



# ВЫБОРОЧНЫЙ МЕТОД

Авторы: Л. Н. Большев

**ВЫБОРОЧНЫЙ МЕТОД**, статистич. метод исследования общих свойств совокупности к.-л. объектов на основе изучения свойств лишь части этих объектов, называемой выборкой. Математич. теория В. м. опирается на два раздела математич. статистики – теорию выбора из конечной совокупности и теорию выбора из бесконечной совокупности. Осн. отличие В. м. для конечных и бесконечных совокупностей заключается в том, что в первом случае В. м. применяется, как правило, к объектам неслучайной, детерминированной природы (напр., число дефектных изделий в данной партии готовой продукции не является случайной величиной, это число – неизвестная постоянная, которую надлежит оценить по выборочным данным). Во втором случае В. м. обычно применяется для изучения свойств случайных объектов (напр., для исследования свойств непрерывно распределённых случайных ошибок измерений, каждое из которых может быть истолковано как реализация одного из бесконечного множества возможных результатов).

## Выбор из конечной совокупности

Выбор из конечной совокупности и его теория являются основой статистич. контроля качества, а также применяются в социологич. исследованиях. Считается, что выборка будет правильно отражать свойства всей совокупности, если выбор производится случайно, т. е. так, что любая из возможных выборок заданного объёма  $n$  из совокупности объёма  $N$  имеет одинаковую вероятность быть фактически выбранной.

На практике наиболее часто используется выбор без возвращения (бесповторная выборка), когда каждый отобранный объект в исследуемую совокупность не возвращается (такой выбор применяется, напр., при статистич. контроле качества, а также при демографич. исследованиях). Выбор с возвращением (выборка с повторением) рассматривается обычно лишь в теоретич. исследованиях (примером выбора с возвращением является регистрация числа частиц, коснувшихся в течение данного времени стенок сосуда, внутри которого совершается броуновское движение). Если  $n$  существенно меньше  $N$ , то повторный и бесповторный выборы дают практически эквивалентные результаты.

Свойства совокупности, исследуемые с помощью В. м., могут быть качественными и количественными. В первом случае задача выборочного обследования заключается в определении числа  $M$  объектов совокупности, обладающих к.-л. признаками (напр., при статистич. контроле часто интересуются числом  $M$  дефектных изделий в партии объёма  $N$ ). Оценкой для  $M$  служит отношение  $Nm/n$ , где  $m$  – число объектов с данным признаком в выборке объёма  $n$ . В случае количественного признака имеют дело с определением среднего значения  $\bar{X} = (X_1 + X_2 + \dots + X_N)/N$  совокупности. Оценкой  $\bar{X}$  для является выборочное среднее  $\bar{x} = (x_1 + x_2 + \dots + x_n)/n$ , где  $x_1, x_2, \dots, x_n$  – те значения признака из исследуемой совокупности  $X_1, X_2, \dots, X_N$ , которые принадлежат  $n$  выбранным объектам. Первый случай можно свести ко второму, полагая величины  $X_i$  равными единице, если  $i$ -й объект обладает заданным признаком, и равными нулю для остальных объектов; в этой ситуации  $\bar{X} = M/N$  и  $\bar{x} = m/n$ .

В математич. теории В. м. за характеристику изменчивости признака обычно принимают  $D$  – квадратный корень из дисперсии  $D^2 = \frac{(X_1 - \bar{X})^2 + \dots + (X_N - \bar{X})^2}{N}$ , представляющей собой среднее значение квадратов отклонений  $X_i$  от их среднего значения  $\bar{X}$ . В случае качественного признака  $D^2 = M(N-M)/N^2$ . О точности оценок  $m/n$  и  $x$  судят по их дисперсиям  $D^2_{m/n} = \mathbf{E} \left( \frac{m}{n} - \frac{M}{N} \right)^2$ ,  $D^2_{\bar{x}} = \mathbf{E} (\bar{x} - \bar{X})^2$ , которые в терминах дисперсии конечной совокупности  $D^2$  выражаются в виде отношений  $D^2/n$  и  $D^2/n$  (в случае выборок с повторением) и  $D^2(N-n)/(n(N-1))$  (в случае бесповторных выборок); здесь  $\mathbf{E}$  – знак математич. ожидания. Т. к. во многих практически интересных задачах случайные величины  $m/n$  и  $\bar{x}$  при  $n \geq 30$  приближённо подчиняются нормальному распределению, то отклонения  $m/n$  от  $M/N$  и  $\bar{x}$  от  $\bar{X}$ , превышающие по абсолютной величине  $2D_{m/n}$  и  $2D_{\bar{x}}$  соответственно, могут при  $n \geq 30$  осуществляться в среднем приблизительно в одном случае из двадцати.

Более полную информацию о распределении количественного признака в данной совокупности можно получить с помощью эмпирич. распределения этого признака в выборке.

