю частично», «согласен в основном», «пожалуй, не согласен», «нейтра-

лен» (равнодушен), «воздерживаюсь от ответа» и т.д. В целом это приводит к увеличению рассеивания исследуемых признаков (через увеличение показателя дисперсии  $\delta$  для каждой исследуемой подвыборки).

В силу этого обстоятельства можно предположить, что более адекватным способом оценки репрезентативности выборки проводимого или проведенного социального исследования могут служить формулы расчета  $\bar{x}, \delta$  и  $\mu$ , в которых возможные позиции выбора условно сведены к двум оценкам (состояниям): «положительно» или «отрицательно» воспринят респондентом поставленный в целом анкетой вопрос.

В-третьих, практика эмпирических измерений (в том числе методом опроса) показывает, что отношение респондентов к одному и тому же вопросу анкеты имеет различную степень «солидарности» (и, как следствие, репрезентативности) в зависимости от качества социальных параметров, например, региона, типа поселения, профессионально-квалификационных различий и т.д.

Кроме того практика социальных измерений показывает, что в пределах подвыборок нарушается и «нормальность» распределения выбора предпочтений респондентами социальных подгрупп, вплоть до эксцесса, который, кстати, также содержит в себе информацию о состоянии приоритетов в выборке. Кроме того, он содержит информацию о наличии в социальной подгруппе дополнительных, может быть не учтенных в модели исследования, представительств, имеющих существенные различия. Например, программой исследования предполагается опросить рабочих, но при этом не учитываются различия рабочих квалифицированных и неквалифицированных, работающих на крупных предприятиях и мелких.

Численное значение дисперсии как показателя «солидарности» мнений респондентов на поставленные в анкете вопросы является, на наш взгляд, важным количественным подтверждением как общей, так и частных задач исследования. В нашем случае – это задача выявления степени различия в характере восприятия населением различных провинциальных городов Владимирской области элементов социальной инноватики, а также выявления наличия дисфункций элементов инноватики, их личностной и общественной значимости.

Чем меньше показатель дисперсии, тем выше степень «солидарности» в восприятии каждого феномен инноватики социальной группой области. Показателем приоритетности в выборе из возможных позиций каждого вопроса анкеты социальной подгруппой, а также и его направленность, служит математическое ожидание  $\bar{x}_{\text{соц. подгр.}}$ , численно выраженное в единицах выбора [3].

В ходе самого исследования и анализе результатов часто возникают задачи сопоставления промежуточных или окончательных данных на значимость или незначимость различий. Особенно тогда, когда процентные доли на первый взгляд различаются друг от друга всего лишь на 3-5 %, и, следуя интуиции или иному основанию, можно было бы поставить между сопоставляемыми результатами знак приближенного равенства. Однако, с учетом объемов сопоставляемых выборок (социальных групп, участвующих в исследовании), критерии статистической достоверности на уровне 5%-й ошибки ( $p \leq 0,05$ ) позволяют внести большую определенность в значимости различий. В нашем исследовании в качестве инструмента сопоставления выступает непараметрический критерий  $\varphi^*$  - угловое преобразование Фишера.

В Таблице 1 приведены полученные эмпирические значения ошибки выборки (в результате применения расчетных методов по всем конечным генеральным совокупностям и в целом по области) применительно к исследованию восприятия феноменов социальной инноватики населением области. При этом выяснено:

1. Эмпирические показатели ошибки выборки в целом по области не вышли за планируемые пределы по всем основным задачам ( $\mu \leq 0,05$ ), что указывает на правомерность применения принятых программой исследования расчетных методов выборки.
2. Показатели математического ожидания указывают на усредненные, приоритетные выборы респондентов по области.

**Обобщенные эмпирические показатели распределения и ошибки выборки  
(по Владимирской области)**

Вопросы анкеты	Среднее значение $\delta$ по всем группам (по области)	Среднее значение выбора позиции вопроса $\bar{x}$ по всем группам (по области)	Полученное эмпирическое значение ошибки выборки по всем группам по области (репрезентативность) $\mu = \frac{\delta}{\sqrt{1557}}$
1.	1,39	3,40	0,035; (3,5%)
2.	0,535	7,39	0,013; (1,3%)
3.	0,928	10,98	0,024; (2,4%)
4.	0,88	15,03	0,022; (2,2%)
5.	0,78	19,07	0,019; (1,9%)
6.	1,01	22,32	0,026; (2,6%)
7.	1,05	26,4	0,027; (2,7%)
8.	0,78	30,14	0,020; (2,0%)
9.	1,15	33,37	0,029; (2,9%)
10.	0,968	37,6	0,025; (2,5%)
11.	0,73	41,08	0,019; (1,9%)
12.	1,0	44,5	0,025; (2,5%)
13.	0,89	48,77	0,022; (2,2%)
14.	0,66	51,94	0,016; (1,6%)
15.	0,92	55,11	0,023; (2,3%)
16.	0,82	59,0	0,021; (2,1%)
17.	1,15	62,0	0,029; (2,9%)
18.	0.6	66	0,015; (1,5%)
Общая средняя ошибка выборки			0,0228; (2,28%)

Общие краткие выводы:

1. Расчетный метод подхода к объемам выборки в интересах социальных исследований, с учетом особенностей применения известных формул теории вероятности и математической статистики, оправдывает себя на практике, обеспечивая при этом необходимые границы достоверности эмпирических результатов. При этом использование табличных данных института Гэллапа в условиях современной и статистически слабо предсказуемой российской действительности не всегда оправдано [1].

2. Математические методы контроля выборки в ходе эмпирических измерений открывают возможности выявления дополнительных значимо различных подвыборок (выявляемым по значениям

эксцесса в процессе пилотных опросов), приоритетности выборов респондентов в среднем (по значению математического среднего в единицах выбора), степени «солидарности» в выборе приоритетов и др. Это является немаловажным условием исследования социальных процессов и состояний массового сознания.

### Литература

1. *Добреньков В.И., Кравченко А.И.* Методы социологического исследования: Учебник. – М.: ИНФРА-М, 2008. – 768 с. – (Классический университетский учебник).

2. *Карасев А.И.* Теория Вероятностей и математическая статистика: Учебник для экономических специальностей вузов. – 4-е изд. – М.: Статистика, 1979. – 279 с., ил.

E-MAIL: [SPD\\_MIVLGU@MAIL.RU](mailto:SPD_MIVLGU@MAIL.RU)