Статистические оценки параметров распределения

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое именно распределение имеет признак. Естественно возникает задача оценки параметров, которыми определяется это распределение. Например, если наперед известно, что изучаемый признак распределен в генеральной совокупности нормально, то необходимо оценить (приближенно найти) математическое ожидание и среднее квадратическое отклонение, так как эти два параметра полностью определяют нормальное распределение:

Обычно в распоряжении исследователя имеются лишь данные выборки, например значения количественного признака $x_1, x_2, \ldots, x_n$, полученные в результате n наблюдений (здесь и далее наблюдения предполагаются независимыми). Через эти данные и выражают оцениваемый параметр. Рассматривая $x_1, x_2, \ldots, x_n$ как независимые случайные величины $X_1, X_2, \ldots, X_n$, можно сказать, что найти статистическую оценку неизвестного параметра теоретического распределения — это значит найти функцию от наблюдаемых случайных величин, которая и дает приближенное значение оцениваемого параметра. Например, как будет показано далее, для оценки математического ожидания нормального распределения служит функция (среднее арифметическое наблюдаемых значений признака)

$$\bar{X} = (X_1 + X_2 + \ldots + X_n)/n.$$

Итак, статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых случайных величин.

Несмещенные, эффективные и состоятельные оценки

Для того чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определенным требованиям. Ниже указаны эти требования.

Пусть Θ^* — статистическая оценка неизвестного параметра Θ теоретического распределения. Допустим, что по выборке объема n найдена оценка Θ_1^*. Повторим опыт, т. е. извлечем из генеральной совокупности другую выборку того же объема и по ее данным найдем оценку Θ_2^*. Повторяя опыт многократно, получим числа $\Theta_1^*, \Theta_2^*, \ldots, \Theta_k^*$, которые, вообще говоря, различны между собой. Таким образом, оценку Θ^* можно рассматривать как случайную величину, а числа Θ_1^*, Θ_2^*, ..., Θ_k^* — как ее возможные значения.

Несмещенной называют статистическую оценку Θ^*, математическое ожидание которой равно оцениваемому параметру Θ при любом объеме выборки, т. е.

$$M(\Theta^*) = \Theta.$$

Иными словами, соблюдение требований $M(\Theta^*) = \Theta$ гарантирует от получения систематических ошибок.

Смещенной называют оценку, математическое ожидание которой не равно оцениваемому параметру.

Эффективной называют статистическую оценку, которая (при заданном объеме выборки n) имеет наименьшую возможную дисперсию.

Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру. Например, если дисперсия несмещенной оценки при $n \to \infty$ стремится к нулю, то такая оценка оказывается и состоятельной.