

УДК 330.43 DOI: 10.14451/1.221.380

Робастная альтернатива тесту Стьюдента при исследовании маркетинговых данных

© 2023 **Трушникова Ирина Олеговна**

Кандидат экономических наук, доцент кафедры маркетинга. Санкт-Петербургский Государственный Экономический Университет. Россия, Санкт-Петербург.

E-mail: itrushnikova@mail.ru

© 2023 **Красильников Алексей Борисович**

Кандидат экономических наук, доцент кафедры маркетинга. Санкт-Петербургский Государственный Экономический Университет. Россия, Санкт-Петербург.

E-mail: AKRAC_64@mail.ru

© 2023 **Кузьменкова Вероника Николаевна**

Кандидат экономических наук, доцент. Военный институт (Железнодорожных войск и военных сообщений) Военной академии МТО им. А. В. Хрулева. Россия, Санкт-Петербург.

E-mail: veronicakuzmenkova@rambler.ru

Ключевые слова: робастные методы, критерий знаков, тест Стьюдента, тест Уилкоксона, мощность теста, случайно парный эксперимент.

В статье рассматриваются три теста: критерий знаков, t-тест Стьюдента и непараметрический тест Уилкоксона. Тест Уилкоксона предлагается в качестве альтернативы, сочетающей в себе преимущества теста знаков и t-теста Стьюдента. С помощью компьютерного моделирования показано, как решить вопрос о выборе между тестами, если распределение исходных данных известно. Текст статьи сопровождается расчетами, выполненными в программной среде языка R.

1. В практике обработки результатов маркетинговых наблюдений часто распределение генеральной совокупности неизвестно, либо (для непрерывных случайных величин) отличается от нормального распределения, так что применение классических параметрических статистических методов необоснованно и может приводить к ошибочным выводам. В этом случае применяют методы, не зависящие (или свободные) от распределения генеральной совокупности исходных данных – непараметрические методы.

В статье, после определения понятия математической модели для параметра положения генеральной совокупности и в ее контексте, с единой точки зрения рассматриваются три однопараметрических теста: тест знаков, t-тест Стьюдента и тест Уилкоксона. Тест знаков – это непараметрический тест с минимумом требований к распределению исходных данных, но без учета их амплитуды. Параметрический t-тест Стьюдента используется в случае, когда исходные данные распределены по нормальному закону. Непараметрический тест Уилкоксона предлага-

ется автором в качестве разумной альтернативы, сочетающей в себе преимущества теста знаков и t-теста Стьюдента. Мощность теста Уилкоксона сравнима с мощностью t-теста в случае нормально распределенной выборки, и превышает мощность t-теста в случае, когда распределение выборки имеет более тяжелые «хвосты» по сравнению с нормальным распределением.

Для каждого из трех рассмотренных тестов приведены соответствующие результаты для оценки и доверительного интервала для параметра положения в рамках выбранной математической модели. С помощью компьютерного моделирования показано как в терминах мощности можно решать вопрос о выборе между t-тестом и тестом Уилкоксона в каждом конкретном случае, когда распределение исходных данных известно.

Текст статьи сопровождается расчетами, выполненными в программной среде языка R, иллюстрирующими применение предлагаемых автором процедур для решения практических маркетинговых задач: выявление предпочтений между двумя брендами на однотипную продукцию и сравнение величины спроса на однотипный товар в двух торговых точках.

2. Классический тест (критерий) знаков [1] использует минимум предварительных предположений о характере исследуемых данных. Например, при сравнении двух объектов, тест знаков использует информацию только о том, что один объект лучше другого в каком-либо оговоренном смысле.

В маркетинге тест знаков широко применяется для сравнения популярности двух брендов (пусть это бренд А и бренд Б) одного и того же по функциональности товара (например, мороженого).

В таких случаях эксперимент организуется следующим образом. Группе экспертов в случайном порядке предлагается выбрать (дать предпочтение) тот или иной сорт мороженого при условии, что его бренд им неизвестен. Для иллюстрации примем, что было выбрано 12 экспер-

тов, каждый из которых в случайном порядке попробовал оба бренда мороженого, и бренд А был выбран 10 раз, бренд Б был выбран 1 раз и в одном случае выявить предпочтение не удалось. Эти результаты вполне убедительно говорят в пользу предпочтения экспертами бренда А. Чтобы подтвердить эту догадку статистически, можно оценить вероятность того, что такой результат получен в случае, когда верна нулевая гипотеза H_0 , состоящая в том, что между брендами различий нет, то есть каждый бренд выбирается равновероятно (с вероятностью 1/2). Для этого в условиях теста знаков рассчитывают следующую статистику:

$$S = \sum_{i=1}^n \text{sign}(X_i),$$

где $\text{sign}(t) = -1, 0, 1$ для $t < 0, t = 0, t > 0$ соответственно.

