

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ КЛАССИЧЕСКИХ КРИТЕРИЕВ МНОЖЕСТВЕННЫХ СРАВНЕНИЙ, НА НЕНОРМАЛЬНЫХ НЕОДНОРОДНЫХ РАСПРЕДЕЛЕНИЯХ, МЕТОДОМ МОНТЕ-КАРЛО

© 2021 В. Н. Алдобаев^{1✉}, А. Д. Артемьева², А. А. Масликов²

¹Научно-исследовательский центр токсикологии и гигиенической регламентации биопрепаратов — филиал Государственного научного центра «Институт иммунологии» ул. Ленина, 102А, 142253 пос. Большевик, Московская обл., Российская федерация

²Филиал «Протвино» государственного университета «Дубна» Северный проезд, 9, 142281 Протвино, Российская федерация

Аннотация. Традиционные аналитические методы исследования применимости методов множественных сравнений эффективны только при весьма жёстких ограничениях на соответствующие генеральные совокупности. В то же время для решения этого вопроса с успехом можно применять компьютерные симуляции и метод Монте-Карло. Методом Монте-Карло мы симулируем проведение тестов, выполняемых при множественных сравнениях на выборках малого объёма из искаженных (по сравнению с нормальным) распределений. Исследуется возможность применения классических критериев дисперсионного анализа (ANOVA) и непараметрического теста Краскела — Уоллиса для выборок малого объёма с ненормальным распределением и/или неоднородных по дисперсии. В качестве критерия однородности выборок по дисперсиям используется тест Левене. Нормальность (Гауссовость) выборок проверяется с помощью теста Шапиро – Уилка. Для искажения нормальности выборок используются генеральные совокупности, распределённые по хи-квадрат и t -распределению Стьюдента с малым числом степеней свободы. Также ненормальность распределений отслеживается с помощью параметров: эксцесс (коэффициент островершинности) и асимметрия. В качестве генератора псевдослучайных чисел применяется так называемый вихрь Мерсенна реализованный в рамках пакета программ Wolfram Mathematica. Число испытаний для каждого набора параметров доведено до миллиона. Вычисляются эффективные вероятности ошибок 1-го рода и делаются выводы о влиянии негомогенности дисперсий, «ненормальности» эксцесса и асимметрии на эффективность исследуемых критериев. В результате можно сказать, что зачастую нет оснований использовать непараметрические методы вместо параметрических в ущерб мощности соответствующих критериев.

Ключевые слова: математическая статистика, метод Монте-Карло, множественные сравнения, дисперсионный анализ, тест Краскела — Уоллиса.

ВВЕДЕНИЕ

В медицине, фармакологии и токсикологии, в частности при проведении доклинических и клинических испытаний фармацевтической продукции приходится иметь дело с множественными сравнениями усреднённых

по группам различных физиологических параметров (показателей), например, биохимические параметры крови и мочи, гематологические показатели крови, показатели функции дыхания и кровообращения, различные поведенческие показатели и массовые индексы внутренних органов в случае животных и т. д.

В соответствии с различными форматами экспериментов группы для сравнения формируются с учётом различных внешних факторов, априори влияющих на величины срав-

✉ Алдобаев Владимир Николаевич
e-mail: aldobaev@toxicbio.ru



Контент доступен под лицензией Creative Commons Attribution 4.0 License.
The content is available under Creative Commons Attribution 4.0 License.

ниваемых параметров, и зачастую в исследовании параллельно участвуют три и более групп. Обычно в таких случаях для сопоставления сразу всех средневыворочных параметров принято использовать классические методы множественного сравнения средних дисперсионный анализ (ANOVA) и его непараметрический аналог критерий Краскела — Уоллиса (Kruskal — Wallis) [1–3].

Как правило, при проведении доклинических и клинических испытаний в математико-статистических исследованиях приходится иметь дело с выборками небольшого объема, для которых снижается мощность критериев, устанавливающих нормальность распределения элементов выборок и/или однородность дисперсий соответствующих генеральных совокупностей. В тоже время общепринято [1, 2, 4], что условием надёжной работы большинства классических статистических тестов, в том числе дисперсионного анализа и критерия Краскела — Уоллиса, является выполнение этих условий. В данной работе, мы постарались изучить насколько сильно отклонение от нормальности и негомогенность влияют на работу популярных в математической статистике критериев. Исследовалась возможность применения классических критериев дисперсионного анализа (ANOVA) и непараметрического теста Краскела — Уоллиса для выборок малого объёма с ненормальным распределением и/или неоднородных по дисперсии.

