Материалы Российской НТК “Информатика и проблемы телекоммуникаций”, Новосибирск. 2006. – Т.1. – С.190-194.

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ СТАТИСТИКИ, ИСПОЛЬЗУЕМОЙ В Т-МЕТОДЕ МНОЖЕСТВЕННЫХ СРАВНЕНИЙ ПРИ НАРУШЕНИИ ПРЕДПОЛОЖЕНИЙ О НОРМАЛЬНОСТИ

Пономаренко В.М., Лемешко Б.Ю.
НГТУ, Новосибирск
E-mail: headrd@fpm.ami.nstu.ru Тел. (383) 346-37-54

Подавляющее большинство статистических критериев, связанных с проверкой гипотез о средних, оказываются весьма устойчивыми по отношению к нарушениям предположений о нормальности ошибок наблюдений [1]. Это подтверждается и проводившимися ранее исследованиями [2-3].

Однако среди таких критериев выделяются методы множественных сравнений. Эти методы предназначены для одновременного проведения большого количества парных сравнений средних так, чтобы вероятность хотя бы одного неверного заключения оставалась на первоначально выбранном уровне значимости α. Относительно одного из наиболее известных методов множественного сравнения – метода Тьюкки или Т-метода в [1] делается предположение об отсутствии устойчивости к нарушению предположений о нормальности.

В данной работе с помощью методики статистического моделирования, хорошо зарекомендовавшей себя при исследовании статистических закономерностей [4,5], исследуется устойчивость Т-метода к нарушению предположений о нормальности.

Рассматривалась модель однофакторного анализа вида

$$y_{ij} = \mu_i + e_{ij},\ i = 1,\ldots,I,\ j = 1,\ldots,n, \tag{1}$$

где $\mu_1,\ldots,\mu_I$ – средние отклика y на I уровнях фактора, n – число наблюдений на уровне. В классической постановке относительно модели (1) делается предположение о том, что ошибки наблюдения $\{e_{ij}\}$ независимы и распределены по нормальному закону с нулевым математическим ожиданием и одинаковой дисперсией σ^2.

В этом случае статистика

$$R = \frac{\max_i u_i - \min_i u_i}{s\sqrt{\frac{1}{n}}}, \quad где \quad u_i = \hat{\mu}_i - \mu_i, \quad i = 1,\ldots,I, \tag{2}$$

$$\hat{\mu}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}, \quad s = \sqrt{\frac{1}{I(n-1)}\sum_{i=1}^{I}\sum_{j=1}^{n}\left(y_{ij} - \hat{\mu}_i\right)^2}, \tag{3}$$

подчиняется [1] распределению стьюдентизированного размаха $q_{I,I(n-1)}$ со степенями свободы I и $I(n-1)$.

Знание распределения статистики (2) позволяет получить совместные доверительные интервалы для $\frac{1}{2}I(I-1)$ разностей вида

$$\{\mu_{i_1} - \mu_{i_2}\}, \quad i_1, i_2 = 1, \ldots, I. \tag{4}$$

Поэтому устойчивость Т-метода к нарушению предположений о нормальности определяется, в частности, тем, как будет изменяться распределение статистики (2) в случае, когда $\{e_{ij}\}$ распределены по закону, отличному от нормального. Исследования поведения распределения статистики (2) проводились для ряда значений I и n, при различных законах распределения ошибок наблюдений модели (1). В частности, при распределениях симметричного семейства $De(\lambda)$, которые имеют более "тяжелые" хвосты по сравнению с нормальным законом при $\lambda < 2$ и менее "тяжелые" хвосты при $\lambda > 2$. Нормальное распределение представляет собой частный случай при $\lambda = 2$. Для выявления влияния на распределения статистик асимметричности закона использовалось рас-

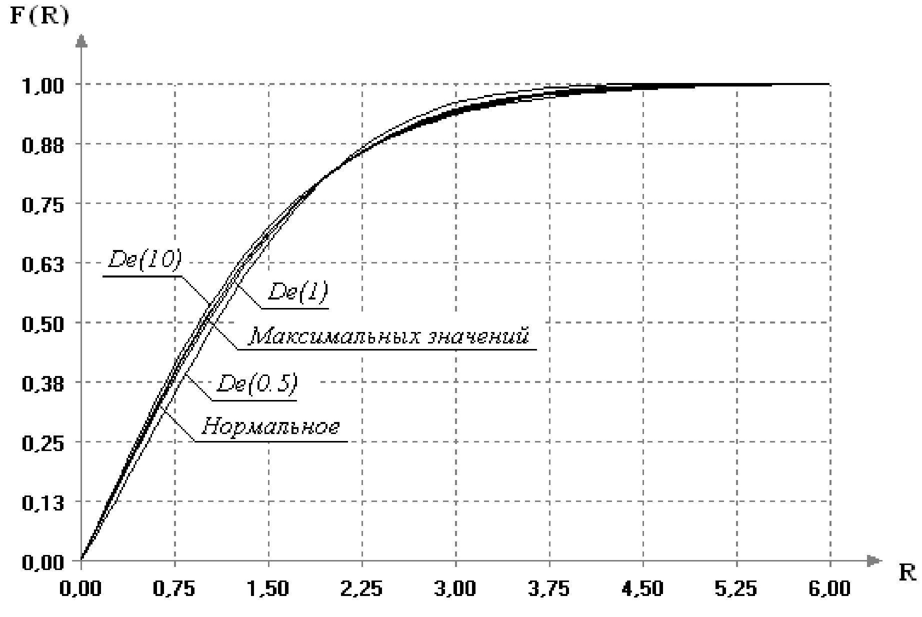
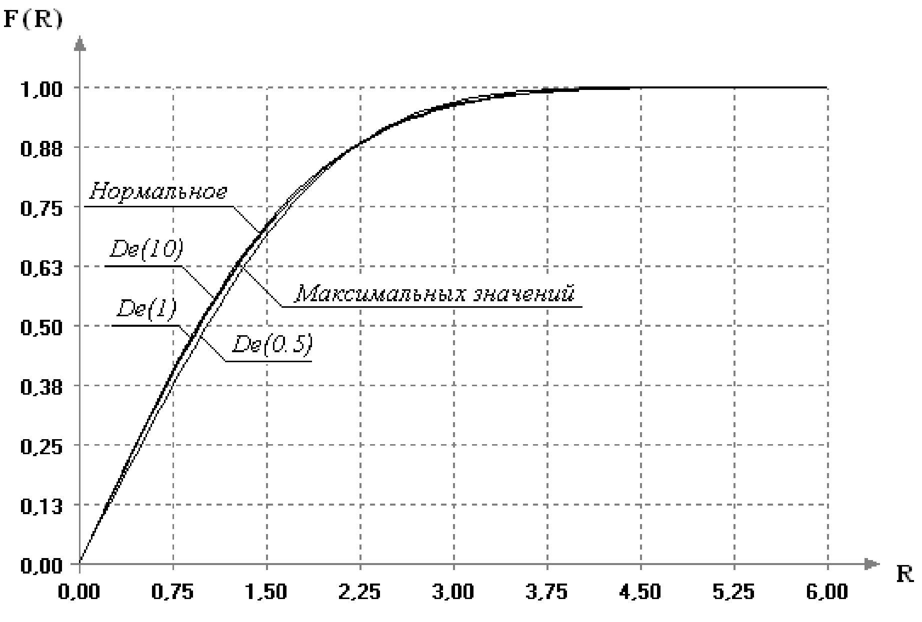


