

УДК 311.2

ПРИМЕНЕНИЕ ВЫБОРОЧНОГО МЕТОДА В СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ИССЛЕДОВАНИЯХ

Е.С. Шмарихина

Новосибирский государственный университет
экономики и управления «НИНХ»
E-mail: Stat2004@inbox.ru

В докладе обосновывается применение выборки в социально-экономических исследованиях. Рассматривается сущность прямой и обратной задачи выборочного метода, определяются термины, приводятся формулы расчета средней и предельной ошибки выборочной средней величины и доли. При решении обратной задачи выборки указывается важность поправки на конечность совокупности в социально-экономических исследованиях. Обсуждается порядок расчета численности выборки и оценка необходимых для этого показателей при подготовке выборочного обследования. Отмечается необходимость пересчета коэффициента доверия при обработке собранных данных для уточнения доверительной вероятности, с которой результаты выборочного наблюдения распространяются на генеральную совокупность.

Апробируется решение обратной задачи выборочного метода в рамках обследования населения Новосибирской области на тему: «Социальная поддержка населения: необходимость и направления помощи», проведенного ООО «Институт прикладной статистики и стратегического анализа». «Ремонт» выборки из-за ее смещения по возрасту респондентов увеличил фактическую численность до 460 чел. Значения пересчитанных коэффициентов доверия по признакам, заложенным при определении численности выборки, подтверждают высокую достоверность информации о генеральной совокупности.

Ключевые слова: бесповторный отбор, выборочный метод, генеральная совокупность, коэффициент доверия, повторный отбор, предельная ошибка выборки, средняя ошибка выборки, численность выборки.

APPLICATION OF SAMPLING METHOD IN THE SOCIO-ECONOMIC STUDY

E.S. Shmarikhina

Novosibirsk State University of Economics and Management
E-mail: Stat2004@inbox.ru

In the report some aspects of quality of statistical information are outlined. Sampling in social-economic researches is given ground to. The essence of direct and return problem of selection method is considered, terms are specified, formulas of calculation of an average and limit error of a sample average and a share are given. The importance of correction for finitude of population in social and economic researches when solving return problem of sampling is indicated. The procedure of payments of sampling number and assessment of indicators necessary for this purpose at preparation of sampling are discussed. The work points out the need for recalculation of coefficient of trust when processing collected data for specification of confidential probability with results of sampling extending on main population. Solution of return problem of sampling method is approved within research of population of the Novosibirsk region – Social support of population: the need for help and directions of support. «Repair» of sampling because of its shift on the age of respondents increased the actual number to 460 people. Values of the recounted coefficients of trust ac-

cording to the features which were based upon at determination of the sampling number confirm high reliability of information on population.

Key words: sampling method, population, quality of statistical information, coefficient of trust, limiting error of sampling, average sampling error, number of sampling.

В системе российской государственной статистики наряду со сплошной переписью хозяйствующих субъектов и населения практикуются и их выборочные обследования, проводимые в межпереписной период. Кроме того, различные социологические службы, центры, институты занимаются изучением мнения населения с использованием выборки. Применение выборочного наблюдения обусловлено, прежде всего, особенностями развития социально-экономических явлений и процессов, состоящими в сложности, динамичности и подверженности сильному влиянию субъективного фактора в отличие от фактов природы, что делает их в большей степени неопределенными. Поэтому в социально-экономических исследованиях повторные измерения одного и того же объекта практически невозможны. Кроме того, наличие научно обоснованной теоретической базы и экономия временных, финансовых и трудовых ресурсов по сравнению со сплошными наблюдениями обеспечивают привлекательность выборочного метода для исследователей. Но в выборочной информации возможно возникновение большего числа ошибок по сравнению со сплошным учетом.

Согласно теории выборочного метода измеряются только случайные ошибки репрезентативности. Прямая задача выборочного метода «состоит в использовании данных о составе генеральной совокупности для вероятностного суждения о предполагаемом составе различных выборок» [2, с. 3] (рис. 1).

Параметром выборки (статистикой) называют показатель, вычисленный на основе данных выборки. Статистика является случайной величиной, так как выборка сформирована случайным отбором. Параметр генеральной совокупности (параметр) – это показатель, вычисленный для всей генеральной совокупности. Параметр является фиксированным числом, так как при его вычислении отсутствует случайность.