Пусть

$$S^+ = \# \{X_i > 0\}.$$

Тогда $S = 2S^+ - n$. Здесь предполагается, что ни одно из значений X_i не равно нулю (на практике, равные нулю значения из выборки исключают, а объем выборки корректируют).

При условии нулевой гипотезы H_0 , статистика S^+ имеет биномиальное распределение с числом испытаний n (в нашем случае $n = 12 - 1 = 11$) и вероятностью успеха 1/2. Пусть s^+ – выборочное значение статистики S^+ (в нашем случае $s^+ = 10$), тогда величину p-value для правостороннего критерия можно рассчитать в виде

$$P_{H_0}(S^+ \geq s^+) = 1 - F_B(s^+ - 1; n; 0.5),$$

где $F_B(t; n; p)$ – функция биномиального распределения с параметрами n и p .

```
> pbinom(q = 9, size = 11, prob = 0.5,
+       lower.tail = F)
```

[1] 0.005859375

На основании полученного значения p -value нулевую гипотезу следует отклонить на уровне значимости менее 1% и, таким образом, подтвердить, что в результате опроса мороженное бренда А оказалось предпочтительнее мороженого бренда .

Тест знаков – это пример непараметрического теста, свободного от вида распределения генеральной совокупности из которой получены исходные данные. Чтобы понять это допустим, что эксперты измеряли вкус мороженого при помощи чисел. Очевидно, распределение статистики теста знаков при нулевой гипотезе H_0 (а это биномиальное распределение) не зависит от данных экспертами числовых оценок (в этом смысле употребляется термин «свободный от вида распределения»).

3. Сформулируем математическую модель для только что составленной практической задачи. Пусть $X_1, X_2, \dots, X_n$ – случайная выборка, полученная по следующему закону

$$X_i = \theta + e_i, \quad (1)$$

где предполагается, что ошибки $e_1, e_2, \dots, e_n$ – это независимые, одинаково распределенные случайные величины с непрерывной и симметричной относительно нуля плотностью распределения $f(t)$.

При условии симметрии такие параметры положения как математическое ожидание и медиана X_i очевидно равны θ . Модель (1) является моделью о параметре положения.

Таким образом, в рассмотренном в пункте 1 примере с мороженым с помощью критерия знаков была протестирована в рамках модели (1) гипотеза

$$H_0 : \theta = 0, H_a : \theta > 0. \quad (2)$$

4. Наряду с тестом знаков для тестирования

гипотезы (2) на практике широко применяется классический t -тест Стьюдента [4], в основе которого заложена идея о сумме наблюдений. По аналогии со статистикой S теста знаков статистику t -теста можно записать в виде:

$$T = \sum_{i=1}^n \text{sign}(X_i) \cdot |X_i|.$$

Ясно, что распределение статистики T зависит от вида плотности распределения $f(t)$, то есть t -тест Стьюдента является параметрическим.

Обычно t -тест записывают в форме t -отношения

$$t = \frac{\bar{X}}{s/\sqrt{n}},$$

где $\bar{X}$ и s – выборочное среднее арифметическое и стандартное отклонение соответственно.

Если выборка получена из нормального распределения, то статистика t имеет t -распределение Стьюдента с $n - 1$ степенями свободы. Пусть t_0 – наблюдаемое по выборке значение t . Тогда p -value для t -теста для тестирования гипотезы (2) равно

$$P_{H_0}(t \geq t_0) = 1 - F_T(t_0, n - 1),$$

где $F_T(t; v)$ – функция t -распределения Стьюдента с v степенью свободы. В случае нормального распределения найденное p -value будет точным, в других случаях – это приближенное значение.

5. Мы рассмотрели два теста: непараметрический тест знаков и параметрический t -тест Стьюдента. Преимущество первого – минимум требований к виду распределения генеральной совокупности, преимущество второго – значительно большая мощность теста по сравнению с первым в случае, если данные распределены нормально. Маркетинговые данные, в виду высокой стоимости проведения исследований, представляют собой выборки небольшого объема. Это также не позволяет проверить гипотезу о нормальности генеральной совокупности.

