Исследование критериев однородности средних

А. Ю. Новикова, Б. Ю. Лемешко[1]
Новосибирский государственный технический университет

Методами статистического моделирования исследованы распределения статистик критериев однородности средних. Приводятся результаты сравнительного анализа мощности критериев по отношению к конкурирующим гипотезам, делаются выводы о предпочтительности использования тех или иных критериев. Реализована возможность применения и исследования распределений статистик критериев, в условиях нарушения стандартных предположений. Полученные результаты должны способствовать корректному применению критериев в приложениях.

Ключевые слова: проверка гипотезы, статистическое моделирование, мощность критерия.

1. Введение

Критерии проверки гипотез об однородности используются во многих приложениях. При этом речь может идти о проверке гипотез об однородности законов распределения, соответствующих анализируемым выборкам, или об однородности математических ожиданий, или об однородности дисперсий. Естественно, что более подробные выводы могут быть получены в первом случае.

Критерии проверки гипотез об однородности математических ожиданий (об однородности средних) используются в задачах статистического контроля качества для анализа наличия (отсутствия) возмущения в значениях контролируемого показателя процесса. Подобные задачи проверки гипотез возникают в различных измерительных задачах.

С применением параметрических критериев проверки однородности средних часто связывают вопросы: корректности получаемых выводов в конкретной ситуации. Дело в том, что одним их основных предположений при построении этих критериев является принадлежность наблюдаемых случайных величин (погрешностей измерений) нормальному закону распределения. Поэтому исследователей часто критикуют за использование параметрических критериев однородности средних без проверки принадлежности результатов измерений нормальному закону, высказывая опасения возможной некорректности выводов вследствие того, что при нарушении данного предположения условные распределения статистик критериев могут сильно изменяться.

В то же время, существуют результаты, показывающие, что ряд параметрических критериев однородности [1-3] обладает определенной устойчивостью, а распределения статистик таких критериев устойчивы к нарушению стандартного предположения о нормальности. Это значит, что при определённых нарушениях можно применять такие критерии, и выводы будут оставаться корректными, если наблюдаемый закон не будет очень уж сильно отличаться от нормального.

Непараметрические критерии однородности средних не требуют выполнения предположения о нормальности, но могут иметь свои недостатки. В частности, они несколько уступают в мощности параметрическим, условные распределения их статистик при ограниченных объёмах выборок (при справедливости проверяемой гипотезы) могут существенно отличаться от асимптотических. В некоторых ситуациях у непараметрических критериев может исче-

[1] Исследования выполнены при поддержке Министерства образования и науки РФ в рамках государственной работы «Обеспечение проведения научных исследований» и проектной части государственного задания (проект № 1.1009.2017/ПЧ).

зать свойство непараметричности. Поэтому использование непараметрических критериев требует знания достоинств и недостатков этих критериев.

В данной работе рассматриваются непараметрические критерии однородности средних: ранговый знаковый критерий Уилкоксона, критерий Фишера–Йейтса–Терри–Гефдинга, критерий Ван дер Вардена и многовыборочный критерий Ван дер Вардена.

2. Критерии проверки гипотез об однородности средних

Проверяемая гипотеза об однородности математических ожиданий имеет вид

$$H_0\colon \mu_1 = \mu_2 = \dots = \mu_k ,$$

а конкурирующая с ней гипотеза

$$H_1\colon \mu_{i_1} \neq \mu_{i_2} ,$$

где неравенство выполняется, хотя бы для одной пары индексов i_1, i_2.

Для проверки гипотез строятся статистические критерии. В процессе проверки на основании вычисляемого по анализируемым выборкам значения статистики критерия принимается решение: проверяемая гипотеза принимается или отклоняется. С проверкой статистических гипотез связывают ошибки двух видов. Ошибка первого рода заключается в том, что гипотеза H_0 отвергается, когда она верна. Вероятность такой ошибки (уровень значимости), как правило, обозначают α. Ошибка второго рода связана с тем, что гипотеза H_0 не отклоняется при справедливости некоторой конкурирующей гипотезы H_1. Вероятность ошибки второго рода обозначают β. При заданной конкурирующей гипотезе H_1 можно говорить о мощности критерия $1-\beta$ относительно этой гипотезы при заданном уровне значимости α.

Очевидно, что чем выше мощность критерия, тем лучше он различает гипотезы H_0 и H_1, тем он предпочтительней

В данном разделе кратко рассматривается ряд непараметрических критериев, исследованных в данной работе.

