УДК 621.391.00

НЕПАРАМЕТРИЧЕСКИЙ РАНДОМИЗАЦИОННЫЙ ТЕСТ: ИДЕЯ, АЛГОРИТМ, ПРЕИМУЩЕСТВА, НЕДОСТАТКИ И ПРИМЕР ПРИМЕНЕНИЯ

А. П. Сергеев [1,2], А. С. Буторова [1,2], Е. А. Корюкин [2], В. С. Бобаков [2], С. В. Павлова [2]

[1] *Институт промышленной экологии УрО РАН, г. Екатеринбург, Россия*
[2] *Уральский федеральный университет им. первого Президента России Б. Н. Ельцина, г. Екатеринбург, Россия*

Пермутационный подход имеет ряд преимуществ по сравнению с традиционными параметрическими статистическими методами, что делает его более предпочтительным для обработки данных в некоторых исследовательских ситуациях. Пермутационные методы не используют предположений о виде распределения вероятностей; подходят для небольших наборов данных и дают исследователю большую свободу выбора тестовой статистики. Настоящая статья посвящена применению одного из вариантов пермутационных методов – непараметрическому рандомизационному тесту. Авторы излагают предпосылки рандомизационного теста и приводят некоторые формализации для случая сравнения двух групп наблюдений. В работе рассматривается пример применения рандомизационного теста для проверки статистической гипотезы о равенстве нулю межгрупповой разности средних значений в случаях правостороннего и левостороннего критических множеств. В примере сравнивается производительность двух способов очистки – стандартного и нового. В правостороннем случае в качестве тестовой статистики выступает сумма времен очистки в группе стандартного способа, в левостороннем случае – межгрупповая разность средних времен. С помощью рандомизационного теста показано, что новый способ очистки не дает статистически значимого преимущества во времени очистки перед стандартным способом.

Ключевые слова: пермутационный подход; рандомизационный тест; перестановка; проверка статистических гипотез; ошибка первого рода; ошибка второго рода; *p*-значение.

1. Введение

Непараметрический рандомизационный тест – один из представителей пермутационных методов – был впервые применен в 1935 г. Рональдом Эйлмером Фишером (1890–1962 гг.) для построения точного статистического вывода в задаче оценки связи между двумя дихотомическими переменными [1–3]. Пермутационный подход требует большого объема вычислений, характерного для комбинаторных объектов. Такие объемы вычислений практически невозможно провести вручную, а Алан Тьюринг построит электромеханический компьютер только через пять лет в 1940 г.

Современные компьютеры обладают вычислительной мощностью, достаточной, чтобы успешно работать с комбинаторными объектами и применять пермутационный подход для статистического вывода [4–6]. Этот подход применим практически в любой предметной области, которая имеет дело с вероятностной формализацией процедуры принятия решений в условиях неопределенности [7–11]. Пермутационные методы являются не только инструментом принятия решений, но также могут использоваться исследователями для построения метрик их уверенности в обоснованности уже принятого решения [8].

Основная идея пермутационных методов заключается в том, что выполняются все перестановки или заранее определенное число перестановок данных, для каждой из которых вычисляется некоторая выбранная исследователем из предметных соображений статистика, наблюдаемое распределение которой принимается за референтное распределение.

Пермутационый подход предполагает две вероятностных модели статистического вывода: рандомизационную и популяционную [12–14]. Согласно рандомизационной модели (рандомизационные тесты) множество доступных объектов разбивается на несколько непересекающихся групп случайным образом. Согласно популяционной модели (перестановочные тесты) объекты случайным образом отбираются из некоторой популяции или из нескольких популяций (субпопуляций). В рандомизационной модели используются процедуры, которые обычно называют рандомизационными тестами (тестами рандомизации).

Любая процедура статистического вывода допускает возможность ошибки в принятии решения. С помощью уровня значимости процедура статистического вывода позволяет контролировать и количественно оценивать вероятности ошибок. Вероятностями ошибок можно управлять с помощью приближенных методов вывода, однако успешность их применения будет варьироваться. Точные методы статистического вывода гарантируют контроль ошибок. При этом точные методы наиболее полезны, когда они позволяют контролировать ошибки при относительно широких допущениях. Эта идея будет проиллюстрирована ниже с помощью одновыборочного теста Стьюдента и доверительного интервала, которые являются точными только при ограничениях, накладываемых на распределение, из которого получены данные.

Любая процедура проверки статистической гипотезы предусматривает контроль вероятности ошибок первого и второго рода. Ошибка первого рода есть отклонение истинной нулевой гипотезы, ошибка второго рода – неотклонение ложной нулевой гипотезы. Заданный исследователем уровень значимости, обозначаемый обычно α, контролирует вероятность ошибки первого рода. Заметим, что вероятность ошибки первого рода не обязательно равна заданному уровню значимости. Вероятность ошибки первого рода есть вероятность получить *p*-значение меньшее или равное α при справедливости нулевой гипотезы.

Цель настоящей работы – описать идею и показать внутреннее устройство рандомизационного теста, а также продемонстрировать его применение на примере тестирования статистической гипотезы об отсутствии межгруппового различия в выраженности признака.

2. Материалы и методы

Ниже покажем, как работает параметрический тест, на примере одновыборочного теста Стьюдента.

Пусть $X(1)$,..., $X(n)$ – случайная выборка из нормального распределения с параметром положения m.

Нулевая гипотеза H_0 утверждает, что параметр положения m некоторой нормально распределенной переменной X равен m_0. Проверим нулевую гипотезу H_0: $m = m_0$ против двусторонней альтернативы H_1: $m \neq m_0$.

