УДК 519.2 Вестник СПбГУ. Сер. 1. Т. 3 (61). 2016. Вып. 3

ЧИСЛЕННОЕ СРАВНЕНИЕ ПЕРЕСТАНОВОЧНЫХ И КЛАССИЧЕСКИХ МЕТОДОВ ПРОВЕРКИ СТАТИСТИЧЕСКИХ ГИПОТЕЗ*

В. Б. Мелас, Д. И. Сальников, А. О. Гудулина

Санкт-Петербургский государственный университет,
Российская Федерация, 199034, Санкт-Петербург, Университетская наб., 7–9

Статья посвящена классической задаче проверки статистической гипотезы о равенстве двух распределений. Для нормальных распределений во многих смыслах оптимальным является критерий Стьюдента. Но на практике сравниваемые распределения часто не являются нормальными и, вообще говоря, неизвестны. В случае, когда ничего не известно относительно сравниваемых распределений, для решения этой задачи обычно применяется непараметрический критерий Колмогорова—Смирнова. В статье рассматриваются методы, основанные на перестановках, которые в последние годы привлекают внимание своей простотой, универсальностью и достаточно высокой эффективностью. Методами стохастического моделирования проведено сравнительное исследование мощности нескольких перестановочных тестов и классических методов (тесты Колмогорова—Смирнова, Стьюдента и Манна—Уитни) для широкого класса функций распределения. Рассматриваются нормальные распределения, распределения Коши и их смеси, а также экспоненциальные распределения, распределения Вейбулла, Фишера и Стьюдента.

Установлено, что для многих типичных распределений наибольшую мощность имеет перестановочный метод, основанный на сумме абсолютных величин разностей. Особенно велико преимущество этого метода перед остальными в случае, когда сравниваются симметричные распределения с совпадающими центрами. Таким образом, указанный перестановочный метод можно рекомендовать к применению в тех случаях, когда сравниваемые распределения отличны от нормальных. Библиогр. 9 назв. Табл. 5.

Ключевые слова: статистические гипотезы, перестановочные методы.

1. Введение. Задача проверки гипотезы о равенстве двух распределений является классической задачей математической статистики и имеет большой теоретический и практический интерес. Хорошо известно (см., например, [1]), что в случае, когда оба распределения являются нормальными и имеют одинаковые дисперсии, классический тест Стьюдента обладает рядом оптимальных свойств. Но на практике распределения часто не являются нормальными и, вообще говоря, неизвестны. При этом сильную конкуренцию тесту Стьюдента составляют непараметрические тесты, важным классом которых являются тесты, основанные на перестановках.

В работе представлены результаты исследования мощности нескольких перестановочных тестов, а также тестов Стьюдента, Колмогорова—Смирнова и Манна—Уитни.

2. Постановка задачи и описание перестановочных тестов. Рассмотрим классическую задачу проверки нулевой гипотезы

$$H_0 : F_1 = F_2 \tag{1}$$

против альтернативной гипотезы

$$H_1 : F_1 \neq F_2, \tag{2}$$

*Работа выполнена при поддержке СПбГУ (проект 6.38.435.2015).

DOI: 10.21638/11701/spbu01.2016.309

где F_1, F_2 — функции распределения общего вида, по результатам наблюдений

$$Y_1 = (y_{1,1}, y_{1,2}, \ldots, y_{1,n_1}), \ \ Y_2 = (y_{2,1}, y_{2,2}, \ldots, y_{2,n_2}). \tag{3}$$

Для простоты обозначений и без потери общности предположим сбалансированность выборок, т. е. справедливость равенств $n_1 = n_2 = n$ (в случае несбалансированной выборки аргументы очень похожи). Определим векторы

$$Z(\pi_0) = \{y_{11}, \ldots, y_{1n}, y_{21}, \ldots, y_{2n}\}, \tag{4}$$

$$Z(\pi_k) = \{\tilde{y}_{11}, \ldots, \tilde{y}_{1n}, \tilde{y}_{21}, \ldots, \tilde{y}_{2n}\}, \tag{5}$$

$$\begin{aligned} &\tilde{y}_{1i_l} = y_{2j_l}, \quad \tilde{y}_{2i_l} = y_{1j_l}, \quad l = 1, \ldots, k, \\ &\tilde{y}_{1j} = y_{1j}, \quad \tilde{y}_{2j} = y_{2j}, \quad j \neq j_1, \ldots, j_k, \end{aligned} \tag{6}$$

