О мощности и применении критериев однородности дисперсий в нестандартных условиях

Б. Ю. Лемешко[1]
Новосибирский государственный технический университет

Исследованы свойства параметрических и непараметрических критериев, используемых при проверке гипотез об однородности дисперсий. Проведен сравнительный анализ мощности критериев. Рассмотрено применение параметрических критериев в условиях нарушения стандартных предположений.

Ключевые слова: критерий однородности дисперсий, мощность критерия, статистика, статистическое моделирование

1. Введение

Критерии проверки гипотез об однородности дисперсий при анализе результатов измерений применяются в различных приложениях, когда пытаются убедиться в неизменности (или наоборот в изменении) характеристик рассеяния, соответствующих анализируемых выборок.

В критериях проверки однородности дисперсий проверяемая гипотеза о постоянстве дисперсий k выборок имеет вид

$$H_0\colon \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2,$$

а конкурирующая с ней гипотеза

$$H_1 : \sigma_{i_1}^2 \neq \sigma_{i_2}^2,$$

где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2.

Для проверки такого рода гипотез может использоваться значительный перечень классических параметрических критериев. Обоснованное применение этих критериев, гарантирующее корректность выводов, требует выполнения стандартного предположения о принадлежности анализируемых выборок случайных величин нормальному закону. Основная масса параметрических критериев чрезвычайно чувствительна к нарушению предположения о нормальности (изменяются распределения статистик при справедливости H_0). Очевидно, что в приложениях это предположение зачастую не выполняется, что не позволяет опереться на классические результаты о предельных распределениях статистик критериев или на существующие таблицы критических значений (процентных точек).

В случае непараметрических аналогов речи о принадлежности выборок нормальным законам не идёт. Однако предполагается, что анализируемые выборки принадлежат пусть неизвестному, но одному и тому же виду закона с одинаковыми математическими ожиданиями. Именно тогда обеспечивается корректность применения непараметрических критериев. Очевидно, что и это более слабое предположение на практике может не выполняться. Следует

[1] Исследования выполнены при поддержке Министерства образования и науки РФ в рамках государственной работы «Обеспечение проведения научных исследований» и проектной части государственного задания (проект № 1.1009.2017/ПЧ)

упомянуть и то, что статистики непараметрических критериев представляют собой дискретные величины и распределения некоторых статистик при ограниченных объёмах выборок существенно отличаются от их непрерывных асимптотических (предельных) распределений. А это сказывается на корректности выводов на практике.

Одной из основных характеристик статистического критерия является его мощность относительно заданной конкурирующей гипотезы H_1, которая представляет собой разность $1-\beta$, где β – вероятность ошибки 2-го рода (принять гипотезу H_0 при справедливости H_1) при заданной вероятности α ошибки 1-го рода (отклонить H_0 при её справедливости).

Относительно одной и той же конкурирующей гипотезы H_1 мощности существующих параметрических критериев значительно различаются. При этом мощность лучших параметрических критериев, как правило, заметно превосходит мощность непараметрических (в том числе, в условиях нарушения стандартного предположения о нормальности).

Множество имеющихся критериев ставит перед исследователем нетривиальную задачу выбора наиболее мощного критерия однородности дисперсий и обеспечения корректности вывода с его использованием в условиях конкретного приложения, когда стандартные предпосылки применения критерия могут не выполняться.

Данная работа посвящена сравнительному анализу мощности критериев однородности дисперсий и корректности применения имеющегося множества критериев в условиях нарушения стандартных предположений. Работа является продолжением исследований [1-10], опирается на развиваемый подход [11], предполагающий интенсивное использование статистического моделирования при исследовании вероятностных закономерностей.

2. Параметрические критерии однородности дисперсий

В данном разделе кратко описаны статистики ряда исследованных параметрических критериев. Более подробное описание критериев с указанием положительных и отрицательных моментов, характеризующих достоинства и недостатки критериев, результаты исследований сходимости распределений статистик к предельным, таблицы критических значений и рекомендации по применению приведены в [12]. Там же приводятся ссылки на первоисточники, связанные с построением соответствующих критериев.

Статистика **критерия Бартлетта** (Б) вычисляется в соответствии с соотношением

$$\chi^2 = M\left[1+\frac{1}{3(k-1)}\left(\sum_{i=1}^{k}\frac{1}{\nu_i}-\frac{1}{N}\right)\right]^{-1},$$

где $M = N\ln\left(\frac{1}{N}\sum_{i=1}^{k}\nu_i S_i^2\right)-\sum_{i=1}^{k}\nu_i \ln S_i^2$; k – количество выборок; n_i – объемы выборок; $\nu_i = n_i$, если математическое ожидание известно, и $\nu_i = n_i - 1$, если не известно; $N=\sum_{i=1}^{k}\nu_i$; S_i^2 – оценки выборочных дисперсий. При неизвестном математическом ожидании оценки $S_i^2 = \frac{1}{n_i-1}\sum_{j=1}^{n_i}\left(X_{ji}-\overline{X}_i\right)^2$, $\overline{X}_i = \frac{1}{n_i}\sum_{j=1}^{n_i}X_{ji}$, где X_{ij} – j-е наблюдение в i-й выборке.