Таким образом, для практики целесообразно иметь приемлемую альтернативу: непараметрический тест, сравнимый по мощности с t-тестом в случае нормального распределения исходных данных и имеющий, в общем случае, более высокую мощность по сравнению с t-тестом если распределение исходных данных имеет «тяжелые» хвосты по сравнению с нормальным распределением.

Такой предлагаемой в статье альтернативной является ранговый тест Уилкоксона [6]. Данный тест выгодно использует отличие t-теста от теста знаков, состоящее в том, что статистика t-теста является функцией расстояний элементов выборки относительно нуля в дополнение к их знакам. В статистике теста Уилкоксона используются ранги этих расстояний.

Обозначим $R|X_i|$ ранг $|X_i|$ среди всех $|X_1|, |X_2|, \dots, |X_n|$, упорядоченных от меньшего значения к большему. По аналогии с тестом знаков и t-тестом статистику Уилкоксона можно записать в виде

$$W = \sum_{i=1}^n \text{sign}(X_i) \cdot R|X_i|.$$

Статистика W , также как и рассмотренная выше статистика S , при условии нулевой гипотезы H_0 не зависит от вида распределения $f(t)$.

Хотя распределение W не может быть найдено явно в виде аналитической формулы для ее расчета в программных пакетах используется итерационный алгоритм.

Также как для теста знаков наряду со статистикой W составляют сумму рангов положительных элементов выборки W^+ , то есть

$$W^+ = \sum_{X_i > 0} R|X_i| = \frac{1}{2}W + \frac{n(n+1)}{4}.$$

Пусть w^+ – выборочное значение статистики W^+ . Тогда для гипотезы (2) p-value для теста Уилкоксона равно

$$P_{H_0}(W^+ \geq w^+) = 1 - F_{W^+}(w^+ - 1; n),$$

где $F_{W^+}(x; n)$ – функция распределения статистики Уилкоксона для выборки размера n .

6. Каждый из трех рассмотренных тестов имеет соответствующие оценку и доверительный интервал для параметра θ в модели (1).

Основу рассматриваемых ниже доверительных интервалов составляют порядковые статистики. Будем использовать их стандартные обозначения, а именно: $X_{(1)}$ – наименьший элемент выборки, $X_{(2)}$ – следующий по величине, ..., $X_{(n)}$ – наибольший элемент выборки. Таким образом, $X_{(1)} < X_{(2)} < \dots < X_{(n)}$. Например, для выборки $x_1 = 51, x_2 = 64, x_3 = 43$ упорядоченная выборка имеет вид $x_{(1)} = 43, x_{(2)} = 51, x_{(3)} = 64$.

6.1. Оценкой параметра положения θ , связанной с тестом знаков является выборочная медиана

$$\hat{\theta} = \text{med}\{X_1, X_2, \dots, X_n\}.$$

Для $0 < \alpha < 1$ соответствующий доверительный интервал для θ с доверительной вероятностью $(1 - \alpha)100\%$ задается в виде

$$(X_{(c_1+1)}, X_{(n-c_1)}),$$

где $X_{(i)}$ – i -ая порядковая статистика выборки, $c_1 - \alpha/2$ – квантиль биномиального распределения с параметрами n и $p = 1/2$, то есть $F_B(c_1; n; 0.5) = \alpha/2$.

Данный доверительный интервал не зависит от вида распределения ошибок e_i и, следовательно, имеет точную доверительную вероятность $(1 - \alpha)100\%$ при любом распределении ошибок в модели (1). Отметим, что из-за дискретности биномиального распределения, при каждом значении n существует лишь ограниченный набор значений для α .

6.2. Оценкой параметра положения θ , связанной с t-тестом Стьюдента является выборочное среднее

$$\hat{\theta} = \bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

Классический доверительный интервал в этом случае имеет вид

$$\bar{X} \pm t_{\frac{\alpha}{2}, n-1} \cdot \frac{s}{\sqrt{n}},$$

где $t_{\frac{\alpha}{2}, n-1}$ – $\alpha/2$ -квантиль t -распределения Стьюдента с $n - 1$ степенями свободы и $F_T(-t_{\frac{\alpha}{2}, n-1}; n - 1) = \alpha/2$.

Данный доверительный интервал зависит от вида распределения ошибок e_i и имеет точную доверительную вероятность $(1 - \alpha)100\%$ в случае их нормального распределения и, если их распределение отлично от нормального, то доверительные коэффициенты приблизительно равны $(1 - \alpha)100\%$.

