

*Andrew L. Gusev,  
ScD, professor,*

*Elena M. Oshchepkova,  
IV year student,  
Perm State National Research University;*

## The Method of Indirect Questions in Assessing the Accuracy of Sociological Research

**Key words:** *statistical evaluation, the accuracy of research, sociological survey, an indirect question.*

**Annotation:** *the article examines the possibility of verifying the accuracy of the results conducted sociological research. The known method for producing statistical evaluation at public opinion polls. The authors propose to use in conjunction with this method, additional (indirect) method of questions. In the illustrated example of this method.*

При проведении социологических опросов обрабатывается статистическая информация. При достаточном размере выборки (объем опроса), сформированной с помощью простого случайного отбора, в ней будут представлены все категории статистических данных, присутствующие в генеральной совокупности, и примерно в тех же самых пропорциях. Такая выборка является репрезентативной.

На практике реализация вероятностных выборок, сформированных в соответствии со всеми предъявляемыми к ним требованиями, как правило, невозможна из-за лимита ресурсов, как человеческих и временных, так и финансовых. Это касается как простого случайного отбора, так и стратифицированных и многоступенчатых выборок.

В отдельности сами по себе возникающие погрешности не велики, но наложение допущений может в результате давать недопустимые ошибки. Действительно, этим недостатком обладают все стохастические процессы выбора наблюдений – в данном случае реализации репрезентативной выборки. В связи с этим, после проведения исследований, возникает задача оценки адекватной вероятностной модели при практической её реализации (1).

Для оценки точности результатов социологических исследований в (2) предлагался весьма эффективный пошаговый способ построения оценки, позволяющих получить оценку с заданной точностью или, в крайнем случае, при невозможности получить заданную точность, определить оценку полученной точности.

Опишем этот способ. Сначала сформулируем математическую постановку задачи. Затем представим задачу и её решение в общем виде. И потом приведем простой конкретный пример.

Пусть статистика  $T$  может быть найдена в общем случае как  $\bar{T} = F(x_1, \dots, x_n)$ ,

например,  $\bar{T} = (n^{-1}) \sum_{i=1}^n x_i$ , где  $n$  – количество наблюдений случайной величины  $X$ .

Зададим некоторое  $n$  и получим следующую статистику по  $n$  наблюдениям (первый шаг):  $T_{1,j} = F(x_{1,1}, x_{1,2}, \dots, x_{1,n})$ , где  $j=1$  (первый шаг оценки точности), например,

$$T_{1,j} = (jn)^{-1} \sum_{i=1}^{jn} x_{1,i}$$

Далее по другим  $n$  наблюдениям получим статистику:

$$T_{2,j} = (x_{2,1}, x_{2,2}, \dots, x_{2,n}), \text{ например, } T_{2,j} = (jn)^{-1} \sum_{i=1}^{jn} x_{2,i}$$

Двойная индексация указывает на то, что для построения статистик  $T_{1,j}$  и  $T_{2,j}$  одинаковое количество наблюдений, но они не пересекаются.

Вычислим статистику  $T_{3,j} = (T_{1,j} + T_{2,j})/2$ . Далее для построения оценок (второй шаг) добавляется ещё по  $n$  наблюдений и т.д. То есть построение оценок становится последовательной процедурой. На первом шаге процедуры берется по  $n$  наблюдений для построения статистик  $T_{1,j}$  и  $T_{2,j}$ , на втором шаге –  $2n$  наблюдений и т.д.

В принятых обозначениях  $x_{1,i}$  и  $x_{2,i}$  ( $i = \overline{1, jn}$ ) – наблюдения, по которым строится оценка,  $j=1, 2, \dots$  – число шагов для достижения заданной точности  $\varepsilon$  (допустимой погрешности). Число шагов  $j$  принимает значение  $j_0$ , при котором впервые выполняется неравенство  $|T_{1,j} - T_{2,j}|/T_{3,j} < \varepsilon$ . Тогда за  $\bar{T}$  может быть принята статистика  $T_{3,j_0}$ , т.е.  $\bar{T} = T_{3,j_0}$ .