Тогда представленная следующая T-статистика имеет распределение Стьюдента с числом степеней свободы $df = n - 1$:

$$T = \frac{\bar{X} - m_0}{S/\sqrt{n}};\ \bar{X} = \frac{1}{n}\sum_{i=1}^{i=n} X(i)\ ;\ S^2 = \frac{\sum_{i=1}^{i=n}(X(i) - \bar{X})^2}{n-1}.$$

Перед тестированием гипотезы исследователь должен задать уровень значимости α.

Правило принятия решения: если наблюдаемое значение тестовой статистики попадает в критическое множество, тогда нулевую гипотезу следует отклонить:

$$«\text{Если } |T_{obs}| \geq qT\left(df = n - 1, 1 - \frac{\alpha}{2}\right), \text{тогда } H_0 \text{ отклонить}».$$

Здесь $qT(n-1{,}1-\alpha/2)$ – квантиль порядка $1-\alpha/2$ распределения Стьюдента с числом степеней свободы $df = n-1$.

Ошибка первого рода возникает в случае, если модуль статистики теста равен или превышает квантиль $qT(n-1{,}1-\alpha/2)$ при справедливости нулевой гипотезы. Тогда вероятность ошибки первого рода в точности равна выбранному уровню значимости α.

Однако если предположение о нормальности распределения случайной переменной X в популяции неверно, то и T-статистика не имеет t-распределения. Отсюда следует, что вероятность ошибки первого рода может быть не равна α, а иногда и значительно отличаться от α. В этом случае контролировать вероятность ошибки первого рода становится затруднительно, и исследователю остается только надеяться, что он ошибается не слишком сильно. Сформулируем определение точного теста: точный тест есть тест, для которого вероятность ошибки первого рода в точности равна заданному уровню значимости α. Например, одновыборочный тест Стьюдента будет точным только тогда, когда реализация выборки получена из нормального распределения. Аналогичная ситуация возникает при построении с использованием T-статистики доверительного интервала, например, для математического ожидания. Если реализация выборки получена не из нормального распределения, вероятность того, что построенный доверительный интервал содержит (накрывает) истинное значение оцениваемоего параметра, может и, скорее всего, будет отличаться от заданной доверительной вероятности $1-\alpha$. Поэтому здесь одновыборочный доверительный интервал гарантированно будет точным только тогда, когда выборка получена из нормального распределения, что возможно только в идеальном случае, когда нормальное распределение намеренно сгенерировано исследователем. Другими словами, в условиях реального эксперимента у экспериментатора никогда не будет уверенности в нормальности распределения.

В случае рандомизационной модели случайное распределение доступных объектов по группам – единственная основа для статистического вывода. Необязательно иметь случайную выборку из некоторой популяции с заданным распределением. Однако при этом любые выводы в рандомизационной модели будут ограничены объектами исследования.

Пример. Предположим, что сравниваются стандартный и новый способы очистки некоторых объектов. При этом наблюдается время в секундах, в течение которого объект очищается до некоторого допустимого уровня. Нас интересует, действительно ли новый способ дает статистически значимое преимущество во времени очистки перед стандартным способом.

Из n объектов, доступных для исследования, n_1 случайным образом распределяются для очистки новым способом, в то время как оставшиеся $n_0 = n - n_1$ очищают стандартным способом.

Сформулируем гипотезы. Нулевая гипотеза H_0: «разницы между новым и стандартным способами очистки нет» против односторонней альтернативы H_1: «новый способ сокращает время очистки объекта». Таким образом, мы имеем дело с односторонним (левосторонним) критическим множеством. Протестируем нулевую гипотезу на уровне значимости $\alpha = 0{,}05$.

Выберем в качестве тестовой статистики разность групповых средних времен очистки. Введем следующие обозначения:

r – номер рандомизации ($r = 0$ – нулевая рандомизация, т. е. наблюдаемое распределение объектов по группам);

$x(r, i)$ – группа, в которую попадет i-й объект при r-й рандомизации ($x = 0$ – группа объектов, очищаемых стандартным способом, $x = 1$ – группа объектов, очищаемых новым способом);

$x(r = 0, i)$ – группа, в которой действительно очищался i-й объект ($x = 0$ – группа объектов, очищаемых стандартным способом, $x = 1$ – группа объектов, очищаемых новым способом);

y(*i*) – время, в течение которого *i*-й объект очищается до некоторого допустимого уровня (в секундах).

Заметим, что

$$n_0 = n - \sum_{i=1}^{i=n} x(0,i)\,;\; n_1 = \sum_{i=1}^{i=n} x(0,i).$$

Суммарное время очистки всех объектов в предположении истинности нулевой гипотезы не зависит от рандомизации и равно

$$\sum_{i=1}^{i=n} y(i).$$

Среднее время очистки в группе стандартного способа при *r*-й рандомизации равно

$$\frac{\sum_{i=1}^{i=n} y(i) - \sum_{i=1}^{i=n} y(i)x(r,i)}{n - \sum_{i=1}^{i=n} x(0,i)} = \frac{\sum_{i=1}^{i=n} y(i)\big(1 - x(r,i)\big)}{n_0}.$$

Среднее время очистки в группе нового способа при *r*-й рандомизации равно

$$\frac{\sum_{i=1}^{i=n} y(i)x(r,i)}{\sum_{i=1}^{i=n} x(0,i)} = \frac{\sum_{i=1}^{i=n} y(i)x(r,i)}{n_1}.$$