где $\pi_k = \pi_k(s)$, $s = 1, 2, \ldots, (C_n^k)^2$ — различные способы замены k элементов из первой половины на k элементов из второй половины. Обозначим через $Z = Z(\pi_0)$ совокупность векторов (4), $\bar{Y}$ выборочное среднее, Y_{med} медиану и определим на множестве Z критерии $K_i = K_i(Z)$, $i = 1, 2, \ldots, 6$:

$$K_1(Z) = (\bar{Y}_1 - \bar{Y}_2)^2, \quad K_2(Z) = \sum_{i,j=1}^{n} \left(X_{1i}(t) - X_{2j}(t)\right)^2,$$

$$K_3(Z) = \frac{nK_1(Z)}{S^2(Z)}, \quad K_4(Z) = (Y_{1med} - Y_{2med})^2,$$

$$K_5(Z) = \left(\sum_{i=1}^{n} |Y_{1i} - Y_{1med}| - \sum_{i=1}^{n} |Y_{2i} - Y_{2med}|\right)^2, \quad K_6(Z) = \sum_{i,j=1}^{n} |Y_{1i} - Y_{2j}|,$$

где использованы обозначения

$$S^2(Z) = S_1^2(Z) + S_2^2(Z),$$

$$S_1^2(Z) = \frac{1}{N}\sum_{t=1}^{N}\left(\sum_{i=1}^{n}\frac{(X_{1i}(t) - \bar{X}_1(t))^2}{n}\right),$$

$$S_2^2(Z) = \frac{1}{N}\sum_{t=1}^{N}\left(\sum_{i=1}^{n}\frac{(X_{2i})(t) - \bar{X}_2(t)^2}{n}\right).$$

Для $Z = Z(\pi)$, $\pi = \pi_k(s)$, $s = 1, \ldots, (C_n^k)^2$, $k = 1, 2, \ldots, n$ функции $K_1, K_2, \ldots, K_6$ определяются теми же формулами с заменой $Z = Z(\pi_0)$ на $Z = Z(\pi)$. Под перестановочным K_i-тестом проверки гипотезы H_0 будем понимать следующий алгоритм.

Пусть имеем $r_2 = (C_n^k)^2$, $k = n/2$, n — четное, и пусть r_1 — число перестановок π_k, для которых выполняется $K_6(Z(\pi_k)) > K_6(Z(\pi_0))$. Тогда, если справедливо $r_1/r_2 \geq \alpha$ для $K_1, \ldots, K_4, K_6$ и $r_1/r_2 \leq (1 - \alpha)$ для K_5, где α — заданный уровень значимости, нулевая гипотеза не отвергается. Иначе нулевая гипотеза отвергается в пользу альтернативной.

В работе [2] для специального случая задачи проверки гипотез были предложены критерии на основе норм L_1 и L_2, они и составили основу рассматриваемых критериев. В недавней работе [3] было показано, что три метода перестановок, основанные на норме L_2, эквивалентны друг другу.

Мощность критерия K_1 была изучена численными методами в работе [4]. Критерий K_2 был введен в [2]. Критерий K_3 является естественным обобщением классического t-критерия и аналогичен критерию перестановок, который был предложен в [5] и [6]. Критерии K_4 и K_6 также рассматривались в [2]. Критерий K_5, по информации авторов настоящей статьи, является новым. Применение критерия K_1 в задачах регрессионного анализа рассматривалось также в работах [7–9].

В качестве альтернатив будем рассматривать тест Стьюдента (*t.test*), тест Колмогорова—Смирнова (*ks.test*) и тест Манна—Уитни (*wilcox.test*). Тест Стьюдента рассматривается как обладающий оптимальными свойствами при сравнении нормальных распределений с одинаковыми дисперсиями. Тест Колмогорова—Смирнова — непараметрический тест, основанный на выборочной функции распределения и поэтому наиболее универсальный из возможных тестов. Тест Манна—Уитни — непараметрический тест, основанный на рангах и, по сообщениям стандартных руководств, наиболее мощный непараметрический тест в случае распределений, отличающихся только сдвигом. Задача заключается в сравнительном анализе мощности рассматриваемых тестов для типичных распределений F_1 и F_2. Сначала мы приведем результаты об эквивалентности некоторых из рассмативаемых перестановочных тестов, а затем сравним оставшиеся тесты с помощью статистического моделирования.