Понятно, что при принятых обозначениях, для вычисления с заданной точностью  $\varepsilon$  (допустимой погрешностью) необходимо наблюдать  $N=2j_0n$  раз случайную величину  $X$ . В случае, когда исследователь ограничен временными или материальными рамками, может случиться такое, что требуемая точность оценок не будет достигнута. В этом случае, целесообразно оценить погрешность полученных оценок, зафиксировав максимально возможное число шагов процедуры, равное  $j_1$ . Для этого, устанавливается минимальное значение  $\varepsilon_0$ , для которого справедливо неравенство:

$$|T_{1,j_1} - T_{2,j_1}|/T_{3,j_1} \leq \varepsilon_0. \text{ Здесь } \varepsilon_0 - \text{ достигнутая точность (допущенная погрешность).}$$

Метод косвенных вопросов. При проведении социологического опроса, по мнению авторов, целесообразно в анкеты вводить дополнительные (косвенные) вопросы, никак не связанные с целью социологического исследования. Причем, если на интересующий вопрос социологического исследования даны  $k$  возможных вариантов ответа, то и на дополнительный (косвенный) вопрос целесообразно предлагать  $k$  вариантов ответа (здесь возможны вариации).

При подведении результатов социологического исследования при статистической оценке того или иного ответа на интересующий вопрос всегда можно найти близкий по статистической оценке ответ на дополнительный (косвенный) вопрос. Это

обстоятельство поможет исследователю более точно оценить качество проведенного социологического исследования. Поясним это на примере.

Авторами были организованы и проведены социологические исследования среди школьников на предмет их отношения к спорту. Возможные варианты ответов были следующими: «Никогда не занимаюсь спортом», «Редко занимаюсь спортом», «Часто занимаюсь спортом» и «Я – спортсмен». Кроме вопроса об отношении к спорту был включен вопрос о наличии домашнего животного со следующими вариантами ответов: «Нет домашнего животного», «Собака или кошка», «Грызун» и «Другое животное».

В результате проведенных исследований были получены следующие результаты. По группе респондентов «Школьники» из 300 человек оценивалась вероятность ответа респондента «Я - спортсмен». В принятых обозначениях  $n$  равнялось 50. На первом шаге были получены следующие оценки:

$T_{1,1} = 0,1$ ,  $T_{2,1} = 0,3$ ,  $T_{3,1} = 0,2$  и допущенная погрешность оценки  $|T_{1,1} - T_{2,1}|/T_{3,1} = 1$ . На втором шаге были получены следующие оценки:  $T_{1,2} = 0,15$ ,  $T_{2,2} = 0,24$ ,  $T_{3,2} = 0,195$  и допущенная погрешность оценки  $|T_{1,2} - T_{2,2}|/T_{3,2} = 0,46$ . Наконец, на 3 шаге были получены следующие оценки:  $T_{1,3} = 0,17$ ,  $T_{2,3} = 0,21$ ,  $T_{3,3} = 0,19$  и допущенная погрешность оценки  $|T_{1,3} - T_{2,3}|/T_{3,3} = 0,21$ .

На основании этого можно утверждать, что в группе респондентов «Школьники» с вероятностью 0,19 респондент будет отвечать «Я - спортсмен» и погрешность этой вероятности не превышает 21%, т.е. вероятность ответа «Я - спортсмен» лежит в интервале от 0,15 до 0,23.

Однако наиболее близкий по оценке вероятности ответ «Нет домашнего животного» на косвенный вопрос имеет оценку вероятности равную 0,17 с погрешностью 0,24. Следовательно, исследователь обязан предположить, что вероятность ответа на интересующий вопрос «Я - спортсмен» лежит в интервале от 0,144 до 0,236.

В заключении отметим, что ещё более «аккуратными» интервальные статистические оценки могут быть при рассмотрении вероятности одновременно двух событий. А именно, когда респондент на основной вопрос отвечает «Я - спортсмен», а на косвенный вопрос отвечает «Нет домашнего животного».

### **References:**

1. Gusev AL. *Evaluation of the accuracy of the results in the study of the perception of the health risks: Bulletin of Perm sociology 2009-2010 study*, 2010; 39-43.
2. Gusev AL. *The accuracy of statistical evaluation; Basic research*, 2013, № 8 (4); 828-831.