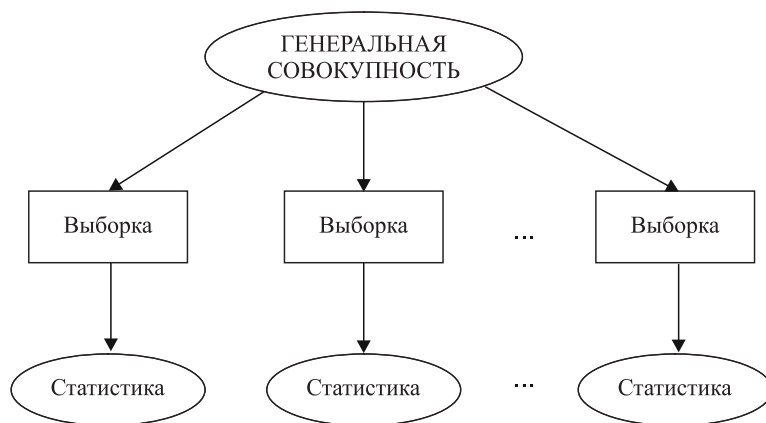


Рис. 1. Прямая задача выборочного метода

По данным многократного извлечения выборки из генеральной совокупности можно построить выборочное распределение статистики [8, с. 319–334], которое близко нормальному распределению даже тогда, когда распределение объектов в генеральной совокупности не является таковым. Этот результат называют центральной предельной теоремой, который упрощает статистический вывод. Вычисления производятся по формулам (1), (2) и (3). По выборочному распределению может быть рассчитана средняя квадратическая ошибка репрезентативности:

$$s = \sqrt{\frac{\sum \varepsilon_i^2 f_i}{\sum f_i}}, \quad (1)$$

где ε_i^2 – квадрат ошибки репрезентативности для i -й выборки; f_i – число выборок с одинаковым значением выборочной средней величины (доли).

Среднее квадратическое отклонение выборочных средних величин от генеральной средней называется средней ошибкой выборочной средней величины и рассчитывается по формуле

$$s_{\bar{x}} = \sqrt{\frac{\sum (\bar{x}_i - \mu)^2 f_i}{\sum f_i}}, \quad (2)$$

где $\bar{x}_i$ – средняя величина в i -й выборке; μ – генеральная средняя величина.

Тогда найдем среднюю ошибку выборочной доли

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad (3)$$

где p_i – доля в i -й выборке; π – генеральная доля.

В социально-экономических исследованиях из генеральной совокупности извлекается одна выборка и решается обратная задача выборочного метода – вынесение суждения о генеральной совокупности по составу выборки (рис. 2).

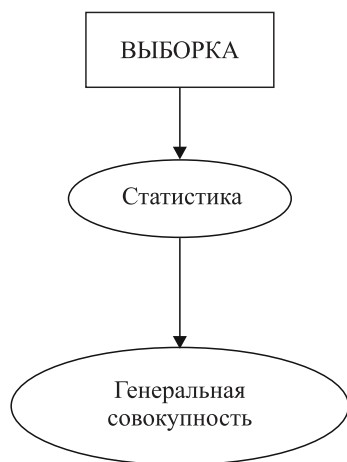


Рис. 2. Обратная задача выборочного метода

При решении этой задачи статистика становится известной величиной, параметр – неизвестной. При достаточно большой численности выборки статистики выступают в качестве оценок генеральных параметров. Большинство часто используемых оценок ($\bar{x}$, s^2 , p) являются несмещенными или почти несмещенными (s). Средняя ошибка является оценкой среднего квадратического отклонения выборочного распределения статистики (4), в расчетах средних ошибок применяются формулы (5) и (6).

Из центральной предельной теоремы известно среднее квадратическое отклонение выборки

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}, \quad (4)$$

тогда средняя ошибка выборочной средней величины

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}. \quad (5)$$

Исходя из тех же рассуждений, средняя ошибка выборочной доли

$$s_p = \sqrt{\frac{p(1-p)}{n}}. \quad (6)$$

Когда объем генеральной совокупности настолько мал, что выборка составляет большую ее часть, среднюю ошибку можно уменьшить, введя в формулу поправку для конечной генеральной совокупности $\sqrt{(N-n)/N}$ [6, с. 363]. Эта процедура вызвана тем, что при отсутствии возврата отобранных единиц в генеральную совокупность ограничивается возможность выбора и тем самым нарушается случайность процесса выборки, что может увеличить ошибку репрезентативности. В социально-экономических исследованиях поправка на конечность совокупности становится актуальной, так как выборка обычно формируется из конечных основ, представляющих собой перечень единиц наблюдения. Поправка на конечность совокупности, или корректирующий множитель, табулирована следующим образом:

$$\sqrt{\frac{N-n}{N-1}} \approx \sqrt{\frac{N-n}{N}} = \sqrt{1 - \frac{n}{N}} = \sqrt{1-f}, \quad (7)$$

где $f = n/N$ – доля отбора» [2, с. 269]; n – численность выборочной совокупности; N – численность генеральной совокупности.

