УДК 519.24

О свойствах и мощности параметрических критериев однородности дисперсий

Б.Ю. Лемешко, Т.С. Сатаева[1]
Новосибирский государственный технический университет

Исследованы распределения статистик параметрических критериев проверки однородности дисперсий (критериев Неймана–Пирсона, О'Брайена, Линка, Ньюмана, Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Z-критерия Оверолла–Вудворда и модифицированного Z-критерия), в том числе при нарушении стандартного предположения о нормальности. Проведен сравнительный анализ мощности множества параметрических критериев. Предложена и апробирована методика применения критериев в условиях нарушения стандартного предположения, предусматривающая интерактивное моделирование распределений статистик.

Ключевые слова: критерий однородности дисперсий, критерий Неймана–Пирсона, критерий О'Брайена, критерий Линка, критерий Ньюмана, критерий Блисса–Кокрена–Тьюки, критерий Кадуэлла–Лесли–Брауна, Z-критерий Оверолла–Вудворда, модифицированный Z-критерий Оверолла–Вудворда, мощность критерия

1. Введение

Применение критериев проверки гипотез об однородности дисперсий достаточно востребовано в различных приложениях. Не оказываются исключением и задачи обработки результатов измерений. Возможно, наиболее ярким примером востребованности такого рода критериев в области метрологии являются задачи сличения лабораторных испытаний.

По литературным источникам можно сформировать достаточно представительный перечень параметрических и непараметрических критериев, которые можно использовать для проверки гипотезы об однородности дисперсий, которая для случая анализа m выборок имеет вид

$$H_0:\ \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2. \tag{1}$$

В качестве конкурирующей, как правило, рассматривают гипотезу

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

где неравенство выполняется, по крайней мере, для одной пары индексов i_1, i_2. Часть из этих критериев может использоваться только при $m = 2$.

Качество статистических выводов, осуществляемых по результатам анализа, обеспечивается корректностью применения соответствующих критериев и использованием критериев, обладающих лучшими свойствами (большей мощностью). Поэтому специалист, столкнувшийся с необходимостью статистического анализа результатов измерений, должен выбирать среди тех критериев, которые способны обеспечить корректность принимаемого решения о результатах проверки гипотезы в условиях предположений, характеризующих анализируе-

[1] Работа выполнена при поддержке Министерства образования и науки РФ в рамках проектной части государственного задания (№ 2.541.2014/К).

мый измерительный процесс. И при этом отдать предпочтение критерию, обладающему в этих условиях большей мощностью.

Стандартным предположением, обуславливающим возможность применения классических параметрических критериев однородности дисперсий, является принадлежность анализируемых выборок нормальному закону распределения. Это условие резко ограничивает область применения параметрических критериев, так как делает невозможным использование классических результатов, связанных с распределениями статистик критериев при справедливости H_0, которые были получены именно при данном предположении.

На непараметрические критерии, в которых по существу проверяется гипотеза о равенстве параметров масштаба, такого ограничения не накладывается.

В работах [1-4] был проведен сравнительный анализ мощности и исследованы свойства ряда параметрических (Бартлетта, Кокрена, Фишера, Хартли, Левене) и непараметрических (Ансари–Бредли, Муда, Сижела–Тьюки, Кейпена, Клотца) критериев, в том числе в условиях нарушения стандартных предположений. Было показано, что при $m = 2$ параметрические критерии Бартлетта, Кокрена, Фишера и Хартли являются эквивалентными, а при $m > 2$ преимущество оказывается за критерием Кокрена. При этом мощность параметрических критериев существенно выше непараметрических аналогов. Оказалось, что свойство "непараметричности" непараметрических критериев очень ограничено. Да, распределения статистик непараметрических критериев при справедливости H_0 не зависят от вида закона, которому принадлежат анализируемые выборки, но при этом выборки должны принадлежать одному виду закона. Например, это означает, что при справедливости $H_0: \sigma_1^2 = \sigma_2^2$ асимптотическим распределением нормализованных статистик этих критериев стандартный нормальный закон будет только в том случае, если выборки принадлежат одному и тому же виду закона (то есть, при однородности законов сравниваемых выборок). Необходимость выполнения такой предпосылки также существенно сокращает область корректного применения непараметрических критериев.

Явное преимущество в мощности параметрических критериев заставило нас рассмотреть возможность их применения в условиях нарушения классического предположения о нормальности (в условиях принадлежности выборок различным законам [3, 5]).

В данной работе мы дополняем выводы работ [1-4] результатами сравнительного анализа ещё ряда параметрических критериев однородности дисперсий (Неймана–Пирсона [6], О'Брайена [7], Линка (отношения размахов) [8], Ньюмана (стьюдентизированного размаха) [9], Блисса–Кокрена–Тьюки [10], Кадуэлла–Лесли–Брауна [11], Z-критерия Оверолла–Вудворда [12] и модифицированного Z-критерия [13]). Цель работы заключалась в исследовании распределений статистик критериев и расширении таблиц процентных точек, в сравнительном анализе мощности критериев, в реализации возможности применения критериев в условиях нарушения стандартного предположения о нормальности случайных величин.

Исследование распределений статистик и оценка мощности критериев относительно различных конкурирующих гипотез осуществлялась методами статистического моделирования [14] при использовании программной системы ISW [15]. Число статистических экспериментов при моделировании выборок статистик составляло величину порядка $N = 10^6$. При таких величинах N разность между истинным законом распределения статистики и смоделированным эмпирическим по модулю, как правило, не превышает величины 10^{-3}.