При справедливости гипотезы H_0 и принадлежности выборок нормальной генеральной совокупности статистика приближенно подчиняется χ^2_{k-1}-распределению.

Статистика **критерия Кокрена** (К) выражается формулой

$$Q = \frac{S_{\max}^2}{S_1^2+S_2^2+\cdots+S_k^2},$$

где $S_{\max}^2 = \max\left(S_1^2,\ S_2^2,...,\ S_k^2\right)$. Проверяемая гипотеза отклоняется при больших значениях статистики.

Статистика **критерия Хартли** (Х) имеет вид

$$F = \frac{S_{\max}^2}{S_{\min}^2},$$

где $S_{\max}^2 = \max\left(S_1^2,\ S_2^2,...,\ S_k^2\right)$; $S_{\min}^2 = \min\left(S_1^2,\ S_2^2,...,\ S_k^2\right)$.

Статистика **критерия Левене** (Л) задается выражением

$$W = \frac{N-k}{k-1}\sum_{i=1}^{k} n_i\left(\overline{Z}_{i\bullet} - \overline{Z}_{\bullet\bullet}\right)^2 \Big/ \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(Z_{ij} - \overline{Z}_{i\bullet}\right)^2,$$

где $N = \sum_{i=1}^{k} n_i$; X_{ij} – j-е наблюдение в i-й выборке; $Z_{ij} = \left|X_{ij} - \overline{X}_{i\bullet}\right|$, в котором $\overline{X}_{i\bullet}$ – среднее в i-й выборке; $\overline{Z}_{i\bullet}$ – среднее Z_{ij} по i-й выборке; $\overline{Z}_{\bullet\bullet}$ – среднее Z_{ij} по всем выборкам.

Критерий Фишера (Ф) со статистикой вида

$$F = \frac{s_1^2}{s_2^2},$$

используется для проверки гипотезы о равенстве дисперсий *двух* выборок. Критерий двусторонний.

Статистика правостороннего **критерия Неймана–Пирсона** (НП) определяется отношением

$$h = \frac{1}{k}\sum_{i=1}^{k} s_i^2 \Big/ \left(\prod_{i=1}^{k} s_i^2\right)^{\frac{1}{k}}.$$

Статистика **критерия О'Брайена** (ОБ) имеет вид

$$V = \frac{1}{k-1}\sum_{i=1}^{k} n_i\left(\overline{V_i} - \overline{\overline{V_i}}\right)^2 \Big/ \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(V_{ij} - \overline{V_i}\right)^2,$$

где $V_{ij} = \dfrac{(n_i - 1.5)n_i\left(X_{ij} - \bar{X}_i\right)^2 - 0.5 s_i^2 (n_i - 1)}{(n_i - 1)(n_i - 2)}$, $\bar{X}_i$ – среднее значение, s_i^2 – оценка дисперсии i-й выборки, $\overline{V_i} = \dfrac{1}{n_i}\sum_{j=1}^{n_i} V_{ij}$, $\overline{\overline{V_i}} = \dfrac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} V_{ij}$.

Двусторонний **критерий Линка** (ЛИ) является аналогом критерия Фишера со статистикой вида:

$$F^* = \frac{\omega_{n_1}}{\omega_{n_2}},$$

где $\omega_{n_1} = x_{1,\max} - x_{1,\min}$, $\omega_{n_1} = x_{2,\max} - x_{2,\min}$ – размахи, а $x_{1,\max}$, $x_{2,\max}$, $x_{1,\min}$, $x_{2,\min}$ – максимальные и минимальные элементы сравниваемых выборок.

Статистика **критерия Ньюмана** (Н) задаётся выражением:

$$q = \frac{\omega_{n_1}}{s_{n_2}},$$

где $\omega_{n_1} = x_{1,\max} - x_{1,\min}$, $s_{n_2} = \sqrt{\frac{1}{n_2 - 1}\sum_{i=1}^{n_2}(x_{2i} - \bar{x}_2)^2}$.

Статистика критерия **Блисса–Кокрена–Тьюки** (БКТ) представляет собой аналог статистики критерия Кокрена:

$$c = \max_{1\le i\le k}\omega_i \Big/ \sum_{i=1}^{k}\omega_i ,$$

где ω_i – размах i-й выборки.