Теперь мы можем записать выражение для тестовой статистики:

$$T(r) = \frac{\sum_{i=1}^{i=n} y(i)x(r,i)}{n_1} - \frac{\sum_{i=1}^{i=n} y(i)\big(1 - x(r,i)\big)}{n_0} =$$

$$= \frac{\sum_{\{i|x(r,i)=1\}} y(i)}{n_1} - \frac{\sum_{\{i|x(r,i)=0\}} y(i)}{n_0}$$

$$= \frac{\sum_{\{i|x(r,i)=1\}} y(i)}{n_1} - \frac{\sum_{\{i|x(r,i)=0\}} y(i)}{n_0} + \frac{\sum_{\{i|x(r,i)=0\}} y(i)}{n_1}$$

$$- \frac{\sum_{\{i|x(r,i)=0\}} y(i)}{n_1} =$$

$$= \frac{\sum_{\{i|x(r,i)=0\}} y(i) + \sum_{\{i|x(r,i)=1\}} y(i)}{n_1} - \frac{(n_0 + n_1)\sum_{\{i|x(r,i)=0\}} y(i)}{n_0 n_1} =$$

$$= \frac{\sum_{i=1}^{i=n} y(i)}{n_1} - \frac{(n_0 + n_1)}{n_0 n_1}\boxed{\sum_{\{i|x(r,i)=0\}} y(i)}^{\text{Статистика 0}} =$$

$$= \frac{\sum_{\{i|x(r,i)=1\}} y(i)}{n_1} - \frac{\sum_{\{i|x(r,i)=0\}} y(i)}{n_0} + \frac{\sum_{\{i|x(r,i)=1\}} y(i)}{n_0}$$

$$- \frac{\sum_{\{i|x(r,i)=1\}} y(i)}{n_0}$$

$$= \frac{(n_0 + n_1)\sum_{\{i|x(r,i)=1\}} y(i)}{n_0 n_1} - \frac{\sum_{\{i|x(r,i)=0\}} y(i) + \sum_{\{i|x(r,i)=1\}} y(i)}{n_0} =$$

$$= \frac{(n_0+n_1)}{n_0 n_1} \overbrace{\boxed{\sum_{\{i|x(r,i)=1\}} y(i)}}^{\text{Статистика 1}} - \frac{\sum_{i=1}^{i=n} y(i)}{n_0} =$$

$$= \frac{n_0 \sum_{i=1}^{i=n} y(i) x(r,i) - n_1 \sum_{i=1}^{i=n} y(i)\big(1 - x(r,i)\big)}{n_0 n_1}$$

$$= \frac{\overbrace{\boxed{\sum_{i=1}^{i=n} y(i)\Big(n_0 x(r,i) - n_1\big(1 - x(r,i)\big)\Big)}}^{\text{Статистика 2}}}{n_0 n_1}.$$

Рассмотрим три выражения в этом равенстве, обозначенные соответственно: Статистика 0, Статистика 1 и Статистика 2. В этих выражениях выделены элементы, которые зависят от рандомизации. В случае Статистика 0 в качестве статистики, эквивалентной для $T(r)$, может выступать сумма времени очистки в группе стандартного способа, но при этом тест станет правосторонним. В случае Статистика 1 в качестве эквивалентной статистики для $T(r)$ может выступать сумма времени очистки в группе нового способа, при этом тест останется левосторонним. Наконец, в случае Статистика 2 в качестве эквивалентной статистики для $T(r)$ может выступать числитель дроби, при этом тест также останется левосторонним.

3. Результаты

Рассмотрим пример применения рандомизационного теста. Пусть имеется $n = 7$ объектов, которые случайно разбили на две группы: группу стандартного способа очистки $x(r = 0,\ i) = 0$ и группу нового способа очистки $x(r = 0,\ i) = 1$. В группе стандартного способа $n_0 = 3$ объекта, в группе нового способа – $n_1 = 4$ объекта, $n = n_0 + n_1$. Наблюдаемое (при нулевой рандомизации) время очистки объектов приведено в табл. 1.

Таблица 1. Нулевая (наблюдаемая) и первая рандомизации

Номер рандомизации, r	Номер объекта, i	Группа способа очистки, $x(r, i)$	Время очистки, $y(i)$
0	1	0	33
0	2	0	40
0	3	0	25
0	4	1	23
0	5	1	19
0	6	1	22
0	7	1	26
Ниже идут первая и следующие за ней рандомизации			
1	1	0	33
1	2	0	40
1	3	1	25
1	4	0	23
1	5	1	19
1	6	1	22
1	7	1	26
...	...	...	...

Наблюдаемое (при нулевой рандомизации) значение статистики равно

$$T_{obs} = T(0) = \frac{\sum_{i=1}^{i=n} y(i)\left(n_0 x(0,i) - n_1\big(1 - x(0,i)\big)\right)}{n_0 n_1} =$$

$$= \frac{1}{12}(33 \cdot (-4) + 40 \cdot (-4) + 25 \cdot (-4) + 23 \cdot (3) + 19 \cdot (3) + 22 \cdot (3) +$$

$$+26 \cdot (3)) = -10{,}17.$$

Нулевая гипотеза утверждает, что новый способ не меняет время очистки. Следовательно, объект под номером 1, очищенный стандартным способом, например за 23 секунды, был бы очищен новым способом за те же 23 секунды. Таким образом, время очистки при нулевой гипотезе есть константа, а случайной переменной является номер группы по способу очистки, в которую может попасть объект. Поскольку разбиение на две группы с заранее фиксированными размерами групп проводилось случайно, любое разбиение имеет равную вероятность произойти. Назовем каждое такое разбиение рандомизацией. Число таких рандомизаций, включая наблюдаемую (нулевую рандомизацию), в нашем случае равно

$$\binom{n}{n_0} = \binom{n}{n_1} = C_n^{n_0} = C_n^{n_1} = \frac{n!}{(n-n_0)!\,n_0!} = \binom{7}{3} = \binom{7}{4} = \frac{7!}{(7-3)!\,3!} = 35.$$

Следовательно, каждая такая рандомизация может произойти с вероятностью 1/35.

