Лекция 21

Метод моментов

При заданном виде закона распределения случайной величины ***X*** неизвестные параметры этого распределения можно оценить, т.е. выразить как функцию вариант выборки, на основе метода моментов.

Этот метод состоит в том, что приравниваются соответствующие теоретические и эмпирические моменты и из полученных уравнений находятся оценки параметров. В случае одного параметра в теоретическом распределении для его оценки достаточно составить одно уравнение. Если имеются два параметра в теоретическом распределении, то нужно приравнять соответственно два теоретических и эмпирических момента и т.д.

1. Оценка одного параметра.

Пусть задана плотность распределения $f(x,\theta)$ с одним параметром. Согласно методу моментов приравниваем, например, соответствующие начальные моменты первого порядка, т.е. среднюю выборки $\overline{x}_в$ и математическое ожидание распределения $M[X]$. Здесь достаточно одного уравнения относительно этого параметра

$$M[X] = \overline{x}_в \tag{1}$$

Поскольку математическое ожидание является функцией параметра θ

$$M[X] = \int_{-\infty}^{+\infty} x \cdot f(x,\theta)\,dx = \phi(\theta) \tag{2}$$

соотношение (1) можно рассматривать как уравнение с одним неизвестным, которое определяет точечную оценку параметра θ, являющуюся функцией.

Пример 1. Методом моментов по выборке $x_1, x_2, \ldots, x_n$ найти точечную оценку неизвестного параметра λ показательного распределения с известной функцией плотности распределения $f(x) = \lambda e^{-\lambda x}$ $(x \geq 0)$.

Решение. Формула (2) при помощи интегрирования по частям даёт: $M[X] = \dfrac{1}{\lambda}$. Далее из формулы (1) получаем, что $\lambda = \dfrac{1}{\overline{x}_в}$, т. е. искомая точечная оценка параметра λ показательного распределения равна обратной выборочной средней: $\lambda^* = \dfrac{1}{\overline{x}_в}$.

2. Оценка двух параметров.

Пусть задана функция плотности распределения $f(x,\theta_1,\theta_2)$. Для нахождения неизвестных параметров нужно иметь два уравнения, поэтому приравняем друг другу соответственно начальные теоретические и эмпирические моменты первого и второго порядков:

$$M[X] = \overline{x}_в,\ D[X] = D_в. \tag{3}$$

Поскольку $M[X]$ и $D[X]$ есть функции от θ_1 и θ_2, соотношения (3) определяют точечные оценки этих параметров как функции от выборки:

$$\theta_1^* = \theta_1(x_1, x_2, \ldots, x_n), \quad \theta_2^* = \theta_2(x_1, x_2, \ldots, x_n).$$

Пример 2. Найти методом моментов по выборке $x_1, x_2, \ldots, x_n$ точечные оценки неизвестных параметров a и σ нормального распределения с функцией плотности распределения вероятности

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Решение. Из курса теории вероятностей известно, что для нормального распределения $M[X] = a$, $D[X] = \sigma^2$. Используя формулы (3), получаем искомые точечные оценки параметров: $a^* = \overline{x}_в$ и $\sigma^* = \sqrt{D_в}$.

Пример 3. На предприятии изготавливается определенный вид продукции. Ежемесячный объем выпуска этой продукции является случайной величиной, для характеристики которой принят показательный закон распределения $f(x) = \lambda e^{-\lambda x}$ $(x \ge 0)$.

В течение шести месяцев проводился замер объемов выпуска продукции, Получены следующие данные:

Месяц	1	2	3	4	5	6
Объем выпуска	20	24	25	28	27	32

Найти оценку параметра λ.

Решение. Так как закон распределения содержит лишь один параметр λ, то для его оценки требуется составить одно уравнение.

Находим выборочную среднюю:

$$\overline{x}_в = (20 + 24 + 25 + 28 + 27 + 32)/6 = 26.$$

Определяем математическое ожидание:

$$M[X] = \int_0^{+\infty} x \cdot f(x)dx = \lambda \int_0^{+\infty} x \cdot e^{-\lambda x} dx.$$

Интегрируя по частям, получаем $M[X] = \frac{1}{\lambda}$, откуда

$$\frac{1}{\lambda} = \overline{x}_в. \tag{4}$$

Равенство (4) является приближенным, так как правая часть его является случайной величиной. Таким образом, из уравнения (4) получается не точное значение λ, а его оценка λ^*:

$$\frac{1}{\lambda^*} = \overline{x}_в$$

Итак, $\frac{1}{\lambda^*} = 26$, откуда $\lambda^* = \frac{1}{26} = 0{,}0384615384615384615384615384615384615384615384 6154$.

Дисперсионный анализ

Дисперсионный анализ — метод в математической статистике, направленный на поиск зависимостей в экспериментальных данных путём исследования значимости различий в средних значениях. В отличие от t-критерия, позволяет сравнивать средние значения трёх и более групп. Разработан Р. Фишером для анализа результатов экспериментальных исследований. В литературе также встречается обозначение ANOVA (от англ. *ANalysis Of VAriance*).