В условиях нарушения стандартного предположения о нормальности распределения статистик критериев исследовались в случае принадлежности выборок обобщённому нормальному закону с плотностью

$$f(x) = \frac{\theta_0}{2\theta_1 \Gamma(1/\theta_0)} \exp\left(-\left(\frac{|x-\theta_2|}{\theta_1} \right)^{\theta_0} \right) \quad (3)$$

при различных значениях параметра формы θ_0. Частными случаями этого семейства являются нормальный закон при $\theta_0 = 2$ и распределение Лапласа при $\theta_0 = 1$. Далее в тексте и на рисунках обозначение $De(\theta_0)$ соответствует распределению вида (3) при соответствующем значении параметра формы θ_0. Отметим, что чем меньше значение параметра формы θ_0, тем «тяжелее» хвосты распределения $De(\theta_0)$, чем больше параметр, тем хвосты «легче».

2. Критерий Неймана-Пирсона

Статистика критерия [6] определяется отношением арифметического среднего всех оценок дисперсий s_i^2 к их геометрическому среднему.

$$h = \frac{1}{m}\sum_{i=1}^{m} s_i^2 \Bigg/ \left(\prod_{i=1}^{m} s_i^2\right)^{\frac{1}{m}}, \qquad (1)$$

где m – количество выборок, $s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}\left(x_{ij} - \bar{x}_i\right)^2$ – оценки выборочных дисперсий, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ – выборочное среднее значение, x_{ij} – j-й элемент i-й выборки. Предполагается, что $n_1 = n_2 = \ldots = n_m = n$. Критерий правосторонний. Проверяемая гипотеза H_0 отклоняется при больших значениях статистики (4), когда $h > h_{1-\alpha}$.

Распределения статистики (4) зависят от n и от числа выборок m. В данной работе уточнены значения процентных точек распределений статистики.

Критерий Неймана-Пирсона крайне чувствителен к отклонениям от предположения о нормальности (см. рис. 1).

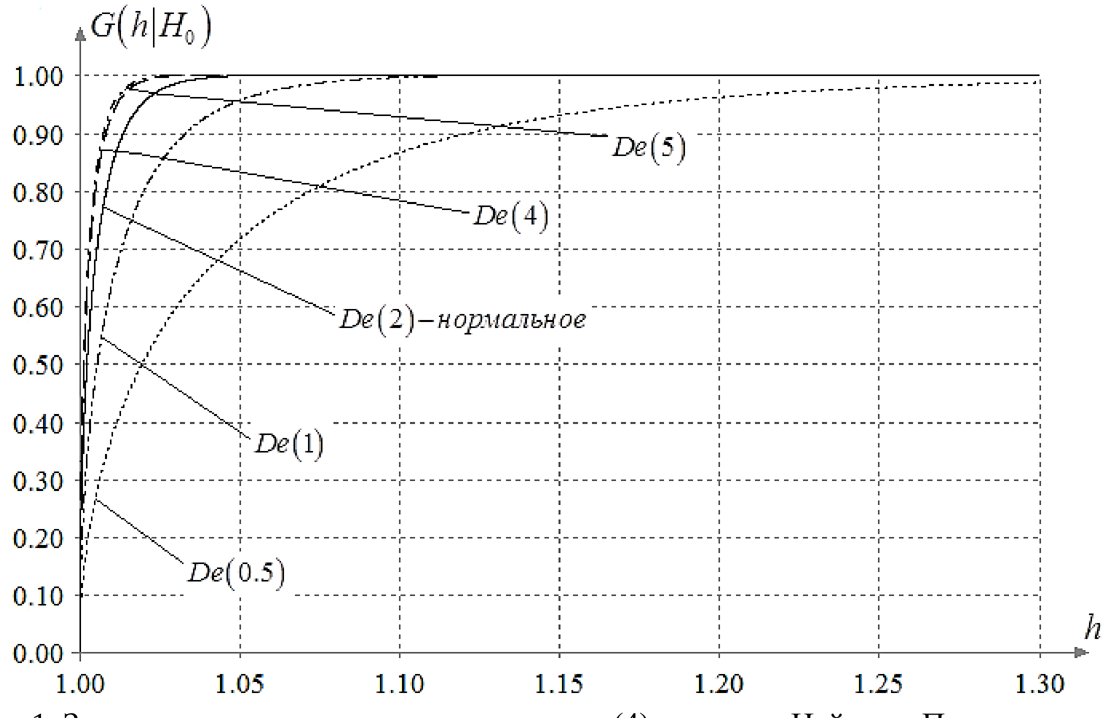

Рис. 1. Зависимость распределения статистики (4) критерия Неймана-Пирсона от вида закона (при $n = 100$, $m = 2$)

На этом рисунке на примере семейства (3) иллюстрируется характер зависимости распределений статистик (4) от вида наблюдаемого закона (при различных значениях параметра формы θ_0 семейства законов (3)), которым принадлежат анализируемые выборки при справедливости H_0 (при $n=100$ и $m=2$).

Естественно, что критерий со статистикой (4) может использоваться и при неравных n_i, однако следует учитывать, что в этом случае распределения статистик при справедливости H_0 будут отличаться от распределений, имеющих место при равных n_i.

3. Критерий О`Брайена

При формировании статистики критерия [7] каждый j-й элемент i-й выборки x_{ij} преобразуется в соответствии с формулой

$$V_{ij} = \frac{(n_i - 1.5) n_i \left(x_{ij} - \bar{x}_i\right)^2 - 0.5 s_i^2 (n_i - 1)}{(n_i - 1)(n_i - 2)}, \tag{5}$$

где n_i – объём, $\bar{x}_i$ – среднее значение, s_i^2 – оценка дисперсии i-й выборки.

