6 Однофакторный дисперсионный анализ

1. Однофакторный дисперсионный анализ статистических данных.
2. Проверка гипотезы о значимости статистического влияния фактора на математические ожидания исследуемых случайных величин.

Дисперсионный анализ – статистический метод, предназначенный для оценки влияния различных факторов на результат эксперимента, а также для последующего планирования аналогичных экспериментов.

По числу факторов, влияние которых исследуется, различают *однофакторный* и *многофакторный* дисперсионный анализ.

Пусть на случайную величину о воздействует фактор F, который имеет k постоянных уровней $F_1, F_2, \ldots, F_k$. На каждом уровне произведено по n_i испытаний. Результаты наблюдений – числа x_{ij} $(i=\overline{1,k};\ j=\overline{1,n_i})$.

Будем рассматривать результаты измерений x_{ij} как выборки из нормальных распределений: $\xi_i \sim N(a_i, \sigma^2)$ $(i=\overline{1,k})$.

По выборочным данным вычисляются:

– групповые средние

$$\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}, \quad i=\overline{1,k};$$

– общая средняя

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}, \quad n=\sum_{i=1}^{k} n_i.$$

Проверяется нулевая гипотеза

$$H_0 : a_1 = a_2 = \ldots = a_k.$$

Для проверки нулевой гипотезы вычисляется наблюдаемое значение критерия

$$F_{\text{набл}} = \frac{\dfrac{1}{k-1}\sum_{i=1}^{k}(\bar{x}_i-\bar{x})^2 n_i}{\dfrac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2}.$$

Критическая область имеет вид

$$K = \left[F_{\alpha;k-1,n-k}; +\infty\right),$$

где $F_{\alpha;k-1,n-k} = F(100\alpha\%; k-1, n-k)$ – 100α-процентная точка распределения Фишера с $(k-1)$ и $(n-k)$ степенями свободы.

Если $F_{\text{набл}} \notin K$, т. е. $F_{\text{набл}} < F_{\alpha;k-1,n-k}$, то нет оснований отвергнуть нулевую гипотезу; если $F_{\text{набл}} \in K$, т. е. $F_{\text{набл}} \geq F_{\alpha;k-1,n-k}$, то нулевую гипотезу отвергают.

Формулы, используемые для расчета наблюдаемого значения критерия, приведены в таблице 8.

Таблица 8

Источник изменчивости	Суммы квадратов отклонений	Число степеней свободы	Дисперсия	Наблюдаемое значение критерия
Фактор F (между группами)	$\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2 n_i$ (факторная)	$k-1$	$S_1^2 = \frac{1}{k-1}\sum_{i=1}^{k}(\bar{x}_i - \bar{x})^2 n_i$ (факторная)	$F_{\text{набл}} = \frac{S_1^2}{S_2^2}$
Остаточная (внутри групп)	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$ (остаточная)	$n-k$	$S_2^2 = \frac{1}{n-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_i)^2$ (остаточная)	
Общая изменчивость	$\sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x})^2$ (общая)	$n-1$	$S^2 = S_1^2 + S_2^2$ (общая)	

Отклонение нулевой гипотезы является статистическим доказательством влияния фактора F на математические ожидания исследуемых случайных величин.

Пример 6.1 Произведено по шесть испытаний на каждом из пяти уровней фактора F. Методом дисперсионного анализа при уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний приведены в таблице 9.

Таблица 9

№ испытания	Уровни фактора				
i	F_1	F_2	F_3	F_4	F_5
1	1,37	–0,13	1,16	–1,16	0,93
2	0,11	–0,64	–0,3	2,65	0,01
3	1,56	–0,46	–0,31	1,55	–1,56
4	–0,11	–0,88	1,13	0,29	1,59
5	0,23	–0,56	–0,17	–2,16	–1,13
6	–0,76	1,28	0,6	–0,77	–1,74

Вычислим вспомогательные величины:

– групповые средние

$$\bar{x}_1 = \frac{1,37 + 0,11 + 1,56 - 0,11 + 0,23 - 0,76}{6} \approx 0,4;$$

$$\bar{x}_2 = \frac{-0,13 - 0,64 - 0,46 - 0,88 - 0,56 + 1,28}{6} \approx -0,23;$$

$$\bar{x}_3 = \frac{1,16 - 0,3 - 0,31 + 1,13 - 0,17 + 0,6}{6} \approx 0,35;$$

$$\bar{x}_4 = \frac{-1,16 + 2,65 + 1,55 + 0,29 - 2,16 - 0,77}{6} \approx 0,07;$$

$$\bar{x}_5 = \frac{0,93 + 0,01 - 1,56 + 1,59 - 1,13 - 1,74}{6} \approx -0,32;$$

– общая средняя

$$\bar{x} = \frac{1,62}{30} \approx 0,05.$$

$$\sum_{j=1}^{6} \left(x_{1j} - \bar{x}_1\right)^2 \approx 4; \quad \sum_{j=1}^{6} \left(x_{2j} - \bar{x}_2\right)^2 \approx 3,04; \quad \sum_{j=1}^{6} \left(x_{3j} - \bar{x}_3\right)^2 \approx 2,46;$$

$$\sum_{j=1}^{6} \left(x_{4j} - \bar{x}_4\right)^2 \approx 16,11; \quad \sum_{j=1}^{6} \left(x_{5j} - \bar{x}_5\right)^2 \approx 9,53; \quad \sum_{i=1}^{5}\sum_{j=1}^{6} \left(x_{ij} - \bar{x}_i\right)^2 \approx 35,14.$$

$$\sum_{i=1}^{5} \left(\bar{x}_i - \bar{x}\right)^2 n_i \approx 2,56.$$

Результаты расчетов приведены в таблице 10.

Таблица 10

Источник изменчивости	Суммы квадратов отклонений	Число степеней свободы	Дисперсия	Наблюдаемое значение критерия
Фактор F (между группами)	2,56	4	0,64	$F_{набл} = \dfrac{0,64}{1,41} \approx 0,45$
Остаточная (внутри групп)	35,14	25	1,41	
Общая изменчивость	37,70	29	1,3	

По таблице процентных точек распределения Фишера найдем $F_{0,05;4,25} = 2,76$. Так как $F_{набл} \notin K$, то нет оснований отвергнуть нулевую гипотезу. Другими словами, групповые средние не различаются значимо.

Вопросы для самоконтроля

1. Сформулируйте основную задачу дисперсионного анализа.
2. Как найти факторную и остаточную дисперсии?
3. Как построить критическую область?