Критерий Кадуэлла–Лесли–Брауна (КЛБ) предложен в качестве аналога критерия Хартли с заменой в статистике отношений оценок дисперсий на отношения размахов

$$K = \max_{1\le i\le k}\omega_i \Big/ \min_{1\le i\le k}\omega_i .$$

Статистика **Z–критерия Оверолла–Вудворда** (Z) имеет вид:

$$Z = \frac{1}{k-1}\sum_{i=1}^{k} Z_i^2 ,$$

где $Z_i = \sqrt{\frac{c_i\left(n_i - 1\right)s_i^2}{MSE}} - \sqrt{c_i\left(n_i - 1\right) - \frac{c_i}{2}}$, $MSE = \frac{1}{N-k}\sum_{i=1}^{k}\sum_{j=1}^{n_i}\left(X_{ij} - \bar{X}_i\right)^2$, $c_i = 2 + 1/n_i$, $N = \sum_{i=1}^{k} n_i$.

Модифицированный **Z–критерий Оверолла–Вудворда** (MZ) отличается вычислением величин c_i:

$$c_i = 2.0\left[\frac{1}{K_i}\left(2.9 + \frac{0.2}{n_i}\right)\right]^{\frac{1.6(n_i - 1.8K_i + 14.7)}{n_i}},$$

где $K_i = \frac{1}{n_i - 2}\sum_{j=1}^{n_i} G_{ij}^4$, $G_{ij} = \left(x_{ij} - \bar{x}_i\right)\Big/\sqrt{\frac{n_i - 1}{n_i}s_i^2}$.

3. Непараметрические критерии однородности характеристик рассеяния

Статистика **критерия Ансари–Бредли** (АБ) может быть вычислена следующим образом:

$$S = \sum_{i=1}^{n_1}\left\{\frac{n_1 + n_2 + 1}{2} - \left|R_i - \frac{n_1 + n_2 + 1}{2}\right|\right\},$$

где R_i – ранги элементов первой выборки в общем вариационном ряду. Как правило, используется нормированная статистика

$$S^* = (S - E[S]) / \sqrt{D[S]},$$

дискретное распределение которой достаточно хорошо приближается стандартным нормальным законом.

Статистика **критерия Муда** (М) имеет вид

$$M = \sum_{i=1}^{n_1}\left(R_i - \frac{n_1 + n_2 + 1}{2}\right)^2 ,$$

где R_i – ранги элементов первой выборки в общем вариационном ряду двух выборок. Чаще используется нормализованная статистика

$$M^* = \frac{\left(M - E[M] + \frac{1}{2}\right)}{\sqrt{D[M]}}.$$

Статистика **критерия Сижела–Тьюки** (СТ) строится по вариационному ряду объединенной выборки объёмом $n = n_1 + n_2$, для чего ряд преобразуется в последовательность вида

$$x_1, x_n, x_{n-1}, x_2, x_3, x_{n-2}, x_{n-3}, x_4, x_5, \ldots,$$

т. е. оставшийся ряд «переворачивается» каждый раз после приписывания рангов паре крайних значений. В качестве статистики критерия используется сумма рангов элементов первой выборки

$$R = \sum_{i=1}^{n_1} R_i ,$$

или её нормированный вид $R^* = (R - E[R]) / \sqrt{D[R]}$.

Статистика **критерия Клотца** (Кл) имеет вид

$$L = \sum_{i=1}^{n_1} u^2_{R_i/(n_1+n_2+1)} ,$$

где R_i – ранг i-го элемента первой выборки в общем упорядоченном по возрастанию ряду $(n_1 + n_2)$ значений объединенной выборки, u_γ – γ-квантиль стандартного нормального распределения. Обычно используется нормализованная статистика.

В **k-выборочном критерии Флайне–Киллина** (ФК) по исходным выборкам вычисляются абсолютные значения $z_{ji} = \left| X_{ji} - \tilde{X}_i \right|$, где $\tilde{X}_i$ – выборочная медиана i-й выборки, $i = \overline{1,k}$. Далее строится вариационный ряд объединённой выборки z_{ji}, $j = \overline{1,n_i}$, $i = \overline{1,k}$. Для элементов i-й выборки на основании рангов R_{ji} её элементов z_{ji} в объединённой выборке строятся метки

$$a_{n,R_{ji}} = \Phi^{-1}\left(\frac{1 + R_{ji} / (n+1)}{2} \right), \; j = \overline{1,n_i} ,$$

где $n = \sum_{i=1}^{k} n_i$, находятся $\overline{A}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} a_{n,R_{ji}}$. Статистика критерия имеет вид:

$$\chi_0^2 = \frac{\sum_{i=1}^{k} n_i \left(\overline{A}_i - \overline{a} \right)^2}{V^2},$$

где $\overline{a} = \frac{1}{n} \sum_{j=1}^{n} a_{n,j}$, $V^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(a_{n,j} - \overline{a} \right)^2$. Асимптотическим распределением этой статистики при справедливости H_0 и больших объёмах выборок является χ^2_{k-1}-распределение.

