

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЙ СТАТИСТИК ДИСПЕРСИОННОГО АНАЛИЗА В МОДЕЛЯХ СО СЛУЧАЙНЫМИ ФАКТОРАМИ¹

Пономаренко В.М., Лемешко Б.Ю.
НГТУ, Новосибирск, E-mail: headrd@fpm.ami.nstu.ru

Статистические критерии, связанные с проверкой гипотез о средних значениях, обычно являются весьма устойчивыми по отношению к нарушениям предположений о нормальности ошибок наблюдений [1]. Это подтверждается и проводившимися ранее исследованиями [2-3]. С другой стороны [1], критерии, связанные с проверкой гипотез о дисперсиях, весьма чувствительны к нарушению предположений о нормальности случайных величин, входящих в соответствующие модели дисперсионного анализа.

Очевидно, что в реальных ситуациях нормальный закон далеко не всегда является наилучшей моделью для ошибок наблюдений или случайных уровней факторов. В таких случаях использование распределений статистик критериев, применение которых было обосновано в предположениях нормальности, может приводить к некорректным статистическим выводам.

Цель данной работы состояла в исследовании методами компьютерного моделирования распределений статистик дисперсионного анализа при различных законах распределения ошибок и уровней случайных факторов.

Сбалансированная модель дисперсионного анализа с одним случайным фактором имеет вид

$$y_{ij} = \mu + a_i + e_{ij}, \quad (1)$$

где I - число уровней фактора A , выбранных случайным образом, $i = 1, \dots, I$; J - число опытов на каждом уровне, $j = 1, \dots, J$; μ - среднее значение эффекта фактора A ; $\{a_i\}$, $\{e_{ij}\}$ независимы в совокупности и имеют нулевые средние.

В классической постановке выдвигаются предположения о нормальности эффектов уровней фактора и ошибок наблюдений:

$$\Omega: \begin{cases} \{a_i\} \text{ распределены } N(0, \sigma_A^2), \\ \{e_{ij}\} \text{ распределены } N(0, \sigma_e^2), \end{cases} \quad (2)$$

где σ_A^2 - дисперсия, с которой распределены $\{a_i\}$ - эффекты уровней случайного фактора A , σ_e^2 - дисперсия ошибок наблюдений $\{e_{ij}\}$.

В самом общем случае относительно случайного фактора A проверяется гипотеза, имеющая вид

$$H_A: \sigma_A^2 \leq \theta_0 \sigma_A^2, \quad \theta_0 \geq 0, \quad (3)$$

¹ Работа выполнена при поддержке Министерства образования и науки РФ (код проекта 15378)

где θ_0 - некоторая заданная константа.

Если в результате применения соответствующего критерия проверяемая гипотеза при $\theta_0 = 1$ не отклоняется, то это свидетельствует о том, что влияние эффектов уровней фактора A на отклик y не превышает влияния на отклик y имеющихся ошибок наблюдения.

В условиях предположения (2) и справедливости

$$\sigma_A^2 = \theta_0 \sigma_e^2 \quad (4)$$

статистика вида

$$S = \frac{1}{(1 + J\theta_0)} \frac{\overline{SS}_A}{\overline{SS}_e} = \frac{1}{(1 + J\theta_0)} \frac{\frac{J}{I-1} \sum_{i=1}^I (\bar{y}_{i\cdot} - \bar{y}_{..})^2}{\frac{1}{I(J-1)} \sum_{i=1}^I \sum_{j=1}^J (\bar{y}_{ij} - \bar{y}_{i\cdot})^2}, \quad (5)$$

где $\bar{y}_{i\cdot} = \frac{1}{J} \sum_{j=1}^J y_{ij}$, $\bar{y}_{..} = \frac{1}{I} \sum_{i=1}^I \bar{y}_{i\cdot}$, подчиняется F -распределению Фишера со степенями свободы $I-1$ и $I(J-1)$.

В данной работе распределения статистики (5) исследовались при различных законах ошибок наблюдений и случайного фактора модели (1), в частности при распределении максимального значения

$$f(x, \theta_1, \theta_2) = \frac{1}{\theta_2} \exp \left\{ -\frac{x - \theta_1}{\theta_2} - \exp \left(-\frac{x - \theta_1}{\theta_2} \right) \right\}$$

и семействе симметричных распределений с плотностью

$$De(\lambda) = f(x, \theta_1, \theta_2, \lambda) = \frac{\lambda}{2\sqrt{2\theta_2}\Gamma(1/\lambda)} \exp \left\{ -\left(\frac{|x - \theta_1|}{\sqrt{2\theta_2}} \right)^\lambda \right\}$$

при различных значениях параметра формы λ . Частными случаями семейства $De(\lambda)$ являются распределения Лапласа ($\lambda = 1$) и нормальное ($\lambda = 2$).

