# СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ КАК СРЕДСТВО ОБЕСПЕЧЕНИЯ КОРРЕКТНОСТИ ВЫВОДОВ ПРИ ИСПОЛЬЗОВАНИИ КРИТЕРИЕВ ОДНОРОДНОСТИ ДИСПЕРСИЙ В НЕСТАНДАРТНЫХ УСЛОВИЯХ

Б. Ю. Лемешко, И. В. Веретельникова

*Новосибирский государственный технический университет, 630073, Новосибирск*

УДК 519.24

Даже в условиях классического предположения о принадлежности анализируемых выборок нормальному закону распределения статистик ряда параметрических критериев проверки однородности дисперсий представлены лишь ограниченными таблицами критических значений. Вследствие этого исключается возможность по результатам проверки гипотезы оценить достигнутый уровень значимости $p_{value}$, снижается обоснованность вывода о результатах проверки. При нарушении стандартного предположения о нормальности возможность использования классических результатов полностью исключается. В то же время параметрические критерии однородности имеют явное преимущество в мощности перед непараметрическими аналогами, в том числе в условиях нарушений стандартного предположения о нормальности. Опираясь на статистическое моделирование распределений статистик критериев, в работе решаются проблемы корректного использования критериев однородности дисперсий в нестандартных условиях с оценкой $p_{value}$.

**Ключевые слова:** статистическое моделирование, проверка однородности дисперсий, распределение статистики, мощность, уровень значимости.

## Введение

Одной из наиболее часто решаемых в приложениях задач является проверка гипотез об однородности дисперсий $k$ выборок, имеющая вид $H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_n^2$ .

Для проверки такой гипотезы предложено множество параметрических (Бартлетта, Кокрена, Фишера, Хартли, Левене, Неймана — Пирсона, О'Брайена, Ньюмана, Z–критерий Оверолла — Вудворда и его модификация, Блисса — Кокрена — Тьюки, Кадуэлла — Лесли — Брауна, Линка) и непараметрических (Ансари — Бредли, Муда, Сижела — Тьюки, Кейпена, Клотца, Флайне — Киллина) критериев [1].

При выборе критерия ориентируются на его мощность и на возможность обоснованного принятия решения, в том числе, при возможном нарушении стандартных предположений. При этом следует учитывать следующие известные факты [1].

1. Параметрические критерии (по крайней мере, лучшие их представители) имеют явное преимущество в мощности по сравнению с непараметрическими.

2. Стандартным предположением, обуславливающим возможность применения параметрических критериев однородности дисперсий, является принадлежность анализируемых выборок нормальному закону распределения. В случае его нарушении распределения статистик критериев, соответствующие справедливости $H_0$, существенно изменяются.

3. 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 нельзя оценить достигнутый уровень значимости $p_{value}$.

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).

4. При ограниченных объёмах выборок распределения статистик параметрических критериев зачастую существенно отличаются от известных асимптотических распределений этих статистик.

5. 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.

6. Распределения нормализованных статистик непараметрических критериев являются дискретными и при малых объёмах выборок существенно отличаются от асимптотического стандартного нормального закона.

7. Параметрические критерии имеют очевидное преимущество в мощности перед непараметрическими и в условиях нарушения стандартного предположения (в случае законов, отличающихся от нормального). Это заставляет задуматься о возможности их корректного применения при использовании в условиях нарушения стандартного предположения.

Таким образом, для построения корректного статистического вывода по результатам проверки гипотезы 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. Реализация такой же возможности в случае использования критериев в условиях нарушения стандартного предположения о нормальности случайных величин существенно расширяет сферу применения параметрических (и непараметрических) критериев однородности дисперсий.

## 1 Вычисление достигнутого уровня значимости

Принятие решения о результатах проверки гипотезы $H_0$ на основании значения $p_{value}$ всегда более обосновано, чем в результате сравнения полученного значения статистики $S^*$ с заданными критическими значениями, извлекаемыми из соответствующей таблицы процентных точек.

В случае правостороннего критерия достигнутый уровень значимости определяется соотношением

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

где $G(S \mid H_0)$ — функция распределения вероятностей статистики применяемого критерия при справедливости $H_0$.

В случае двустороннего критерия критическая область состоит из двух частей. А $p_{value}$ определяется соотношением

$$p_{value} = 2\min\{G(S^* \mid H_0), 1 - G(S^* \mid H_0)\}. \tag{2}$$

Вычисление $p_{value}$ в соответствии с соотношениями (1) для правостороннего критерия или (2) для двустороннего не вызывает труда при известном распределении статистики критерия. Если информация о распределении статистики соответствующего критерия отсутствует и представлена лишь таблицей процентных точек, либо объёмы выборок относительно невелики и таковы, что распределение статистики существенно отличается от предельного (асимптотического), то корректное вычисление $p_{value}$ представляет собой некоторую проблему.