Формулы расчета средней ошибки репрезентативности дополняются указанным множителем (7) при бесповторном отборе.

На величину случайной ошибки репрезентативности влияет численность выборочной совокупности – чем она больше, тем меньше ошибка. Формулы расчета численности выборки выводятся из формул предельных ошибок и имеют следующий вид:

– при повторном отборе:

$$n = \frac{t^2 \sigma^2}{\Delta^2}, \quad (8)$$

– при бесповторном отборе:

$$n = \frac{t^2 \sigma^2 N}{t^2 \sigma^2 + \Delta^2 N}. \quad (9)$$

Особенности применения указанных формул обусловлены тем, что численность выборки рассчитывается при подготовке выборочного наблюдения (и статистического исследования в целом), когда нет данных даже о выборочной совокупности.

При любом методе отбора (бесповторный, повторный) планируемой выборки расчет ее численности начинается по формуле повторного отбора (8). В этой формуле Δ – предельная ошибка репрезентативности, которая

задается, исходя из требуемой точности результатов выборки. Предельная и средняя ошибки связаны между собой равенством:

– предельная ошибка выборочной средней величины:

$$\Delta_{\bar{x}} = t \cdot s_{\bar{x}}, \quad (10)$$

– предельная ошибка выборочной доли:

$$\Delta_p = t \cdot s_p. \quad (11)$$

Величина предельной ошибки выборочной доли (11) берется равной 5 % традиционно [3, с. 114; 4, с. 84]. Величина предельной ошибки выборочной средней величины (10) задается в абсолютных единицах измерения или переводится в относительную величину путем деления ошибки в абсолютных единицах измерения на среднее значение признака, вариацию которого репрезентирует выборка. Тогда в формулах расчета численности выборки (8), (9) вместо дисперсии подставляется квадрат коэффициента вариации [5, с. 154].

В формулах (8) и (9) t – табличная величина, соответствующая заданной доверительной вероятности $F(t)$, с которой будут гарантированы оценки параметров генеральной совокупности по данным выборочного наблюдения, так называемый коэффициент доверия. Согласно теории выборочного метода коэффициент доверия t является нормированным отклонением и рассчитывается как отношение ошибки конкретной выборки к средней квадратической ошибке репрезентативности. При решении обратной задачи выборочного метода коэффициент доверия t выражается следующим образом:

– для выборочной средней величины:

$$t = \frac{\bar{x} - \mu}{s_{\bar{x}}}, \quad (12)$$

– для выборочной доли:

$$t = \frac{p - \pi}{s_p}. \quad (13)$$

Когда $n \geq 30$ по таблице значений интеграла вероятностей Лапласа–Гаусса, зная нормированное отклонение t , можно определить вероятность или на основе заданной вероятности установить величину t . Когда $n < 30$, следует находить $F(t)$ по таблицам распределения Стьюдента. В t -критерии Стьюдента вероятность зависит от числа степеней свободы, которое равно числу единиц n за вычетом числа закрепленных параметров, например, вариация около средней имеет $n - 1$ степени свободы, парная корреляция имеет $n - 2$ степени свободы [1, с. 115–116].

В формулах расчета численности выборки (8) и (9) требуется значение генеральной дисперсии (σ^2), которая неизвестна при подготовке обследования, поэтому существуют следующие способы ее оценки [5]:

1) результаты прошлых исследований той же совокупности, если ее структура и условия развития достаточно стабильны;

2) исходя из примерного значения средней величины, дисперсию находят из соотношения: $\sigma \approx \frac{1}{3} \bar{x}$;

- 3) исходя из минимального и максимального значений признака:
- если распределение признака подчиняется закону нормального распределения: $\sigma = \frac{1}{6}(x_{\max} - x_{\min})$;
 - если распределение признака асимметричное: $\sigma \approx \frac{1}{5}(x_{\max} - x_{\min})$;
- 4) для относительной величины принимают максимальную величину дисперсии, равную 0,25.

