

В.С. Еременко, Ю.В. Куц, В.М. Мокийчук

ОЦЕНКА ОДНОРОДНОСТИ ВЫБОРОК МАЛОГО ОБЪЕМА

Приведены результаты исследований критериев нормальности и однородности выборок малого объема. Показано, что применяемые при обработке результатов измерений критерии согласия не позволяют однозначно определить тип закона распределения, а мощности критериев однородности Фишера и Стьюдента для малых выборок практически не зависят от типа закона распределения.

статистические критерии, однородность, результат измерения, неопределенность, воспроизводимость, имитационное моделирование

Стандарт ISO 17025 [1], регламентирующий требования к компетентности испытательных лабораторий, устанавливает также и ряд требований предъявляемых к самой процедуре измерения и оценке её неопределенности. Среди этих требований следует отметить:

- обеспечение качества результатов измерений;
- оценку пригодности результатов измерений;
- воспроизводимость результатов измерений;
- оценку компетентности персонала, выполняющего измерения.

Соответствие испытательной лабораторией приведенным требованиям должно сопровождаться документальным подтверждением, базирующимся на научно обоснованных и достоверных критериях. Наиболее общим подходом, обеспечивающим выполнение вышеприведенных требований, является доказательство однородности результатов измерений, полученных в различных лабораториях, в разное время, в различных условиях, различными методами и разными операторами. Особенностью проверки однородности результатов измерений в испытательных лабораториях (особенно при аналитических испытаниях) является ограниченное количество результатов (до 10–15), что приводит к определенным трудностям выбора и использования критериев однородности.

Одними из наиболее мощных критериев проверки однородности независимых выборок являются критерий Фишера для проверки однородности дисперсий и критерий Стьюдента для проверки однородности математических ожиданий. Из литературных источников [2, 3] известно, что мощность критерия Фишера определяется степенью приближения закона распределения выборок к гауссовскому. Поэтому, использование данного критерия сопряжено с дополнительной проблемой идентификации закона распределения выборки и, в случае отличия закона распределения от гауссовского, использования модифицированных, робастных критериев.

Существует ряд критериев для оценки законов распределения выборок, таких как, например, со-

ставной критерий, W-критерий Шапиро-Уилкса [4], RS-критерий [3] и критерий, предложенный в работе [5], однако их адекватность для выборок малого объема (до 15 значений) не достаточно исследована.

Целью статьи является:

1. Исследование существующих критериев проверки «нормальности» закона распределения применительно к выборкам малого объема.
2. Исследование критериев проверки однородности результатов измерений по выборкам малого объема.

В статье [6] проведен анализ составного критерия применительно к малым выборкам и показано, что данный критерий не дает удовлетворительных результатов.

Для приведенного в работе [5] критерия не определены его статистические характеристики, а приведенные данные не достаточно полно отражают достоверность идентификации закона распределений.

Критерии согласия использующие статистики, определяемые на основе коэффициентов асимметрии и эксцесса, не применяются для анализа выборок малого объема из-за значительных дисперсий оценок этих коэффициентов [3], и поэтому не рассматриваются в данной статье.

Применение W-критерия регламентировалось ГОСТ 11.006-74, и, хотя, он был отменен в 1987 г., W-критерий продолжает встречаться в отечественной и зарубежной литературе по обработке результатов измерений для малых выборок.

Исходя из вышесказанного, в статье рассматриваются только W- и RS-критерии оценки законов распределения.

Результаты исследований. Экспериментальное исследование критериев проводилось с помощью имитационного моделирования, которое основывается на методе статистических испытаний Монте-Карло.

При исследовании в качестве тестовых генерировались 10000 выборок случайной величины из генеральных совокупностей с гауссовским, равно-

мерным и треугольным законами распределения и объемом $N = 4, 10, 16$ и 34 значения.

W-критерий. На рис. 1 приведены гистограммы оценки W-статистики для центрированных выборок объемом $N = 10$ значений и среднеквад-

ратическим отклонением (СКО) равным 1, а на рис. 2 приведены гистограммы для аналогичных выборок объемом 34 значения (W_α – граничное значение статистики при 5-% уровне значимости для данного N).