К сожалению, распределения большинства параметрических критериев однородности дисперсий (даже в классическом случае принадлежности выборок нормальному закону) существенно зависят от объемов выборок. Поэтому при формировании решения о результатах проверки гипотезы $H_0$ приходится опираться на таблицы процентных точек, которые сформированы для ограниченных наборов $n_i$ и, часто, подразумевают равенство объёмов сравниваемых выборок.

Каким же образом можно повысить качество статистических выводов? 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 данном объёме выборки $n$), появляется возможность исследования распределения статистики в реальном времени проверки гипотезы (в интерактивном режиме) [1]. 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 $p_{value}$.

При таком подходе необходимое для проверки гипотезы эмпирическое распределение $G_N(S_n \mid H_0)$ статистики соответствующего критерия строится в результате статистического моделирования с точностью, зависящей от числа экспериментов $N$ в методе Монте — Карло [2]. Затем по эмпирическому распределению $G_N(S_n \mid H_0)$ и вычисленному по анализируемой выборке значению статистики $S^*$ критерия в соответствии с соотношениями (1) или (2) 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 (как в [3]) в целях ускорения распараллеливать процессы моделирования и привлекать доступные вычислительные ресурсы.

## 2 Применение критериев в "нестандартных" условиях

Использование интерактивного режима для исследования распределений статистик открывает возможность применения критериев в условиях нарушения стандартного предположения о принадлежности результатов измерений нормальному закону. Отклонение от нормальности приводит к существенным изменениям распределений $G(S \mid H_0)$ статистик критериев. В меньшей степени это касается критерия О'Брайена, модифицированного Z–критерия Оверолла — Вудворда и модификаций критерия Левене. Однако за эту устойчивость данные критерии платят некоторым снижением мощности (при выполнении стандартных предположений). В то же время распределения $G(S \mid H_0)$ статистик этих трёх критериев всё-таки настолько отклоняются от имеющих место при стандартных предположениях (в случае принадлежности выборок законам с тяжелыми хвостами), что пренебрегать этим нельзя. Поэтому корректность выводов зависит от того, насколько точно знания о распределении $G(S \mid H_0)$ соответствуют реальным условиям, характеризующим анализируемые результаты измерений.

Для рассмотренных критериев однородности дисперсий проиллюстрируем использование интерактивного режима исследования $G(S \mid H_0)$ и точности оценивания $p_{value}$ в зависимости от числа экспериментов 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 [2].

**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:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.205 | 0.232 | -0.219 | 0.829 | 0.127 | 0.939 | 0.995 | 0.706 | -0.450 | -0.361 |
| -0.364 | -0.107 | 1.054 | -0.095 | -2.188 | 0.453 | -1.052 | 0.640 | -0.417 | -2.144 |
| -3.473 | -0.857 | -0.678 | 0.070 | -1.139 | 0.574 | 0.409 | 0.206 | 0.184 | 1.273 |
| -0.326 | -1.245 | 0.227 | 0.185 | 0.383 | 0.126 | 0.255 | 1.110 | -0.310 | -0.178 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.269 | -0.187 | -0.013 | -1.248 | -0.247 | -0.541 | 1.209 | -2.814 | 0.575 | -0.452 |
| -0.427 | 0.337 | 1.138 | -1.090 | -0.858 | -0.006 | -1.212 | -0.180 | 1.751 | -0.485 |
| -0.779 | -0.752 | 0.342 | -0.175 | 0.509 | 0.209 | 0.596 | 1.869 | 1.764 | 1.084 |
| 0.995 | 0.633 | 0.003 | -0.642 | -1.225 | -0.115 | -1.543 | 0.137 | -1.290 | 2.189 |

В таблице 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 законе Лапласа |
| | | Теоретическая | $N = 10^4$ | $N = 10^6$ | $N = 10^6$ |
| Бартлетта | 0.268028 | 0.604658 | 0.605 | 0.6045 | 0.734 |
| Кокрена | 0.541643 | – | 0.605 | 0.6045 | 0.734 |
| Фишера | 0.846236 | 0.604671 | 0.596 | 0.6045 | 0.734 |
| Хартли | 1.1817 | – | 0.605 | 0.6045 | 0.734 |
| Неймана — Пирсона | 1.00349 | 0.607 | 0.605 | 0.6045 | 0.734 |
| Z–критерий Оверолла — Вудворда | 0.279266 | 0.597183 | 0.605 | 0.6045 | 0.734 |
| Модифицированный Z–критерий | 0.115111 | 0.734398 | 0.741 | 0.7348 | 0.732 |
| О'Брайена | 0.162623 | 0.687856 | 0.702 | 0.6971 | 0.722 |
| Левене (со средним) | 0.604953 | – | 0.454 | 0.4451 | 0.459 |
| Ньюмана | 4.56411 | – | 0.780 | 0.7777 | 0.730 |
| Линка | 0.948631 | – | 0.806 | 0.8079 | 0.877 |
| Блисса — Кокрена — Тьюки | 0.513181 | – | 0.814 | 0.8084 | 0.878 |
| Кадуэлла — Лесли — Брауна | 1.05415 | – | 0.814 | 0.8084 | 0.878 |