Пример 1

Срок хранения продукции, изготовленной по технологии А, составил:

Срок хранения	x_i	5	6	7
Число единиц	n_i	2	4	4

а изготовленной по технологии B:

Срок хранения	y_i	5	6	7	8
Число единиц	m_i	1	8	7	1

Предположив, что случайные величины X и Y распределены по нормальному закону, проверить гипотезу $H_0 : \sigma_x^2 = \sigma_y^2$ при уровне значимости 0,1 и альтернативной гипотезе $H_1 : \sigma_x^2 \neq \sigma_y^2$.

Р е ш е н и е . Вычислим «исправленные» выборочные дисперсии s_x^2, s_y^2.

Для этого найдем $\overline{x}$, $\overline{y}$:

$$\overline{x} = \frac{5\cdot 2 + 6\cdot 4 + 7\cdot 4}{10} = 6{,}2; \quad \overline{y} = \frac{5\cdot 1 + 6\cdot 8 + 7\cdot 7 + 8\cdot 1}{17} = 6{,}5.$$

Тогда

$$s_x^2 = \left[\frac{25\cdot 2 + 36\cdot 4 + 49\cdot 4}{10} - 6{,}2^2\right]\cdot\frac{10}{9} = 0{,}62;$$

$$s_y^2 = \left[\frac{25\cdot 1 + 36\cdot 8 + 49\cdot 7 + 64\cdot 1}{17} - 6{,}5^2\right]\cdot\frac{17}{16} = 0{,}11.$$

Учитывая, что $s_x^2 > s_y^2$, определим $f_{расч}$:

$$f_{расч} = \frac{0,62}{0,11} = 5,64.$$

Критическое значение $f_{кр}^n$ находим из условия

$$P\left(F\left(l_1 = 10-1, l_2 = 17-1\right) > f_{кр}^n\right) = \frac{\alpha}{2} = 0,05.$$

По таблице F - распределения определяем $f_{кр}^n = 2,54$.

Так как число $f_{расч} = 5,64$ попадает в критическую область (2,54; ∞), то гипотезу о равенстве дисперсий среднего срока хранения продукции, изготовленной по технологиям А и В, отвергаем.

Пример 2

Для проверки влияния внутрицехового оформления на качество продукции рассмотрены три участка по производству однотипной продукции и проведена выборочная проверка процента брака за пять месяцев. Результаты см. в таблице.

Методом дисперсионного анализа при уровне значимости $\alpha = 0,05$ проверить нулевую гипотезу о существенном влиянии оформления участка на качество продукции.

Таблица

Номер измерения	Уровни фактора		
	F_1	F_2	F_3
1	2	3	1
2	4	5	4
3	3	4	5
4	2	3	10
5	1	6	3
Групповая средняя	2,4	4,2	4,6

Р е ш е н и е . Находим общую среднюю:

$$\bar{x} = \frac{1}{3}(2,4 + 4,2 + 4,6) = 3,73.$$

Для расчета $SS_{общ}$ по упрощенной формуле $SS_{общ} = \sum_{j=1}^{m}\sum_{i=1}^{n} x_{ij}^2 - nm(\bar{x})^2$

составляем таблицу квадратов вариант:

Номер измере	Уровни фактора		
	F_1	F_2	F_3
1	4	9	1
2	16	25	16
3	9	16	25
4	4	9	100
5	1	36	9
$\sum$	34	95	151

Вычисляем $SS_{общ}$:

$$SS_{общ} = 34 + 95 + 151 - 3 \cdot 5 \cdot 3,73^2 = 71,3.$$

Находим SS_x по формуле $SS_x = n\left[\sum_{i=1}^{m}(\bar{x}_j)^2 - m(\bar{\bar{x}})^2\right]$

$$SS_x = 5(2,4^2 + 4,2^2 + 4,6^2 - 3 \cdot 3,37^2) = 14,1.$$

Получаем SS_ε:

$$SS_\varepsilon = SS_{общ} - SS_x = 71,3 - 14,1 = 57,2.$$

Определяем факторную и остаточную дисперсии:

$$S_{общ}^2 = \frac{SS_{общ}}{m-1} = \frac{14,1}{2} = 7,05;$$

$$S_\varepsilon^2 = \frac{SS_\varepsilon}{m(n-1)} = \frac{57,2}{12} = 4,77$$

Находим $F_{расч} = \frac{7,05}{4,77} = 1,48$.

Для уровня значимости $\alpha = 0,05$, чисел степеней свободы 2 и 12 находим $F_{табл}$ из таблицы распределения Фишера — Снедекора:

$$F_{табл}(0,05; 2; 12) = 3,89.$$

В связи с тем, что $F_{расч} < F_{табл}$, нулевую гипотезу о существенном влиянии внутрицехового оформления на процент брака отвергаем.