Лекция №12. Тема 8.9. Элементы теории оценок.

Понятие оценки, свойства оценок. Оценка математического ожидания и дисперсии по выборке. Статистическое оценивание параметров распределения.

§ 4. Числовые характеристики статистического распределения: выборочное среднее, дисперсия

4.1. Параметры распределения

Одна из задач математической статистики: по имеющейся выборке оценить значения числовых характеристик исследуемой случайной величины. Основными числовыми характеристиками выборки являются выборочное среднее $\bar{x}_в$ и выборочная дисперсия $D_в$.

Определение. *Выборочным средним* называется среднее арифметическое значений случайной величины, принимаемых в выборке:

$$\bar{x}_в = \frac{x_1 + \ldots + x_n}{n} \text{ или } \bar{x}_в = \frac{\sum_{i=1}^{l} x_i n_i}{n}.$$

Выборочное среднее $\bar{x}_в$ служит для точечной оценки математического ожидания $M(X)$ исследуемой случайной величины

$$\bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

Определение. *Выборочной дисперсией* называется

$$D_в(X) = \frac{\sum_{i=1}^{l} \left(x_i - \bar{x}_в\right)^2 \cdot n_i}{n} = \frac{\sum_{i=1}^{l} {x_i}^2 \cdot n_i}{n} - {\bar{x}_в}^2.$$

Выборочная дисперсия $D_в(X)$ служит для точечной оценки дисперсии $D(X)$.

4.2. Несмещенность, состоятельность, эффективность параметров распределения

Получив статистические оценки параметров распределения (выборочное среднее, выборочную дисперсию и т.д.), нужно убедиться, что они в достаточной степени служат приближением соответствующих характеристик генеральной совокупности. Определим требования, которые должны при этом выполняться.

Пусть θ^* - статистическая оценка неизвестного параметра θ теоретического распределения. Извлечем из генеральной совокупности несколько выборок одного и того же объема n и вычислим для каждой из них оценку параметра θ: ${\theta_1}^*, {\theta_2}^*, {\theta_3}^*, \ldots, {\theta_n}^*$.

Тогда оценку θ^* можно рассматривать как случайную величину, принимающую возможные значения ${\theta_1}^*, {\theta_2}^*, {\theta_3}^*, \ldots, {\theta_n}^*$.

1. Если математическое ожидание θ^* не равно оцениваемому параметру, мы будем получать при вычислении оценок систематические ошибки одного знака (с избытком, если $M(\theta^*) > 0$, и с недостатком, если $M(\theta^*) < 0$). Следовательно, необходимым условием отсутствия систематических ошибок является требование $M(\theta^*) = 0$.

Оценка θ^* называется *несмещенной,* если ее математическое ожидание совпадает с истинным значением параметра $\theta : M\left[\theta^*\right] = \theta$.

2. *Определение.* Оценка θ^* называется *эффективной,* если она несмещенная и при этом имеет наименьшую дисперсию (наименьший разброс относительно θ) по сравнению с другими несмещенными оценками параметра θ.

3. *Определение.* Оценка θ^* называется *состоятельной,* если при неограниченном увеличении объема выборки θ^* сходится по вероятности к истинному значению параметра $\theta : \theta^* \to \theta$ при $n \to \infty$.

Состоятельной несмещенной оценкой математического ожидания *a* является

$$\bar{x}_в = \frac{\sum_{i=1}^{l} x_i n_i}{n}.$$

В отличие от выборочного среднего, выборочная дисперсия является смещенной оценкой дисперсии генеральной совокупности. Пользуясь оценкой $D_в$ вместо D, мы будем совершать некоторую систематическую ошибку, так как её математическое ожидание несколько меньше истинного значения. Чтобы её ликвидировать, достаточно ввести поправку, умножив $D_в$ на $\frac{n}{n-1}$. Таким образом, можно предложить другую оценку дисперсии – *исправленную дисперсию* s^2, вычисляемую по формуле $s^2 = \frac{n}{n-1} D_в = \frac{\sum_{i=1}^{l} (x_i - \bar{x}_в)^2 n_i}{n-1}$. Эта оценка является состоятельной несмещенной оценкой дисперсии D.

Такая оценка будет являться *несмещенной.* Ей соответствует *исправленное среднее квадратическое отклонение (СКО)*

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{l} (x_i - \bar{x}_в)^2 n_i}{n-1}}.$$

При больших n поправка $\frac{n}{n-1}$ становится близкой к единице и её применение теряет смысл.

Пример 1. Оптический пирометр установлен на светящуюся нить накала, различными операторами было произведено несколько измерений температуры. Получены следующие результаты (табл.9).

Таблица 9

Температура, °С	925	950	975	1000	1025	500
Число измерений	1	9	6	18	10	2

Требуется найти среднюю квадратическую и вероятную ошибки в предположении, что эта выборка взята из нормально распределенной совокупности.

Решение. Найдем сначала выборочное среднее значение $\bar{x}_в$.

Проще находить среднее для 1000 – x, а не для x (табл. 10).

Таблица 10

$1000–x_i$	Число измерений n	$n\cdot(1000–x_i)$	$(1000–x_i)^2$	$n\cdot(1000–x_i)^2$
1	2	3	4	5
75 50 25 0 -25 -50	1 9 6 18 10 2	75 450 150 0 -250 -100	5625 2500 625 0 625 2500	5625 22500 3750 0 6250 5000
$\sum_{i=1}^{l}$	46	325	–	43125

$$1000-\bar{x}_в=\frac{\sum_{i=1}^{l}x_i n_i}{n}=\frac{325}{46}=7{,}1 \text{ или } \bar{x}_в=1000-7{,}1=992{,}9°.$$

$$D_в(X)=\frac{\sum_{i=1}^{l}x_i^{\,2}\cdot n_i}{n}-\bar{x}_в^{\,2}=\frac{43125}{46}-7{,}1^2=887.$$

Найдем исправленную дисперсию s^2,

$$s^2=\frac{n}{n-1}D_в=\frac{46}{45}\cdot 887=906{,}7\ ^\circ\mathrm{C}.$$