Следует отметить, что критерии, опирающиеся на различные метки, подобные критериям Клотца или Флайне–Киллина, где при выборе меток используются квантили нормального закона, свойствами непараметричности (а, следовательно, устойчивости) обладают в меньшей степени по сравнению с ранговыми критериями Муда, Ансари–Бредли, Сижела–Тьюки.

Более подробное описание статистик непараметрических критериев с указанием достоинств и недостатков конкретных критериев, сходимости распределений статистик к асимптотическим (непрерывным) можно найти в [12].

4. Сравнительный анализ мощности критериев

Исследования распределений статистик проводились при различных наблюдаемых законах, в частности в случае принадлежности моделируемых выборок обобщённому нормальному закону с плотностью

$$De(\theta_2) = f(x; \theta_0, \theta_1, \theta_2) = \frac{\theta_2}{2\theta_1 \Gamma(1/\theta_2)} \exp\left(-\left(\frac{|x-\theta_0|}{\theta_1} \right)^{\theta_2} \right)$$

при различных значениях параметра формы θ_2. Это семейство может быть хорошей моделью для законов распределения погрешностей различных измерительных систем. Распределение $\mathrm{De}(\theta_2)$ включает в качестве частных случаев распределение Лапласа $(\theta_2 = 1)$ и нормальное $(\theta_2 = 2)$. Это семейство позволяет задавать различные симметричные законы распределения, в той или иной мере отличающиеся от нормального: чем меньше значение параметра формы θ_2, тем «тяжелее» хвосты распределения $\mathrm{De}(\theta_2)$, чем больше параметр, тем хвосты «легче».

При сравнительном анализе мощности критериев, как правило, рассматривалась ситуация (конкурирующая гипотеза), когда $k-1$ выборка принадлежала закону с некоторым $\sigma = \sigma_0$, в то время как одна из выборок, например с номером k, имела некоторую отличную дисперсию. Проверяемой гипотезе соответствует ситуация H_0: $\sigma_1^2 = \sigma_2^2 = ... = \sigma_k^2 = \sigma_0^2$.

Результаты исследования мощности критериев показали, что в случае принадлежности анализируемых выборок нормальному закону при $k = 2$, все исследуемые критерии можно расположить по убыванию мощности следующим образом:

(Б ~ К ~ Х ~ Ф ~ НП ~ Z) $\succ$ (ОБ ~ MZ) $\succ$ Кл $\succ$ Л $\succ$ ФК $\succ$ М $\succ$ Н $\succ$ (АБ ~ СТ) $\succ$
$\succ$ (БКТ ~ КЛБ ~ ЛИ),

где знак «~» означает эквивалентность критериев по мощности, а «$\succ$» – преимущество в мощности. Критерии Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна и Линка очень существенно уступают в мощности всем остальным.

В случае принадлежности выборок законам с более лёгкими (по сравнению с нормальным законом) хвостами критерии упорядочиваются по мощности практически так же, как и при нормальном законе.

При (симметричных) законах с более тяжелыми хвостами по сравнению с нормальным законом порядок предпочтения меняется. В случае тяжёлых хвостов критерии располагаются в следующем порядке:

ФК $\succ$ Кл $\succ$ М $\succ$ Л $\succ$ (СТ ~ АБ) $\succ$ ОБ $\succ$ MZ $\succ$ (Б ~ К ~ Х ~ Ф ~ НП ~ Z) $\succ$ Н $\succ$
$\succ$ (БКТ ~ КЛБ ~ ЛИ).

С увеличением числа сравниваемых выборок ситуация меняется. Практически исчезают группы эквивалентных критериев. Исключение составляет лишь пара критериев Бартлетта и Неймана–Пирсона, которые в любой ситуации остаются эквивалентными по мощности. Многовыборочные критерии в случае законов с хвостами, более легкими чем у нормального закона, могут быть упорядочены по убыванию мощности практически так же как и при нормальном законе:

К $\succ$ ОБ $\succ$ Z $\succ$ MZ $\succ$ (Б ~ НП) $\succ$ Х $\succ$ Л $\succ$ ФК $\succ$ (БКТ ~ КЛБ).

В случае тяжёлых хвостов критерии по убыванию мощности располагаются в другом порядке:

ФК $\succ$ Л $\succ$ ОБ $\succ$ MZ $\succ$ (Б ~ НП) $\succ$ Z $\succ$ Х $\succ$ К $\succ$ КЛБ $\succ$ БКТ.

5. Применение критериев в «нестандартных» условиях

Как можно заметить, параметрические критерии (по крайней мере, лучшие их представители) имеют явное преимущество в мощности по сравнению с непараметрическими.