Построим все рандомизации и вычислим для каждой такой рандомизации значение статистики *T*(*r*) (табл. 2).

Таблица 2. Все возможные рандомизации

Номер рандомизации, *r*	Стандартный способ			Новый способ				*T*(*r*)	Вероятность
0	33	40	25	23	19	22	26	–10,17	1/35
1	23	33	19	40	22	25	26	3,25	1/35
2	23	33	22	40	19	25	26	1,50	1/35
3	23	33	25	40	19	22	26	–0,25	1/35
4	23	33	26	40	19	22	25	–0,83	1/35
5	23	40	19	33	22	25	26	–0,83	1/35
6	23	40	22	33	19	25	26	–2,58	1/35
7	23	40	25	33	19	22	26	–4,33	1/35
8	23	40	26	33	19	22	25	–4,92	1/35
9	23	19	22	33	40	25	26	9,67	1/35
10	23	19	25	33	40	22	26	7,92	1/35
11	23	19	26	33	40	22	25	7,33	1/35
12	23	22	25	33	40	19	26	6,17	1/35
13	23	22	26	33	40	19	25	5,58	1/35
14	23	25	26	33	40	19	22	3,83	1/35
15	33	40	19	23	22	25	26	–6,67	1/35
16	33	40	22	23	19	25	26	–8,42	1/35
17	23	33	40	19	22	25	26	–9,00	1/35
18	33	40	26	23	19	22	25	–10,75	1/35
19	33	19	22	23	40	25	26	3,83	1/35
20	33	19	25	23	40	22	26	2,08	1/35
21	33	19	26	23	40	22	25	1,50	1/35
22	33	22	25	23	40	19	26	0,33	1/35
23	33	22	26	23	40	19	25	–0,25	1/35

Окончание табл. 2

Номер рандомизации, *r*	Стандартный способ			Новый способ				*T*(*r*)	Вероятность
24	33	25	26	23	40	19	22	–2,00	1/35
25	40	19	22	23	33	25	26	–0,25	1/35
26	40	19	25	23	33	22	26	–2,00	1/35
27	40	19	26	23	33	22	25	–2,58	1/35
28	40	22	25	23	33	19	26	–3,75	1/35
29	40	22	26	23	33	19	25	–4,33	1/35
30	40	25	26	23	33	19	22	–6,08	1/35
31	19	22	25	23	33	40	26	8,50	1/35
32	19	22	26	23	33	40	25	7,92	1/35
33	19	25	26	23	33	40	22	6,17	1/35
34	22	25	26	23	33	40	19	4,42	1/35

P-значение теста рандомизации H_0 *pvalue* может быть рассчитано как вероятность получить статистику теста столь же экстремальную или более экстремальную (в пользу H_1), чем наблюдаемая статистика теста *T*(0).

Выбор способа очистки случаен, следовательно, все рандомизации равновероятны:

$$pvalue = \mathbf{P}\{T(r) \le T(0) | H_0\} = \frac{\sum_{r=0}^{\binom{n}{n_0}-1} I(T(r) \le T(0))}{\binom{N}{n_0}} = \frac{\sum_{r=0}^{\binom{n}{n_0}-1} I(T(r) \le T(0))}{\frac{n!}{(n-n_0)!\, n_0!}},$$

где *I*() – индикаторная функция.

Вычислим эту вероятность.

$$pvalue = \mathbf{P}\{T(r) \le T(0) | H_0\} = \frac{\sum_{r=0}^{\binom{n}{n_0}-1} I(T(r) \le T(0))}{\binom{n}{n_0}} = \frac{2}{35} = 0{,}057.$$

Поскольку *p*-значение = 0,057 > α = 0,05, тест не отклоняет нулевую гипотезу. Этот результат, в частности, можно интерпретировать так: новый способ не дает статистически значимого преимущества во времени очистки перед стандартным способом.

Построим функцию распределения тестовой статистики *T* при условии H_0:

$$F(t|H_0) = \mathbf{P}\{T \le t | H_0\}.$$

Поскольку все рандомизации попарно несовместны, можно сгруппировать одинаковые значения статистики *T* и сложить соответствующие им вероятности. Проделав это, получим таблицу распределения вероятностей по отсортированным значениям тестовой статистики (табл. 3).

Таблица 3. Распределение вероятностей по значениям тестовой статистики

Значение тестовой статистики, *t*	Вероятность, *p*(*t*)	Функция распределения, *F*(*t*)
–10,750	0,0286	0,0286
–10,167	0,0286	0,0571
–9,000	0,0286	0,0857
–8,417	0,0286	0,1143
–6,667	0,0286	0,1429

Окончание табл. 3

Значение тестовой статистики, t	Вероятность, $p(t)$	Функция распределения, $F(t)$
–6,083	0,0286	0,1714
–4,917	0,0286	0,2000
–4,333	0,0571	0,2571
–3,750	0,0286	0,2857
–2,583	0,0571	0,3429
–2,000	0,0571	0,4000
–0,833	0,0571	0,4571
–0,250	0,0857	0,5429
0,333	0,0286	0,5714
1,500	0,0571	0,6286
2,083	0,0286	0,6571
3,250	0,0286	0,6857
3,833	0,0571	0,7429
4,417	0,0286	0,7714
5,583	0,0286	0,8000
6,167	0,0571	0,8571
7,333	0,0286	0,8857
7,917	0,0571	0,9429
8,500	0,0286	0,9714
9,667	0,0286	1,0000