6.3. Оценкой параметра положения θ , связанной с тестом Уилкоксона является оценка Ходжеса-Лемана [3]:

$$\hat{\theta}_W = \text{med}_{i \leq j} \left\{ \frac{X_i + X_j}{2} \right\}.$$

Здесь парные средние $A_{ij} = (X_i + X_j)/2$, $i \leq j$ – средние Уолша.

Пусть $A_{(1)} < \dots < A_{(\frac{n(n+1)}{2})}$ – упорядоченный набор средних Уолша. Тогда $(1 - \alpha)$, 100%-доверительный интервал для θ , имеет вид

$$\left(A_{(c_2+1)}, A_{((n(n+1)/2)-c_2)} \right),$$

где $c_2 - \alpha/2$ – квантиль распределения Уилкоксона.

При условии симметрии относительно нуля распределения ошибок e_i данный доверительный интервал не зависит от вида распределения $f(t)$. Отметим также, что размах значений W^+ – это множество $\{0, 1, \dots, n(n+1)/2\}$, которое имеет порядок n^2 . Таким образом, для небольших по

объему выборок тест Уилкоксона менее зависит от дискретного характера распределения статистики критерия, чем тест знаков и значит найденный уровень значимости α в случае теста Уилкоксона, ближе к ожидаемому.

7. Какой из двух тестов, тест Уилкоксона или t -тест Стьюдента более мощный? Теоретический ответ на этот вопрос зависит от вида распределения случайных ошибок e_i . Однако на него можно легко ответить практически, в случае, когда распределение ошибок заранее известно. Представленные ниже расчеты, выполненные на языке R [5], являются примером компьютерного моделирования, проведенного для конкретных условий.

Пусть в модели (1) случайные ошибки имеют t -распределение с двумя степенями свободы (здесь отметим важность использовать именно стандартизованные распределения). Рассмотрим двусторонний тест $H_0 : \theta = 0$, $H_A : \theta \neq 0$ на уровне значимости $\alpha = 0.05$.

Представленный ниже код генерирует 10000 выборок по 30 элементов в каждой. Для каждой выборки запоминается p -value для двух тестов. Тогда эмпирической мощностью теста рассчитывается как доля тех p -value, которые меньше или равны 0.05.

```
n = 30
df = 2
nsims = 10000
mu = .5

collwil = rep(0, nsims)
collt = rep(0, nsims)
for (i in 1:nsims) {
  x = rt(n, df) + mu
  wil = wilcox.test(x)
  collwil[i] = wil$p.value
  ttest = t.test(x)
  collt[i] = ttest$p.value
}

powwil = rep(0, nsims)
powwil[collwil <= .05] = 1
powerwil = sum(powwil) / nsims

powt = rep(0, nsims)
powt[collt <= .05] = 1
powert = sum(powt) / nsims
```

Код был запущен трижды: при $\theta = 0$ (в ситуации

нулевой гипотезы) и двух альтернативных ситуациях $\theta = 0.5$ и $\theta = 1$. Эмпирические мощности тестов приведены в следующей таблице:

Таблица 1.

Тест	$\theta = 0$	$\theta = 0.5$	$\theta = 1$
Уилкоксона	0.0489	0.4671	0.9211
t-test	0.0385	0.2855	0.6938

Анализ таблицы показывает, что эмпирический уровень значимости α при $\theta = 0$ для теста Уилкоксона близок к номинальному значению 0.05, что не удивительно, так как этот тест свободен от вида распределения ошибок в модели (1). При этом t-тест немного более консервативный, то есть даёт более широкий доверительный интервал для параметра θ . В двух других ситуациях при $\theta = 0.5$ и $\theta = 1$ в терминах мощности тест Уилкоксона существенно превосходит t-тест и поэтому ему следует отдать предпочтение.

8. В качестве заключительного примера рассмотрим применение тестов Уилкоксона и t-теста Стьюдента для решения практической маркетинговой задачи: по данным об объеме продаж (в штуках) восьми товарных позиций в двух магазинах А и В за неделю постараемся ответить на вопрос: в каком магазине спрос на товары выше?