Статистика критерия имеет вид

$$V = \frac{\frac{1}{m-1} \sum_{i=1}^{m} n_i \left(\overline{V_i} - \overline{\overline{V_i}}\right)^2}{\frac{1}{N-m} \sum_{i=1}^{m} \sum_{j=1}^{n_i} \left(V_{ij} - \overline{V_i}\right)^2}, \tag{6}$$

где – $\overline{V_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} V_{ij}$, $\overline{\overline{V_i}} = \frac{1}{N} \sum_{i=1}^{m} \sum_{j=1}^{n_i} V_{ij}$, $N = \sum_{i=1}^{m} n_i$. Критерий правосторонний, и проверяемая гипотеза H_0 отклоняется при больших значениях статистики (6).

Предельным распределением статистики критерия О’Брайена при справедливости H_0 является $F_{m-1,N-m}$-распределение Фишера с $m-1$ и $N-m$ степенями свободы [7]. Однако проведенные исследования показывают, что распределение статистики (6) критерия О’Брайена достаточно медленно сходится к соответствующему распределению Фишера. Например, в случае $m=2$ отличием реального распределения $G(V|H_0)$ статистики (6) от соответствующего $F_{1,N-m}$-распределения Фишера можно пренебречь лишь при $n_1 = n_2 = n \ge 80$. При малых объёмах выборок существенное отличие распределения статистики $G(V|H_0)$ от $F_{m-1,N-m}$-распределения Фишера наблюдается при больших значениях V, поэтому использование процентных точек $F_{m-1,N-m}$-распределения приводит к увеличению вероятности β ошибок второго рода (вследствие уменьшения уровня значимости по сравнению с заданным α).

Чтобы обеспечить возможность корректного применения критерия и при малых объемах выборок, в работе получены верхние процентные точки распределений статистики критерия О’Брайена при различном количестве m сравниваемых выборок для $n_1 = n_2 = n \le 80$.

Исследования показали, что при $N-m \le 80$ распределения $G(V|H_0)$ при значениях V таких, что $1-G(V|H_0) < 0.1$, оказываются ближе к $F_{m-1,\infty}$-распределению, чем к $F_{m-1,N-m}$-распределению Фишера. Поэтому в таких ситуациях корректность выводов можно повысить, используя $F_{m-1,\infty}$-распределение для оценки достигнутого уровня значимости (p_{value}) или выбирая в соответствии с $F_{m-1,\infty}$-распределением критические значения $V_{1-\alpha}$.

Распределения статистики критерия О'Брайена (как и в случае критерия Левене [2]) достаточно устойчивы к нарушению предположения о принадлежности выборок нормальному закону. Отклонения в сторону законов с более "лёгкими" чем у нормального закона хвостами практически не влияют на распределение статистики. А при законах с более "тяжелыми" хвостами отклонения от распределений, имеющих место при нормальном законе, не настолько велики, как в случае других параметрических критериев.

Подобную же устойчивость к нарушению стандартного предположения о нормальности среди критериев, рассмотренных в данной работе, демонстрирует только модифицированный Z-критерий.

4. Критерий Линка

Критерий Линка (критерий отношения размахов) является аналогом критерия Фишера и используется только при анализе 2-х выборок ($m=2$). Статистика критерия имеет вид [8]:

$$F^* = \frac{\omega_{n_1}}{\omega_{n_2}}, \tag{7}$$

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.

Критерий двусторонний. Проверяемая гипотеза отклоняется с уровнем значимости α, если $F^* > F^*_{1-\alpha/2}$ или $F^* < F^*_{\alpha/2}$, где $F^*_{1-\alpha/2}$ и $F^*_{\alpha/2}$ – верхнее и нижнее критические значения статистики.

Распределение статистики критерия существенно зависит от объемов сравниваемых выборок. Критерий отношения размахов крайне чувствителен к любым отклонениям от нормальности.

В ходе исследований методами статистического моделирования были уточнены нижние и верхние процентные точки для статистики (7) критерия отношения размахов в случае принадлежности выборок нормальному закону.

5. Критерий Ньюмана

Статистика критерия Ньюмана (стьюдентизированного размаха) имеет вид [9]:

$$q = \frac{\omega_{n_1}}{s_{n_2}}, \tag{8}$$

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

Как и предыдущий, этот критерий также является двусторонним. Проверяемая гипотеза H_0 о равенстве дисперсий отклоняется, если $q < q_{\alpha/2}$ или $q > q_{1-\alpha/2}$, где $q_{\alpha/2}$ и $q_{1-\alpha/2}$ – нижнее и верхнее критические значения статистики при заданном уровне значимости α.

Распределения статистики (8) критерия стьюдентизированного размаха при справедливости H_0 зависят от объёмов анализируемых выборок и от вида закона, которому принадлежат выборки. В работе уточнены нижние и верхние критические значения статистики критерия стьюдентизированного размаха, выход за которые приводит к отклонению нулевой гипотезы.

6. Критерий Блиса–Кокрена–Тьюки

Статистика критерия [10], предложенного в качестве аналога критерия Кокрена, имеет вид

$$c = \frac{\max\limits_{1 \le i \le m} \omega_i}{\sum\limits_{i=1}^{m} \omega_i}, \tag{9}$$

где m – количество сравниваемых выборок, $\omega_i = \max\limits_{1 \le j \le n_i} x_{ij} - \min\limits_{1 \le j \le n_i} x_{ij}$ – размах i-й выборки.