Теперь, используя построенную функцию рандомизационного распределения тестовой статистики в предположении H_0, вычислим одностороннее (в нашем случае – левостороннее) p-значение (рис. 1):

$$pvalue = F(T_{obs}|H_0) = 0{,}057.$$

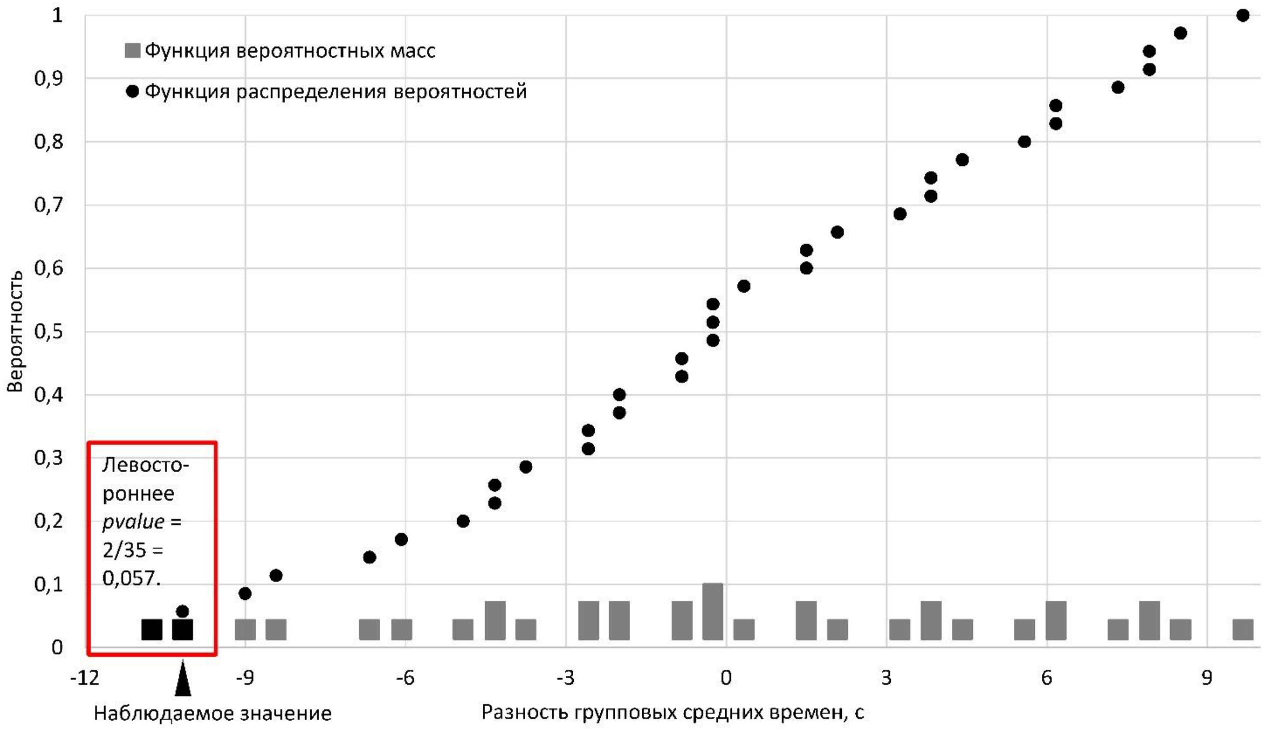

Рис. 1. Левосторонний тест

Рассмотрим случай, когда в качестве тестовой статистики $T(r)$ выступает сумма времен очистки в группе стандартного способа (рис. 2). В этом случае тест станет правосторонним, а функция вероятностных масс будет зеркально симметрична той, что мы наблюдали в левостороннем случае (рис. 1). При этом *p*-значение также станет правосторонним.