3. Эквивалентность некоторых перестановочных тестов. Следующая теорема устанавливает эквивалентность трех критериев, так как каждый их них характеризуется одной и той же функцией мощности.

Теорема 3.1. *Для любых функций распределения F_1, F_2 критерии перестановок K_1, K_2 и K_3 для проверки нулевой гипотезы H_0, заданной формулой (1), против альтернативы H_1, заданной (2), эквивалентны для любой перестановки и для любого произвольно заданного уровня значимости α.*

Доказательство теоремы можно найти в работе [3].

Рассмотрим теперь два симметричных распределения с общим центром. Из вида рассматриваемых тестов вытекает, что в этом случае t-критерий и критерий K_1 полностью бесполезны.

Теорема 3.2. *Для любых симметричных относительно одного и того же центра функций распределения F_1, F_2 для проверки нулевой гипотезы H_0, заданной формулой (1), против альтернативы H_1, заданной (2), мощность теста K_1, а также теста Стьюдента совпадает с уровнем значимости для любого произвольно заданного уровня значимости α.*

Мощность теста Манна—Уитни, как показывают численные эксперименты, может быть чуть больше уровня значимости, но он также бесполезен в этой ситуации. Несколько лучшие (но также незначительные) возможности имеют в этой ситуации тесты K_4 и K_5. Однако, как мы увидим на численных примерах, критерий K_6 позволяет эффективно проверять гипотезу для распределений, описываемых этой теоремой.

4. Сравнение мощности перестановочных тестов. Проведем сравнительный анализ мощности тестов K_1, K_4, K_5 и K_6. Начнем со случая, когда сраниваемые распределения являются нормальными. Для нормальных распределений с одинаковыми дисперсиями следует ожидать, что наилучшим окажется перестановочный аналог t-критерия, т. е. критерий K_1. Надо сразу сказать, что этот вывод подтверждается результатами статистического моделирования (см. табл. 1), но выигрыш K_1 по сравнению с K_6 составляет всего 3–5% мощности. С другой стороны, критерий K_1 совершенно бесполезен, когда сравниваемые распределения имеют одинаковые средние и отличаются только дисперсиями, так как в этом случае мощность критерия тождественно равна уровню значимости α. Рассмотрим следующие распределения:

(1) нормальные распределения с одинаковыми дисперсиями;

(2) нормальные распределения с одинаковыми средними;

(3) составные распределения, 95% которых составляет нормальное рапределение и 5% — распределение Коши;

(4) распределения Коши с одинаковыми центрами;

(5) распределения Коши с одинаковой шириной;

(6) распределения Стьюдента со сдвигом и без;

(7) распределения Вейбулла;

(8) распределения Фишера с различными параметрами.

Мы провели моделирование этих распределений для $n = 10$ и $n = 30$. Каждый эксперимент повторялся 10 000 раз. Для перестановочных тестов было выбрано 1600 случайных перестановок (это количество выбрано на основе предварительных экспериментов). Приведем результаты только для случая размера выборки $n = 30$, так как для $n = 10$ результаты вполне аналогичны. В табл. 1 представлены результаты для случая, когда распределения отличаются только сдвигом.