В действительности обе анализируемые выборки моделировались в соответствии с распределением Лапласа 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 закону Лапласа.

С одной стороны, можно заметить существенное различие в оценках $p_{value}$ при законах Лапласа и нормальном. С другой стороны, можно отметить, что в случае устойчивых критериев Левене, О'Брайена и модифицированного Z–критерия различие в таких оценках минимально.

Необходимо подчеркнуть 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.

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

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 0.254 | -0.254 | -0.017 | 0.002 | 1.937 | -2.476 | -0.092 | -0.543 | 2.588 | 1.970 |
| 1.869 | 0.453 | -0.616 | -2.806 | 2.382 | 0.476 | 0.641 | -2.581 | -0.659 | -0.027 |
| 1.775 | 2.154 | -1.801 | -0.774 | -0.522 | 1.413 | -0.042 | -0.175 | -0.929 | 0.664 |
| -0.298 | 0.409 | 0.040 | 0.418 | 0.478 | -0.052 | -4.354 | 1.521 | -2.126 | 1.177 |

Эта выборка также смоделирована в соответствии с распределением Лапласа, но при $\sigma = 1.5$.

В таблице 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 представлены теоретические оценки $p_{value}$ и оценки, полученные по результатам статистического моделирования при количестве экспериментов $10^6$. В предположении о принадлежности выборок закону Лапласа приведены полученные оценки $p_{value}$ при $N = 10^6$.

В данном случае 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 законе Лапласа. 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.

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 законе Лапласа |
| | | Теоретическая | $N = 10^6$ | $N = 10^6$ |
| Бартлетта | 9.72943 | 0.0077 | 0.0079 | 0.1198 |
| Кокрена | 0.534165 | – | 0.0032 | 0.0667 |
| Хартли | 2.50172 | – | 0.0140 | 0.1508 |
| Неймана — Пирсона | 1.08774 | – | 0.0079 | 0.1198 |
| Z–критерий Оверолла — Вудворда | 5.0054 | 0.0067 | 0.0065 | 0.1098 |
| Модифицированный Z–критерий | 2.31571 | 0.0987 | 0.0953 | 0.0897 |
| О'Брайена | 3.2241 | 0.0434 | 0.0396 | 0.0336 |
| Левене (со средним) | 2.82473 | – | 0.0661 | 0.0713 |
| Блисса — Кокрена — Тьюки | 0.415913 | – | 0.0861 | 0.3301 |
| Кадуэлла — Лесли — Брауна | 1.46271 | – | 0.1899 | 0.5106 |

Таким образом, параметрические критерии однородности дисперсий, обладающие наибольшей мощностью, можно корректно применять как при выполнении стандартного предположения о нормальности, так и в условиях его нарушения. И в том и другом случае возможность вычисления оценок $p_{value}$ повышает информативность статистических выводов.

Реализация такой процедуры применения критериев возможно только с опорой на программное обеспечение. Обязательным предварительным условием перехода к ней является идентификация вида закона распределения, наилучшим образом описывающего анализируемые выборки [4].

Решение о наиболее предпочтительной модели закона может лежать за рамками задачи проверки гипотезы об однородности, а при отсутствии такой возможности, приниматься в процессе анализа совокупности исследуемых выборок (или объединённой выборки). Если при этом в качестве наилучшей модели оказывается некоторое семейство распределений (например, обобщённый нормальный закон, семейства гамма- и бета-распределений и т.п.), в случае которого конкретный вид закона определяется значением параметра (или параметров) формы, то модель должна быть идентифицирована с точностью до значения этого параметра (его оценка должна быть найдена и зафиксирована).

## 3 Замечания о мощности критериев

В случае анализа двух выборок и выполнении стандартного предположения о нормальности наибольшей и одинаковой мощностью относительно тех же конкурирующих гипотез обладают 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, которая уступает 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, потом критерий Муда, который уже заметно уступает критериям Фишера, Бартлетта, Кокрена, Хартли, Неймана — Пирсона, Z–критерию Оверолла — Вудворда, О'Брайена, модифицированному Z-критерию и Левене.