С другой стороны, в случае нарушения стандартного предположения о нормальности невозможно использовать классические результаты, связанные с их применением и полученные именно при данном предположении. Исключение составляют группа устойчивых критериев (О'Брайена, Левене и модифицированный Z-критерий Оверолла–Вудворда), но и в этом случае зависимость от вида закона, которым принадлежат анализируемые выборки, также прослеживается.

Но даже в случае выполнения стандартного предположения возможность корректного применения ряда параметрических критериев ограничена тем, что не известны распределения статистик, и имеются лишь таблицы критических значений статистик для некоторого ряда объёмов выборок. Поэтому при проверке гипотезы нельзя оценить достигнутый уровень значимости p_{value}.

На непараметрические критерии однородности характеристик рассеяния, в которых по существу проверяется гипотеза о равенстве параметров масштаба, не накладывается предположения о нормальности. Однако требуется выполнение не менее сильного предположения об однородности законов анализируемых выборок. Кроме того, распределения нормализованных статистик непараметрических критериев (Ансари–Бредли, Муда, Сижела–Тьюки) являются дискретными и при малых объёмах выборок существенно отличаются от асимптотического стандартного нормального закона.

Для построения корректного статистического вывода по результатам проверки гипотезы об однородности дисперсий необходимо выбрать наиболее мощный критерий и в соответствии с этим критерием оценить достигнутый уровень значимости p_{value}. Принятие решения на основе оценки p_{value} всегда более информативно, чем в результате сравнения вычисленной статистики с некоторым критическим значением.

В случае применения критериев, распределения статистик которых (при выполнении стандартного предположения) неизвестны или отличаются от известных асимптотических вследствие ограниченности объёмов анализируемых выборок, оценивание p_{value} представляет собой некоторую проблему. Однако эта проблема вполне решаема при наличии соответствующего программного обеспечения, позволяющего найти оценку p_{value} по результатам статистического моделирования.

Более того, реализация такой же возможности [7, 8] в случае использования критериев в условиях нарушения стандартного предположения о нормальности случайных величин существенно расширяет сферу применения параметрических (и непараметрических) критериев однородности дисперсий.

В настоящее время благодаря резкому увеличению возможностей вычислительной техники в программных системах статистического анализа существенно возрастает роль использования компьютерных методов исследования закономерностей. Например, когда распределение статистики критерия, используемого для проверки некоторой гипотезы, к моменту начала проверки (в силу разных причин) оказывается неизвестным (при данных объёмах выборок n_i), появляется возможность исследования распределения статистики в реальном времени проверки гипотезы (в интерактивном режиме) [7, 8, 12]. Например, в интерактивном режиме можно исследовать неизвестное распределение статистики любого критерия однородности дисперсий, зависящее от объема выборки, при тех значениях n_i, которые соответствуют анализируемым выборкам, и оценить по найденному в результате моделирования эмпирическому распределению статистики достигнутый уровень значимости p_{value}.

При таком подходе необходимое для проверки гипотезы эмпирическое распределение $G_N(S_n|H_0)$ статистики соответствующего критерия строится в результате статистического моделирования с точностью, зависящей от числа экспериментов N в методе Монте-Карло [11]. Затем по эмпирическому распределению $G_N(S_n|H_0)$ и вычисленному по анализируемой выборке значению статистики S^* критерия в соответствии с соотношением

$$p_{value} = P\{S > S^* | H_0\} = 1 - G(S^* | H_0)$$

для правостороннего критерия или по соотношению

$$p_{value} = 2\min\left\{G(S^*|H_0),\, 1 - G(S^*|H_0)\right\}$$

для двустороннего критерия определяется оценка p_{value}.

Таким образом, результаты статистического моделирования, осуществляемого в интерактивном режиме (непосредственно в процессе проводимого анализа) используются при формировании вывода по итогам проверки гипотезы. Реализация такого интерактивного режима требует наличия развитого программного обеспечения, позволяющего в целях ускорения распараллеливать процессы моделирования и привлекать доступные вычислительные ресурсы. В условиях распараллеливания время построения распределения $G_N(S_n|H_0)$ статистики критерия оказывается не очень заметным на фоне полного решения задачи статистического анализа.

Для рассмотренных критериев однородности дисперсий проиллюстрируем использование интерактивного режима исследования $G_N(S_n|H_0)$ и точности оценивания p_{value} в зависимости от числа экспериментов N моделируемых эмпирических распределений статистик, в том числе при нарушении стандартного предположения.

Напомним, что для того чтобы погрешность оценивания p_{value} с доверительной вероятностью 0.99 не превышала величины 0.01, количество экспериментов имитационного моделирования N должно быть порядка 16 600, для того, чтобы погрешность не превышала 0.001 – количество экспериментов должно быть порядка 1 660 000.