Таблица 1. **Мощность тестов при наличии сдвига,** $n = 30$

Распределение	F_1	F_2	K_1	K_4	K_5	K_6
Нормальное	(0, 1)	(0, 1)	0.045	0.049	0.045	0.046
		(0.25, 1)	0.16	0.132	0.129	0.148
		(0.5, 1)	0.475	0.385	0.372	0.446
		(0.75, 1)	0.815	0.707	0.689	0.787
		(1, 1)	0.969	0.915	0.904	0.96
Составное (95% нормального и 5% Коши)	(0, 1)	(0, 1)	0.056	0.055	0.054	0.055
		(0.25, 1)	0.134	0.123	0.122	0.137
		(0.5, 1)	0.376	0.352	0.345	0.409
		(0.75, 1)	0.642	0.659	0.64	0.723
		(1, 1)	0.823	0.887	0.872	0.929
Коши	(0, 1)	(0, 1)	0.049	0.048	0.048	0.049
		(0.5, 1)	0.074	0.218	0.223	0.122
		(1, 1)	0.129	0.611	0.643	0.36
		(1.5, 1)	0.217	0.888	0.913	0.668
		(2, 1)	0.299	0.979	0.986	0.874
Стьюдента	(1, 0)	(1, 0)	0.049	0.05	0.05	0.049
		(1, 0.5)	0.2	0.35	0.353	0.275
		(1, 1)	0.502	0.856	0.853	0.719
		(1, 1.5)	0.698	0.984	0.987	0.922
		(1, 2)	0.794	0.999	0.999	0.975
Вейбулла	(1, 3)	(1, 3)	0.05	0.05	0.051	0.049
		(1, 2.5)	0.104	0.08	0.08	0.097
		(1, 2)	0.323	0.226	0.211	0.297
		(1, 1.5)	0.731	0.548	0.514	0.702
		(1, 1)	0.979	0.898	0.877	0.974

Как видно из табл. 1, в случае нормальных распределений, отличающихся только сдвигом, а также в случае распределения Вейбулла наилучшим оказывается критерий K_1, но критерий K_6 уступает ему лишь одну-две сотых по мощности. Остальные критерии значительно уступают в мощности. Для распределений Коши и Стьюдента наилучшим оказывается критерий K_5. А для составного распределения критерий K_6 значительно — на 0.1 мощности превосходит остальные в большинстве случаев.

Таблица 2. **Мощность тестов при отсутствии сдвига,** $n = 30$

Распределение	F_1	F_2	K_1	K_4	K_5	K_6
Нормальное	(0, 1)	(0, 1)	0.048	0.047	0.045	0.046
		(0, 1.5)	0.05	0.065	0.059	0.139
		(0, 2)	0.052	0.103	0.082	0.464
		(0, 2.5)	0.053	0.155	0.111	0.795
		(0, 3)	0.053	0.202	0.136	0.944
Составное (95% нормального и 5% Коши)	(0, 1)	(0, 1)	0.05	0.047	0.048	0.049
		(0, 1.5)	0.05	0.063	0.057	0.125
		(0, 2)	0.05	0.103	0.083	0.406
		(0, 2.5)	0.053	0.142	0.107	0.709
		(0, 3)	0.054	0.184	0.128	0.893
Коши	(0, 1)	(0, 1)	0.047	0.048	0.048	0.049
		(0, 3)	0.058	0.175	0.127	0.419
		(0, 5)	0.052	0.324	0.217	0.743
		(0, 7)	0.052	0.446	0.294	0.877
		(0, 9)	0.056	0.532	0.36	0.935
Фишера	(100, 2)	(100, 2)	0.048	0.05	0.049	0.047
		(100, 1.6)	0.084	0.064	0.06	0.085
		(100, 1.2)	0.24	0.13	0.113	0.244
		(100, 0.8)	0.643	0.357	0.302	0.654
		(100, 0.4)	0.98	0.869	0.813	0.98
Вейбулла	(5, 1)	(5, 1)	0.055	0.053	0.051	0.053
		(4, 1)	0.057	0.06	0.058	0.072
		(3, 1)	0.069	0.107	0.098	0.211
		(2, 1)	0.078	0.242	0.198	0.722
		(1, 1)	0.059	0.565	0.454	1

Табл. 2 показывает, что в отсутствие сдвига для симметричных распределений все критерии, кроме K_6, полностью бесполезны, что можно рассматривать просто как иллюстрацию теоремы 2. Но критерий K_6 работает эффективно. В случае распределения Фишера с отличающимся вторым параметром (число степеней свободы распределения χ-квадрат в знаменателе) наиболее эффективными оказываются критерии K_1 и K_6, которые имеют приблизительно одинаковые мощности.

5. Сравнение наилучшего перестановочного и неперестановочных критериев. На основе анализа, проведенного в предыдущем разделе, будем сравнивать критерий K_6, который в большинстве случаев оказывается наилучшим среди перестановочных критериев, с t-критерием и с критериями Колмогорова—Смирнова и Манна—Уитни. Начнем опять с распределений, отличающихся сдвигом.