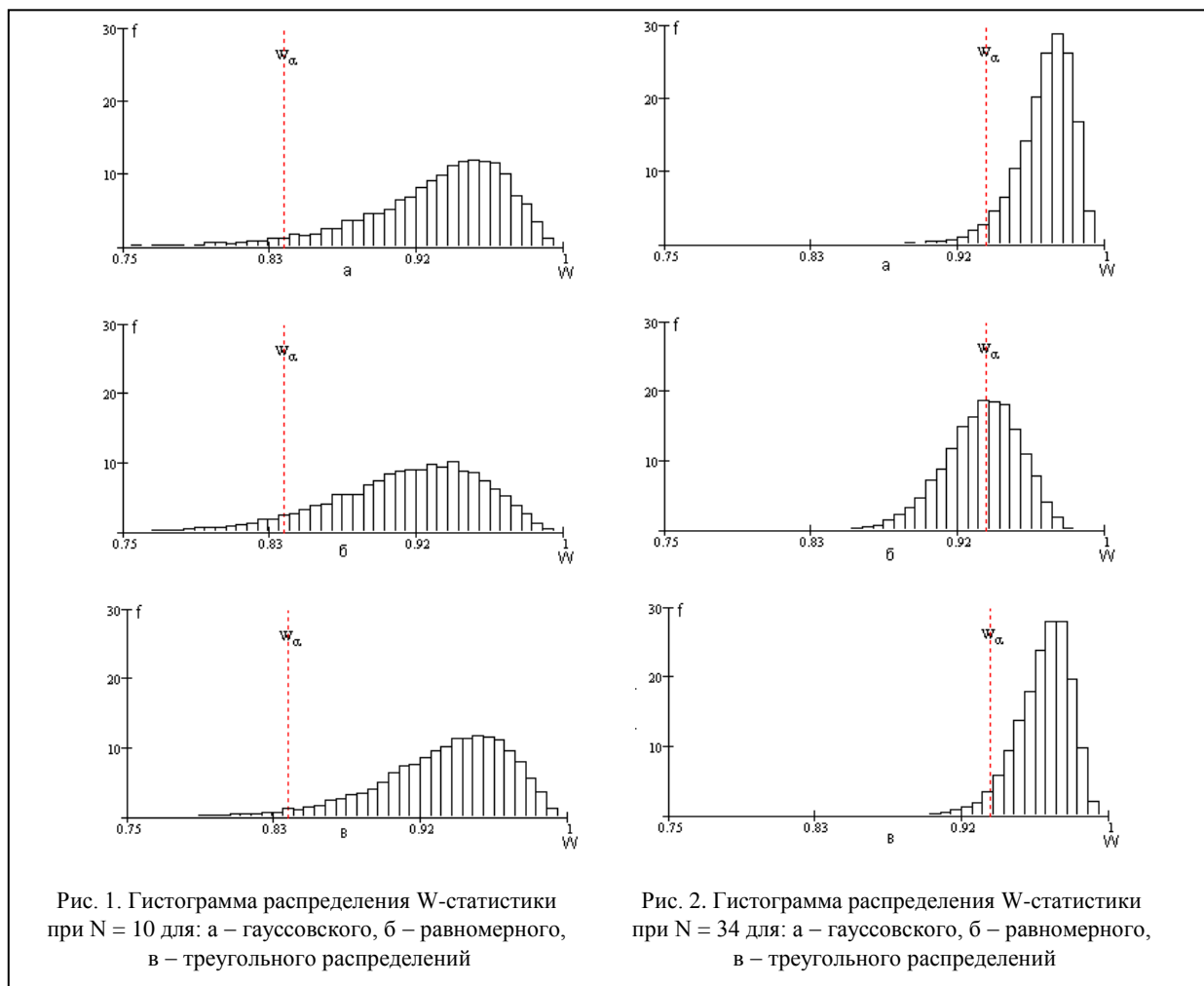


Рис. 1. Гистограмма распределения W-статистики при $N = 10$ для: а – гауссовского, б – равномерного, в – треугольного распределений

Рис. 2. Гистограмма распределения W-статистики при $N = 34$ для: а – гауссовского, б – равномерного, в – треугольного распределений

Как видно из приведенных рисунков, для выборок малого объема критические области W-статистик перекрываются, что не позволяет однозначно принимать гипотезу о «нормальности» закона распределения исследуемой выборки. Этот вывод подтверждается рассчитанными значениями вероятности ошибки второго рода, приведенными в табл. 1. Аналогичные результаты были получены для значений СКО, равных 2, 3, 4, и по этому в статье не приводятся.