Пример 1. Пусть проверяется гипотеза о равенстве дисперсий 2-х следующих выборок объемом $n_i = 40$, $i = \overline{1,2}$, в предположении о принадлежности их нормальному закону:

0.669	3.434	-1.534	-2.116	-1.502	0.938	-1.058	-0.812	0.291	0.881
0.915	-2.166	-0.017	2.525	-3.825	1.887	-0.002	0.199	-1.144	0.237
-3.745	-0.620	-1.097	2.755	3.617	0.947	6.307	2.375	-8.381	-3.718
-1.425	0.041	1.508	0.265	1.329	0.945	-1.076	4.099	0.312	5.299

0.488	-2.837	-3.356	5.679	-0.096	-0.881	-1.317	-2.911	-5.250	1.147
1.103	4.606	-0.063	-2.394	7.522	-7.021	0.080	6.549	0.446	-0.805
5.147	-0.765	1.882	-10.886	-1.574	2.508	3.611	-0.030	0.871	-3.127
0.962	2.007	3.550	8.576	-0.655	2.612	3.501	-1.761	2.229	-0.681

В табл. 1 приведены значения статистик, вычисленные при проверке однородности дисперсий, соответствующих этим 2-м выборкам. В таблице представлены оценки p_{value}, полученные по смоделированным распределениям статистик рассматриваемых критериев при количестве экспериментов $N = 10^4$ и $N = 10^6$ в предположении о принадлежности случайных величин нормальному закону. Для критериев, относительно которых известны асимптотические распределения статистик, в таблице приведены также теоретические оценки p_{value}, вычисленные в соответствии с этими распределениями.

Таблица 1. Оценки p_{value}, полученные при проверке однородности первых 2-х выборок (при справедливости H_0)

Критерий	Значение статистики	p_{value}			
		При нормальном законе			При обобщ. норм.
		Теоретич.	$N=10^4$	$N=10^6$	$N=10^6$
Бартлетта	4.81269	0.028251	0.0258	0.0284	0.2458
Кокрена	0.671384	–	0.0258	0.0284	0.2458
Фишера	0.489462	0.0282466	0.0264	0.0286	0.2461
Хартли	2.04306	–	0.0258	0.0284	0.2458
Неймана–Пирсона	1.0449	–	0.0258	0.0284	0.2458
Z–критерий Оверолла–Вудворда	4.79659	0.0285162	0.0258	0.0284	0.2458
Модифицированный Z–критерий	2.024061	0.154825	0.1548	0.1523	0.1603
О'Брайена	2.53901	0.115109	0.1175	0.1155	0.1086
Левене (со средним)	3.31882	–	0.0740	0.0757	0.0838
Ньюмана	3.87145	–	0.5538	0.5653	0.3583
Линка	0.754701	–	0.1988	0.1985	0.4977
Блисса–Кокрена–Тьюки	0.569898	–	0.2001	0.1976	0.4987
Кадуэлла–Лесли–Брауна	1.32503	–	0.2001	0.1976	0. 4987
Флигнера–Киллина	2.94386	–	0.1566	0.1579	0.1675

В действительности обе анализируемые выборки моделировались в соответствии с обобщённым нормальным законом с параметром формы $\theta_2 = 0.75$, масштаба $\theta_1 = 1.414213562$ со значением $\sigma = 1$. Поэтому в последней колонке таблицы представлены оценки p_{value}, полученные по смоделированным распределениям статистик рассматриваемых критериев при количестве экспериментов $N = 10^6$ в предположении о принадлежности случайных величин данному обобщённому закону.

С одной стороны, можно заметить существенное различие в оценках p_{value} при обобщённом нормальном законе и нормальном. В то же время, очевидно, что в случае устойчивых критериев Левене, О'Брайена и модифицированного Z–критерия различие в таких оценках минимально.

Необходимо обратить внимание на следующее. Если реальный закон распределения обладает более "тяжелыми" хвостами по сравнению с нормальным законом, а мы, применяя параметрический критерий однородности дисперсий, опираемся на классические результаты, связанные с выполнением предположения о нормальности, то это приводит к увеличению (по сравнению с заданной) вероятности ошибки 1-го рода и к уменьшению вероятности ошибки 2-го рода. Если же реальный закон обладает более "лёгкими" хвостами, то в аналогичной ситуации это приводит к уменьшению вероятности ошибки 1-го рода и к увеличению вероятности ошибки 2-го рода. В данном случае мы имеем закон с существенно более тяжёлыми хвостами по сравнению с нормальным.