Критерий правосторонний. Если статистика $c > c_{1-\alpha}$, где $c_{1-\alpha}$ – верхнее критическое значение при заданном уровне значимости α, то проверяемая гипотеза H_0 о равенстве дисперсий отклоняется.

Распределение статистики критерия сильно зависит от объема выборки.

Значения верхних процентных точек для статистики критерия при некоторых объёмах выборок $n_1 = n_2 = n$ приведены в таблице А.7. Формально критерий Блисса–Кокрена–Тьюки можно применять и при неравных объёмах n_i анализируемых выборок, однако следует иметь ввиду, что в этом случае критические значения статистики будут отличаться от приведенных в таблице А.7.

Как и в случае критерия Кокрена, распределения статистики данного критерия сильно зависят от закона, которому принадлежат анализируемые выборки.

7. Критерий Кадуэлла–Лесли–Брауна

Этот критерий [11] был предложен в качестве аналога критерия Хартли с заменой в статистике отношений оценок дисперсий на отношения размахов

$$K = \frac{\max\limits_{1 \le i \le m} \omega_i}{\min\limits_{1 \le i \le m} \omega_i}, \tag{10}$$

где m – количество выборок, ω_i – размах i-й выборки.

Критерий правосторонний. При $K > K_{1-\alpha}$, где $K_{1-\alpha}$ – верхнее критическое значения статистики при заданном уровне значимости α, проверяемая гипотеза H_0 отклоняется.

Распределения статистики критерия Кадуэлла–Лесли–Брауна так же, как и статистики критерия Блисса–Кокрена–Тьюки, существенно зависят и от объема выборок, и от закона распределения, которому подчиняются выборки.

В работе уточнены критические значения $K_{1-\alpha}$ для различного количества выборок m при равных объемах выборок $n_i = n,\ i = \overline{1, m}$, полученные методами статистического моделирования. Исследования показали, что критерий Кадуэлла–Лесли–Брауна при больших объемах выборок значительно уступает по мощности большинству рассмотренных критериев. Относительно критерия Кадуэлла–Лесли–Брауна также можно отметить возможность его применения при неравных объёмах n_i анализируемых выборок. В этом случае критические значения статистики будут отличаться от полученных.

8. Z–критерий Оверолла–Вудворда

Статистика критерия имеет вид [1212]:

$$Z=\frac{1}{m-1}\sum_{i=1}^{m} Z_i^2\ , \tag{11}$$

где $Z_i=\sqrt{\frac{c_i\left(n_i-1\right)s_i^2}{MSE}}-\sqrt{c_i\left(n_i-1\right)-\frac{c_i}{2}}$, $MSE=\frac{1}{N-m}\sum_{i=1}^{m}\sum_{j=1}^{n_i}\left(x_{ij}-\bar{x}_i\right)^2$, m – количество выборок, $c_i=2+1/n_i$, n_i – размер i-й выборки, s_i^2 – несмещенная оценка дисперсии i-й выборки, $N=\sum_{i=1}^{m} n_i$, x_{ij} – j-й элемент i-й выборки, $\bar{x}_i$ –среднее значение i-й выборки.

При справедливости проверяемой гипотезы H_0 об однородности дисперсий и принадлежности анализируемых выборок нормальному закону предельное распределение статистики (11) не зависит от размера выборки и подчиняется $F_{m-1,\infty}$-распределению Фишера. Но при малых объемах n_i распределение статистики Z–критерия Оверолла–Вудворда заметно отличается от предельного $F_{m-1,\infty}$-распределения Фишера.

Проведенные исследования показали, что различием между реальным распределением статистики критерия и предельным $F_{m-1,\infty}$-распределением можно пренебречь при объемах выборок $n\geq 50$. Поэтому в работе (в предположении о принадлежности анализируемых выборок нормальному закону) для объемов выборок $n\leq 50$ методами статистического моделирования (для разного количества сравниваемых выборок m и для различных $n_i=n$ объемов выборок) были вычислены верхние критические значения $Z_{1-\alpha}$.

Как и в случае большинства параметрических критериев однородности дисперсий, распределение статистики Z–критерия очень чувствительно к нарушению предположения о нормальности.

9. Модифицированный Z–критерий Оверолла–Вудворда

В работе [13] в целях построения критерия, более устойчивого к нарушению стандартного предположения о нормальности, Оверолл и Вудворд предложили модификацию Z–критерия, статистика которого отличается вычислением величин c_i.

Новое значение c_i связано с размером выборки n_i и средним значением коэффициентов эксцесса и вычисляется следующим образом:

$$c_i=2.0\left[\frac{1}{K_i}\left(2.9+\frac{0.2}{n_i}\right)\right]^{\frac{1.6\left(n_i-1.8K_i+14.7\right)}{n_i}}, \tag{11}$$

где $K_i=\frac{1}{n_i-2}\sum_{j=1}^{n_i} G_{ij}^4$ – оценка коэффициента эксцесса i-й выборки, $G_{ij}=\left(x_{ij}-\bar{x}_i\right)\Big/\sqrt{\frac{n_i-1}{n_i}s_i^2}$, $\bar{K}$ – среднее значение оценок коэффициентов эксцесса всех выборок

Модифицированный Z–критерий создавался с целью уменьшения зависимости распределения статистики критерия от законов распределения, которым подчиняются сравниваемые выборки.

