ДОКЛАДЫ АН ВШ РФ

2005 июль – декабрь № 2 (5)

ЕСТЕСТВЕННЫЕ НАУКИ

УДК 519.2

ПРОВЕРКА ГИПОТЕЗ В МОДЕЛЯХ ДИСПЕРСИОННОГО АНАЛИЗА СО СЛУЧАЙНЫМИ ФАКТОРАМИ ПРИ НАРУШЕНИИ ПРЕДПОЛОЖЕНИЙ О НОРМАЛЬНОСТИ[1]

Член–корреспондент СО АН ВШ **Б.Ю. Лемешко,**
В.М. Пономаренко

Методами статистического моделирования исследуются распределения статистик дисперсионного анализа при нарушении предположений о нормальных законах ошибок измерений и случайного фактора. Исследуется мощность критерия в данных условиях.

Введение

В основе множества критериев проверки статистических гипотез, применяемых в задачах статистического анализа данных, лежит предположение о принадлежности наблюдаемых величин или ошибок измерений нормальному закону. К сожалению, далеко не всегда в реальных задачах удается убедиться в том, что мы имеем дело с нормальными случайными величинами. Более того, реальный закон может существенно отличаться от нормального.

В таких ситуациях встает вопрос о правомерности применения классического аппарата статистического анализа или о необходимости исследования поведения статистического критерия и построения распределения статистики критерия в этих конкретных условиях (при конкретных законах случайных величин и ошибок измерений, учитываемых статистикой).

Практика показывает, что применение одних статистических критериев остается корректным и тогда, когда предположения о нормальности серьезно нарушаются. Примером могут служить критерии, связанные с проверкой гипотез о средних [1, 2]. Применение других оказывается невозможным даже при незначительных отклонениях от нормальности. Например, все критерии, предназначенные для проверки гипотез о дисперсиях, чувствительны к нарушению предположений о нормальности [3, 4-7].

Классический дисперсионный анализ [3, 8], широко используемый в различных приложениях, также опирается на предположения о нормальности ошибок наблюдений и эффектов уровней случайных факторов. Причем подчеркивается [3], что используемые критерии весьма требова-

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

тельны к выполнению этого предположения. Следовательно, для того, чтобы быть уверенным в выводах и результатах, получаемых на основании дисперсионного анализа, необходимо убедиться в том, что случайные величины, фигурирующие в модели, подчиняются нормальным законам (что у нас нет оснований, для отклонения гипотез о нормальности).

А что делать, если на основании результатов измерений (на выборках случайных величин) мы убедились в том, что закон или законы распределения существенно отличаются от нормального? Как правило, если данных хватило для того, чтобы с уверенностью отклонить гипотезу о нормальности, можно подобрать модель закона, наилучшим образом описывающую наблюдаемую случайную величину. Однако тогда, если речь идет о возможности применения дисперсионного анализа, возникает вопрос о том, насколько сильно изменилось распределение статистики соответствующего критерия, или о построении распределения статистики критерия в этих реальных условиях. Если мы определим характер изменения, а лучше, построим распределение статистики критерия, то сможем корректно применять соответствующие критерии при данных законах распределения случайных величин, входящих в модель дисперсионного анализа.

Очевидно, что для решения всего множества задач чисто аналитическими методами не хватает интеллектуальных ресурсов. Накопленный опыт показывает, что для исследования статистических закономерностей наиболее эффективно применение компьютерных технологий моделирования [9, 10]. К тому же, на распределения статистик различных критериев в отличной от классической ситуациях, как правило, влияет множество факторов, в том числе, виды законов распределения случайных величин (ошибок измерений), методы оценивания параметров [11–17] и т.д. Все это увеличивает число вариантов требующих разрешения задач.

Цель данной работы состояла в исследовании методами статистического моделирования распределений статистик однофакторного дисперсионного анализа в ситуациях, когда законы распределения ошибок наблюдений и случайного фактора отличаются от нормального. Построение моделей распределений статистик при различных законах распределения ошибок и уровней случайного фактора расширяет аппарат и область применения дисперсионного анализа.

Постановка задачи и условия проведения экспериментов

Однофакторная сбалансированная модель дисперсионного анализа имеет вид

$$y_{ij} = \mu + a_i + e_{ij}\,, \tag{1}$$

где I - число уровней фактора A, выбранных случайным образом, $i = 1,\ldots,I$; J - число опытов на каждом уровне, $j = 1,\ldots,J$; μ - среднее значение эффекта фактора A; $\{a_i\}$, $\{e_{ij}\}$ в совокупности независимы и имеют нулевые средние. В случае независимости $\{a_i\}$ и $\{e_{ij}\}$ дисперсия отклика $\sigma_Y^2 = \sigma_A^2 + \sigma_e^2$, где σ_A^2 - дисперсия эффектов уровней $\{a_i\}$ случайного фактора A, σ_e^2 - дисперсия ошибок наблюдений $\{e_{ij}\}$.

