

Математическая постановка задачи проверки различий между группами (сравнение моментов первого или второго порядка)

Математическая постановка задачи сформулирована следующим образом: пусть имеется множество групп студентов $\{G_1, G_2, \dots, G_m\}$, изучающих дисциплину. В каждой группе G_l множество студентов описывается как $S_l = \{s_1^{(l)}, s_2^{(l)}, \dots, s_n^{(l)}\}$, где каждый студент $s_i^{(l)}$ характеризуется вектором признаков, отражающих его траекторию обучения $s_i^{(l)} = (x_{i1}^{(l)}, x_{i2}^{(l)}, \dots, x_{ik}^{(l)})$, где $x_{ij}^{(l)}$ — метрика активности (например, количество выполненных заданий, балл за тест). Обозначим успеваемость i -го студента через $y_i^{(l)}$ (итоговый балл). Проверку различий между группами с использованием t -критерия проведем следующим образом: для каждой группы G_l рассчитаем метрики активности $A_j = \{a_1^{(l)}, a_2^{(l)}, \dots, a_p^{(l)}\}$, где $a_j^{(l)}$ — моменты первого или второго порядка. Для каждой метрики и каждой пары групп (G_l, G_r) поставим нулевую H_0 и альтернативную H_1 гипотезы:

$$H_0: \mu_{a_j}^{(l)} = \mu_{a_j}^{(r)}, H_1: \mu_{a_j}^{(l)} \neq \mu_{a_j}^{(r)}, \quad (1)$$

где $\mu_{a_j}^{(l)}$ — среднее значение параметра a_j в группе G_l ,

Критерий успешности для статистических тестов (гипотеза H_0 отвергается при $p\text{-value} < \alpha$ ($\alpha=0.05$)). Обобщенный t -критерий для нескольких групп для всех пар (l, r) запишем как:

$$t = \frac{a_j^{(l)} - a_j^{(r)}}{\sqrt{\frac{s_l^2}{n_l} + \frac{s_r^2}{n_r}}}, \quad (2)$$

где s_l^2, s_r^2 - выборочные дисперсии, n_l, n_r — размеры выборок