Пример 2. Проверим гипотезу о равенстве дисперсий 3-х выборок, 2 из которых взяты из предыдущего примера, а третья приведена ниже:

0.491	-0.644	17.738	1.647	1.583	9.204	-2.428	8.187	0.258	2.188
1.968	1.199	-0.166	-0.001	3.963	-1.867	0.255	-0.856	6.808	-5.456
2.567	-1.386	-1.649	-3.656	0.488	15.338	3.550	-3.967	-1.208	-3.985
-4.683	-5.998	-11.521	-0.683	12.323	-26.835	-0.084	0.149	-7.759	2.377

Эта выборка также смоделирована в соответствии с указанным обобщённым законом, но при $\sigma = 1.5$, что соответствует параметру масштаба $\theta_1 = 2.121320344$. Эмпирические распределения, соответствующие 3-м анализируемым выборкам, приведены на рис.1. Результаты анализа представлены в табл. 2.

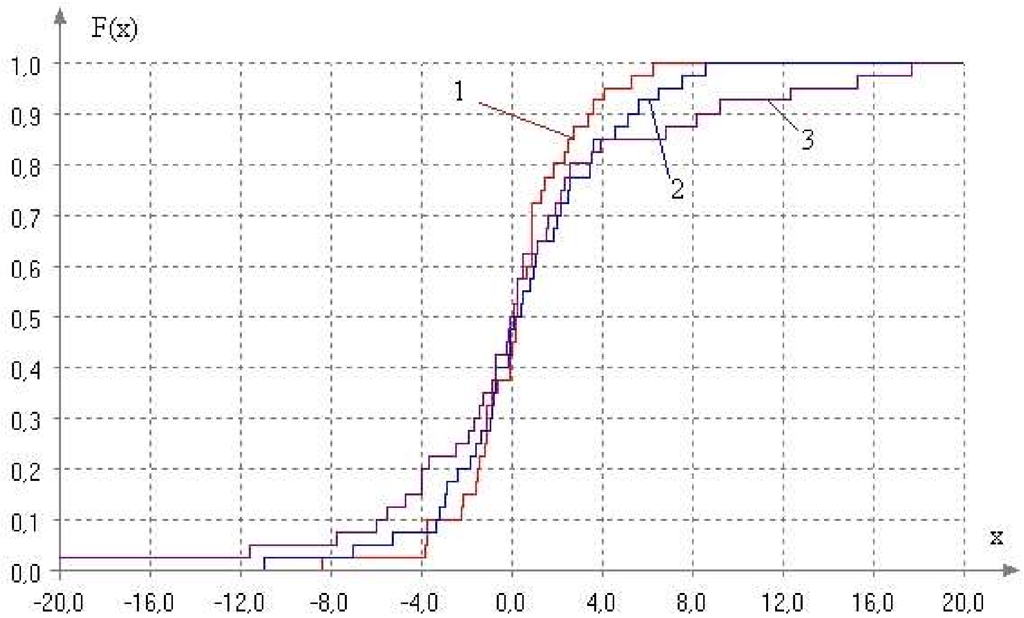

Рис.1. Эмпирические распределения, соответствующие выборкам

Таблица 2. Оценки p_{value}, полученные при проверке однородности 3-х выборок (отношение стандартных отклонений 1:1:1.5)

Критерий	Значение статистики	p_{value}		
		При нормальном законе		При обобщ. норм.
		теоретическая	$N = 10^6$	$N = 10^6$
Бартлетта	39.2341	0.0000	0.0000	0.0045
Кокрена	0.707025	–	0.0000	0.0050
Хартли	7.34369	–	0.0000	0.0043
Неймана–Пирсона	1.40376	–	0.0000	0.0045
Z–критерий Оверолла–Вудворда	18.6803	0.0000	0.0000	0.0044
Модифицированный Z–критерий	5.6877	0.0034	0.0045	0.0047
О’Брайена	3.70905	0.2458	0.0245	0.0167
Левене (со средним)	4.79491	–	0.0108	0.0132
Блисса–Кокрена–Тьюки	0.5662	–	0.0000	0.0122

В табл. 2 приведены значения статистик, вычисленные при проверке гипотезы об однородности дисперсий, соответствующих 3-м рассматриваемым выборкам. В предположении о принадлежности выборок нормальному закону в таблице указаны теоретические оценки p_{value} и оценки, полученные по результатам статистического моделирования при количестве экспериментов 10^6. В предположении о принадлежности выборок обобщённому нормальному закону с параметром формы $\theta_2 = 0.75$ приведены оценки p_{value} при $N = 10^6$.

Видно, что если мы проигнорируем факт нарушения стандартного предположения о нормальности, то по всем критериям (за исключением модифицированного Z-критерия и критерия О'Брайена) получим значения p_{value} меньшие по сравнению с истинными, имеющими место при обобщённом нормальном. Если бы реальный закон обладал более лёгкими хвостами по сравнению с нормальным, то значения p_{value}, получаемые в предположении о нормальности, превышали бы истинные значения.