Заметим, что для малых выборок, которые рассматриваются в табл. 3, перестановочный тест K_6 превосходит остальные тесты по мощности. Особенно малоэффективным является критерий Стьюдента для распределений Коши и Стьюдента. Тест Колмогорова—Смирнова заметно уступает тесту K_6, особенно он малоэффективен для распределения Вейбулла. В табл. 4 рассматриваются те же самые распределения, но для выборки втрое большего размера.

Для этого случая тест K_6 немного уступает критериям Манна—Уитни и Колмогорова—Смирнова в случае распределений Коши и Стьюдента. Тест Колмогорова—Смирнова для таких выборок уже не показывает низкую эффективность — его мощность близка к мощности K_6 или превосходит ее, кроме случаев нормального и близ-

Таблица 3. **Мощность тестов при наличии сдвига,** $n = 10$

Распределение	F_1	F_2	K_6	*t.test*	*ks.test*	*wilcox.test*
Нормальное	(0, 1)	(0, 1)	0.052	0.05	0.013	0.045
		(0.5, 1)	0.17	0.176	0.056	0.156
		(1, 1)	0.533	0.556	0.24	0.511
		(1.5, 1)	0.866	0.886	0.579	0.857
		(2, 1)	0.982	0.988	0.862	0.98
Составное (95% нормального и 5% Коши)	(0, 1)	(0, 1)	0.051	0.044	0.013	0.044
		(0.5, 1)	0.154	0.146	0.045	0.143
		(1, 1)	0.481	0.447	0.211	0.452
		(1.5, 1)	0.801	0.739	0.507	0.772
		(2, 1)	0.956	0.871	0.792	0.941
Коши	(0, 1)	(0, 1)	0.05	0.018	0.012	0.042
		(1, 1)	0.187	0.068	0.106	0.191
		(2, 1)	0.481	0.183	0.383	0.468
		(3, 1)	0.739	0.317	0.652	0.684
		(4, 1)	0.872	0.419	0.806	0.808
Стьюдента	(1, 0)	(1, 0)	0.05	0.018	0.012	0.044
		(1, 0.75)	0.238	0.092	0.105	0.254
		(1, 1.5)	0.602	0.266	0.454	0.689
		(1, 2.25)	0.792	0.376	0.75	0.897
		(1, 3)	0.876	0.457	0.894	0.963
Вейбулла	(1, 5)	(1, 5)	0.053	0.039	0.011	0.044
		(1, 4)	0.072	0.054	0.018	0.061
		(1, 3)	0.164	0.128	0.049	0.142
		(1, 2)	0.428	0.34	0.156	0.354
		(1, 1)	0.872	0.736	0.517	0.776

Таблица 4. **Мощность тестов при наличии сдвига,** $n = 30$

Распределение	F_1	F_2	K_6	*t.test*	*ks.test*	*wilcox.test*
Нормальное	(0, 1)	(0, 1)	0.049	0.049	0.034	0.048
		(0.25, 1)	0.151	0.158	0.098	0.154
		(0.5, 1)	0.445	0.475	0.315	0.456
		(0.75, 1)	0.778	0.81	0.638	0.788
		(1, 1)	0.958	0.97	0.883	0.961
Составное (95% нормального и 5% Коши)	(0, 1)	(0, 1)	0.05	0.043	0.033	0.05
		(0.25, 1)	0.136	0.117	0.094	0.141
		(0.5, 1)	0.407	0.342	0.29	0.41
		(0.75, 1)	0.726	0.61	0.584	0.734
		(1, 1)	0.93	0.783	0.844	0.932
Коши	(0, 1)	(0, 1)	0.051	0.02	0.037	0.05
		(0.5, 1)	0.118	0.033	0.164	0.171
		(1, 1)	0.369	0.073	0.554	0.51
		(1.5, 1)	0.665	0.134	0.864	0.794
		(2, 1)	0.874	0.21	0.974	0.935
Стьюдента	(1, 0)	(1, 0)	0.051	0.02	0.036	0.048
		(1, 0.5)	0.279	0.102	0.29	0.376
		(1, 1)	0.732	0.298	0.821	0.887
		(1, 1.5)	0.929	0.466	0.988	0.995
		(1, 2)	0.971	0.541	1	1
Вейбулла	(1, 3)	(1, 3)	0.057	0.054	0.038	0.056
		(1, 2.5)	0.101	0.102	0.061	0.092
		(1, 2)	0.305	0.314	0.184	0.268
		(1, 1.5)	0.7	0.716	0.483	0.619
		(1, 1)	0.973	0.975	0.878	0.937

ких к нему рапределений. Критерий Стьюдента, наоборот, наиболее эффективен в случае нормального распределения и распределения Вейбулла.