Знаковый ранговый критерий Уилкоксона. Этот критерий является непараметрическим аналогом *t*-критерия для сравнения двух средних значений непрерывных распределений и представляет собой развитие критерия Манна и Уитни [4].

Для вычисления статистики Уилкоксона строится ряд разностей $|x_i - y_i|$ двух выборок случайных величин X и Y одинакового объема n, который затем ранжируется по возрастанию. В упорядоченном ряду значений $|x_i - y_i|$ определяют сумму рангов T величин $x_i - y_i > 0$:

$$T = \sum_{i=1}^{n} R_i ,\ x_i - y_i > 0 .$$

При $T_1(\alpha) \le T \le T_2(\alpha)$ гипотеза о наличии сдвига отклоняется, где $T_1(\alpha)$, $T_2(\alpha)$ - нижнее и верхнее критические значения [4].

Можно использовать нормализованную статистику данного критерия [4]:

$$T^* = \frac{T - \dfrac{n(n+1)}{4}}{\sqrt{n(n+1)(2n+1)/24}},$$

где n – объем выборок. Как показали наши исследования распределения нормализованной статистики, стандартным нормальным законом можно пользоваться для вычисления достигнутого уровня значимости p_{value} при объемах выборок $n \ge 9$.

Ещё одна модификация статистики предложена в [5]

$$K = \frac{T^*}{2}\left\{1 + \sqrt{\frac{n-1}{n-(\mathrm{T}^*)^2}}\right\},$$

для которой критические значения вычисляются на основании критических значений стандартного нормального закона и t-распределения Стьюдента с $n-1$ степенями свободы

$$K(\alpha) = \frac{1}{2}u_\alpha + \frac{1}{2}t_{n-1,\alpha},$$

где u_α – α-квантиль стандартного нормального распределения, $t_{n-1,\alpha}$ – α-квантиль распределения Стьюдента.

Критерий Фишера-Йейста-Терри-Гефдинга. Критерий рассмотрен в работах Фишера и Йейтса [6], Терри [7] и Гефдинга [8]. Статистика критерия имеет вид:

$$S = \sum_{i=1}^{n_1} a_{n_1+n_2}(\mathrm{R}_i),$$

где $a_{n_1+n_2}(i) = E\left[z^i_{n_1+n_2}\right]$ – математическое ожидание i-й порядковой статистики в выборке объема n_1+n_2 из стандартного нормального распределения; R_i - ранг значений y_i в объединенной ранжированной выборке величин X и Y (или ранг x_i в объединенной выборке, тогда суммирование ведётся по $i=1,...,n_2$). Для вычисления $a_n(i)$ может быть использована аппроксимация $a_n(i) = 4.91\left[p^{0.14} - (1-p)^{0.14}\right]$, где $p = (i-3/8)/(n+1/4)$.

Если $|S| < S(\alpha)$, то гипотеза H_0 не отклоняется, где $S(\alpha)$ – критическое значение статистики [4].

Данные исследования подтвердили, что при $n_1, n_2 \ge 8$ распределение статистики S критерия Фишера–Йейтса–Терри–Гефдинга удовлетворительно аппроксимируется нормальным законом со средним $E[S]=0$ и дисперсией

$$D[\mathrm{S}] = \frac{n_1 n_2}{(n_1+n_2)(n_1+n_2-1)} \sum_{i=1}^{n_1+n_2} a^2_{n_1+n_2}(\mathrm{i}).$$

В этом случае гипотеза H_0 не отклоняется, если $|S^*| = \left|\frac{S}{\sqrt{D[\mathrm{S}]}}\right| < u_{(1+\alpha)/2}$, где u_γ – γ-квантиль стандартного нормального распределения.

Многовыборочный критерий Ван дер Вардена. Статистика критерия Ван дер Вардена для проверки гипотезы о равенстве средних по k выборкам имеет вид [9]:

$$T = (n-1)\sum_{i=1}^{k}\frac{1}{n_i}\left(\sum_{j=1}^{n_i} u_{R_{ij}/(n+1)}\right)^2 \Bigg/ \sum_{i=1}^{n} u^2_{i/(n+1)},$$

где $n = \sum_{i=1}^{k} n_i$, u_γ – γ-квантиль стандартного нормального закона, R_{ij}, $j=\overline{1,n_i}$, – ранг j-го элемента i-й выборки в вариационном ряду объединённой выборки объёма n.

При справедливости проверяемой гипотезы H_0 статистика хорошо описывается χ^2_{k-1}-распределением. Было показано, что отклонением распределения статистики от χ^2_{k-1}-распределения можно пренебречь при $n_i > 30$.