В классической постановке предполагается нормальность эффектов уровней фактора и ошибок наблюдений:

$$\Omega : \begin{cases} \{a_i\} \ \textit{распределены} \ \mathrm{N}(0, \sigma_{\mathrm{A}}^2), \\ \{e_{ij}\} \ \textit{распределены} \ \mathrm{N}(0, \sigma_{\mathrm{e}}^2). \end{cases} \tag{2}$$

В самом общем случае в модели компонент дисперсии относительно некоторого случайного фактора A проверяется гипотеза вида

$$H_A : \sigma_A^2 \le \theta_0 \sigma_e^2, \quad \theta_0 \ge 0, \tag{3}$$

где θ_0 - некоторая заданная константа.

Наиболее часто гипотеза (3) проверяется при $\theta_0 = 0$:

$$H_A : \sigma_A^2 = 0. \tag{4}$$

Если в результате применения соответствующего критерия проверяемая гипотеза (4) не отклоняется, то это свидетельствует о незначимом влиянии эффектов уровней фактора A на отклик y.

Аналогичный результат при $\theta_0 = 1$ говорит о том, что влияние эффектов уровней фактора A на отклик y не превышает влияния на него имеющихся ошибок наблюдений.

Другие рассматриваемые в данной работе ситуации ($\theta_0 = 0.5$ и $\theta_0 = 2$), по-видимому, могут быть проинтерпретированы по аналогии со случаем для $\theta_0 = 1$.

Для проверки гипотезы (3) используется статистика вида [3]

$$S = \frac{1}{(1 + J\theta_0)} \frac{\overline{SS}_A}{\overline{SS}_e} = \frac{1}{(1 + J\theta_0)} \frac{SS_A / I - 1}{SS_e / I(J-1)}, \tag{5}$$

где

$$SS_A = J \sum_{i=1}^{I} (y_{i\bullet} - y_{\bullet\bullet})^2, \tag{6}$$

$$SS_e = \sum_{i=1}^{I} \sum_{j=1}^{J} (y_{ij} - y_{i\bullet})^2, \tag{7}$$

$$y_{i\bullet} = \frac{1}{J}\sum_{j=1}^{J} y_{ij}\ ,\ y_{\bullet\bullet} = \frac{1}{I}\sum_{i=1}^{I} y_{i\bullet}\ .$$

При выполнении условий (2) и справедливости гипотезы вида

$$H_A:\ \sigma_A^2 = \theta_0\sigma_e^2 \tag{8}$$

статистика (5) подчиняется F-распределению Фишера с числом степеней свободы $I-1$ и $I(J-1)$ [3].

Используемые при вычислении статистики (5) суммы квадратов SS_A (6) и SS_e (7) при выполнении (2) подчиняются распределениям χ^2_{I-1} и $\chi^2_{I(J-1)}$ соответственно. Исследование распределений статистик SS_A и SS_e при нарушении (2) может помочь в понимании поведения распределения статистики (5) в этих условиях.

Распределения статистики (5) исследовались методами статистического моделирования при различных значениях I и J. Однако в данной работе результаты иллюстрируются на частном случае при значениях $I=5$, $J=6$ и $n=IJ=30$. Объем моделируемых выборок исследуемых статистик и оценок составлял $N=10000$.

Дисперсия ошибок наблюдения σ_e^2 (без потери общности) задавалась равной единице, дисперсия случайных эффектов уровней фактора A определялась в соответствии с условиями проверяемой гипотезы (8): $\sigma_A^2 = \theta_0\sigma_e^2$. В рамках рассматриваемой модели (1) случайные эффекты $\{e_{ij}\}$ и $\{a_i\}$ предполагались независимыми. Поэтому ошибки $\{e_{ij}\}$ и эффекты $\{a_i\}$ моделировались независимо друг от друга. Выборочные значения наблюдений y_{ij} формировались в соответствии с видом модели (1).

Исследования распределений статистик проводились при различных законах ошибок наблюдений и случайного фактора модели (1). В данной работе приводятся результаты исследований, когда ошибки $\{e_{ij}\}$ и эффекты $\{a_i\}$ подчинялись законам: нормальному, с функцией плотности

$$f(x,\theta_1,\theta_2) = \frac{1}{\theta_2\sqrt{2\pi}}\exp\left(-\frac{(x-\theta_1)^2}{2\theta_2^2}\right);$$

распределению максимальных значений –

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

семейству симметричных распределений с плотностью

$$De(\lambda) = f(x, \theta_1, \theta_2, \lambda) = \frac{\lambda}{2\sqrt{2}\theta_2 \Gamma(1/\lambda)} \exp\left(-\left(\frac{|x - \theta_1|}{\sqrt{2}\theta_2} \right)^{\lambda} \right) \qquad (9)$$

при различных значениях параметра формы λ. Распределения семейства (9) далее обозначаются как $De(\lambda)$. В работе приводятся результаты в основном при значениях параметра формы λ, равном 1 (соответствует распределению Лапласа) и равном 10.