Если в результате расчета численности выборки по формуле повторного отбора (8) доля отбора превысит 0,05, то проводят расчет вторично по формуле бесповторного отбора (9), которая дает минимальный объем планируемой выборки. В ходе выборочного обследования численность выборки может меняться, поэтому по окончании, рассчитав среднюю ошибку выборочной статистики, следует пересчитать коэффициент доверия t по формуле $t = \Delta/s$. Таким образом, уточняется доверительная вероятность, с которой результаты выборочного наблюдения распространяются на генеральную совокупность.

Решение обратной задачи выборочного метода осуществляется при апробации методики расчета итоговой оценки качества [7, с. 68–69] выборочного обследования населения Новосибирской области на тему: «Социальная поддержка населения: необходимость и направления помощи», проведенного ООО «Институт прикладной статистики и стратегического анализа» в сентябре–октябре 2009 г.

Генеральную совокупность представляют все слои населения Новосибирской области без учета детей в дорабочем возрасте, численность, которой на начало 2009 г. составила 2234815 человек. При подготовке обследования была собрана информация о составе и структуре генеральной совокупности (табл. 1–3).

Таблица 1

Состав и структура населения Новосибирской области по полу на начало 2009 г.

Показатель	Все население	В том числе:	
		Мужчины	Женщины
Численность населения, чел.	2639857	1214454	1425403
Структура населения, %	100,0	46,0	54,0

Таблица 2

Состав и структура населения Новосибирской области по возрасту на начало 2009 г.

Показатель	Численность населения, чел.	Структура населения, %
Все население, в том числе:	2639857	100,0
моложе трудоспособного возраста	405042	15,4
в трудоспособном возрасте	1671520	63,3
старше трудоспособного возраста	563295	21,3

Таблица 3

Состав и структура населения Новосибирской области по муниципальным образованиям на начало 2009 г.

Показатель	Численность населения, чел.	Структура населения, %
Все население, в том числе:	2639857	100,0
Городские округа	1592442	60,3
Муниципальные районы	1047415	39,7

Объем выборки рассчитывался по формуле повторного отбора (8), многократно – по полу, возрасту и месту поселения. Значение ошибки выборки бралось традиционно, равное 0,05, задавалась вероятность, с которой результаты выборки будут распространяться на генеральную совокупность, 0,95.

Расчет численности выборки производился по указанным признакам:

– пол

$$n_1 = \frac{1,96^2 \cdot 0,46(1 - 0,46)}{0,05^2} = 382 \text{ чел.};$$

– возраст

$$n_2 = \frac{1,96^2 \cdot 0,154(1 - 0,154)}{0,05^2} = 200 \text{ чел.};$$

– место поселения

$$n_3 = \frac{1,96^2 \cdot 0,603(1 - 0,603)}{0,05^2} = 368 \text{ чел.}$$

Дисперсия первого признака дала наибольший объем выборки, который оказался равным 382 чел., что составило 0,02 % от генеральной совокупности. Процент отбора оказался меньше 5, поэтому численность выборки по формуле бесповторного отбора (9) не пересчитывалась.

По завершении опроса в структуре выборочной совокупности было обнаружено смещение по возрасту респондентов. Поэтому был произведен «ремонт» выборки методом «взвешивания» вручную (табл. 4).

Таблица 4

Распределение населения по возрасту выборочного обследования населения Новосибирской области «Социальная поддержка населения: необходимость и направления помощи»

Показатель	Население в трудоспособном возрасте	Население старше трудоспособного возраста	Всего
Планируемая выборка, чел.	286	96	382
Структура планируемой выборки, %	74,8	25,2	100,0
Фактическая выборка, чел.	266	116	382
Структура фактической выборки, %	69,6	30,4	100,0
«Ремонт» методом взвешивания, чел.	344	116	460

Необходимый «ремонт» выборки увеличил ее фактическую численность до 460 чел. Коэффициент доверия был пересчитан и опрошенная выборка населения Новосибирской области репрезентирует генеральную совокупность по признакам, которые закладывались при расчете численности выборочной совокупности:

- по полу с вероятностью 0,9684;
- по возрасту – 0,9970;
- по месту поселения – 0,9715.

Таким образом, увеличение объема выборки повысило достоверность информации о генеральной совокупности.