В данном случае гипотеза об однородности 3-х выборок должна быть отклонена, так как можно видеть существенное (случайное) различие между эмпирическими распределениями 2-х первых выборок и явное отличие распределения 3-й выборки (см. рис 1).

6. Заключение

В случае анализа двух выборок и выполнении стандартного предположения о нормальности наиболее предпочтительными и эквивалентными по мощности относительно тех же конкурирующих гипотез являются критерии Фишера, Бартлетта, Кокрена, Хартли, Неймана–Пирсона и Z–критерий Оверолла–Вудворда.

Далее в порядке убывания мощности следует группа устойчивых критериев: О`Брайена, модифицированный Z-критерий, Левене.

Наименее перспективна для применения группа параметрических критериев Ньюмана, Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Линка, которые уступают в мощности даже непараметрическим критериям, имея перед последними некоторое преимущество в мощности лишь при очень малых объёмах выборок.

Среди рассмотренных непараметрических критериев наибольшей мощностью обладает критерий Клотца, затем идёт критерий Флайне–Киллина, далее критерий Муда, который уже заметно уступает критериям Фишера, Бартлетта, Кокрена, Хартли, Неймана–Пирсона, Z–критерию Оверолла–Вудворда, О`Брайена, модифицированному Z-критерию и критерию Левене.

Литература

1. *Лемешко Б. Ю.*, *Миркин Е. П.* Критерии Бартлетта и Кокрена в измерительных задачах при вероятностных законах, отличающихся от нормального // Измерительная техника. 2004. № 10. С. 10-16.
2. *Lemeshko B., Mirkin E.* Bartlett and Cochran tests in measurements with probability laws different from normal // Measurement Techniques, 2004, Vol. 47, № 10. P. 960-968.
3. *Лемешко Б. Ю., Лемешко С. Б., Горбунова А. А.* О применении и мощности критериев проверки однородности дисперсий. Ч. I. Параметрические критерии // Измерительная техника. 2010. № 3. С. 10-16.
4. *Lemeshko B. Yu., Lemeshko S. B., Gorbunova A. A.* Application and power of criteria for testing the homogeneity of variances. Part I. Parametric criteria // Measurement Techniques, Vol. 53, No. 3, 2010. P. 237-246.

5. *Лемешко Б. Ю., Лемешко С. Б., Горбунова А. А.* О применении и мощности критериев проверки однородности дисперсий. Ч. II. Непараметрические критерии // Измерительная техника. 2010. № 5. С. 11-18.
6. *Lemeshko B. Yu., Lemeshko S. B., Gorbunova A. A.* Application and power of criteria for testing the homogeneity of variances. Part II. Nonparametric criteria // Measurement Techniques, Vol. 53, No. 5, 2010. P. 476-486.
7. *Gorbunova A. A., Lemeshko B. Yu.* Application of Variance Homogeneity Tests Under Violation of Normality Assumption // Proceedings of the International Workshop "Applied Methods of Statistical Analysis. Simulations and Statistical Inference" – AMSA'2011, Novosibirsk, Russia, 20-22 September, 2011. P. 28-36.
8. *Лемешко Б.Ю., Сатаева Т.С.* О свойствах и мощности параметрических критериев однородности дисперсий // Обработка информации и математическое моделирование : материалы Рос. науч.-техн. конф. Новосибирск : СибГУТИ, 2016. С. 85–98.
9. *Boris Yu. Lemeshko, Tatyana S. Sataeva*. On the Properties and Application of Tests for Homogeneity of Variances in the Problems of Metrology and Control // Series on Advances in Intelligent Systems and Computing, Vol. 543, 2017. P. 784-798.
10. *Лемешко Б. Ю., Сатаева Т. С.* Применение и мощность параметрических критериев проверки однородности дисперсий. Ч. 3. // Измерительная техника. 2017. № 1. С. 8-13.
11. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. Новосибирск : Изд-во НГТУ, 2011. 888 с.
12. *Лемешко Б. Ю.* Критерии проверки гипотез об однородности. Руководство по применению. М. : ИНФРА-М, 2017. 208 с.

Лемешко Борис Юрьевич
главный научный сотрудник кафедры прикладной и теоретической информатики НГТУ, д.т.н., профессор (630073, Новосибирск, пр. Карла Маркса, 20), тел. (383) 346-06-00, e-mail: Lemeshko@ami.nstu.ru, http://www.ami.nstu.ru/~headrd/

About power and application of homogeneity tests of variances in non-standard conditions

B. Yu. Lemeshko

Novosibirsk State Technical University

The properties of parametric and nonparametric tests used for testing hypotheses of homogeneity of variances have been studied. The comparative analysis of power of these tests has been carried out. The application of parametric tests has been considered in the case of non-standard conditions.

Keywords: homogeneity test of variances, power of test, statistic, statistical simulating