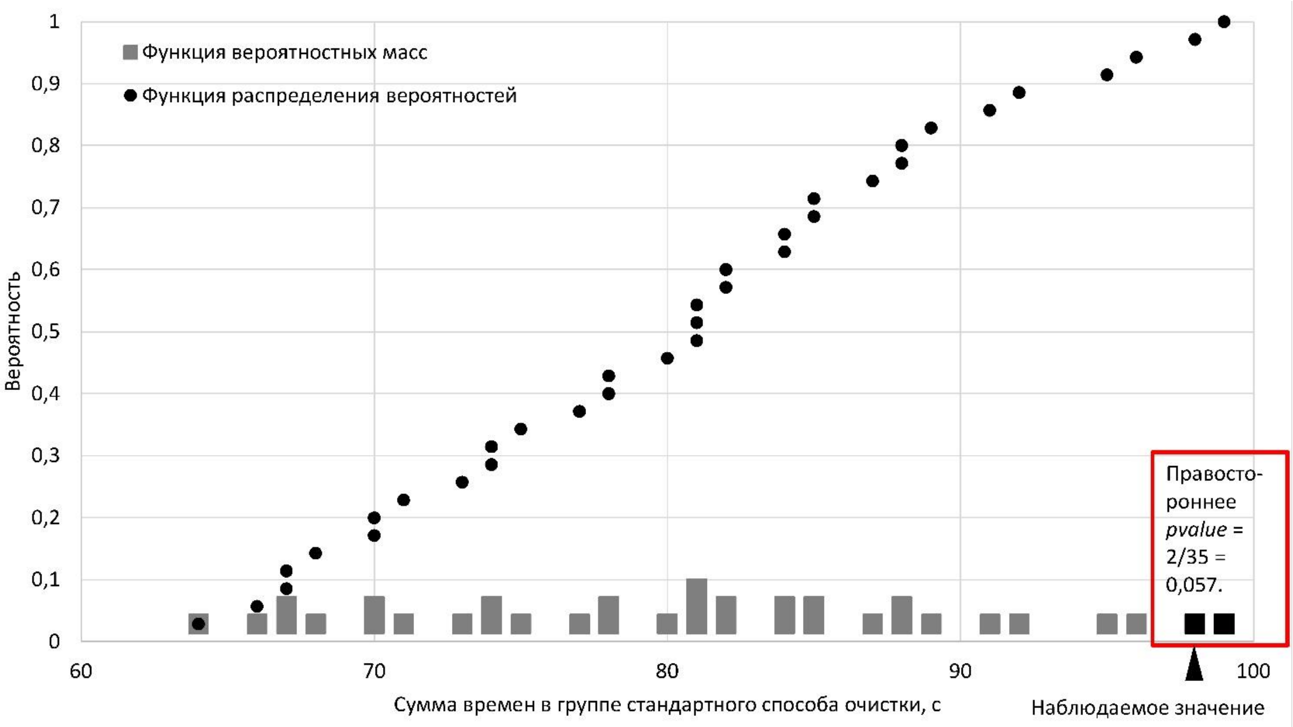

Рис. 2. Правосторонний тест

Рандомизационное распределение дискретно, поэтому *p*-значение кратно 1/35, но не каждое кратное значение возможно. В нашем примере достижимые *p*-значения равны $k/35$, где k = 1, 2, 3, 4, 5, 6, 7, 9, 10, 12, 14, 16, 19, 20, 22, 23, 24, 26, 27, 28, 30, 31, 33, 34, 35 (их всего 25 штук). Если уровень значимости равен одному из достижимых *p*-значений $\alpha = k/35$, тогда

$$\mathbf{P}\{Type\ I\ error\} = \mathbf{P}\{pvalue \le \alpha | H_0\} = \mathbf{P}\left\{\sum_{r=0}^{\binom{N}{n_0}-1} I\big(T(r) \le T(0)\big) \le k \,\middle|\, H_0\right\} =$$

$$= \frac{k}{\binom{N}{n_0}} = \alpha.$$

Рандомизационный тест H_0 есть точный тест, если выбранный уровень значимости равен одному из достижимых *p*-значений.

Рандомизационный тест H_0 есть консервативный тест, если выбранный уровень значимости не равен ни одному из достижимых *p*-значений, а

$$\mathbf{P}\{Type\ I\ error\} = \frac{k}{\binom{N}{n_0}} = \frac{k}{\frac{n!}{(n-n_0)!\,n_0!}} = k\frac{(n-n_0)!\,n_0!}{n!} < \alpha$$

является наибольшим достижимым *p*-значением меньшим α.

Рандомизационный тест гарантированно контролирует вероятность ошибки первого рода при условии рандомизации объектов по способу очистки. Другими словами, каждому объекту случайно должен быть назначен один из двух способов очистки.

4. Обсуждение

Пермутационные методы обладают рядом свойств, которые делают их привлекательной альтернативой традиционным статистическим методам при решении многих исследовательских задач [15, 16].

Во-первых, пермутационные методы не требуют предварительных знаний о распределениях и допущений, связанных с параметрическим тестом, таким как, например, дисперсионный анализ, который предполагает нормальность распределения и однородность дисперсии. Это, однако, не означает, что неоднородность дисперсии не влияет на результаты применения пермутационных методов. Также сложность, связанная с применениями традиционных статистических тестов, возникает, когда меньшая из выборок обладает большей дисперсией. Это увеличивает риск отклонения нулевой гипотезы, когда она на самом деле верна, иными словами, потенциально увеличивает вероятность ошибки первого рода.

Во-вторых, пермутационные методы могут быть построены как аналоги большого количества традиционных параметрических тестов. Это могут быть, например, все разновидности *t*- и *F*-тестов, простая и множественная регрессия, а также всевозможные меры ассоциации [17, 18].

В-третьих, пермутационные методы успешно применяются в исследованиях, в которых случайно или в соответствии с авторским экспериментальным дизайном получены небольшие наборы экспериментальных данных. Например, небольшое число наблюдений характерно для лонгитюдного исследования, в начале которого число участников было достаточно большим, а к концу – большая доля по каким-либо причинам выбыла из эксперимента. Еще один пример – пилотное исследование с небольшим количеством участников для установления работоспособности методики исследования до проведения основного экспериментального этапа. Пермутационные методы продуктивны в исследовательских ситуациях с небольшим числом наблюдений, в то время как традиционные статистические тесты в таких случаях оказываются ненадежными.

В настоящей работе применение пермутационного подхода показано на примере непараметрического рандомизационного теста для сравнения выраженности признака в двух группах. Недостаток рандомизационного теста заключается в том, что он полностью зависит от набора наблюдаемых данных, для которых он применяется. Таким образом, полученные выводы могут быть сделаны только относительно рассматриваемой группы наблюдений и не могут быть автоматически распространены на популяцию в целом. Тем не менее рандомизационный тест остается полезным, так как позволяет сделать оценку на малом количестве наблюдений, а также обосновать и сформулировать план дальнейших исследований уже на случайной выборке из популяции.

5. Выводы

1. Рассмотренный в настоящей работе непараметрический рандомизационный тест дает исследователю большую свободу выбора тестовой статистики, с помощью которой оцениваются экспериментальные условия.

2. Непараметрический рандомизационный тест не требует предварительных знаний о распределениях и допущений, характерных для параметрических тестов.

3. Непараметрический рандомизационный тест может быть применен к самым различным данным, в т. ч. к данным, содержащим малое число наблюдений, а также к данным, не являющимся реализацией случайной выборки.

6. Благодарности

Исследование выполнено при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках Программы развития

Уральского федерального университета им. первого Президента России Б. Н. Ельцина в соответствии с программой стратегического академического лидерства «Приоритет-2030».