При $n_i = 10$ в интервале значений функции распределения статистики от 0.9 до 1 функция χ^2_{k-1}-распределения максимально отклоняется от реального распределения статистики в сторону меньших значений на величину порядка 0.005. То есть на такую величину может быть завышен достигнутый уровень значимости p_{value}, вычисляемый по χ^2_{k-1}-распределению.

3. Сравнительный анализ мощности критериев

В процессе исследования распределений статистик и оценивания мощности критериев однородности средних были зафиксированы некоторые особенности, которые могут влиять на формирование выводов при использовании соответствующих критериев.

Методами статистического моделирования (для вероятностей ошибок первого рода $\alpha = 0.1; 0.05; 0.01$) мощность рассматриваемых критериев однородности средних для случая 2-х выборок исследовалась относительно конкурирующих гипотез:

H_1^1: $\mu_2 = \mu_1 + 0.1\sigma$; H_1^2: $\mu_2 = \mu_1 + 0.2\sigma$; H_1^3: $\mu_2 = \mu_1 + 0.3\sigma$; H_1^4: $\mu_2 = \mu_1 + 0.4\sigma$;

H_1^5: $\mu_2 = \mu_1 + 0.5\sigma$; H_1^6: $\mu_2 = \mu_1 + \sigma$.

В табл.1 рассматриваемые критерии упорядочены по убыванию мощности.

Таблица 1. Оценки мощности критериев относительно конкурирующей гипотезы H_1^1: $\mu_2 = \mu_1 + 0.1\sigma$

Критерий	α	Объёмы выборок				
		$n=10$	$n=20$	$n=30$	$n=50$	$n=100$
Критерий Фишера–Йейтса–Терри–Гефдинга	0.1	0.143	0.166	0.184	0.216	0.281
	0.05	0.076	0.090	0.103	0.125	0.173
	0.01	0.017	0.022	0.026	0.033	0.052
Ранговый знаковый критерий Уилкоксона	0.1	0.137	0.155	0.180	0.210	0.276
	0.05	0.064	0.086	0.098	0.122	0.169
	0.01	0.016	0.020	0.024	0.032	0.050
Многовыборочный критерий Ван дер Вардена	0.1	0.110	0.116	0.127	0.141	0.179
	0.05	0.028	0.032	0.034	0.043	0.064
	0.01	0.012	0.014	0.015	0.020	0.037
Критерий Ван дер Вардена	0.1	0.108	0.116	0.125	0.141	0.183
	0.05	0.055	0.061	0.067	0.078	0.108
	0.01	0.012	0.013	0.015	0.020	0.031

В случае выполнения предположений о нормальном законе и при близкой конкурирующей гипотезе наибольшую мощность показали критерий Фишера–Йейтса–Терри–Гефдинга и ранговый знаковый критерий Уилкоксона. Однако, при уровне значимости 0.01 критерий Ван дер Вардена начинает уступать по мощности многовыборочному критерию Ван дер Вардена. Подобную картину можно наблюдать и при относительно далекой конкурирующей гипотезе H_1^3 когда $n > 50$. Подробнее результаты исследований представлены в [9].

Многовыборочный критерий Ван дер Вардена может применяться при числе выборок больше 2-х. В табл. 2 и 3 оценки мощности данного критерия сравниваются с оценками мощности критерия Краскела–Уалиса [9] (в случае принадлежности выборок нормальному закону) при числе выборок $m = 3$ и $m = 5$ и их объемах $n = 50$.

Таблица 2. Оценки мощности многовыборочных критериев однородности средних относительно конкурирующей гипотезы $H_1^2:\ \mu_2 = \mu_1 + 0.2\sigma$

Критерий	$m=3$			$m=5$		
	0.1	0.05	0.01	0.1	0.05	0.01
Многовыборочный критерий Ван дер Вардена	0.251	0.160	0.054	0.232	0.140	0.045
H-критерий Краскела–Уаллиса	0.244	0.155	0.054	0.224	0.134	0.041

Таблица 3. Оценки мощности многовыборочных критериев однородности средних относительно конкурирующей гипотезы $H_1^6:\ \mu_2 = \mu_1 + \sigma$

Критерий	$m=3$			$m=5$		
	0.1	0.05	0.01	0.1	0.05	0.01
Многовыборочный критерий Ван дер Вардена	1.000	1.000	0.997	1.000	1.000	0.998
H-критерий Краскела–Уаллиса	1.000	0.999	0.996	1.000	1.000	0.997

В многовыборочном варианте в случае выполнения предположений о нормальности предпочтение в мощности можно отдать критерию Ван дер Вардена.