Таблица 1
Значения ошибки второго рода W-критерия

| Объем выборки | Закон распределения | | |
|---------------|---------------------|-------------|-------------|
| | Гауссовский | Равномерный | Треугольный |
| 4 | 0,042 | 0,959 | 0,961 |
| 10 | 0,044 | 0,922 | 0,967 |
| 16 | 0,046 | 0,862 | 0,966 |
| 34 | 0,042 | 0,477 | 0,956 |

RS-критерий. Двусторонний RS-критерий основан на определении отношения оценки размаха (R) к оценке СКО (S) выборки и определяет критические границы интервала, принадлежность к которому RS-статистики свидетельствует о гауссовом распределении генеральной совокупности, к которой принадлежит выборка. На рис. 3 и 4 приведены гистограммы RS-статистики для центрированных выборок разного объема N. Критические границы статистики RS_H и RS_B даны для уровня значимости равного 0,05. Рассчитанные значения вероятности ошибки второго рода для данного критерия приведены в табл. 2.

Экспериментальные исследования показали, что RS-критерий является более мощным, чем W-критерий и при объеме выборки $N > 30$ позволяет с вероятностью более 0,87 разделить выборки с гауссовским и равномерным законами распределения. Однако для выборок с треугольным законом рас-

предела даже при выборке $N = 34$ значение вероятности ошибки второго рода остается значительным. Исходя из этого, можно сделать вывод о

непригодности данного критерия для проверки гипотезы о «нормальности» генеральной совокупности при малых объемах выборки.

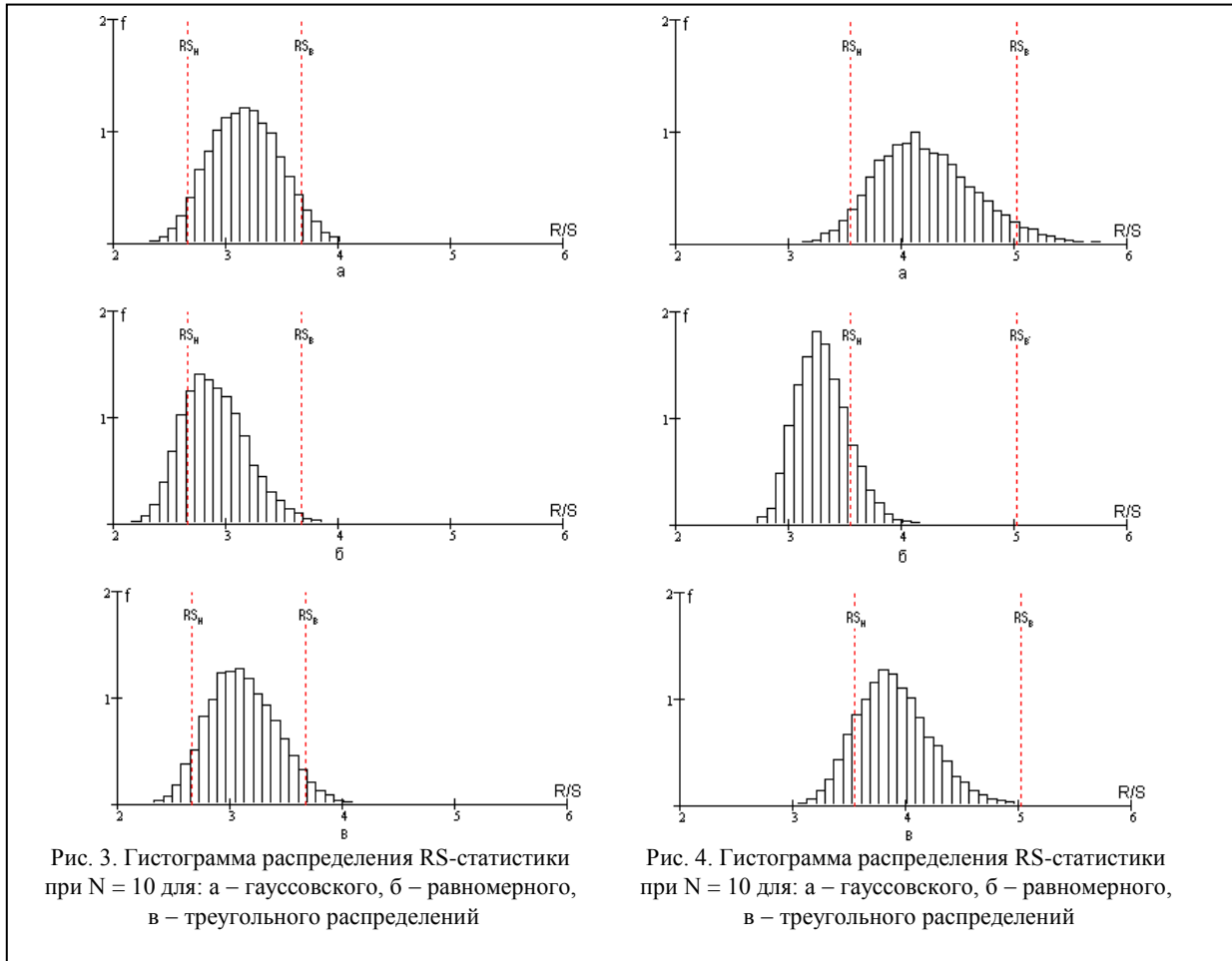


Рис. 3. Гистограмма распределения RS-статистики при $N = 10$ для: а – гауссовского, б – равномерного, в – треугольного распределений