Использование распределений из семейства $De(\lambda)$ при моделировании $\{e_{ij}\}$ и $\{a_i\}$ позволяет широко варьировать симметричные законы распределения, изменяя значения параметра λ: чем меньше параметр λ, тем "тяжелее" хвосты распределения $De(\lambda)$, чем больше параметр λ, тем хвосты "легче". Нормальное распределение является частным случаем распределения $De(\lambda)$ при $\lambda = 2$.

При моделировании значения параметров масштаба θ_2 и сдвига θ_1 соответствующего закона задавались в соответствии с требуемыми значениями математического ожидания M и дисперсии σ^2 моделируемых случайных величин.

Распределение максимальных значений использовалось для выявления влияния на распределения статистик асимметричности закона.

Исследование распределений статистики SS_A и SS_e

Иллюстрацией работоспособности методики моделирования является совпадение (близость) смоделированных эмпирических распределений статистик известным теоретическим, когда ситуация соответствует известной классической: ошибки наблюдений и эффекты уровней фактора подчиняются нормальному закону.

Количественной мерой близости эмпирических распределений суммы квадратов SS_A соответствующим теоретическим распределениям этой статистики при проверке согласия служат достигнутые уровни значимости $P\{S > S^*\}$ для применяемых критериев согласия, где S^* - значение статистики S используемого критерия, вычисленное по конкретной выборке исследуемых статистик. Проверка, осуществляемая по критериям χ^2 Пирсона, Колмогорова, ω^2 Крамера-Мизеса-Смирнова, Ω^2 Андерсона-Дарлинга [16, 17], показала очень высокую степень близости эмпирических распределений статистик к теоретическим.

Рис. 1 иллюстрирует полученные в результате моделирования распределения статистики SS_A при различных законах распределения эффектов $\{a_i\}$ и $\{e_{ij}\}$. Здесь же для сравнения приведено теоретическое χ^2_4-

распределение этой статистики, которое имеет место при выполнении предположений о нормальности (2).

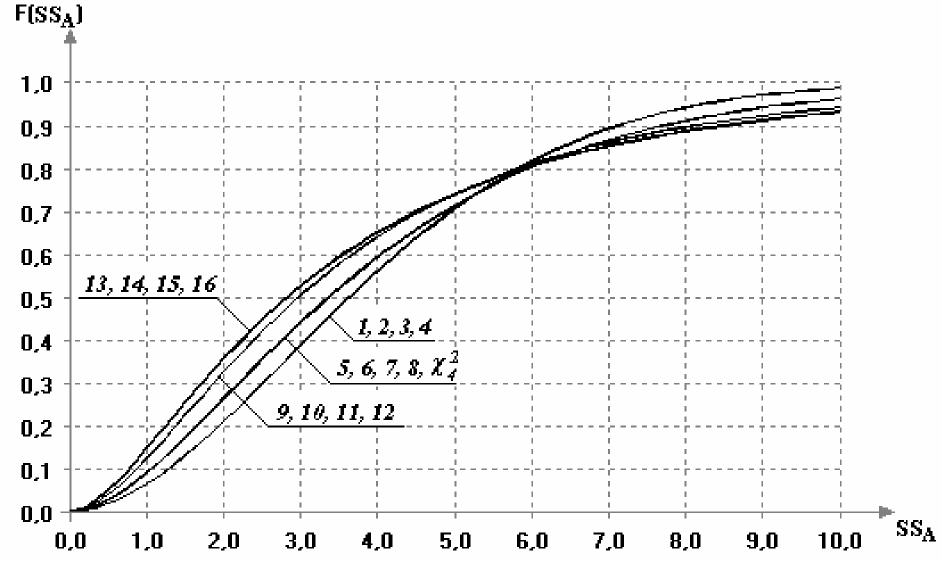

Рис. 1. Распределения статистики SS_A вида (6) при справедливости гипотезы H_A вида (8) и значении параметра $\theta_0 = 1$: 1 – *F(SS$_A$(De(10), De(10)))*, 2 – *F(SS$_A$(Norm, De(10)))*, 3 – *F(SS$_A$(Max, De(10)))*, 4 – *F(SS$_A$(De(1), De(10)))*, 5 – *F(SS$_A$(De(10), Norm))*, 6 – *F(SS$_A$(Norm, Norm))*, 7 – *F(SS$_A$(Max, Norm))*, 8 – *F(SS$_A$(De(1), Norm))*, 9 – *F(SS$_A$(De(10), Max))*, 10 – *F(SS$_A$(Norm, Max))*, 11 – *F(SS$_A$(Max, Max))*, 12 – *F(SS$_A$(De(1), Max))*,13 – *F(SS$_A$(De(10), De(1)))*, 14 – *F(SS$_A$(Norm, De(1)))*, 15 – *F(SS$_A$(Max, De(1)))*, 16 – *F(SS$_A$(De(1), De(1)))*