В ходе настоящих исследований было показано, что распределение статистики модифицированного Z–критерия с ростом объема выборок очень медленно сходится к $F_{m-1,\infty}$-распределению Фишера. Даже при больших объемах выборок распределение статистики критерия не согласуется с $F_{m-1,\infty}$-распределением, хотя в области больших значений статистики отличие её распределения от $F_{m-1,\infty}$-распределения не играет существенного значения. Для корректного применения критерия при малых объёмах выборок найдены критические значения, полученные с использованием методов статистического моделирования.

В то же время следует отметить, что распределение статистики модифицированного Z–критерия действительно обладает большей стабильностью к нарушения стандартного предположения о нормальности. При симметричности наблюдаемых законов явное отличие распределения модифицированной статистики, от имеющей место при нормальном законе, проявляется лишь при тяжёлых хвостах.

Отметим, что борьба за устойчивость привела не только к ухудшению сходимости распределения статистики модифицированного Z–критерия к $F_{m-1,\infty}$-распределению, но и к некоторому снижению мощности критерия.

10. Сравнительный анализ мощности критериев

В связи с применяемыми критериями однородности дисперсий давно и неизменно волнуют две основные (и связанные между собой) проблемы. Первая заключается в крайней неустойчивости большей части существующих параметрических критериев однородности дисперсий, а вторая касается анализа мощности этих критериев. В частности, этим вопросам (среди многих) посвящены работы [16-18], и эти же вопросы рассматривались в [1-3].

В данном случае при анализе мощности критериев проверяемая гипотеза имела вид $H_0 : \sigma_1^2 = \sigma_2^2 = ... = \sigma_m^2 = \sigma_0^2$. В качестве конкурирующих гипотез рассматривались ситуации, когда $m-1$ выборка принадлежала закону с некоторым $\sigma = \sigma_0$, а одна из выборок, например, с номером m принадлежала закону с отличающимся значением σ ($H_1 : \sigma_m = 1.1\sigma_0$, $H_2 : \sigma_m = 1.2\sigma_0$, $H_3 : \sigma_m = 1.5\sigma_0$). В сравнительном анализе кроме вышерассмотренных приняли участие также критерии Бартлетта, Кокрена, Левене, Хартли, Фишера, оценки мощности которых были взяты из [2].

Полученные оценки мощности всех рассматриваемых критериев в случае принадлежности выборок нормальному закону для уровней значимости $\alpha = 0.1$, 0.05, 0.01 при $m = 2$ представлены в таблицах 1–3, где критерии расположены по убыванию мощности. В этих же таблицах для сравнения представлены оценки мощности непараметрических критериев Муда, Ансари–Бредли и Сижела–Тьюки [3].

Таблица 1. Мощность критериев однородности дисперсий относительно конкурирующей гипотезы $H_1 :\ \sigma_2 = 1.1\sigma_1$

Критерий	α	Объем выборки				
		$n=10$	$n=20$	$n=40$	$n=60$	$n=100$
Бартлетта, Кокрена, Хартли, Фишера, Неймана–Пирсона, Z–критерий Оверолла–Вудворда	0.1	0.112	0.128	0.157	0.188	0.246
	0.05	0.058	0.068	0.090	0.111	0.156
	0.01	0.012	0.016	0.023	0.032	0.051
Модифицированный Z–критерий, О’Брайена	0.1	0.109	0.125	0.154	0.184	0.243
	0.05	0.056	0.066	0.087	0.108	0.153
	0.01	0.012	0.015	0.022	0.030	0.049
Левене	0.1	0.110	0.123	0.150	0.176	0.228
	0.05	0.056	0.065	0.084	0.103	0.141
	0.01	0.012	0.014	0.021	0.028	0.044
Ньюмана	0.1	0.111	0.123	0.143	0.159	0.186

	0.05	0.057	0.066	0.080	0.091	0.112
	0.01	0.012	0.015	0.020	0.025	0.033

Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Линка	0.1	0.111	0.119	0.133	0.141	0.154
	0.05	0.057	0.063	0.072	0.078	0.087
	0.01	0.012	0.014	0.018	0.019	0.023
Муда	0.1	0.111	0.120	0.143	0.166	0.211
	0.05	0.057	0.064	0.080	0.096	0.128
	0.01	0.012	0.014	0.020	0.026	0.039
Ансари–Бредли	0.1	0.101	0.125	0.135	0.154	0.190
	0.05	0.052	0.064	0.074	0.087	0.113
	0.01	0.011	0.014	0.019	0.023	0.033
Сижела–Тьюки	0.1	0.106	0.121	0.135	0.154	0.190
	0.05	0.055	0.062	0.075	0.087	0.113
	0.01	0.011	0.010	0.018	0.023	0.033

Таблица 2. Мощность критериев однородности дисперсий относительно конкурирующей гипотезы $H_2: \ \sigma_2 = 1.2\sigma_1$