1. МАТЕРИАЛЫ И МЕТОДЫ

В качестве инструмента мы использовали программы, написанные в пакете Wolfram Mathematica (WM). В основе исследования лежал метод Монте-Карло, использованный для генерации случайных выборок из различных распределений. Далее к выборкам применяли исследуемые тесты и проверяли, насколько частоты попадания результатов в критическую область критериев близки к теоретически ожидаемым частотам. WM по умолчанию для генерации высококачественных псевдослучайных чисел применяет клеточный автомат «ExtendedCA». Этот генератор исполь-

зует определенное правило пяти соседей, и каждая новая ячейка зависит от пяти несмежных ячеек из предыдущего шага. Также использовался генератор «Mersenne Twister» следуя алгоритму Мацумото и Нисимуре [5, 6]. Вихрь Мерсенна — это генератор сдвигового регистра обобщенной обратной связи с огромным периодом ($2^{19937} - 1$), гарантирующим высочайшую степень случайности.

С помощью пакета WM формировались компактные программы для генерации представительного числа случайных выборок из широкого набора библиотечных распределений. Мы использовали количество генераций $10^5 - 10^6$ для каждого набора параметров. Далее, к этим выборкам применялись интересующие тесты и методом Монте-Карло получали статистически достоверные результаты. Исследовались малые выборки объёмом 5–10 элементов. Отклонения исследуемых распределений от нормального (Гауссова) распределения описываются величинами: эксцесс (коэффициент островершинности) (E) и асимметрия (A). Для нормального распределения $E = 3$, $A = 0$.

2. РЕЗУЛЬТАТЫ ИССЛЕДОВАНИЯ

2.1. Проверка теста ANOVA на критичность применения к неоднородным по дисперсиям выборкам

В классической формулировке тест ANOVA (множественные сравнения выборок на основе дисперсионного анализа) требует нормальности и однородности по дисперсиям выборок [1,4]. Исследовалось, насколько критична для работоспособности ANOVA будет утрата выборками однородности.

Мы исследовали случаи 3-х и 4-х выборок с кардинально нарушенной гомогенностью дисперсий. Рассматривались выборки из нормальных распределений с одинаковым математическим ожиданием (МО) и геометрически растущими среднеквадратическими отклонениями (СКО): $\sigma, 3\sigma, 9\sigma$ для случая 3-х выборок, и $\sigma, 2\sigma, 4\sigma, 8\sigma$ для случая 4-х выборок. К наборам выборок применялись тесты Левене (Levene) и проводился дисперсион-

ный анализ. Результаты, полученные для объёмов выборок от 5 до 8, представлены в табл. 1.

Метод Монте-Карло, состоящий на первом этапе в применении теста Левене к 10^6 троек или четвёрок таких выборок позволяет сделать вывод, что 70–95 % этих выборок не проходят тест на гомогенность дисперсий на уровне значимости $\alpha = 0.05$. Естественно, с ростом объёма выборок увеличивается достоверность обнаружения негомогенности. Таким образом, налицо сильное нарушение однородности по дисперсиям. На втором этапе на этом же материале проводился дисперсионный анализ (ANOVA) в стандартной форме с $\alpha = 0.05$. В результате получили, что неоднородность даже при столь малых объёмах выборок как 5 элементов приводит к эффективному увеличению вероятности ошибки 1-го рода максимум до 10 % с 5 % в случае гомогенности дисперсий. При этом наблюдается устойчивая тенденция к уменьшению вероятности ошибки 1-го рода с ростом объёмов выборок.

2.2. Проверка тестов ANOVA и Краскела — Уоллиса на эффективность в применении к ненормальным по распределению выборкам

Были проведены исследования для ANOVA и теста Краскела — Уоллиса с контролем нормальности распределения элементов выборок тестом Шапиро — Уилка (Shapiro — Wilk). Исследовались выборки с отклонениями по эксцессу и асимметрии по сравнению с нормальным распределением. Для этого

в качестве тестовых использовали по 4 выборки различного объёма из распределений Стьюдента (t -распределение), хи-квадрат и логнормального.