Обозначения в подрисуночных надписях указывают на законы распределения ошибок и уровней случайного фактора, при которых строились распределения статистики SS_A. Например, обозначение $F(SS_A(Norm, De(10)))$ соответствует распределению статистики SS_A, построенному при справедливости (8) для случая, когда ошибки $\{e_{ij}\}$ подчиняются нормальному закону, а эффекты уровней фактора $\{a_i\}$ – распределению с плотностью (9) и параметром формы, равным 10 (распределению $De(10)$). Как показали исследования, существенное влияние на распределение статистики SS_A оказывает закон распределения эффектов уровней фактора $\{a_i\}$. В то же время, закон распределения ошибок наблюдений $\{e_{ij}\}$ на распределение статистики SS_A значимого влияния не оказывает.

Нечувствительность распределения SS_A к отклонению распределения ошибок наблюдений от нормального закона кроется в структуре рассматриваемой суммы квадратов. Статистика SS_A имеет вид (6) и, учитывая вид модели (1), SS_A можно записать как

$$SS_A = J\sum_{i=1}^{I}(a_i - a_{\bullet} + e_{i\bullet} - e_{\bullet\bullet})^2 , \qquad (10)$$

где $a_{\bullet} = \frac{1}{J}\sum_{j=1}^{J} a_j$, $e_{i\bullet} = \frac{1}{J}\sum_{j=1}^{J} e_{ij}$, $e_{\bullet\bullet} = \frac{1}{I}\sum_{i=1}^{I} e_{i\bullet}$. Из (10) видно, что в SS_A учитывается только межгрупповое рассеяние ошибок наблюдений, где под группой понимается совокупность наблюдений на различных уровнях фактора A .

Распределение случайной величины $\{e_{i\bullet} - e_{\bullet\bullet}\}$, по-видимому, очень быстро стремится к нормальному при любом законе распределения ошибок $\{e_{ij}\}$. Вследствие этого вид закона распределения ошибок $\{e_{ij}\}$ не оказывает значимого влияния на распределение статистики SS_A .

По аналогии с SS_A, сумма квадратов SS_e при подстановке в (7) выражения (1) может быть представлена в виде

$$SS_e = \sum_{i=1}^{I}\sum_{j=1}^{J}(e_{ij} - e_{i\bullet})^2 .$$

Отсюда следует, что распределение SS_e не зависит от распределения эффектов $\{a_i\}$. Это же подтверждают результаты моделирования распределений статистики. Распределения статистики SS_e были промоделированы при ошибках наблюдений $\{e_{ij}\}$, принадлежащих нормальному закону, но при различных законах распределения $\{a_i\}$ эффектов уровней фактора. Близость полученных эмпирических распределений статистик теоретическому χ^2_{25}-распределению этой статистики, имеющему место при выполнении предположений о нормальности (2), была проверена по совокупности критериев согласия. Во всех случаях были достигнуты высокие уровни значимости, подтверждающие близость эмпирических распределений теоретическому как при выполнении предположений (2), так и в ситуациях, когда распределение $\{a_i\}$ существенно отклоняется от нормального закона.

Но если распределение статистики SS_e не зависит от закона распределения эффектов уровней фактора $\{a_i\}$, оно существенно зависит от вида распределения ошибок $\{e_{ij}\}$. Рис. 2 иллюстрирует поведение распределения статистики SS_e (7) в зависимости от закона распределения $\{e_{ij}\}$ в случае принадлежности $\{a_i\}$ нормальному закону. При отклонении распределения $\{e_{ij}\}$ от нормального закона происходит значимое отклонение распределения статистики SS_e от χ^2_{25}-распределения. Причем характер отклонения распределения статистики SS_e относительно χ^2_{25}-распределения такой же, как и характер отклонения распределения статистики SS_A относительно χ^2_4-распределения.