7. Список литературы

1. *Fisher, R. A.* Statistical methods for research workers / R. A. Fisher. – Edinburgh, Scotland. – Oliver and Boyd, 1925. – 232 p.
2. *Fisher, R. A.* The arrangement of field experiments / R. A. Fisher // Journal of the American Statistical Association. – 1926. – Vol. 33. – P. 503–513.
3. *Fisher, R. A.* The design of experiments / R. A. Fisher. – Edinburgh, Scotland England. – Oliver and Boyd, 1935. – 252 p.
4. *Ernst, M. D.* Permutation Methods: A Basis for Exact Inference / M. D. Ernst // Statistical Science. – 2004. – Vol. 19, № 4. – P. 676–685. – DOI: 10.1214/088342304000000396.
5. *Fişek, M. H.* Permutation tests for goodness-of-fit testing of mathematical models to experimental data / M. H. Fişek, Z. Barlas // Social Science Research. – 2013. – Vol. 42, Iss. 2. – P. 482–495. – DOI: 10.1016/j.ssresearch.2012.09.010.
6. *Ross, S. M.* Simulation, bootstrap statistical methods, and permutation tests / S. M. Ross // Introduction to Probability and Statistics for Engineers and Scientists. – Amsterdam. – Academic Press; Elsevier, 2021. – P. 619–647. – DOI: 10.1016/B978-0-12-824346-6.00024-7.
7. *Taylor, A. B.* Four applications of permutation methods to testing a single-mediator model / A. B. Taylor, D. P. MacKinnon // Behavior Research Methods. – 2012. – Vol. 44, Iss. 3. – P. 806–844. – DOI: 10.3758/s13428-011-0181-x.
8. *Barry, W. T.* Significance analysis of functional categories in gene expression studies: a structured permutation approach / W. T. Barry, A. B. Nobel, F. A. Wright // Bioinformatics. – 2005. – Vol. 21, Iss. 9. – P. 1943–1949. – DOI: 10.1093/bioinformatics/bti260.
9. *Berry, K. J.* A Primer of Permutation Statistical Methods / K. J. Berry, J. E. Johnston, P. W. Mielke. – Cham, Switzerland. – Springer-Verlag, 2019. – 476 p. – DOI: 10.1007/978-3-030-20933-9.
10. An efficient image encryption scheme using gray code based permutation approach / J. Chen, Z. Zhu, C. Fu [et al.] // Optics and Lasers in Engineering. – 2015. – Vol. 67. – P. 191–204. – DOI: 10.1016/j.optlaseng.2014.11.017.
11. Применение перестановочного метода к оценке прогностической способности моделей пространственного распределения концентраций меди и железа в верхнем слое почвы / А. П. Сергеев, А. С. Буторова, А. В. Шичкин [и др.] // Геоинформатика. – 2022. – № 2. – С. 42–53. – DOI: 10.47148/1609-364X-2022-2-42-53.
12. *Edgington, E. S.* Randomization Tests, 3rd ed. / E. S. Edgington. – Dekker, New York, 1995. – 376 p.
13. *Lehmann, E. L.* Nonparametrics: Statistical Methods Based on Ranks / E. L. Lehmann. – Holden–Day, San Francisco, 1975. – 480 p.
14. *Berry, K. J.* Permutation statistical methods: An integrated approach / K. J. Berry, P. W. Mielke, J. E. Johnston. – Cham, Switzerland. – Springer-Verlag, 2016. – 642 p.
15. Permutation methods. Part II / K. J. Berry, J. E. Johnston, P. W. Mielke, L. A. Johnston // Wiley Interdisciplinary Reviews: Computational Statistics. – 2018. – Vol. 10, Iss. 3. – e1429. – DOI: 10.1002/wics.1429.
16. *Mielke, P. W.* Permutation Methods. A Distance Function Approach / P. W. Mielke, K. J. Berry. – New York. – Springer, 2001. – DOI:10.1007/978-1-4757-3449-2.
17. *Goodman, L. A.* Measures of association for cross classifications / L. A. Goodman, W. H. Kruskal // Journal of the American Statistical Association. – 1954. – Vol. 49, № 268. – P. 732–764.
18. *Kończak, G.* Applications of Permutation Methods in the Analysis of Associations / G. Kończak // Argumenta Oeconomica Cracoviensia. – 2020. – No 1(22). – P. 31–45. – DOI: 10.15678/AOC.2020.2203.

Сведения об авторах:

Сергеев Александр Петрович, к. ф.-м. н., ведущий научный сотрудник, и. о. заведующего лабораторией физики и экологии Института промышленной экологии УрО РАН, г. Екатеринбург, Россия; доцент, заведующий лабораторией искусственного интеллекта Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета им. Б. Н. Ельцина, г. Екатеринбург, Россия. Эл. почта: aleksandrpsergeev@gmail.com.

Буторова Анастасия Сергеевна, младший научный сотрудник лаборатории физики и экологии Института промышленной экологии УрО РАН, г. Екатеринбург, Россия; аспирант, младший научный сотрудник лаборатории искусственного интеллекта Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета им. Б. Н. Ельцина, г. Екатеринбург, Россия.

Корюкин Егор Александрович, аспирант, лаборант-исследователь лаборатории искусственного интеллекта Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета им. Б. Н. Ельцина, г. Екатеринбург, Россия.

Бобаков Вениамин Сергеевич, студент, лаборант-исследователь лаборатории искусственного интеллекта Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета им. Б. Н. Ельцина, г. Екатеринбург, Россия.

Павлова Светлана Вячеславовна, младший научный сотрудник лаборатории искусственного интеллекта Института радиоэлектроники и информационных технологий – РТФ Уральского федерального университета им. Б. Н. Ельцина, г. Екатеринбург, Россия.