Множественные сравнения для выборок из асимметричных распределений (логнормальное, хи-квадрат) исследовали в 2-х вариантах: все 4 выборки асимметричны в одну сторону и асимметрия 2-х выборок противоположна 2-м оставшимся. В последнем случае 2 выборки брались отражёнными относительно математического ожидания (МО) или медианы, чтобы МО или медианы (для непараметрического критерия) оставались неизменными. ANOVA и тест Краскела — Уоллиса применяли к 10^6 наборов четвёрок выборок и, в соответствии с методом Монте-Карло, получали эмпирическую вероятность ошибки 1-го рода. Результаты представлены в табл. 2–4.

Для логнормального распределения были выбраны параметры $\mu = 0, \sigma = 1$, тогда $E = 114$, асимметрия $A = 6.18$, математическое ожидание $MO = \sqrt{e}$, медиана $m = 1$. В этом случае (для объёма выборок 5) методом компьютерных симуляций были получены значения, представленные в табл. 2.

По данным табл. 2 критерий Шапиро-Уилка уверенно указывает на значимое отклонение распределений от нормального, однако, в случае сонаправленности асимметрий, это не сказывается критическим образом на работе тестов ANOVA и Краскела — Уоллиса. Только при разнонаправленных асимметриях заметно возрастает вероятность ошибки 1-го рода.

Далее в качестве тестового использовали распределение хи-квадрат (χ^2). В табл. 3

Таблица 1. Результаты применения тестов Левене и ANOVA к наборам выборок, полученных из нормального распределения

[Table 1. Results of applying the Levene and ANOVA tests to sets of samples obtained from a normal distribution]

Объём выборки	3 выборки		4 выборки	
	Левене (%)	ANOVA (%)	Левене (%)	ANOVA (%)
5	71.0	9.74	72.6	10.1
6	83.0	9.37	83.7	9.79
7	90.8	9.13	91.0	9.59
8	95.3	8.89	95.4	9.48

Таблица 2. Наблюдаемая вероятность ошибки 1-го рода для логнормального распределения при объёме выборок 5

[Table 2. Observed probability of type 1 error for a lognormal distribution with a sample size of 5]

Выборки \ Критерии	Шапиро — Уилка	ANOVA	Краскела — Уоллиса
4 одинаковых логнормальных распределения	0.24	0.033	0.055
2 логнормальных распределения, 2 отражённых логнормальных	0.24	0.100	0.09

Таблица 3. Результаты применения теста Шапиро — Уилка для выборок из распределений χ^2

[Table 3. Results of applying the Shapiro — Wilk test for samples from χ^2 distributions]

Эксцесс E; асимметрия A	Объём выборки df	5	6	7	8	9	10
		E = 6; A = 1.41	4	0.097	0.125	0.154	0.185
E = 5; A = 1.15	6	0.078	0.096	0.117	0.137	0.155	0.174
E = 4.5; A = 1	8	0.070	0.083	0.099	0.114	0.127	0.142

Таблица 4. Результаты применения тестов ANOVA, Краскела — Уоллиса для выборок из распределений χ^2

[Table 4. Results of application of ANOVA, Kruskal — Wallis tests for samples from χ^2 distributions]

Эксцесс E; асимметрия A	Тесты	Объём выборки df	5	6	7	8	9	10
			E = 6; A = 1.41	ANOVA	4	0.0449 0.0580	0.045 0.0572	0.0454 0.0571
Краскела — Уоллиса	0.056 0.068	0.054 0.071		0.053 0.073		0.053 0.077	0.052 0.080	0.051 0.083
E = 5; A = 1.15	ANOVA	6	0.0465 0.0552	0.0465 0.0548	0.0468 0.0550	0.0467 0.0540	0.0475 0.0538	0.0474 0.0538
	Краскела — Уоллиса		0.055 0.063	0.053 0.064	0.053 0.066	0.052 0.068	0.052 0.070	0.052 0.071
E = 4.5; A = 1	ANOVA	8	0.0474 0.0535	0.0470 0.0539	0.0480 0.0530	0.0474 0.0528	0.0477 0.0530	0.0480 0.0530
	Краскела — Уоллиса		0.0553 0.061	0.0540 0.061	0.0529 0.062	0.0528 0.063	0.0520 0.065	0.0517 0.067

представлены результаты для выборок разного объёма из распределений χ^2 с различным числом степеней свободы (df).

В табл. 4 для тестов ANOVA и Краскела — Уоллиса верхний ряд показателей соответствует варианту, когда все 4 выборки имеют одинаковую направленность асимметрии,

нижний ряд показателей – когда половина выборок имеет обратную асимметрию.