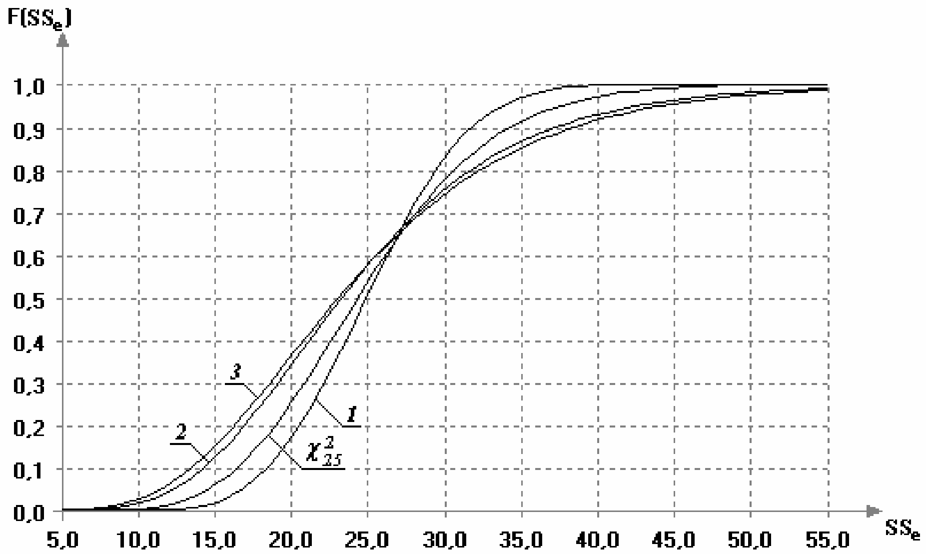

Рис. 2. Распределения SS_e вида (7): 1 – *F(SSe(De(10), Norm))*, 2 – *F(SSe(Max, Norm))*, 3 – *F(SSe(De(1), Norm))*

Распределения статистики (5) при нарушении предположений о нормальности

Для контроля точности моделирования распределения статистики (5) в первую очередь исследовались на классической ситуации при выполнении предположений (2). Результаты статистического моделирования показали, что при выполнении предположений о нормальности (2) и справедливости (8) получаемые эмпирические распределения статистики (5) очень хорошо согласуются с соответствующими теоретическими F-распределениями Фишера при различных значениях параметра θ_0.

При $\theta_0 = 0$ (при отсутствии влияния эффектов уровней факторов) исследовались изменения распределения статистики (5) в зависимости только от распределения ошибок наблюдений $\{e_{ij}\}$. Распределения статистики (5) при проверке гипотезы (4) были промоделированы при различных законах распределения ошибок наблюдений $\{e_{ij}\}$ при значениях $I = 5$ и $J = 6$. Полученные эмпирические распределения сравнивались с $F_{4,25}$-распределением Фишера, которому должно подчиняться распределение статистики в классическом случае

Результаты исследований показали, что даже в случае асимметричности законов распределения ошибок или распределений с “тяжелыми хвостами” (при распределении экстремального значения и при распределении

Лапласа) степень отклонения распределений статистики (5) от классического $F_{4,25}$-распределения Фишера при проверке гипотез вида (4) (при $\theta_0 = 0$) несущественна. Таким образом, распределения статистики (5) при проверке гипотез вида (4) при отсутствии влияния уровней фактора практически не зависят от вида закона распределения ошибок наблюдении $\{e_{ij}\}$. Следовательно, при проверке таких гипотез можно опираться на соответствующее F-распределение Фишера без опасения совершить существенную ошибку при определении достигнутого уровня значимости вследствие значимого отклонения закона распределения ошибок $\{e_{ij}\}$ от нормального.

Исследования распределений статистики (5) при наличии влияния уровней фактора (при $\theta_0 > 0$) показали, что в случае нарушения предположений (2) распределения статистики зависят: от закона распределения ошибок измерений $\{e_{ij}\}$; от распределения уровней случайного фактора $\{a_i\}$; от значения параметра θ_0. Например, рис. 3 иллюстрируют ситуацию, когда эффекты уровней случайного фактора и ошибки наблюдений подчиняются одному и тому же закону.

Результаты исследований при различных законах распределений $\{a_i\}$ и $\{e_{ij}\}$ дают основание говорить о том, что на распределение статистики (5) в большей степени влияет вид распределения $\{a_i\}$ и в несколько меньшей – вид распределения $\{e_{ij}\}$. В то же время, следует отметить, что изменение значения параметра θ_0 не приводит к радикальному изменению распределения статистики (5): оно заметно при малых значениях θ_0. С ростом θ_0 распределение статистики быстро сходится к некоторому предельному.

На основании результатов моделирования были построены модели распределений статистики (5) при различных значениях I и J. В таблице 1 приведены модели распределений статистики (5) в случае $I = 5$, $J = 6$, построенные при различных сочетаниях законов, которым подчиняются ошибки наблюдений и эффекты уровней случайных факторов. Модели строились при значениях параметра θ_0, равных 1, 0.5 и 2, соответственно.

