ОПРЕДЕЛЕНИЕ РАЗМЕРА СЛУЧАЙНОЙ ВЫБОРКИ ПРИ ПРОВЕДЕНИИ МАРКЕТИНГОВЫХ ИССЛЕДОВАНИЙ.

Товстыка А.К.

Белорусский государственный экономический университет
Пинский филиал.

Основной принцип исследований «по выборке» состоит в получении информации о всей популяции по сравнительно небольшой выборке из нее. Размер выборки определяет точность полученных результатов.

Классический метод создания исследуемой выборки - случайный отбор.

При принятии нормального закона распределения среднеквадратичное отклонение определяется так:

$$\sigma = \sqrt{\frac{\rho(100-\rho)}{n}}\ ;$$

где: p - процент популяции, имеющий признак, подлежащий измерению;
n - объем выборки.

Например, в выборку включены 10000 домовладельцев и мы нашли, что 10 % из них соответствуют измеряемому признаку, тогда

$$\sigma = \sqrt{\frac{10(100-10)}{10000}} = 0{,}3\%\ ;$$

Так как вероятность попадания случайной величины X, подчиненной нормальному закону, на заданный интервал определяется как:

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right);$$

где: α и β границы интервала, m – математическое ожидание, то для нахождения вероятности отклонения σ получим:

$$P(m-\sigma < m < m+\sigma) = \Phi\left(\frac{m+\sigma-m}{\sigma}\right) - \Phi\left(\frac{m-\sigma-m}{\sigma}\right) = \Phi(1) - \Phi(-1) = 2 * \Phi(1) \approx 0{,}68\ ;$$

для отклонения в $2 * \sigma$ вероятность будет равна:

$$P(m-2\sigma < m < m+2\sigma) = \Phi\left(\frac{m+2\sigma-m}{\sigma}\right) - \Phi\left(\frac{m-2\sigma-m}{\sigma}\right) = 2 * \Phi(2) \approx 0{,}95$$

Итак, для нашего примера с вероятностью 68 % (одно стандартное отклонение) можно утвердить, что результат лежит между 9,7 и 10,3 %, а при допустимой вероятности 0,95 – между 9,4 и 10,6 % (два σ), а если n = 400, то в последнем случае границы 7-13 % (стандартное отклонение 1,5 %). Поэтому часто принимают необходимый объем выборки в 1000 респондентов (см. таблицу).

Очевидно, что абсолютный уровень ошибки – наивысший при p= 50 %, например, при n = 400 и допустимой вероятности правильного ответа 0,95 σ = 2,5 % и границы действия результатов 45-55 %. Иными словами, чем меньше ясно респондентам решение (50/50 – наихудший вариант), тем меньше точность. В этих случаях надо увеличивать выборку (увеличение выборки в два раза приводит к увеличению точности в $\sqrt{2}$ раза):

$$\sqrt{\frac{p(100-p)}{n} \bigg/ \frac{p(100-p)}{2n}} = \sqrt{2}$$

Еще раз следует подчеркнуть, что эти оценки справедливы для действительно случайной выборки.

Таблица

Диапазоны точности при различных объемах выборки.

Размер выборки	Ожидаемый результат (%) при уровне согласия 0,95		
	10	30	50
50	9	13	14
100	6	9	10
200	4	6	7
500	3	4	4
1000	2	3	3
5000	1	1	1

Собранные статистические данные могут анализироваться различным образом. Например, с использованием многомерного регрессионного анализа, факторного анализа, кластерного анализа и анализа связей.