Из данных табл. 3, 4 видно, что тест Шапиро — Уилка уверенно обнаруживает отклонение элементов выборок от нормального распределения. В свою очередь, очевидно, что отклонение от нормальности распределений

не сказало критическим образом на работоспособности ANOVA — ошибка 1-го рода мало изменилась, причём в сторону уменьшения. Но в случае разнонаправленных асимметрий ситуация несколько хуже, как было и в предыдущем примере в табл. 2. Результаты для теста Краскела — Уоллиса тоже указывают на то, что нет критического роста ошибки 1-го рода, но в целом этот тест ведёт себя менее устойчиво, чем ANOVA.

На последнем этапе были проведены аналогичные тесты для выборок из распределений Стьюдента. Результаты представлены в табл. 5, 6.

Видно, что теоретические эксцессы рассмотренных распределений Стьюдента заметно выше нормального. Для всех выборок из распределений Стьюдента тест Шапиро — Уилка уверенно обнаруживает отклонение

элементов выборок от нормального распределения (см. табл. 5), в то же время ANOVA и тест Краскела — Уоллиса продолжают надёжно работать, почти не искажая вероятность ошибки 1-го рода (см. табл. 6). Следуя ANOVA, мы излишне рискуем посчитать выборки различающимися, например, в случае $\alpha = 0.049$. А непараметрический его аналог, тест Краскела — Уоллиса, наоборот, в данном случае, оказался более консервативен. Иными словами, есть риск не увидеть различий там, где при выполнении условия нормальности распределений они были бы обнаружены. В целом же, изменение эксцесса (в нашем случае в 3 раза) не критично для ANOVA и теста Краскела — Уоллиса. В этом случае можно рекомендовать их совместное применение. При согласии их выводов заключение можно считать вполне достоверным.

Таблица 5. Результаты применения теста Шапиро — Уилка для выборок из распределений Стьюдента

[Table 5. Results of applying the Shapiro — Wilk test for samples from Student's distributions]

Эксцесс E ; асимметрия A	Объём выборки df	5	6	7	8	9	10
		$E = 9$; $A = 0$;	5	0.066	0.077	0.087	0.097
$E = 6$; $A = 0$;	6	0.061	0.070	0.078	0.087	0.092	0.098
$E = 5$; $A = 0$;	7	0.059	0.065	0.072	0.079	0.084	0.089

Таблица 6. Результаты применения тестов ANOVA, Краскела — Уоллиса для выборок из распределений Стьюдента

[Table 6. Results of using ANOVA, Kruskal — Wallis tests for samples from Student's distributions]

Эксцесс E ; асимметрия A	Тесты	Объём выборки df	5	6	7	8	9	10
			$E = 9$; $A = 0$;	ANOVA	4	0.045	0.0450	0.0460
	Краскела — Уоллиса	0.0554	0.0535	0.0530		0.0520	0.0520	0.0520
$E = 6$; $A = 0$;	ANOVA	6	0.0460	0.0462	0.0470	0.0475	0.0470	0.0450
	Краскела — Уоллиса		0.0555	0.0531	0.0530	0.0526	0.0519	0.0518
$E = 5$; $A = 0$;	ANOVA	7	0.0467	0.0470	0.0471	0.0477	0.0480	0.0480
	Краскела — Уоллиса		0.0552	0.0537	0.0528	0.0525	0.0520	0.0516

3. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Наши исследования и расчёты лежали в русле направления, развиваемого школой профессора Лемешко Б. Ю. [7–12]. Существует проблема в определении области применимости ряда известных статистических критериев, направленных на установление (не)однородности центральных тенденций (математических ожиданий, медиан) и дисперсий множества исследуемых выборок. Традиционные аналитические методы исследования этого вопроса эффективны только при весьма жёстких ограничениях на соответствующие генеральные совокупности. В то же время для решения этого вопроса с успехом можно применять компьютерные симуляции и метод Монте-Карло. Наши исследования отчасти были инспирированы работами [9, 13]. Внимание было сосредоточено на проблемах характерных для математико-статистического анализа в медицине, фармакологии и токсикологии. А именно, в отличие от [13], мы исследовали работоспособность тестов не парных, а множественных сравнений применительно к выборкам малого объема (5–10 элементов) в условиях нарушения стандартных предположений. Нами были использованы оригинальные программы, оформленные в пакете WM и, при этом, экстремально увеличено количество симуляций (до 10^5 – 10^6 , ранее речь шла максимум о 10^4) для каждого набора параметров. Полученные результаты позволяют заключить, что дисперсионный анализ слабо чувствителен к негомогенности дисперсий, а искажение нормального вида распределений для ANOVA и Kruskal — Wallis не приводит к существенному росту вероятности ошибок 1-го рода, а при сонаправленной асимметрии в случае ANOVA может приводить к некоторому её уменьшению.