Наилучшей моделью для построенных в результате статистического моделирования эмпирических распределений статистики наиболее часто оказывались законы из семейств бета-распределений второго и третьего рода с плотностями

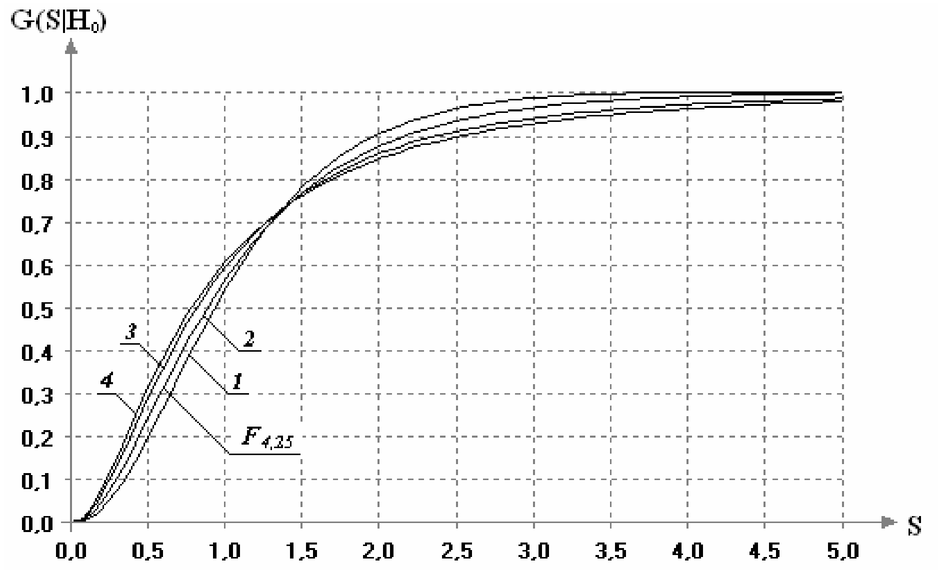

Рис. 3. Распределения статистики (5) при справедливости гипотезы H_A вида (8) и значении параметра $\theta_0 = 1$: 1 – *G(S(De(10),De(10)))*, 2 – *G(S(Norm,Norm))*, 3 – *G(S(Max, Max))*, 4 – *G(S(De(1),De(1)))*

$$BeII(x;\alpha,\beta,\theta_1,\theta_2) = \frac{z^{\alpha-1}}{\theta_2 B(\alpha,\beta)(1+z)^{\alpha+\beta}}$$

и

$$BeIII(x;\alpha,\beta,\theta_1,\theta_2) = \frac{\delta^\alpha z^{\alpha-1}(1-z)^{\beta-1}}{\theta_2 B(\alpha,\beta)[1+(\delta-1)z]^{\alpha+\beta}},$$

где α, β, δ – параметры формы распределений, $z = \dfrac{x-\theta_1}{\theta_2}$, θ_1 – параметр сдвига, θ_2 – параметр масштаба. Заметим, что распределение Фишера является частным случаем бета-распределения второго рода [18]: $F_{k,n} = BeII(\frac{k}{2},\frac{n}{2},\frac{n}{2},0)$. В некоторых случаях наилучшей моделью распределения статистики оказывалось Γ-распределение с плотностью

$$\Gamma(x;\alpha,\beta,\theta_1,\theta_2) = \frac{\alpha z^{\alpha\beta-1}}{\theta_2\Gamma(\beta)}\exp\{-z^\alpha\} \text{ и } z = \frac{x-\theta_1}{\theta_2}.$$

Таблица 1

Модели распределения статистики (5) при различных законах распределения ошибок наблюдений и эффектов уровней случайных факторов, при значении параметра $\theta_0 = 0.5, 1, 2$ **(**$I = 5, J = 6$**)**