Критерий	α	Объем выборки				
		$n=10$	$n=20$	$n=40$	$n=60$	$n=100$
Бартлетта, Кокрена, Хартли, Фишера, Неймана–Пирсона, Z–критерий Оверолла–Вудворда	0.1	0.144	0.199	0.304	0.401	0.564
	0.05	0.079	0.119	0.201	0.283	0.438
	0.01	0.018	0.033	0.071	0.114	0.218
Модифицированный Z–критерий, О’Брайена	0.1	0.134	0.188	0.292	0.389	0.555
	0.05	0.071	0.109	0.189	0.269	0.427
	0.01	0.016	0.029	0.062	0.104	0.205
Левене	0.1	0.135	0.184	0.276	0.363	0.515
	0.05	0.072	0.107	0.177	0.250	0.388
	0.01	0.016	0.028	0.058	0.095	0.180
Ньюмана	0.1	0.140	0.183	0.251	0.304	0.386
	0.05	0.077	0.108	0.161	0.203	0.276
	0.01	0.018	0.030	0.053	0.075	0.116
Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Линка	0.1	0.139	0.171	0.216	0.246	0.289
	0.05	0.075	0.100	0.133	0.156	0.190
	0.01	0.018	0.027	0.042	0.051	0.068
Муда	0.1	0.135	0.173	0.254	0.3300	0.468
	0.05	0.073	0.100	0.161	0.2225	0.344
	0.01	0.016	0.027	0.052	0.082	0.152
Ансари–Бредли	0.1	0.128	0.172	0.226	0.289	0.406
	0.05	0.070	0.097	0.141	0.189	0.287
	0.01	0.015	0.025	0.045	0.066	0.119
Сижела–Тьюки	0.1	0.124	0.165	0.226	0.289	0.405
	0.05	0.066	0.092	0.141	0.190	0.287
	0.01	0.014	0.013	0.044	0.066	0.119

Критерий Неймана–Пирсона и Z–критерий Оверолла–Вудворда по мощности оказались эквивалентны критериям Бартлетта, Кокрена, Хартли и Фишера. Различие в мощности модифицированного Z–критерия и критерия О’Брайена заметно только относительно достаточно далёкой конкурирующей гипотезы H_3. При этом оба эти критерия имеют преимущество в мощности по сравнению с критерием Левене. Заметим, что, как и последний, эти критерии достаточно устойчивы к нарушению стандартного предположения о нормальности анализируемых выборок.

Таблица 3. Мощность критериев однородности дисперсий относительно конкурирующей гипотезы $H_3: \ \sigma_2 = 1.5\sigma_1$

Критерий	α	Объем выборки				
		$n=10$	$n=20$	$n=40$	$n=60$	$n=100$
Бартлетта, Кокрена, Хартли, Фишера, Неймана–Пирсона, Z–критерий Оверолла–Вудворда	0.1	0.312	0.532	0.806	0.926	0.991
	0.05	0.201	0.402	0.705	0.871	0.980
	0.01	0.064	0.182	0.463	0.692	0.924
О'Брайена	0.1	0.266	0.490	0.783	0.917	0.990
	0.05	0.155	0.344	0.664	0.849	0.976
	0.01	0.039	0.127	0.379	0.628	0.903
Модифицированный Z–критерий	0.1	0.265	0.489	0.781	0.916	0.990
	0.05	0.158	0.348	0.666	0.849	0.976
	0.01	0.043	0.138	0.397	0.639	0.906
Левене	0.1	0.269	0.471	0.746	0.888	0.981
	0.05	0.163	0.338	0.628	0.812	0.960
	0.01	0.045	0.131	0.364	0.590	0.866
Ньюмана	0.1	0.296	0.473	0.682	0.796	0.901
	0.05	0.190	0.348	0.566	0.699	0.840
	0.01	0.060	0.153	0.326	0.473	0.667
Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Линка	0.1	0.285	0.425	0.584	0.674	0.776
	0.05	0.181	0.305	0.458	0.554	0.671
	0.01	0.057	0.127	0.237	0.314	0.430
Муда	0.1	0.255	0.425	0.688	0.841	0.964
	0.05	0.158	0.302	0.565	0.751	0.931
	0.01	0.045	0.121	0.319	0.518	0.802
Ансари–Бредли	0.1	0.242	0.393	0.608	0.768	0.926
	0.05	0.150	0.270	0.484	0.659	0.869
	0.01	0.041	0.104	0.254	0.413	0.693
Сижела–Тьюки	0.1	0.246	0.383	0.609	0.768	0.926
	0.05	0.155	0.261	0.484	0.659	0.869
	0.01	0.043	0.056	0.251	0.414	0.693

Критерий Ньюмана с ростом объёмов выборок всё заметнее уступает в мощности критерию Левене. В то же время он имеет явное преимущество в мощности (за исключением $n=10$) по сравнению с критериями Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна и Линка. Три последних эквивалентны по мощности.

Следует отметить, что группа "устойчивых" критериев (модифицированный Z–критерий, О'Брайена и Левене) при малых объёмах выборок (см. при $n=10$) уступает в мощности критериям Ньюмана, Линка, Блисса–Кокрена–Тьюки и Кадуэлла–Лесли–Брауна, но с ростом n имеет явное преимущество перед последними и перед непараметрическими критериями. Внутри этой группы некоторое преимущество остаётся за критерием О'Брайена.

Как можно видеть, критерии Ньюмана, Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна и Линка имеют некоторое преимущество в мощности над непараметрическими критериями только при малых объёмах выборок ($n=10 \div 20$), а при их увеличении заметно уступают непараметрическим.

Критерии Бартлетта, Кокрена, Хартли, Левене, Неймана–Пирсона, О'Брайена, Блисса–Кокрена–Тьюки, Кадуэлла–Лесли–Брауна, Z–критерий Оверолла–Вудворда, модифицированный Z–критерий могут применяться при числе выборок $m > 2$. При этом критерии Бартлетта, Кокрена, Хартли, Неймана–Пирсона и Z–критерий Оверолла–Вудворда уже не образуют группу эквивалентных критериев с одинаковой мощностью. Исключение составляют лишь критерии Бартлетта и Неймана–Пирсона, которые остаются практически эквивалентными по мощности.