В ходе исследования поведения распределения статистики (5) в условиях нарушения предположений нормальности было выявлено, что отклонение за-

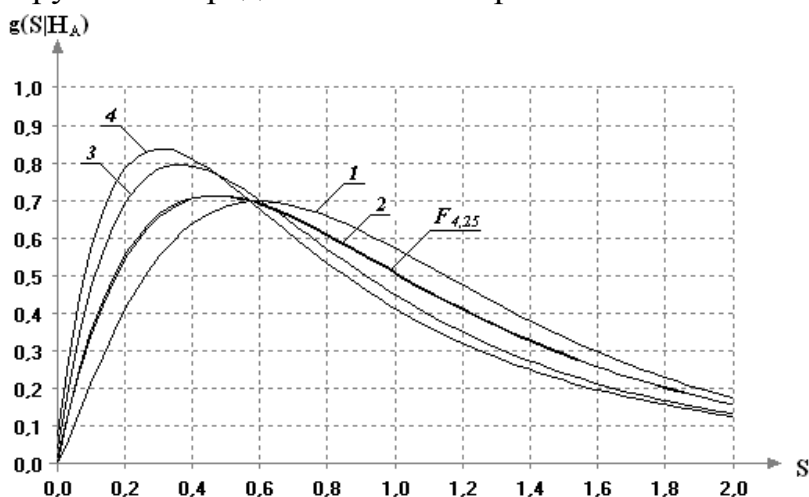


Рис. 1 Плотности распределения статистики (5) при справедливости гипотезы H_A вида (1): 1 - $g(De(2), De(10))$, 2 - $g(S(De(2), De(2)))$, 3 - $g(S(De(2), Max))$, 4 - $g(S(De(2), De(1)))$

кона распределения случайных эффектов уровней фактора A от нормального закона оказывает на распределение статистики существенно большее влияние, чем отклонение от нормального закона случайных ошибок наблюдений. Этот факт иллюстрируют рисунки 1 и 2. На них

изображены некоторые из условных плотностей полученных распределений статистики (5) в условиях нарушения предположений нормальности. Обозначения в подрисуночных надписях указывают на законы распределения ошибок и уровней случайного фактора, при которых строились распределения статистики (5). Например, $g(S(De(2), De(10)))$ соответствует плотности распределения статистики S , построенной при справедливости (4) для случая, когда ошибки $\{e_{ij}\}$ подчиняются нормальному закону, а $\{a_i\}$ – распределению $De(10)$.

Рис. 1 соответствует случаю, когда распределение ошибок наблюдений $\{e_{ij}\}$ оставалось нормальным, а эффекты уровней случайного фактора $\{a_i\}$ подчинялись одному из рассматриваемых распределений. Из рисунка видно, что отклонение распределения статистики (5) от $F_{4,25}$ – распределения статистики в нормальном случае, весьма значительное. Было выявлено, что при фиксированном распределении $\{e_{ij}\}$ распределение статистики тем более островершинно, чем более островершинно распределение $\{a_i\}$.

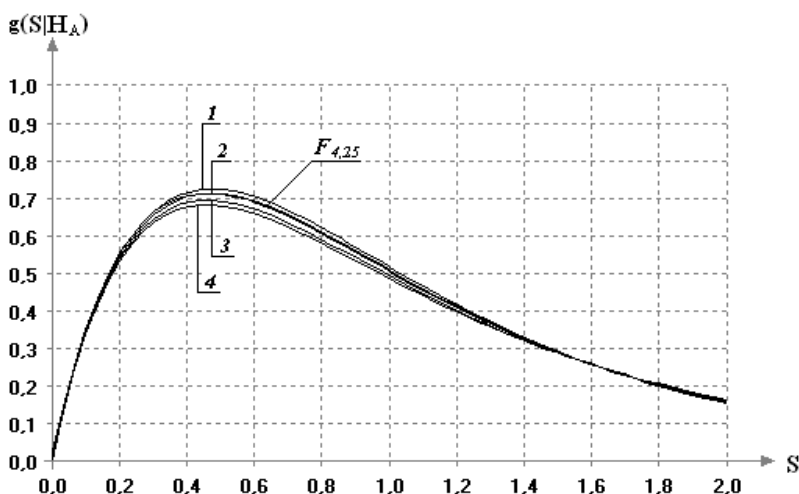


Рис. 2. Плотности распределения статистики (5) при справедливости гипотезы H_A вида (1): 1 – $g(S(De(10), De(2)))$, 2 – $g(S(De(2), De(2)))$, 3 – $g(S(Max, De(2)))$, 4 – $g(S(De(1), De(2)))$

Рис. 2 соответствует случаю, когда распределение $\{a_i\}$ оставалось нормальным, а $\{e_{ij}\}$ подчинялись одному из рассматриваемых распределений. Из представленной рисунке картины следует, что отклонение распределения статистики (5) от распределения $F_{4,25}$ далеко не такое значительное, как в предыдущем случае.

Моделей распределений статистики (5) были построены при различных комбинациях законов распределений статистик и случайных уровней факторов.

1. Шеффе Г. Дисперсионный анализ. М.: Физматгиз, 1980. - 628 с.
2. Лемешко Б.Ю. Пономаренко В.М. Проблемы применения классического аппарата дисперсионного анализа в приложениях технического, экономического и естественно-научного характера // Материалы региональной конференции (с участием иностранных ученых) "Вероятностные идеи в науке и философии". – Новосибирск: Ин-т философии и права СО РАН / НГУ. 2003. – С. 106-109.
3. Lemeshko B.Yu., Ponomarenko V.M. Statistical Hypotheses Testing In Variance Analysis In Case Of Classical Assumptions Failure // Proceedings of the Seventh International Conference "Computer Data Analysis and Modeling: Ro-

bustness and Computer Intensive Methods”, September 6-10, 2004, Minsk. Vol. 1. – P. 110-113.