Распределения ненормализованных статистик рассмотренного множества критериев существенно зависят от объема выборок. Распределения статистик нормализованного рангового знакового критерия Уилкоксона, критерия Фишера–Йейтса–Терри–Гефдинга хорошо описываются стандартным нормальным законом распределения. Это можно объяснить тем что, начиная с объемов выборок $n > 8$ в случае критерия Уилкоксона, и объемов выборок $n > 9$ в случае критерия Фишера–Йейтса–Терри–Гефдинга гипотеза о согласии со стандартным нормальным законом не отклоняется. Про многовыборочный критерий Ван дер Вардена можно сказать следующее: его статистика как и вышеперечисленные, не зависит от объема выборок. Объясняется это тем, что начиная с объемов выборок $n > 21$, проверяемая нулевая гипотеза не отвергается. При таких объёмах выборок распределение данной статистики критерия при справедливости нулевой гипотезы, хорошо описывается χ^2_{k-1}-распределением.

Применение непараметрических критериев однородности средних предполагает, что анализируемые выборки принадлежат одному и тому же виду закона. Именно тогда при справедливости проверяемой гипотезы распределение статистики имеет соответствующее асимптотическое распределение, например, стандартный нормальный закон для большинства нормализованных статистик. Если закон распределения одной из выборок отличается, то распределения статистик непараметрических критериев проверки гипотез об однородности средних (при справедливости проверяемой нулевой гипотезы) будет отличаться от распределения, имеющего место в случае принадлежности выборок одному и тому же виду законов.

4. Заключение

Методами статистического моделирования исследованы распределения статистик непараметрических критериев однородности средних. Проведен сравнительный анализ мощности критериев относительно некоторых конкурирующих гипотез, что позволяет судить о предпочтительности применения тех или других критериев. Отмечены достоинства и недостатки отдельных критериев.

В рамках развиваемой программной системы ISW реализован интерактивный режим исследования распределений статистик, позволяющий оценивать достигнутый уровень значимости p_{value} в ситуации нарушения стандартных предположений или в случае неизвестного распределения статистики, задаваемого лишь таблицей процентных точек. Это делает формируемый статистический вывод о результатах проверки гипотезы более информативным и более обоснованным.

Литература

1. *Лемешко Б. Ю., Лемешко С. Б.* Об устойчивости и мощности критериев проверки однородности средних // Измерительная техника. 2008. № 9. С. 23-28.
2. *Lemeshko B. Yu., Lemeshko S. B.* Power and robustness of criteria used to verify the homogeneity of means // Measurement Techniques. 2008. Vol. 51, № 9. P. 50-959.
3. *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. Новосибирск : Изд-во НГТУ, 2011. 888 с.
4. *Кобзарь А. И.* Прикладная математическая статистика : для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
5. *Iman R. L.* Use of a t-statistic as an approximation to the Wilcoxon signed rank test statistic // Communications in Statistic. 1974. Vol. 3. P. 795-806.
6. *Fisher R. A., Yates F.* Statistical tables for biological, agricultural and medical research. London : Oliver & Boyd, 1938. 90 p.
7. *Terry M. E.* Some rank order test which are most powerful against specific parametric alternatives // The Annals of Mathematical Statistics. 1952. Vol. 23. P. 346-366.
8. *Hoeffding W.* On the distribution of the expected value of the order statistics // The Annals of Mathematical Statistics. 1953. Vol. 24. P. 93-100.
9. *Лемешко Б. Ю.* Критерии проверки гипотез об однородности. Руководство по применению. М. : ИНФРА-М, 2017. 208 с.

Лемешко Борис Юрьевич

д.т.н., профессор, г.н.с. кафедры теоретической и прикладной информатики НГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: lemeshko@ami.nstu.ru.

Новикова Алена Юрьевна

Магистрант кафедры теоретической и прикладной информатикиНГТУ (630073, Новосибирск, просп. Карла Маркса, 20), e-mail: alena.shestakova.92@inbox.ru.

About the trend tests in the mathematical expectation

A. Yu. Novikova, B. Yu. Lemeshko

Novosibirsk State Technical University

The methods of statistical modeling to are used investigate the distribution uniformity of statistics criteria of averages. The result of the criteria averages of the power of criteria with respect to the competing hypotheses are given, conclusions about the use of preference for certain criteria are made. The possibility of the use of statistics and research criteria distributions in a violation of the standard assumptions is presented. The results should contribute to the correct application of the criteria in the applications.

Keywords: checking a hypothesis, statistical modeling, power of the test.