Для распределений с отсутствием сдвига рассмотрим только выборки размера $n = 30$. Результаты представлены в табл. 5.

Таблица 5. **Мощность тестов при отсутствии сдвига,** $n = 30$

Распределение	F_1	F_2	K_6	*t.test*	*ks.test*	*wilcox.test*
Нормальное	(0, 1)	(0, 1)	0.051	0.052	0.034	0.05
		(0, 1.5)	0.141	0.047	0.075	0.051
		(0, 2)	0.464	0.051	0.189	0.062
		(0, 2.5)	0.796	0.05	0.349	0.066
		(0, 3)	0.95	0.048	0.513	0.066
Составное (95% нормального и 5% Коши)	(0, 1)	(0, 1)	0.053	0.042	0.037	0.052
		(0, 1.5)	0.126	0.043	0.068	0.054
		(0, 2)	0.41	0.046	0.166	0.061
		(0, 2.5)	0.712	0.045	0.295	0.065
		(0, 3)	0.894	0.046	0.439	0.064
Коши	(0, 1)	(0, 1)	0.058	0.023	0.037	0.052
		(0, 3)	0.407	0.018	0.231	0.058
		(0, 5)	0.738	0.021	0.507	0.07
		(0, 7)	0.883	0.023	0.69	0.076
		(0, 9)	0.937	0.022	0.808	0.081
Фишера	(100, 2)	(100, 2)	0.051	0.014	0.036	0.052
		(100, 1.6)	0.087	0.022	0.043	0.058
		(100, 1.2)	0.246	0.039	0.095	0.115
		(100, 0.8)	0.645	0.051	0.318	0.297
		(100, 0.4)	0.982	0.013	0.879	0.768
Вейбулла	(5, 1)	(5, 1)	0.048	0.049	0.034	0.048
		(4, 1)	0.07	0.055	0.045	0.053
		(3, 1)	0.216	0.069	0.12	0.079
		(2, 1)	0.718	0.07	0.363	0.133
		(1, 1)	0.999	0.048	0.889	0.248

Из табл. 5 видно, что в этом случае тесты Стьюдента и Манна—Уитни бесполезны. Тест K_6 значительно превосходит тест Колмогорова—Смирнова, особенно в случае нормального распределения и смеси нормального распределения и распределения Коши.

6. Заключение. Стохастическое моделирование является универсальным методом исследования, который позволяет оценивать эффективность статистических процедур в случаях, когда это не удается сделать аналитическими методами. Сравнительная оценка мощности перестановочных тестов и классических тестов Стьюдента, Колмогорова—Смирнова и Манна—Уитни для решения задачи проверки гипотезы о равенстве двух распределений показала, что тест, основанный на сумме модулей разностей элементов двух выборок в большинстве случаев превосходит по мощности все другие рассмотренные тесты. Особенно велико преимущество этого теста, если центры сравниваемых распределений совпадают.

Литература

1. *Leman E.* Testing Statistical Hypotheses. 1979.

2. *Sirsky M.* On the Statistical Analysis of Functional Data Arising from Designed Experiments: Ph.D. thesis. University of Manitoba. 2012.

3. *Corain L., Melas V., Pepelyshev A., Salmaso L.* New insights on permutation approach for hypothesis testing on functional data // Advances in Data Analysis and Classification. 2014. Vol. 8, issue 3. P. 339–356.

4. *Sturino J., Zorych I., Mallick B.* et al. Statistical methods for comparative phenomics using high-throughput phenotype microarrays // The International Journal of Biostatistics. 2010. Vol. 6. P. 3–4.

5. *Cox D., Lee J.* Pointwise testing with functional data using the Westfall—Young randomization method // Biometrika. 2008. Vol. 95. P. 621–634.

6. *Ramsay J., Hooker G., Graves S.* Functional Data Analysis with R and Matlab. 2009.

7. *Keller-McNulty S., Higgins J.* Effect of tail weight and outliers on power and type-i error of robust permutation tests for location // Communications in Statistics — Simulation and Computation. 1987. Vol. 16. P. 17–35.