## Выбор из бесконечной совокупности

В математич. статистике выборкой принято называть результаты к.-л. однородных наблюдений, чаще всего независимых. Предполагается, что принципиально можно осуществить любое число таких наблюдений. Полученные фактич. результаты считают выборкой из бесконечного множества возможных результатов, называемых генеральной совокупностью.

Понятие генеральной совокупности не является логически безупречным и необходимым. Для решения практич. задач нужна не сама бесконечная генеральная совокупность, а лишь те или иные характеристики, которые ей ставятся в соответствие. Эти характеристики с точки зрения теории вероятностей являются числовыми или функциональными характеристиками некоторого распределения вероятностей, а элементы выборки – реализациями случайных величин, подчиняющихся этому распределению. Такое истолкование позволяет распространить на выборочные оценки общую теорию статистич. оценок. По этой причине, напр., в вероятностной теории обработки наблюдений понятие бесконечной генеральной совокупности заменяется понятием распределения вероятностей, содержащего неизвестные параметры. Результаты наблюдений трактуются как экспериментально наблюдаемые значения случайных величин, подчиняющихся этому распределению. Цель обработки – вычисление по результатам наблюдений статистич. оценок для неизвестных параметров распределения, в том или ином смысле оптимальных.

Выше речь шла о выборочном обследовании одной совокупности к.-л. объектов. Однако практич. применение В. м. часто осуществляется во многих однородных совокупностях (напр., при оценке доли бракованных изделий в нескольких партиях готовой продукции). В этой ситуации объектом изучения является не одно число  $M$ , а неск. неизвестных чисел  $M_1, M_2, \dots$ . Пусть, напр., все обследуемые партии готовой продукции содержат по  $N$  изделий, причём  $M_1, M_2, \dots$  – количества дефектных изделий в этих партиях,  $m_1, m_2, \dots$  – соответствующие количества дефектных изделий, обнаруженных в выборках одного и того же объёма  $n$ . Согласно условию т. н. бездефектной приёмки, партия с номером  $i$  передаётся потребителю, если  $m_i = 0$ , в противном случае она бракуется. Предположим, что контроль изделий сопряжён с их уничтожением, и поэтому потребитель получает либо партию объёма  $R_i = 0$  (при  $m_i > 0$ ), либо (при  $m_i = 0$ ) партию объёма  $R_i = N - n$  с количеством дефектных изделий  $D_i = M_i$ , причём значения  $R_1, R_2, \dots$  известны, а значение  $D_1 + D_2 + \dots$

неизвестно. Отношение  $(D_1 + D_2 + \dots) / (R_1 + R_2 + \dots)$  называют долей пропущенного брака, а его математическое ожидание  $g$  – средней долей пропущенного брака. Задача математической статистики заключается в оценке  $g$  по значениям  $R_1, R_2, \dots$ , зафиксированным в результате применения В. м. Если значения  $M_1, M_2, \dots$  можно трактовать как реализации независимых одинаково распределённых случайных величин с известным законом распределения, задаваемым вероятностями  $P\{M_i=r\}=p_r, r=0, 1, \dots, N$ , то, согласно формуле Байеса, статистическая оценка  $\tilde{D}$  среднего числа пропущенных дефектных изделий в принятых партиях выражается формулой  $\tilde{D} = E\{M | \text{vert } m=0\} = \left( \sum_{r=1}^{N-n} \frac{C_{N-r}^n}{C_N^n} p_r \right) / P\{m=0\}$ , где  $C_m^k$  для целых неотрицательных  $m$  и  $k$  означает число сочетаний из  $m$  по  $k$ , при этом  $\tilde{D} \leq \frac{(N-n)P\{m=1\}}{P\{m=0\}}$ , где  $P\{m=k\} = \sum_{r=0}^{N-n} \frac{C_r^k C_{N-r}^{n-k}}{C_N^n} p_r$ ,  $\text{quad } k=0, 1, \dots$ . Поэтому оценка  $\tilde{g} = \frac{\tilde{D}}{(N-n)}$  средней доли пропущенного брака в принятых партиях удовлетворяет неравенству  $\tilde{g} \leq \frac{P\{m=1\}}{P\{m=0\}} \approx \frac{s_1}{s_0}$ , где  $s_0$  – число принятых партий, а  $s_1$  – число тех забракованных партий, в выборках из которых обнаружено ровно одно дефектное изделие.

## Литература

Лит.: Дунин-Барковский И. В., Смирнов Н. В. Теория вероятностей и математическая статистика в технике. М., 1955; Кендалл М. Дж., Стьюарт А. Теория распределений. М., 1966; Беляев Ю. К. Вероятностные методы выборочного контроля. М., 1975; Джессен Р. Методы статистических обследований. М., 1985; Sampling. Amst.; N. Y., 1988.

Processing math: 0%