Рис. 1. Функции распределения статистики (2) в случае распределения $\{e_{ij}\}$ по законам распределения нормальному, Максимальных значений и $De(\lambda)$ с различными значениями параметра формы λ при $I = 2$ и $n = 10$



пределение максимальных значений.

По результатам проведенных исследований можно сделать вывод о том, что устойчивость Т-метода к нарушению предположений о нормальности зависит от числа уровней факторов.

При малом числе уровней факторов, как это видно из рисунков 1 и 2, значительное влияние на распределение статистики оказывают “тяжелые” хвосты ошибок наблюдений. В результате построенные с использованием традиционного аппарата доверительные интервалы для разностей вида (4) будут шире, чем фактические. Но при увеличении числа наблюдений n на уровне фактора степень расхождения границ фактических доверительных интервалов и доверительных интервалов, построенных для случая нормального закона распределения ошибок, значительно снижается.

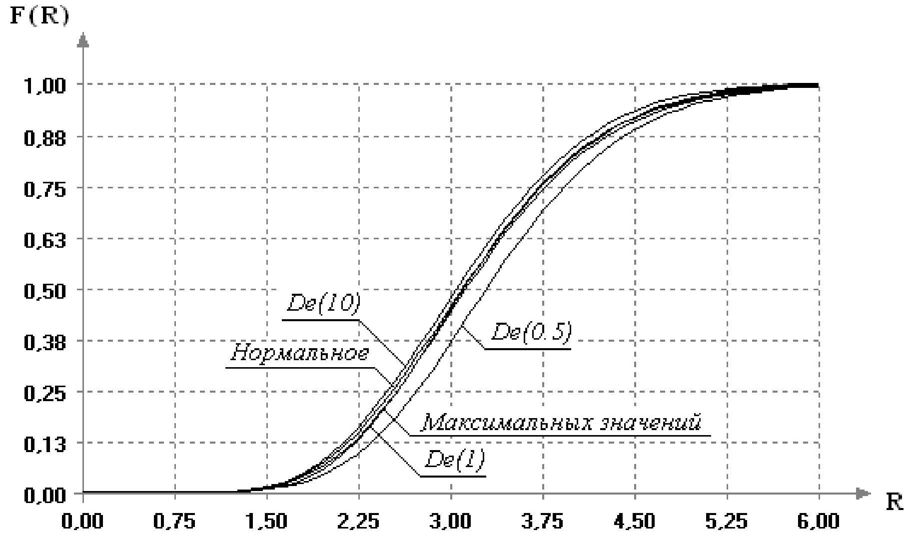


Рис. 3. Функции распределения статистики (2) в случае распределения $\{e_{ij}\}$ по законам распределения нормальному, Максимальных значений и $De(\lambda)$ с различными значениями параметра формы λ при I=10 и n=40

При большом числе уровней факторов наблюдается уже очень высокая степень неустойчивости к нарушению предположений о нормальности, как при малых, так и при больших n. Рисунок 3 иллюстрирует ситуацию, наблюдающуюся при I=10 и n=40.

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 естественнонаучного характера. – Материалы региональной научной конференции (с участием иностранных ученых) «Вероятностные идеи в науке и философии»: Новосибирск, 23-25 сентября 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: Robustness and Computer Intensive Methods”, September 6-10, 2004, Minsk. Vol. 1. – P. 110-113.
4. Лемешко Б.Ю., Постовалов С.Н. О распределениях статистик непараметрических критериев согласия при оценивании по выборкам параметров наблюдаемых законов // Заводская лаборатория. 1998. Т. 64. № 3. С. 61-72.

5. Лемешко Б.Ю., Постовалов С.Н. Применение непараметрических критериев согласия при проверке сложных гипотез // Автометрия. 2001. № 2. С. 88-102.