На основании репрезентативной выборки можно рассчитать простые оценки параметров генеральной совокупности, которые приведены в табл. 5.

Таблица 5

Простая оценка параметров по результатам выборочного обследования населения Новосибирской области «Социальная поддержка населения: необходимость и направления помощи» с вероятностью 0,95

Параметр	Точечная оценка	Интервальная оценка	
		Нижний предел	Верхний предел
1. Доля населения, нуждающегося в социальной помощи	0,482	0,436	0,528
2. Численность населения, нуждающегося в социальной помощи, чел.	1077181	974379	1179982
3. Доля населения, получившего отказ в социальной помощи	0,438	0,393	0,483
4. Численность населения, получившего отказ в социальной помощи, чел.	978849	878282	1079416
5. Доля населения, получающего социальную помощь в настоящее время	0,376	0,332	0,420
6. Численность населения, получающего социальную помощь в настоящее время, чел.	840290	741959	938622
7. Доля населения, неудовлетворенного получаемой социальной помощью	0,173	0,138	0,208
8. Численность населения, неудовлетворенного получаемой социальной помощью, чел.	386623	308404	464842

Таким образом, применение выборочного метода в социально-экономических исследованиях сводится к решению его обратной задачи, что позволяет по результатам обследования выборки получать информацию о генеральной совокупности и управлять ее достоверностью.

Литература

1. *Васильева Э.К., Юзбашев М.М.* Выборочный метод в социально-экономической статистике: учеб. пособие. М.: Финансы и статистика; ИНФРА-М, 2010. 256 с.
2. *Венецкий И.Г., Венецкая В.И.* Основные математико-статистические понятия и формулы в экономическом анализе: справочник / 2-е изд., перераб. и доп. М.: Статистика, 1979. 447 с.

3. Глинский В.В., Ионин В.Г. Статистический анализ: учеб. пособие / 3-е изд., перераб. и доп. М.: ИНФРА-М; Новосибирск: Сибирское соглашение, 2002. 241 с.
4. Давыдов А.А. Объем выборки // Социологические исследования. 1988. № 6. С. 83–89.
5. Елисеева И.И., Юзбашев М.М. Общая теория статистики: учебник для вузов / под ред. И.И. Елисеевой. М.: Финансы и статистика, 1995. 368 с.
6. Сигел Э.Ф. Практическая бизнес-статистика / пер. с англ. М.: Вильямс, 2004. 1056 с.
7. Шмарихина Е.С. Итоговая оценка качества выборочного обследования // Вестник НГУЭУ. 2012. № 1. С. 61–71.
8. Эренберг А. Анализ и интерпретация статистических данных / пер. с англ. Б.И. Клименко; под ред. и с предисл. А.А. Рывкина. М.: Финансы и статистика, 1981. 406 с.

Bibliography

1. Vasil'eva Je.K., Juzbashev M.M. Vyborochnyj metod v social'no-jekonomicheskoy statistike: ucheb. posobie. M.: Finansy i statistika; INFRA-M, 2010. 256 p.
2. Veneckij I.G., Veneckaja V.I. Osnovnye matematiko-statisticheskie ponjatija i formuly v jekonomicheskom analize: spravochnik / 2-e izd., pererab. i dop. M.: Statistika, 1979. 447 p.
3. Glinskij V.V., Ionin V.G. Statisticheskij analiz: ucheb. Posobie / 3-e izd., pererab. i dop. M.: INFRA-M; Novosibirsk: Sibirskoe soglasenie, 2002. 241 p.
4. Davydov A.A. Ob#jom vyborki // Sociologicheskie issledovanija. 1988. № 6. P. 83–89.
5. Eliseeva I.I., Juzbashev M.M. Obshhaja teorija statistiki: uchebnik dlja vuzov / pod red. I.I. Eliseevoj. M.: Finansy i statistika, 1995. 368 p.
6. Sigel Je.F. Prakticheskaja biznes-statistika / per. s angl. M.: Vil'jams, 2004. 1056 p.
7. Shmarihina E.S. Itogovaja ocenka kachestva vyborochnogo obsledovanija // Vestnik NGUJeU. 2012. № 1. P. 61–71.
8. Jerenberg A. Analiz i interpretacija statisticheskikh dannyh / per. s angl. B.I. Klimenko; pod red. i s predisl. A.A. Ryvkina. M.: Finansy i statistika, 1981. 406 p.