Ранее при разработке параметрических статистических критериев парного и множественного сравнения выборочных средних таких как t -критерий и ANOVA со множеством *post-hoc* тестов авторы использовали аналитические подходы, основанные на обязательном допущении нормального распределения элементов сравниваемых выборок.

Предполагалось, что в случае нарушения этого допущения, результаты тестов будут о некоторой степени скомпрометированы. При этом границы применимости каждого из тестов с точки зрения степени отклонений элементов сравниваемых выборок от закона Гауссова распределения или степени негомогенности дисперсий в специальной литературе не обсуждались. Для решения этой проблемы позже возник целый раздел непараметрической статистики, включивший такие критерии как Mann — Whitney U test, Wilcoxon test, Kruskal — Wallis, Dunn test и др., в основу которых были положены ранговые преобразования элементов выборок. Считалось, что при небольшой потере мощности непараметрических критериев можно больше не заботиться о распределениях элементов сравниваемых выборок. Следует отметить, что рассматриваемые проблемы возникают в таких областях науки, где затруднен набор больших объёмов данных для статистической обработки, что, в частности, характерно для медицины, фармакологии и токсикологии по экономическим и этическим соображениям. В этой работе, как и в статье [13] были предприняты попытки оценить границы применимости параметрических критериев парного и множественного сравнения выборочных средних в сравнении с непараметрическими. При этом в большинстве исследованных случаев отклонений параметрические критерии дали более корректные и строгие результаты по сравнению с непараметрическими.

ЗАКЛЮЧЕНИЕ

1. Статистическими исследованиями было установлено, что негомогенность дисперсий малых выборок не оказывает критического влияния на работоспособность ANOVA. В нашем примере с геометрическим ростом СКО наблюдаемое возрастание вероятности ошибки 1-го рода всего с 5 % до 10 %.

2. Тесты множественных сравнений (ANOVA, Краскела — Уоллиса) практически нечувствительны к увеличению эксцесса малых выборок при условии сонаправленности асимметрий. При этом наблюдается рост

ошибки 1-го рода в случае с распределениями, имеющими разнонаправленные асимметрии.

3. В целом, изменение эксцесса сравниваемых выборок в 3 раза не оказывает критического влияния на работоспособность ANOVA и теста Краскела — Уоллиса.

4. Следуя ANOVA, мы излишне рискуем посчитать выборки различающимися. А непараметрический тест Краскела — Уоллиса в исследованных условиях демонстрирует большую консервативность. В этом случае можно рекомендовать их совместное применение.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. Шеффе, Г. Дисперсионный анализ / Г. Шеффе. – Главная редакция физико-математической литературы издательства «Наука», 2011. – 512 с.
2. Любичев, А. А. Дисперсионный анализ в биологии / А. А. Любичев. – Изд-во «Книга по Требованию», 2012. – 101 с.
3. Tabachnick, B. G. Using Multivariate Statistics / B. G. Tabachnick, L. S. Fidell. – Pearson Education, Inc.; 2019. – 816 p.
4. Лапач, С. Н. Статистические методы в медико-биологических исследованиях с использованием Excel. 2-е изд. перераб. и доп. / С. Н. Лапач, А. В. Чубенко, П. Н. Бабич. – МОРИОН, 2001. – 408 с.
5. Matsumoto, M. 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator. ACM Transactions on Modeling and Computer Simulation 8 / M. Matsumoto, T. Nishimura. – 1998. – No 1. – P. 3–30.
6. Nishimura, T. Tables of 64-Bit Mersenne Twisters. ACM Transactions on Modeling and Computer Simulation 10 / T. Nishimura. – 2000. – No 4. – P. 348–357. DOI:10.1145/369534.369540
7. Лемешко, Б. Ю. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б. Ю. Лемешко, С. Б. Лемешко, С. Н. Постовалов, Е. В. Чимитова. – Изд-во НГТУ, 2011. – 888 с.
8. Лемешко, Б. Ю. Критерии проверки гипотез об однородности. 2-е изд., перераб / Б. Ю. Лемешко. – ИНФРА-М, 2021. – 248 с.
9. Лемешко, Б. Ю. Об устойчивости и мощности критериев проверки однородности средних / Б. Ю. Лемешко, С. Б. Лемешко. – Измерительная техника. – 2008. – No 9. – С. 23–28.
10. Lemeshko, B. Y. Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses / B. Y. Lemeshko, S. B. Lemeshko, S. N. Postovalov. – Journal of Applied and Industrial Mathematics. – 2009. – Vol. 3, No. 4. – P. 462–475. DOI: 10.1134/S199047890904005X
11. Lemeshko, B. Y. Features of testing statistical hypotheses under Big Data / B. Y. Lemeshko. – NSTU publ., 2019. – P. 122–137.
12. Лемешко, Б. Ю. Критерии проверки статистических гипотез при анализе больших выборок: проблемы и их решение / Б. Ю. Лемешко. – ИПЦ НГУ, 2019. – С. 50–51. DOI: 10.24411/9999-016A-2019-10044
13. Корнеев, А. А. Условия применимости критериев Стьюдента и Манна–Уитни / А. А. Корнеев, А. Н. Кричевец. – Психологический журнал. – 2011. – Т. 32, № 1. – С. 97–110.