8. *Edgington E. S.* Approximate randomization tests // The Journal of Psychology. 1969. Vol. 72. P. 143–149.

9. *Good P. I.* Resampling Methods: A Practical Guide to Data Analysis. 3 edition. Birkhauser, 2006.

Статья поступила в редколлегию 23 декабря 2015 г.

Сведения об авторах

Мелас Вячеслав Борисович — профессор; vbmelas@post.ru

Сальников Дмитрий Игоревич — студент; mejibkop.ru@mail.ru

Гудулина Анастасия Олеговна — программист; anastasia.gudulina@gmail.com

THE NUMERICAL COMPARING OF CLASSICAL AND PERMUTATION METHODS OF STATISTICAL HYPOTHESIS TESTING

Viatcheslav B. Melas, Dmitrii I. Salnikov, Anastasia O. Gudulina

St. Petersburg State University, Universitetskaya nab., 7–9, St. Petersburg, 199034, Russian Federation; vbmelas@post.ru, anastasia.gudulina@gmail.com, mejibkop.ru@mail.ru

The article is devoted to the classical problem of a statistical hypothesis test for the equality of two distributions. For normal distributions in many ways t-test is the best one. But in practice, compared distributions often are not normal ones and, generally speaking, are not known. The non-parametric Kolmogorov—Smirnov test is usually used to solve this problem in case when the investigator knows nothing about the compared distributions. The article deals with methods based on permutations. In recent years such methods have attracted attention by its simplicity, versatility and relatively high efficiency. A comparative study of power of a few permutation tests and classical methods (such as Kolmogorov—Smirnov test, t-test and Mann—Whitney test) for a wide class of distribution functions was done by using stochastic simulation's methods. Normal and Cauchy distributions and its mixtures are considered as well as exponential, Weibull, Fisher and Student's distributions.

It was found that for many typical distributions the permutation test, based on the sum of the absolute values of the differencesis, is the most powerful one. The advantage of this test over the others is considerably greater for the case, when symmetrical distributions with the same centers are compared. Thus, the permutation test can be recommended for using in cases where the compared distributions are different from normal ones. Refs 9. Tables 5.

Keywords: statistical hypothesis, permutation methods.

References

1. Leman E., *Testing Statistical Hypotheses* (1979).

2. Sirsky M., *On the Statistical Analysis of Functional Data Arising from Designed Experiments*: Ph.D. thesis (University of Manitoba, 2012).

3. Corain L., Melas V., Pepelyshev A., Salmaso L., "New insights on permutation approach for hypothesis testing on functional data", *Advances in Data Analysis and Classification* **8**, Issue 3, 339–356 (2014).

4. Sturino J., Zorych I., Mallick B. et al., "Statistical methods for comparative phenomics using high-throughput phenotype microarrays", *The International Journal of Biostatistics* **6**, 3–4 (2010).

5. Cox D., Lee J., “Pointwise testing with functional data using the Westfall—Young randomization method”, *Biometrika* **95**, 621–634 (2008).

6. Ramsay J., Hooker G., Graves S., *Functional Data Analysis with R and Matlab* (2009).

7. Keller-McNulty S., Higgins J., “Effect of tail weight and outliers on power and type-i error of robust permutation tests for location”, *Communications in Statistics — Simulation and Computation* **16**, 17–35 (1987).

8. Edgington E. S., “Approximate randomization tests”, *The Journal of Psychology* **72**, 143–149 (1969).

9. Good P. I., *Resampling Methods: A Practical Guide to Data Analysis* (3 edition, Birkhauser, 2006).

Для цитирования: Мелас В. Б., Сальников Д. И., Гудулина А. О. Численное сравнение перестановочных и классических методов проверки статистических гипотез // Вестник Санкт-Петербургского университета. Серия 1. Математика. Механика. Астрономия. 2016. Т. 3 (61). Вып. 3. С. 415–423. DOI: 10.21638/11701/spbu01.2016.309

For citation: Melas V. B., Salnikov D. I., Gudulina A. O. The numerical compearing of classical and permutation methods of statistical hypothesis testing. *Vestnik of Saint Petersburg University. Series 1. Mathematics. Mechanics. Astronomy*, 2016, vol. 3 (61), issue 3, pp. 415–423. DOI: 10.21638/11701/spbu01.2016.309