В случае нарушения стандартного предположения и принадлежности двух выборок законам с более легкими хвостами, чем у нормального закона, вышеуказанный порядок предпочтительности сохраняется.

При симметричных законах с более тяжелыми хвостами по сравнению с нормальным законом критерии упорядочиваются следующим образом:

Флайне — Киллина $\succ$ Клотца $\succ$ Муда $\succ$ Левене $\succ$ Сижела — Тьюки $\sim$ Ансари — Бредли $\succ$ О'Брайена $\succ$ Модифицированный Z–критерий $\succ$ **группа критериев** (Бартлетта, Кокрена, Хартли, Фишера, Неймана — Пирсона, Z–критерий Оверолла — Вудворда) $\succ$ Ньюмана $\succ$ **группа критериев** (Блисса — Кокрена — Тьюки, Кадуэлла — Лесли — Брауна, Линка).

Если анализируется большее число выборок, то при выполнении стандартного предположения или в случае принадлежности выборок законам с более лёгкими хвостами по сравнению с нормальным законом критерии по убыванию мощности располагаются в следующем порядке:

Кокрена $\succ$ О'Брайена $\succ$ Z–критерий Оверолла — Вудворда $\succ$ Модифицированный Z–критерий $\succ$ **группа критериев** (Бартлетта, Неймана — Пирсона) $\succ$ Хартли $\succ$ Левене $\succ$ Флайне — Киллина $\succ$ Блисса — Кокрена — Тьюки $\succ$ Кадуэлла — Лесли — Брауна.

Надо отметить, что при симметричных законах с более лёгкими хвостами критерий Левене уступает в мощности критерию Флайне — Киллина.

При симметричных законах с более тяжелыми хвостами ситуация меняется:

Флайне — Киллина $\succ$ Левене $\succ$ О'Брайена $\succ$ Модифицированный Z–критерий $\succ$ **группа критериев** (Бартлетта, Неймана — Пирсона) $\succ$ Z–критерий Оверолла — Вудворда $\succ$ Хартли $\succ$ Кокрена $\succ$ Кадуэлла — Лесли — Брауна $\succ$ Блисса — Кокрена — Тьюки.

## Заключение

Основной недостаток параметрических критериев связан с тем, что классические результаты обеспечивают корректность применения данных критериев лишь при выполнении предположения о принадлежности анализируемых выборок нормальным законам, так как только для этой ситуации известны распределения статистик рассмотренных параметрических критериев или таблицы процентных точек.

Этот недостаток можно преодолеть, используя компьютерные технологии для исследования распределений статистик и оценки $p_{value}$ и распространив действие параметрических критериев на ситуации, когда выборки описываются законами, отличающимися от нормального.

При таком подходе необходимо учитывать, что распределение статистики критерия будет зависеть от вида закона, объема выборок и во многих случаях — от конкретных значений некоторых параметров. При использовании соответствующего программного обеспечения решение таких задач не вызывает принципиальных трудностей, а наличие программного обеспечения позволяет решать эти задачи по мере возникновения потребности [3].

Заметим, что специфика задач по исследованию методами компьютерного моделирования вероятностных закономерностей позволяет эффективно распараллеливать вычислительные процессы, используя ресурсы многоядерных и многопроцессорных компьютеров и компьютерных сетей и получать искомое решение практически в реальном масштабе времени.

## Список литературы

[1] Лемешко Б.Ю. Критерии проверки гипотез об однородности. Руководство по применению. — М. : ИНФРА—М, 2017. — 208 с. DOI: 10.12737/22368

[2] Лемешко Б.Ю., Постовалов С.Н. Компьютерные технологии анализа данных и исследования статистических закономерностей: Учебное пособие. — Новосибирск: Изд-во НГТУ, 2004. — 119 с.

[3] Программная система статистического анализа одномерных случайных величин — ISW. URL: http://ami.nstu.ru/ headrd/ISW.htm (дата обращения 12.03.2017)

[4] Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: Монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. — Новосибирск : Изд-во НГТУ, 2011. — 888 с.

*Борис Юрьевич Лемешко — д.т.н., профессор, главный научный сотрудник*
*Новосибирский государственный технический университет;*
*e-mail: lemeshko@ami.nstu.ru;*
*Ирина Викторовна Веретельникова — аспирант, младший научный сотрудник*
*Новосибирский государственный технический университет;*
*e-mail: ira-veterok@mail.ru.*
*Дата поступления — 28 апреля 2017 г.*