θ_0	Распределение $\{a_i\}$	Распределение $\{e_{ij}\}$				
		De(1)	Max	Norm	De(5)	De(10)
0.5	De(1)	BeII(1.918,4.008,1.823,0)	BeII(1.902,4.259,1.961,0)	BeII(1.899,4.83,2.196,0)	BeIII(2.02,2.85,11.5,13.7,0)	BeII(1.897,5.196,2.341,0)
	Max	BeII(1.991,4.628,2.115,0)	BeII(1.991,4.871,2.217,0)	BeII(1.993,5.42,2.422,0)	BeII(1.997,5.872,2.596,0)	BeII(2.024,5.927,2.566,0)
	Norm	BeII(1.972,8.944,4.687,0)	BeII(1.979,10.295,5.369,0)	BeII(2,12.5,6.25,0)	BeII(2.006,15.857,7.861,0)	BeII(2.021,15.059,7.359,0)
	De(5)	BeII(2.058,16.029,8.527,0)	BeII(2.069,19.735,10.35,0)	BeII(2.164,25.637,12.442,0)	BeII(2.19,33.144,15.612,0)	BeII(2.185,38.433,18.21,0)
	De(10)	BeII(2.123,17.915,9.277,0)	BeII(2.12,24.954,12.918,0)	BeII(2.215,32.47,15.512,0)	Г(2.401,0.933,0.413,0.002)	BeIII(2.32,5.24,3.34,8.21,0)
1	De(1)	BeII(1.809,3.569,1.682,0)	BeII(1.785,3.835,1.832,0)	BeII(1.79,4.359,2.035,0)	BeII(1.801,4.639,2.138,0)	BeII(1.803,4.657,2.132,0)
	Max	BeII(1.957,3.995,1.805,0)	BeII(1.973,4.254,1.894,0)	BeII(2.005,4.716,2.016,0)	BeII(2.001,5.119,2.178,0)	BeII(1.994,5.156,2.193,0)
	Norm	BeII(1.975,8.117,4.24,0)	BeII(1.992,8.99,4.581,0)	BeII(2,12.5,6.25,0)	BeII(2.007,17.484,8.713,0)	BeII(2.008,17.914,8.89,0)
	De(5)	BeII(2.124,15.235,7.91,0)	BeII(2.09,24.459,13.001,0)	Г(2.32,0.947,0.446,0)	Г(2.167,1.012,0.498,0)	Г(2.042,1.046,0.545,0.001)
	De(10)	BeII(2.167,18.99,9.79,0)	BeII(2.175,35.006,18.09,0)	Г(2.297,0.984,0.469,-0.001)	Г(1.957,1.098,0.587,0.001)	Г(1.925,1.107,0.594,0.002)
2	De(1)	BeII(1.71,3.431,1.69,0)	BeII(1.692,3.652,1.811,0)	BeII(1.729,4.058,1.927,0)	BeIII(1.86,2.6,14.15,15.4,0)	BeII(1.727,4.379,2.055,0)
	Max	BeII(1.911,3.82,1.752,0)	BeII(1.935,3.888,1.719,0)	BeII(1.97,4.403,1.875,0)	BeII(1.982,4.733,1.995,0)	BeII(1.987,4.881,2.046,0)
	Norm	BeII(1.978,7.448,3.843,0)	BeII(1.953,9.376,4.92,0)	BeII(2,12.5,6.25,0)	BeII(2.006,15.857,7.861,0)	BeII(2.0421,18.513,8.96,0)
	De(5)	BeII(2.058,16.029,8.53,0)	BeII(2.1,22.494,11.833,0)	Г(2.242,0.985,0.482,-0.002)	Г(1.861,1.128,0.625,0.002)	BeII (2.185,38.43,18.21,0)
	De(10)	BeII(2.211,20.08,10.29,0)	BeII(2.191,27.355,13.85,0)	Г(1.947,1.103,0.607,0)	Г (1.716,1.212,0.706, 0.002)	Г(1.628,1.272,0.753,0)

Исследование мощности критерия со статистикой (5)

Распределение статистики (5) при нарушении предположений о нормальности, как правило, претерпевает существенные изменения. Это приводит и к изменению мощности критерия. В данной работе в рамках модели (1) исследовалась мощность критерия со статистикой (5) при проверке гипотезы вида $H_A: \ \sigma_A^2 = \sigma_e^2$. Конкурирующая гипотеза (альтернатива) задавалась в виде $H_1: \ \sigma_A^2 = C\sigma_e^2$, $C > 1$.

Таблица 2

Мощность критерия со статистикой (5) при различных альтернативах и выполнении предположений нормальности

α	Значение C			
	1.05	1.2	1.5	2
0.1	0.111403	0.148523	0.223937	0.347356
0.05	0.056803	0.080752	0.136446	0.23737

Таблица 3

Мощность критерия, со статистикой (5) при различных законах распределения $\{e_{ij}\}$ и $\{a_i\}$ при $\alpha = 0.05$ и $C = 1.2$

Распределение $\{a_i\}$	Распределение $\{e_{ij}\}$			
	De(1)	Максимального значения	Нормальное	De(10)
De(1)	0.072044	0.071396	0.073892	0.075679
Максимального значения	0.071229	0.076016	0.074033	0.072781
Нормальное	0.078336	0.086783	0.080752	0.083217
De(10)	0.084805	0.090454	0.087468	0.090868

Таблица 4

Мощность критерия, со статистикой (5) при различных законах распределения $\{e_{ij}\}$ и $\{a_i\}$ при $\alpha = 0.1$ и $C = 1.2$

Распределение $\{a_i\}$	Распределение $\{e_{ij}\}$			
	De(1)	Максимального значения	Нормальное	De(10)
De(1)	0.134456	0.134081	0.136636	0.139667
Максимального значения	0.135248	0.141331	0.138363	0.1376
Нормальное	0.144109	0.155027	0.148523	0.151361
De(10)	0.152905	0.161304	0.157872	0.161344

В таблице 2 представлены полученные значения мощности исследуемого критерия при выполнении предположений нормальности (2) в случае различных альтернатив (различных значениях C) при значениях ошибки первого рода $\alpha = 0.1$ и $\alpha = 0.05$. В таблицах 3 и 4 представлены значения мощности рассматриваемого критерия при различных законах распределения $\{e_{ij}\}$ и $\{a_i\}$ при альтернативе заданного вида и значении $C = 1.2$. В таблице 3 приведены значения мощности при вероятности ошибки первого рода $\alpha = 0.05$, в таблице 4 - при $\alpha = 0.1$. Интересно, что при сохранении симметрии и уменьшении величины эксцесса распределений $\{e_{ij}\}$ или $\{a_i\}$ мощность критерия увеличивается.

