§6. Статистические оценки статистических параметров

П.1. Понятие статистической оценки и ее виды

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое распределение имеет признак. Возникает задача оценки параметров, которыми определяется это распределение на основании имеющихся выборочных данных. Например, если изучаемый признак распределен нормально, то необходимо оценить (приближенно найти) математическое ожидание и среднее квадратичное отклонение, так как эти два параметра полностью определяют нормальное распределение.

Оцениваемый параметр выражают через данные выборки, рассматривая их как независимые случайные величины (с.в.).

Опр. Статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых с.в.

Два вида оценок:

Опр. Точечной называют статистическую оценку, которая определяется одним числом (приближенным значением параметра).

Опр. Интервальной называют оценку, которая определяется двумя числами – концами интервала, в котором находится истинное значение параметра с заданной вероятностью.

Для того, чтобы статистические оценки давали «хорошие» приближения оцениваемых параметров, они должны удовлетворять определённым требованиям.

Пусть Q^* - статистическая оценка неизвестного параметра Q теоретического распределения.

Опр. Несмещённой называют статистическую оценку Q^*, математическое ожидание которой равно оцениваемому параметру Q при любом объёме выборки, т. е. $M(Q^*) = Q$.

Опр. Смещённой называют оценку, математическое ожидание которой не равно оцениваемому параметру.

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

При рассмотрении выборок большого объёма (n велико!) к статистическим оценкам предъявляется требование состоятельности.

Опр. Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру.

П.2. Точечные оценки

Опр. Выборочной средней $\bar{x}_в$ называют среднее арифметическое значений признака выборочной совокупности

Пусть дано статистическое распределение выборки

x_i	x_1	…	x_k
n_i	n_1	…	n_k

Тогда $$\bar{x}_B = \frac{(x_1 n_1 + x_2 n_2 + \ldots + x_k n_k)}{n}.$$ При этом $n_1 + \ldots + n_k = n$.

Опр. Выборочной дисперсией D_B называется среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\overline{x}_B$:

$$D_B = \frac{\left(\sum_{i=1}^{k} n_i (x_i - \overline{x}_B)^2\right)}{n}.$$

Рабочая формула для вычисления дисперсии:

Дисперсия равна выборочному среднему квадратов значений признака минус квадрат выборочно средней: $\boldsymbol{D}_{\text{в}} = \overline{\boldsymbol{x}_{\text{в}}^2} - \overline{\boldsymbol{x}_{\text{в}}}^2$, где $\overline{x_{\text{в}}^2} = \frac{n_1 x_1^2 + \cdots + n_k x_k^2}{n}$

Пример.

x_i	1	2	3	4
n_i	20	15	10	5

n= 20+15+10+5=50,

$$\overline{x}_{\text{в}} = \frac{1\cdot20+2\cdot15+3\cdot10+4\cdot5}{50} = \frac{100}{50} = 2,\quad \overline{\text{x}_{\text{в}}^2} = \frac{1^2\cdot20+2^2\cdot15+3^2\cdot10+4^2\cdot5}{50} = \frac{250}{50} = 5,$$

$D_{\text{в}} = 5 - 2^2 = 1.$

Опр. Выборочным средним квадратичным отклонением называют квадратный корень из выборочной дисперсии, т.е. $\sigma_{\text{в}} = \sqrt{D_{\text{в}}}$

Замеч.1 Выборочная средняя и выборочная дисперсия являются точечными оценками математического ожидания и дисперсии соответственно.

Замеч.2 Выборочная средняя является несмещенной, состоятельной оценкой. Если случайная величина распределена по нормальному закону, то оценка является также и эффективной. $D_{\text{в}}$ является смещенной, но состоятельной и эффективной.

Опр. $s^2 = \frac{\text{n}}{\text{n}-\mathbf{1}} \cdot D_{\text{в}}$ называется исправленной дисперсией, а $s = \sqrt{s^2}$ исправленным средним квадратичным отклонением.

Замеч.3 Исправленная дисперсия является несмещенной и состоятельной оценкой.

Замеч.4 При большом объеме выборки $D_{\text{в}}$ и s^2 различаются мало. s^2 чаще используют, когда n<30.