

В. П. Котельников (Ростов-на-Дону, РВИРВ). **О проверке допущений дисперсионного анализа.**

Дисперсионный анализ основан на допущении о нормальных распределениях случайных величин. Виды генеральных распределений случайных величин можно выявлять различными способами, а их параметры определять методом наименьших квадратов. В [1] предлагается отказаться от традиционной проверки согласования генеральных и эмпирических распределений по критериям χ^2 , Колмогорова, ω^2 и их модификациям, а вместо этого при объемах выборок $n > 5$ гипотезу об адекватности распределений проверять методом регрессионного анализа.

Пусть $G(x)$ — генеральное распределение с числом параметров m , $\tilde{G}_i$ — его предсказанное значение, $G_i = i/n$ ($i = 1, \dots, n$) — эмпирическое распределение, оценки дисперсии адекватности $s_a^2 = (n - m)^{-1} \sum_{i=1}^n (G_i - \tilde{G}_i)^2$ и дисперсии эмпирической функции $s^2 = (n - 1)^{-1} \sum_{i=1}^n (G_i - \bar{G})^2$, где $\bar{G} = n^{-1} \sum_{i=1}^n G_i$. Для проверки с уровнем значимости α гипотезы $H_0: \sigma^2 = \sigma_a^2$ против альтернативной гипотезы $H_1: \sigma^2 > \sigma_a^2$ используют статистику $F = s^2/s_a^2$, распределение которой может иметь различный вид. При нормальных распределениях случайных величин для расчета квантилей используют F -распределение Фишера.

Когда распределение статистики F неизвестно, для проверки адекватности различных генеральных распределений предлагается использовать коэффициент детерминации $d = s_a^2/s^2$, $0 \leq d \leq 1$. В решаемой задаче распределение супериндикатора G — равномерное, поэтому $s^2 = \sigma_G^2 = 1/12$ и $d = 12s_a^2$.

При проверке гипотез уровень значимости α принимается по договоренности между заказчиком и исполнителем. Распределение отклонений $G_a = G - \tilde{G}$ обычно не известно, поэтому зададимся допустимыми значениями коэффициента d тоже по договоренности с заказчиком. Проверять согласование генерального распределения (в том числе и нормального) с эмпирическим будем путем сравнения коэффициента d с допустимыми значениями. Например, $d \leq 0,02$ — адекватность отличная, $0,02 < d \leq 0,05$ — хорошая, $0,05 < d \leq 0,10$ — посредственная, $0,10 < d$ — плохая.

В дисперсионном анализе для проверки нулевой гипотезы о незначимости влияния k -го фактора на изучаемую характеристику вычисляются значения F_k , т. е. отношения дисперсий с определенными числами степеней свободы. Если генеральные распределения не нормальные, то F_k нельзя сопоставлять с критическими значениями $F_{1-\alpha}$, найденными по распределению Фишера для заданного уровня значимости α .

Предлагается вычислять коэффициент $d_k = 1/F_k$ и по договоренности с заказчиком задаваться допустимыми значениями этого коэффициента. Степень влияния исследуемого фактора проверять путем сравнения коэффициента d_k с допустимыми значениями. Например, $d_k \leq 0,05$ — влияние сильное, $0,05 < d_k \leq 0,10$ — среднее, $0,10 < d_k \leq 0,15$ — слабое, $0,15 < d_k$ — влияния нет.

СПИСОК ЛИТЕРАТУРЫ

1. Котельников В. П. О проверке согласования генеральных и эмпирических распределений. — Обзорение прикл. и промышл. матем., 2008, т. 15, в. 1, с. 90–91.