NONPARAMETRIC RANDOMIZATION TEST: IDEA, ALGORITHM, ADVANTAGES, DISADVANTAGES AND APPLICATION EXAMPLE

A. P. Sergeev [1], A. S. Butorova [1,2], E. A. Koryukin [2], V. S. Bobakov [2], S. V. Pavlova [2]

[1] *Institute of Industrial Ecology, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, Russia*

[2] *Ural Federal University named after the first President of Russia B. N. Yeltsin, Ekaterinburg, Russia*

The permutation approach has a number of advantages over traditional parametric statistical methods, which makes it preferable for data processing in some research situations. Permutational methods do not use the assumptions about the shape of the probability distribution; they are suitable for small data sets and give the researcher great freedom in choosing test statistics. This paper is devoted to the use of one of the variants of permutation methods – a nonparametric randomization test. The authors outline the premises of the randomization test and provide some formalizations for the case of comparing two groups of observations. An application example of a randomization test to test a statistical hypothesis about the equality of the intergroup difference in means to zero in cases of right- and left-sided critical sets is considered. The example compares the performance of two purification methods – standard and new ones. In the right-sided case, the test statistic is the sum of purification times in the group of the standard method; in the left-sided case, it is the intergroup difference in mean purification times. Using a randomization test, it was shown that the new purification method does not provide a statistically significant advantage in purification time over the standard method.

Key words: permutation approach; randomization test; permutation; statistical hypothesis testing; type I error; type II error; *p*-value.

References

1. *Fisher, R. A.* Statistical methods for research workers / R. A. Fisher. – Edinburgh, Scotland. – Oliver and Boyd, 1925. – 232 p.
2. *Fisher, R. A.* The arrangement of field experiments / R. A. Fisher // Journal of the American Statistical Association. – 1926. – Vol. 33. – P. 503–513.
3. *Fisher, R. A.* The design of experiments / R. A. Fisher. – Edinburgh, Scotland England. – Oliver and Boyd, 1935. – 252 p.
4. *Ernst, M. D.* Permutation Methods: A Basis for Exact Inference / M. D. Ernst // Statistical Science. – 2004. – Vol. 19, № 4. – P. 676–685. – DOI: 10.1214/088342304000000396.
5. *Fişek, M. H.* Permutation tests for goodness-of-fit testing of mathematical models to experimental data / M. H. Fişek, Z. Barlas // Social Science Research. – 2013. – Vol. 42, Iss. 2. – P. 482–495. – DOI: 10.1016/j.ssresearch.2012.09.010.
6. *Ross, S. M.* Simulation, bootstrap statistical methods, and permutation tests / S. M. Ross // Introduction to Probability and Statistics for Engineers and Scientists. – Amsterdam. – Academic Press; Elsevier, 2021. – P. 619–647. – DOI: 10.1016/B978-0-12-824346-6.00024-7.
7. *Taylor, A. B.* Four applications of permutation methods to testing a single-mediator model / A. B. Taylor, D. P. MacKinnon // Behavior Research Methods. – 2012. – Vol. 44, Iss. 3. – P. 806–844. – DOI: 10.3758/s13428-011-0181-x.
8. *Barry, W. T.* Significance analysis of functional categories in gene expression studies: a structured permutation approach / W. T. Barry, A. B. Nobel, F. A. Wright // Bioinformatics. – 2005. – Vol. 21, Iss. 9. – P. 1943–1949. – DOI: 10.1093/bioinformatics/bti260.
9. *Berry, K. J.* A Primer of Permutation Statistical Methods / K. J. Berry, J. E. Johnston, P. W. Mielke. – Cham, Switzerland. – Springer-Verlag, 2019. – 476 p. – DOI: 10.1007/978-3-030-20933-9.
10. An efficient image encryption scheme using gray code based permutation approach / J. Chen, Z. Zhu, C. Fu [et al.] // Optics and Lasers in Engineering. – 2015. – Vol. 67. – P. 191–204. – DOI: 10.1016/j.optlaseng.2014.11.017.

11. Application of the permutation method to the assessment of predictive ability of the models of spatial distribution of copper and iron concentrations in the topsoil / A. P. Sergeev, A. S. Butorova, A. V. Shichkin [et al.] // Geoinformatika. – 2022. – № 2. – P. 42–53. – DOI: 10.47148/1609-364X-2022-2-42-53.
12. *Edgington, E. S.* Randomization Tests, 3rd ed. / E. S. Edgington. – Dekker, New York, 1995. – 376 p.
13. *Lehmann, E. L.* Nonparametrics: Statistical Methods Based on Ranks / E. L. Lehmann. – Holden–Day, San Francisco, 1975. – 480 p.
14. *Berry, K. J.* Permutation statistical methods: An integrated approach / K. J. Berry, P. W. Mielke, J. E. Johnston. – Cham, Switzerland. – Springer-Verlag, 2016. – 642 p.
15. Permutation methods. Part II / K. J. Berry, J. E. Johnston, P. W. Mielke, L. A. Johnston // Wiley Interdisciplinary Reviews: Computational Statistics. – 2018. – Vol. 10, Iss. 3. – e1429. – DOI: 10.1002/wics.1429.
16. *Mielke, P. W.* Permutation Methods. A Distance Function Approach / P. W. Mielke, K. J. Berry. – New York. – Springer, 2001. – DOI:10.1007/978-1-4757-3449-2.
17. *Goodman, L. A.* Measures of association for cross classifications / L. A. Goodman, W. H. Kruskal // Journal of the American Statistical Association. – 1954. – Vol. 49, № 268. – P. 732–764.
18. *Kończak, G.* Applications of Permutation Methods in the Analysis of Associations / G. Kończak // Argumenta Oeconomica Cracoviensia. – 2020. – No 1(22). – P. 31–45. – DOI: 10.15678/AOC.2020.2203.