В таблицах 4-6 приведены полученные оценки мощности многовыборочных критериев при числе выборок $m=3$ и $m=5$ при объёмах выборок $n_i=100$, $i=\overline{1,m}$, относительно конкурирующих гипотез H_1, H_3, H_3 (при отличающейся дисперсии одной из выборок) в случае принадлежности анализируемых выборок нормальным законам. Критерии в таблицах расположены по убыванию мощности. Приводимые оценки мощности позволяют судить о предпочтительности применения тех или иных критериев.

Таблица 4. Мощность многовыборочных критериев относительно конкурирующей гипотезы $H_1: \sigma_m = 1.1\sigma_1$, $n_i=100$, $i=\overline{1,m}$

Критерий	α					
	0.1	0.05	0.01	0.1	0.05	0.01
	$m=3$			$m=5$		
Кокрена	0.250	0.161	0.056	0.241	0.156	0.056
О’Брайена	0.243	0.153	0.051	0.230	0.144	0.048
Z–критерий Оверолла–Вудворда	0.243	0.153	0.051	0.227	0.141	0.046
Неймана–Пирсона, Бартлетта	0.242	0.152	0.049	0.224	0.138	0.044
Модифицированный Z–критерий	0.240	0.150	0.048	0.223	0.137	0.044
Хартли	0.239	0.148	0.046	0.219	0.133	0.040
Левене	0.225	0.139	0.043	0.209	0.127	0.039
Кадуэлла–Лесли–Брауна	0.149	0.083	0.021	0.139	0.075	0.018
Блисса–Кокрена–Тьюки	0.147	0.082	0.021	0.136	0.075	0.019

Таблица 5. Мощность многовыборочных критериев относительно конкурирующей гипотезы $H_2: \sigma_m = 1.2\sigma_1$, $n_i=100$, $i=\overline{1,m}$

Критерий	α					
	0.1	0.05	0.01	0.1	0.05	0.01
	$m=3$			$m=5$		
Кокрена	0.609	0.494	0.286	0.624	0.515	0.316
О’Брайена	0.583	0.461	0.247	0.575	0.460	0.258
Z–критерий Оверолла–Вудворда	0.583	0.461	0.246	0.565	0.445	0.241
Неймана–Пирсона	0.580	0.457	0.240	0.557	0.434	0.228
Бартлетта	0.577	0.459	0.237	0.557	0.434	0.227
Модифицированный Z–критерий	0.574	0.449	0.232	0.554	0.433	0.228
Хартли	0.568	0.443	0.217	0.545	0.418	0.204
Левене	0.530	0.409	0.200	0.513	0.390	0.197
Блисса–Кокрена–Тьюки	0.359	0.187	0.068	0.262	0.170	0.061
Кадуэлла–Лесли–Брауна	0.280	0.180	0.061	0.253	0.158	0.052

На первой позиции с явным преимуществом, как и в [2], находится критерий Кокрена. На второй оказывается критерий О’Брайена, однако в случае анализа 3-х выборок и близких конкурирующих гипотез он не имеет заметного преимущества по сравнению с Z–критерием Оверолла–Вудворда, Неймана-Пирсона и Бартлетта.

В то же время критерий О’Брайена явно мощнее модифицированного Z–критерия и критерия Левене, также являющихся устойчивыми к нарушению стандартного предположения о нормальности.

Таблица 6. Мощность многовыборочных критериев относительно конкурирующей гипотезы H_3: $\sigma_m = 1.5\sigma_1$, $n_i = 100$, $i = \overline{1,m}$

Критерий	α					
	0.1	0.05	0.01	0.1	0.05	0.01
	$m = 3$			$m = 5$		
Кокрена	0.997	0.994	0.974	0.998	0.997	0.987
О’Брайена	0.996	0.990	0.961	0.997	0.994	0.976
Z–критерий Оверолла–Вудворда	0.996	0.991	0.964	0.997	0.993	0.974
Неймана–Пирсона, Бартлетта	0.996	0.990	0.962	0.996	0.992	0.970
Модифицированный Z–критерий	0.995	0.989	0.955	0.996	0.991	0.967
Хартли	0.995	0.988	0.947	0.995	0.989	0.955
Левене	0.990	0.979	0.926	0.991	0.982	0.944
Блисса–Кокрена–Тьюки	0.820	0.728	0.501	0.829	0.742	0.524
Кадуэлла–Лесли–Брауна	0.795	0.691	0.444	0.783	0.675	0.432

11. Заключение

Как следует, в том числе из проведенных исследований, корректность статистических выводов, осуществляемых при проверке гипотез с использованием параметрических критериев однородности дисперсий, напрямую зависит от знания закона распределения статистики при справедливости проверяемой гипотезы H_0. Зачастую, даже при выполнении стандартного предположения о принадлежности анализируемых выборок нормальным законам, необходимое распределение статистики неизвестно или отличается от асимптотического, что не позволяет оценить достигнутый уровень значимости (**p**-value).

При нарушении стандартного предположения о нормальности и принадлежности выборок некоторому другому виду закона распределения статистик рассматриваемых критериев, как правило, неизвестны. Следовательно, в таком случае тем более нельзя сформировать вывод о результатах проверки гипотезы.