Рис. 4. Гистограмма распределения RS-статистики при $N = 10$ для: а – гауссовского, б – равномерного, в – треугольного распределений

Таблица 2

Значения ошибки второго рода RS-критерия

| Объем выборки | Закон распределения | | |
|---------------|---------------------|-------------|-------------|
| | Гауссовский | Равномерный | Треугольный |
| 4 | 0,102 | 0,889 | 0,896 |
| 10 | 0,104 | 0,781 | 0,902 |
| 16 | 0,098 | 0,602 | 0,899 |
| 34 | 0,098 | 0,128 | 0,848 |

Как видно из приведенных рисунков, данный критерий более подходит для проверки гипотезы о равномерности выборки (при соответственном переопределении критических границ), чем для проверки гипотезы о нормальности.

Таким образом, полученные результаты свидетельствуют о том, что ни один из рассмотренных критериев проверки «нормальности» закона распределения не дает удовлетворительных результатов на выборках малого объема, что не позволяет достоверно обосновать выбор классических либо робастных критериев оценки однородности. В связи с этим было проведено дополнительное исследование достовер-

ности критериев однородности на малых выборках при различных законах распределения.

При оценке однородности выборок с гауссовским законом распределения используется составной критерий, основанный на проверке равенства оценок дисперсий выборок согласно F-критерия Фишера и последующей проверке равенства оценок математических ожиданий выборок согласно t-критерия Стьюдента.

F-критерий. F-критерий был выведен Фишером исходя из предположения о гауссовом законе распределения генеральной совокупности, поэтому принято считать, что он чувствителен к отклонению закона распределения от гауссовского [3, 7]. Однако в литературе не приведены результаты исследований, подтверждающие данное утверждение. Авторами был проведен сравнительный анализ мощности критерия для выборок с различными законами распределения (равномерным, треугольным, гауссовским). В табл. 3 приведены полученные экспериментальным путем вероятности принятия решения об однородности дисперсий двух выборок при различных их соотношениях и различных зако-

нах распределения (Г – гауссовском, Р – равномерном, Т – треугольном).

Максимальное отличие рассчитанных вероятностей для различных законов распределения не превышает 12% для выборок $N=4$ и $N=10$ и уменьшается с увеличением объема выборки. Причем на выборках при $N=34$ мощность критерия для данных с равномерным и треугольным распределением даже больше, чем для гауссовского распределения. Это позволяет сделать вывод об возможности применения F-статистики для анализа выборок с рассмотренными законами распределения.

Таблица 3

Вероятность принятия решения о равенстве дисперсий

| | | Отношение СКО выборок σ_1 / σ_2 | | | | | | |
|--------|-------|---|-------|-------|-------|-------|-------|---|
| $N=4$ | | 1 | 1,25 | 1,5 | 1,75 | 2 | 3 | 4 |
| Г | 90,08 | 88,62 | 84,58 | 79,41 | 73,65 | 50,99 | 33,40 | |
| Р | 93,33 | 92,18 | 88,85 | 83,82 | 78,68 | 51,00 | 29,48 | |
| Т | 91,57 | 90,29 | 86,90 | 82,00 | 75,15 | 51,48 | 32,33 | |
| $N=10$ | | | | | | | | |
| Г | 90,06 | 83,35 | 68,89 | 52,24 | 36,56 | 6,76 | 1,22 | |
| Р | 97,09 | 91,87 | 77,53 | 53,76 | 30,47 | 2,11 | 0,2 | |
| Т | 93,39 | 87,53 | 71,85 | 52,37 | 35,39 | 4,79 | 0,68 | |
| $N=16$ | | | | | | | | |
| Г | 89,39 | 79,08 | 54,13 | 33,22 | 16,78 | 0,81 | 0,02 | |
| Р | 97,96 | 89,08 | 57,12 | 24,80 | 7,35 | 0,09 | 0 | |
| Т | 94,20 | 83,10 | 55,83 | 29,53 | 12,99 | 0,45 | 0,01 | |
| $N=34$ | | | | | | | | |
| Г | 89,41 | 63,38 | 25,55 | 6,29 | 1,02 | 0 | 0 | |
| Р | 98,53 | 71,32 | 14,77 | 0,85 | 0,04 | 0 | 0 | |
| Т | 94,08 | 65,45 | 20,88 | 3,45 | 0,36 | 0 | 0 | |

t-критерий. По рассмотренной выше для F-критерия методике были проведены исследования t-критерия Стьюдента для выборок с различным объемом и различными законами распределения. Результаты при СКО выборок, равном 1, приведены в табл. 4. В результате моделирования установлено, что мощность критерия зависит от дисперсии исследуемых выборок, но не зависит от их закона распределения.