С целью решения составим выборку, каждый элемент которой представляет собой разницу в продажах соответствующей товарной позиции в магазинах А и В. Пусть θ характеризует центральное значение образованной выборки. Так как исследуемые товары идентичны, то можно считать их распределение по магазинам равновероятным, например, в соответствии с подбрасыванием правильной монеты. Пусть нулевая гипотеза $H_0 : \theta = 0$ верна, то есть спрос на товары в магазинах одинаковый. Если d — наблюдаемая разница в спросе, то получить значение d или $-d$ зависит от того, орлом или решкой выпала монета при отправке товара в магазин. Следовательно, D и $-D$ имеют одно и то же распределение, то есть D распределено симметрично относительно нуля. Следовательно предположение о симметрии распределе-

ния в модели (1), необходимое для корректного применения теста Уилкоксона, автоматически выполнено.

Следующий код программы показывает результаты применения теста Уилкоксона и t-теста Стьюдента для проверки правосторонней гипотезы (2):

$$H_0 : \theta = 0, H_a : \theta > 0.$$

```
>Store_A<- c(82, 69, 73, 43, 58, 56, 76, 65)
>Store_B<- c(63, 42, 74, 37, 51, 43, 80, 62)
> response <- Store_A - Store_B

>wilcox.test(response, alternative = "greater",
  ~ conf.int = TRUE)

Wilcoxon signed rank exact test

data: response
V = 32, p-value = 0.02734
alternative hypothesis: true location is
  ~ greater than 0
95 percent confidence interval:
  1 Inf
sample estimates:
(pseudo)median
  7.75

>t.test(response, alternative = "greater",
  ~ conf.int = TRUE)

One Sample t-test

data: response
t = 2.3791, df = 7, p-value = 0.02447
alternative hypothesis: true mean is greater
  ~ than 0
95 percent confidence interval:
  1.781971 Inf
sample estimates:
mean of x
  8.75
```

Тест Уилкоксона возвращает статистику W^+ (32), p-value теста (0.02734), оценку Ходжеса-Лемана для θ (7.75) и 95%-правосторонний доверительный интервал для θ ($[1; +\infty)$). Тест t-тест Стьюдента возвращает статистику: отношения (2.3791), p-value теста (0.02447), оценку среднего арифметического для θ (8.75) и 95%-правосторонний доверительный интервал для θ ($[1.782; +\infty)$).

Как видно, обе процедуры отвергают нулевую гипотезу на уровне 0.05, что позволяет говорить

о большем спросе в магазине А, чем в магазине В.

При решении задачи использовался случайно-парный план эксперимента. Для таких планов [2] экспериментальной единицей является блок размером два (в нашем случае пара однотипных товаров), отклик также имеет два уровня (в нашем случае объем продаж в штуках), внутри блока уровни распределяются случайно, например, подбрасыванием правильной монетой (в нашем случае так распределялись товары по торговым точкам).

9. Подведем следующие итоги.

1. В рамках модели о параметре положения при малых и умеренных объемах выборки целесообразно использовать непараметрический знаково-ранговый тест Уилкоксона, единственным требованием которого к рас-

пределению данных является его непрерывность и симметричность относительно нуля.

2. Тест Уилкоксона для небольших по размеру выборок менее зависит от дискретного характера распределения статистики критерия чем тест знаков, сравним по мощности с t-тестом в случае нормального распределения и существенно превышает мощность t-теста в случае непрерывных симметричных распределений с более тяжелыми «хвостами» по сравнению с нормальным распределением.
3. В каждом конкретном случае, когда вид распределения исходных данных известен, вопрос о выборе того или иного теста для их исследования легко решается с помощью компьютерного моделирования.
4. Тест Уилкоксона является оптимальным для задач со случайно парным планом эксперимента.

Библиографический список

1. Conover W. J. Practical nonparametric Statistics. — 3rd ed. — Wiley, 1999. — 584 p.
2. Hinkelmann K., Kempthorne O. Design and Analysis of Experiments. Vol. I and II. — 2nd ed. — Wiley, 2008.
3. Hodges J. L., Lehmann E. L. Estimates of Location Based on Rank Tests // The Annals of Mathematical Statistics. — 1963. — Vol. 34, no. 2. — P. 598–611. — ISSN 00034851. — URL: <http://www.jstor.org/stable/2238406>.
4. Student. The Probable Error of a Mean // Biometrika. — 1908. — Mar. — Vol. 6, no. 1. — P. 1. — DOI: [10.2307/2331554](https://doi.org/10.2307/2331554).
5. The R Project for Statistical Computing. — URL: <https://www.r-project.org> (visited on 03/17/2023).
6. Wilcoxon F. Individual Comparisons by Ranking Methods // Biometrics Bulletin. — 1945. — Dec. — Vol. 1, no. 6. — P. 80–83. — DOI: [10.2307/3001968](https://doi.org/10.2307/3001968).