Алдобаев Владимир Николаевич — канд. биол. наук, начальник отдела аналитической химии и радиобиологии НИЦ ТБП — филиала ФГБУ «ГНЦ Институт иммунологии» ФМБА России.
E-mail: aldobaev@toxicbio.ru
ORCID 0000-0002-7817-8197

Артемьева Анастасия Дмитриевна — выпускница (бакалавр) Филиала «Протвино» государственного университета «Дубна», кафедры технической физики по профилю медицинская физика.
E-mail: artemeva_ad@mail.ru
ORCID 0000-0003-2822-9811,

Масликов Александр Альбертович — канд. физ.-мат. наук, заведующий комплексной лабораторией физики Филиала «Протвино» государственного университета «Дубна», доцент кафедры общеобразовательных дисциплин, доцент кафедры технической физики.

E-mail: masspref@yandex.ru

ORCID 0000-0002-1495-8506

DOI: <https://doi.org/10.17308/sait.2021.3/3737>

ISSN 1995-5499

Received 01.04.2021

Accepted 20.11.2021

© 2021 V. N. Aldobaev¹✉, A. D. Artemyeva², A. A. Maslikov²

¹Research Center for Toxicology and Hygienic Regulation of Biopreparations —
branch of the Institution State Research Center Institute of Immunology
102A, Lenina Street, 142253 Bolshevik, Moscow region, Russian Federation

²Branch «Protvino» of the State University «Dubna»
9, Severniy proezd, 142281 Protvino, Russian Federation

INVESTIGATION OF THE BEHAVIOR OF CLASSICAL CRITERIA FOR MULTIPLE COMPARISONS, ON ABNORMAL HETEROGENEOUS DISTRIBUTIONS, BY THE MONTE-CARLO METHOD

Annotation. Traditional analytical methods for studying the applicability of multiple comparison methods are effective only under very strict restrictions on the corresponding general populations. At the same time, computer simulations and the Monte Carlo method can be successfully applied to solve this issue. Using the Monte Carlo method, we simulate the performance of tests performed in multiple comparisons on small samples from distorted (compared to normal) distributions. We investigate the possibility of using the classical ANOVA tests and the nonparametric Kruskal — Wallis test for small samples with a non-normal distribution and / or heterogeneous in variance. The Levene test is used as a criterion for the homogeneity of samples in terms of variances. Normality (Gaussianity) of samples is checked using the Shapiro — Wilk test. To distort the normality of samples, general populations are used distributed over the chi-square and Student's t-distribution with a small number of degrees of freedom. Also, the non-normality of the distributions is tracked using the parameters: kurtosis (peakedness coefficient) and skewness. The so-called Mersenne Twister implemented within the Wolfram Mathematica software package is used as a pseudo-random number generator. The number of tests for each set of parameters has been brought to a million. The effective probabilities of errors of the 1st kind are calculated and conclusions are drawn about the influence of inhomogeneity of variances, non-normality of kurtosis and asymmetry on the effectiveness of the studied criteria. As a result, we can say that there is often no reason to use nonparametric methods instead of parametric ones, to the detriment of the power of the corresponding criteria.