Заключение

В работе на примере однофакторной сбалансированной модели дисперсионного анализа исследовано влияние на распределения статистик нарушения предположения о нормальности законов распределений ошибок измерений и уровней случайного фактора. Для различных сочетаний законов ошибок измерений и уровней случайного фактора (для различных комбинаций числа наблюдений и количества уровней фактора) построены модели распределений статистик. Построенные модели могут использоваться при решении прикладных задач в ситуациях, когда ошибки измерений или уровни факторов хорошо описываются соответствующими законами. Исследована мощность критерия в зависимости от законов распределения ошибок и случайных факторов.

ЛИТЕРАТУРА

[1] **Лемешко Б.Ю., Помадин С.С.** Проверка гипотез о математических ожиданиях и дисперсиях в задачах метрологии и контроля качества при вероятностных законах, отличающихся от нормального // Метрология. 2004. – № 3. – С.3-15.

[2] **Лемешко Б.Ю., Помадин С.С.** Корреляционный анализ наблюдений многомерных случайных величин при нарушении предположений о нормальности // Сибирский журнал индустриальной математики. 2002. - Т.5. – № 3. – С.115-130.

[3] **Шеффе Г.** Дисперсионный анализ. М.: Физматгиз, 1980. - 628 с.

[4] **Лемешко Б.Ю., Ванюкевич О.Н.** Проверка гипотез о дисперсии при нарушении предположений о нормальности // Сб. научных трудов НГТУ. – 2002. – № 3(29). – С.27-32.

[5] **Лемешко Б.Ю. Пономаренко В.М.** Проблемы применения классического аппарата дисперсионного анализа в приложениях технического, экономического и естественнонаучного характера // Материалы региональной конференции "Вероятностные идеи в науке и философии". – Новосибирск: Ин-т философии и права СО РАН / НГУ. 2003. – С. 106-109.

[6] **Лемешко Б.Ю., Миркин Е.П.** Критерии Бартлетта и Кокрена в измерительных задачах при вероятностных законах, отличающихся от нормального // Измерительная техника. 2004. № 3. – С.10-16.

[7] **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.

[8] **Маркова Е.В., Денисов В.И., Полетаева И.А. Пономарев В.В.** Дисперсионный анализ и синтез планов на ЭВМ. М.: Наука, 1982. – 195 с.

[9] **Лемешко Б.Ю.** Компьютерные методы исследования статистических закономерностей // Информационные системы и технологии: ИСТ`2000: Сб. научн. ст. – Новосибирск. 2001. – С.26-41.

[10] **Лемешко Б.Ю., Постовалов С.Н.** Компьютерное моделирование как способ познания статистических закономерностей в технике, экономике, естествознании // Материалы региональной конференции "Вероятностные идеи в науке и философии". – Новосибирск: Ин-т философии и права СО РАН / НГУ. 2003. – С. 110-113.

[11] **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.

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

[13] **Денисов В.И., Лемешко Б.Ю., Постовалов С.Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа χ^2. Новосибирск: Изд-во НГТУ, 1998. – 126 с.

[14] **Лемешко Б.Ю., Постовалов С.Н.** Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. Новосибирск: Изд-во НГТУ, 1999. – 85 с.

[15] **Лемешко Б.Ю., Пономаренко В.М.** Влияние методов оценивания параметров на распределения статистик дисперсионного анализа при нарушении предположений о нормальности ошибок наблюдений // Тез. док-в МНТК "Информатика и проблемы телекоммуникаций". Новосибирск, 2003. – С.143-146.

[16] **Р 50.1.033-2001.** Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть I. Критерии типа хи-квадрат. – М.: Изд-во стандартов, 2002. – 87 с.

[17] **Р 50.1.037-2002.** Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. – М.: Изд-во стандартов,2002. – 64 с.

[18] **Губарев В.В.** Вероятностные модели: Справочник. В 2-х ч. – Новосибирск: Изд-во НЭТИ, 1992 г. – Т.1. – 198 с. – Т.2. – 188 с.

B.Yu. Lemeshko, V.M. Ponomarenko

HYPOTHESES TESTING IN VARIANCE ANALYSIS MODELS WITH RANDOM FACTORS IN CASE OF NORMALITY ASSUMPTIONS FAILURE

The distributions of variance analysis statistics have been investigated by statistical simulation methods. The case of non-normal distributions of error and random factor has been considered. The power of test has been investigated under these conditions.

Новосибирский государственный технический университет

Статья поступила 07 июля 2005 г.