Однако, если предположения о виде предполагаемого закона можно обосновать, то задача уже не становится неразрешимой. Так как, опираясь на методы статистического моделирования (и соответствующее программное обеспечение [15]), требуемые для принятия решения распределения статистик критериев (при нормальном или “ненормальном” законе) могут быть найдены в процессе проводимого анализа, как это осуществляется в [19-23], в том числе в интерактивном режиме [20-23].

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. *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.
3. *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.
4. *Лемешко Б.Ю., Лемешко С.Б., Постовалов С.Н., Чимитова Е.В.* Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход. – Новосибирск : Изд-во НГТУ, 2011. – 888 с.
5. *Gorbunova A.A., Lemeshko B.Yu.* Application of Parametric Homogeneity of Variances Tests under Violation of Classical Assumption // Proceedings, 2nd Stochastic Modeling Techniques

and Data Analysis International Conference. 5 - 8 June 2012, Chania, Crete, Greece. – P.253-260. (http://www.smtda.net/images/1_SMTDA2012_Proceedings_D-J_119-338.pdf)
6. *Кобзарь А.И.* Прикладная математическая статистика : для инженеров и научных работников. – М. : ФИЗМАТЛИТ, 2006. – 816 с.
7. *O'Brien R.G.* Robust techniques for testing heterogeneity of variance effects in factorial designs // Psychometrika. – 1978. – Vol. 43, No. 3. – P. 327-342.
8. *Link R.F.* The sampling distribution of the ratio of two ranges from independent samples // The annals of mathematical statistics. – 1950. – Vol. 21, No. 1. – P. 112-116.
9. *Newman D.* The distribution of range in samples from a normal population, expressed in terms of an independent estimate of standard deviation // Biometrika. – 1939. – Vol. 31. – P. 20-30.
10. *Bliss C.I., Cochran W.G., Tukey J.W.* A rejection criterion based upon the range // Biometrika. – 1956. – Vol. 43. – P. 418-422.
11. *Leslie R.T., Brown B.M.* Use of range in testing heterogeneity of variance // Biometrika. – 1966. – Vol. 53. – P. 221-227.
12. *Overall J.E., Woodward J.A.* A simple test for heterogeneity of variance in complex factorial design // Psychometrika. – 1974. – Vol. 39. – P. 311-318.
13. *Overall J.E., Woodward J.A.* A robust and powerful test for heterogeneity of variance // University of Texas Medical Branch Psychometric Laboratory. – 1976.
14. *Лемешко Б.Ю., Постовалов С.Н.* Компьютерные технологии анализа данных и исследования статистических закономерностей : учебное пособие. – Новосибирск : НГТУ, 2004. – 120 с.
15. ISW – Программная система статистического анализа одномерных случайных величин. URL: http://ami.nstu.ru/~headrd/ISW.htm (дата обращения 06.01.2016)
16. *Lee H.B., Katz G.S., Restori A.F.* A Monte Carlo Study of Seven Homogeneity of Variance Tests // Journal of Mathematics and Statistics. – 2010. – Vol. 6, No. 3. – P. 359-366.
17. *Conover W.J., Johnson M.E., Johnson M.M.* A comparative study of tests for homogeneity of variances, with applications to the outer continental shelf bidding data // Technometrics. – 1981. – Vol. 23, No. 4. – P. 351–361.
18. *Lim T.-S., Loh W.-Y.* A comparison of tests of equality of variances // Computational Statistics & Data Analysis. – 1996. – Vol. 22, No. 5. – P. 287–301.
19. *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.
20. *Lemeshko B.Yu., Lemeshko S.B., Rogozhnikov A.P.* Real-Time Studying of Statistic Distributions of Non-Parametric Goodness-of-Fit Tests when Testing Complex Hypotheses // Proceedings of the International Workshop “Applied Methods of Statistical Analysis. Simulations and Statistical Inference” – AMSA’2011, Novosibirsk, Russia, 20-22 September, 2011. – P. 19-27.
21. *Лемешко Б.Ю.* Непараметрические критерии согласия. Руководство по применению. – М.: ИНФРА-М, 2014. – 163 с. DOI: 10.12737/11873
22. *Лемешко Б.Ю.* Критерии проверки отклонения распределения от нормального закона. Руководство по применению. – М.: ИНФРА-М, 2015. – 160 с. DOI: 10.12737/6086
23. *Лемешко Б.Ю., Блинов П.Ю.* Критерии проверки отклонения распределения от равномерного закона. Руководство по применению. – М.: ИНФРА-М, 2015. – 183 с. DOI: 10.12737/11304

Лемешко Борис Юрьевич
Д.т.н., профессор, г.н.с. кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: Lemeshko@ami.nstu.ru

Сатаева Татьяна Сергеевна
Магистрант кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: t.sataeva13@gmail.com.

On the properties and power of parametric tests of homogeneity of variances

B. Yu. Lemeshko, T. S. Sataeva

Novosibirsk State Technical University

Distributions of test statistics of classical tests for homogeneity of variance (Neyman–Pearson, O'Brien, Link, Newman, Bliss–Cochran–Tukey, Cadwell–Leslie–Brown, Overall–Woodward Z-variance and modified Overall–Woodward Z-variance tests) are investigated including a case when the standard assumption of the normality is violated. The comparative analysis of power of the classical tests is carried out. Method of application of the tests of violation of the standard assumption that provides an interactive simulation of distributions of the test statistics is proposed and tested.

Key words: test homogeneity variances, Neyman–Pearson test, O'Brien test, Link test, Newman test, Bliss–Cochran–Tukey test, Cadwell–Leslie–Brown test, Overall–Woodward Z-variance test, modified Overall–Woodward Z-variance test, test power