Как видно, данный критерий имеет удовлетворительную мощность на выборках объемом более 30 значений, которая не зависит от формы закона распределения.

Выводы. Проведенные исследования показали необходимость разработки новых или усовершенствование существующих критериев проверки на «нормальность», так как они используются не только как вспомогательные при оценке однородности, но и, в большей степени, при вычислении расширенной неопределенности результата измерений. Более достоверные данные о форме закона распределения на малых выборках можно получить на основе априорной информации о факторах, влияющих на процесс измерения.

Таблица 4

Вероятность принятия решения об однородности выборок

| | | Отношение мат. ожиданий выборок μ_1 / μ_2 | | | | | |
|--------|-------|---|-------|-------|-------|-------|---|
| $N=4$ | | 1 | 1,25 | 1,5 | 1,75 | 2 | 3 |
| Г | 92,82 | 92,11 | 89,22 | 84,12 | 77,12 | 41,56 | |
| Р | 93,06 | 92,10 | 88,52 | 84,83 | 78,78 | 44,47 | |
| Т | 93,08 | 92,10 | 89,45 | 84,47 | 78,31 | 43,94 | |
| $N=10$ | | | | | | | |
| Г | 94,98 | 91,61 | 82,37 | 66,87 | 47,82 | 1,94 | |
| Р | 94,7 | 91,73 | 82,68 | 68,94 | 48,92 | 1,4 | |
| Т | 94,55 | 91,47 | 82,38 | 67,65 | 48,94 | 2,02 | |
| $N=16$ | | | | | | | |
| Г | 95,17 | 89,78 | 74,31 | 48,83 | 24,06 | 0,05 | |
| Р | 94,57 | 89,38 | 73,74 | 50,29 | 25,14 | 0,03 | |
| Т | 94,82 | 89,58 | 73,80 | 48,84 | 24,19 | 0,04 | |
| $N=34$ | | | | | | | |
| Г | 95,41 | 83,57 | 49,53 | 15,53 | 2,38 | 0 | |
| Р | 95,00 | 83,15 | 48,99 | 15,44 | 1,88 | 0 | |
| Т | 95,26 | 83,47 | 48,93 | 14,67 | 2,07 | 0 | |

Рассмотренные критерии однородности обладают достаточной мощностью на выборках объема $N > 30$ и в малой степени зависят от формы законов распределения. Целесообразно разработать критерии однородности устойчивые как к форме закона распределения, так и к объему выборки. Значительный практический интерес представляет проведение сравнительного анализа классических критериев однородности с известными робастными, с целью определения области применения того или иного критерия.

ЛИТЕРАТУРА

1. ДСТУ ISO/IEC 17025-2001. Державний стандарт України. Загальні вимоги до компетентності випробувальних та калібрувальних лабораторій.
2. Шаффе Г. Дисперсионный анализ. – М.: Наука, 1980. – 512 с.
3. Закс Л. Статистическое оценивание. – М.: Статистика, 1976. – 598 с.
4. ГОСТ 11.006-74. Правила проверки согласия опытного распределения с теоретическим.
5. Захаров И.П., Штефан Н. В., Сафарян Г.Г., Сергиенко М.П. Идентификация формы закона распределения погрешностей по выборкам малого объема // Радиотехника. – 2002. – Вып. 129. – С. 123-126.
6. Захаров И.П., Штефан Н.В. Об эффективности применения составного критерия для оценки нормальности распределения // Тр. III Междунар. научн.-технич. конф. “Метрологічне забезпечення в галузі електричних, магнітних та радіовимірювань”. – Х.: ХГНИИМ, 2000. – Т. 2. – С. 164-166. (Метрологія в електроніці – 2000)
7. Мардиа К., Земроц П. Таблицы F-распределений. – М.: Наука, 1984. – 256 с.

Поступила 3.04.2006

Рецензент: канд. техн. наук И.П. Захаров, Харьковский национальный университет внутренних дел.