Keywords: mathematical statistics, Monte Carlo method, multiple comparisons, analysis of variance (ANOVA), Kruskal — Wallis test.

✉ Aldobaev Vladimir N.
e-mail: aldobaev@toxicbio.ru

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Scheffe G.* Dispersion analysis. [Dispersionnyj analiz] (2011). The main editorial office of the physical and mathematical literature of the publishing house «Nauka». 512 p.
2. *Lyubishchev A. A.* (2012) Dispersion analysis in biology. [Dispersionnyj analiz v biologii]. Publishing house «Book on Demand». 101 p.
3. *Tabachnick B. G.* (2019) Using Multivariate Statistics. Pearson Education, Inc. 816 p.
4. *Lapach S. N., Chubenko A. V., Babich P. N.* (2001) Statistical methods in biomedical research using Excel. [Statisticheskie metody v mediko-biologicheskikh issledovaniyah s ispol'zovaniem Excel]. MORION. 408 p.
5. *Matsumoto M.* (1998) 623-Dimensionally Equidistributed Uniform Pseudorandom Number Generator. ACM Transactions on Modeling and Computer Simulation 8. No 1. P. 3–30.
6. *Nishimura T.* (2000) Tables of 64-Bit Mersenne Twisters. ACM Transactions on Modeling and Computer Simulation 10. No 4. P. 348–357. DOI:10.1145/369534.369540
7. *Lemeshko B. Yu., Lemeshko S. B., Postovalov S. N., Chimitova E. V.* (2011) Statistical data analysis, modeling and investigation of probabilistic patterns. Computer approach. [Statisticheskij analiz dannyh, modelirovanie i issledovanie veroyatnostnyh zakonomernostej. Komp'yuternyj podhod]. NSTU Publishing House. 888 p.
8. *Lemeshko B. Yu.* (2021) Criteria for testing hypotheses about homogeneity. [Kriterii proverki gipotez ob odnorodnosti]. INFRA-M. 248 p.
9. *Lemeshko B. Yu., Lemeshko S. B.* (2008) On the stability and power of the criteria for checking the uniformity of averages. [Ob ustojchivosti i moshhnosti kriteriev proverki odnorodnosti srednih]. «Measuring equipment». No 9. P. 23–28.
10. *Lemeshko B. Y.* (2009) Comparative Analysis of the Power of Goodness-of-Fit Tests for Near Competing Hypotheses. I. The Verification of Simple Hypotheses. Journal of Applied and Industrial Mathematics. Vol. 3, No. 4. P. 462–475. DOI: 10.1134/S199047890904005X
11. *Lemeshko B. Y.* (2019) Features of testing statistical hypotheses under Big Data. NSTU publ. P. 122–137.
12. *Lemeshko B. Yu.* (2019) Criteria for testing statistical hypotheses in the analysis of large samples: problems and their solution. [Kriterii proverki statisticheskikh gipotez pri analize bol'shikh vyborok: problemy i ih reshenie]. IPC NGU. P. 50–51. DOI: 10.24411/9999-016A-2019-10044
13. *Korneev A. A., Krichevets A. N.* (2011) Conditions for the applicability of the Student and Mann — Whitney criteria. [Usloviya primenimosti kriteriev St'judenta i Manna — Uitni]. Psychological Journal. V. 32, No 1. P. 97–110.

Aldobaev Vladimir N. — PhD in Biological Sciences, Head of the Division for Analytical Chemistry and Radiobiology of Research Center for Toxicology and Hygienic Regulation of Biopreparations — branch of the Federal State Budgetary Institution State Research Center Institute of Immunology of the Federal Medico-Biological Agency of Russia. E-mail: aldobaev@toxicbio.ru
ORCID 0000-0002-7817-8197

Artemyeva Anastasia D. — graduate (bachelor) of the Department of Technical Physics in the field of medical physics of the Branch «Protvino» of the State University «Dubna» of Russia.
E-mail: artemeva_ad@mail.ru
ORCID 0000-0003-2822-9811

Maslikov Alexander A. — PhD in Physics and Mathematics Sciences, Head of the complex laboratory of Physics, Associate Professor of the Department of General Education Disciplines, Associate Professor of the Department of Technical Physics of the Branch «Protvino» of the State University «Dubna» of Russia. E-mail: masspref@yandex.ru
ORCID 0000-0002-1495-8506