Свойства и мощность некоторых критериев нормальности

Б. Ю. Лемешко[1], В. Н. Белоцерковец
Новосибирский государственный технический университет

Рассмотрено несколько критериев, предназначенных для проверки гипотезы о принадлежности выборки нормальному закону распределения. Исследованы их недостатки и достоинства, проведено сравнение мощности критериев относительно конкурирующих гипотез.

Ключевые слова: критерий нормальности, критерий Мартинеса–Иглевича, критерий Али–Черго–Ревиса, критерий Чен–Шапиро.

1. Введение

Нормальный закон играет очень важную роль в приложениях теории вероятностей и математической статистики. Одна из особенностей, выделяющая нормальный закон среди других законов, состоит в том, что он является предельным законом, к которому приближаются другие законы распределения в типичных условиях. Встречающиеся на практике случайные величины, такие как ошибки измерений, ошибки стрельбы и другие, могут быть представлены как суммы большого числа сравнительно малых слагаемых – элементарных ошибок, которые не зависят друг от друга. Несмотря на то, что эти ошибки могут подчиняться каким-либо другим законам распределения, в сумме большого числа слагаемых особенности распределений нивелируются, и сумма в результате приближенно подчиняется нормальному закону распределения. При этом влияние каждого фактора на конечный результат должно быть существенно меньше суммарного влияния всех остальных факторов.

Проверка принадлежности наблюдаемых данных нормальному закону является важным условием для корректного применения большинства параметрических методов математической статистики, используемых в задачах обработки измерений, стандартизации и контроля качества.

Данные исследования дополняют результаты, представленные в [1-5].

2. Критерий Мартинеса-Иглевича

Критерий применяется против симметричных альтернатив, отличающихся от нормального распределения «хвостами» или эксцессом. Критерий основан на отношении двух оценок дисперсии – обычной и робастной (двухвесовой), имеющей вид [6]:

$$\tilde{s}^2 = n \sum_{|z_i|<1} \left(x_i - \tilde{x} \right)^2 \left(1 - z_i^2 \right)^4 \Big/ \left\{ \sum_{|z_i|<1} \left(1 - z_i^2 \right) \left(1 - 5 z_i^2 \right) \right\}^2 , \tag{1}$$

где

[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)

$$z_i = \begin{cases} \dfrac{x_i - \tilde{x}}{9 med|x_i - \tilde{x}|}, & при \quad |z_i| < 1; \\ 0, & при \quad |z_i| \ge 1. \end{cases} \tag{2}$$

$\tilde{x}$- выборочная медиана, $med(...)$ - медиана ряда в скобках.

В качестве статистики критерия используется отношение дисперсий:

$$I = \frac{s^2}{\tilde{s}^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)\tilde{s}^2} = \frac{1}{n(n-1)} \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \left\{ \sum_{|z_i|<1} \left(1 - z_i^2\right)\left(1 - 5z_i^2\right) \right\}^2}{\sum_{|z_i|<1} (x_i - \bar{x})^2 \left(1 - z_i^2\right)^4} . \tag{3}$$

На рис. 1 показано изменение распределения статистики (1) критерия Мартинеса-Иглевича в зависимости от объема выборки, принадлежащей нормальному закону.

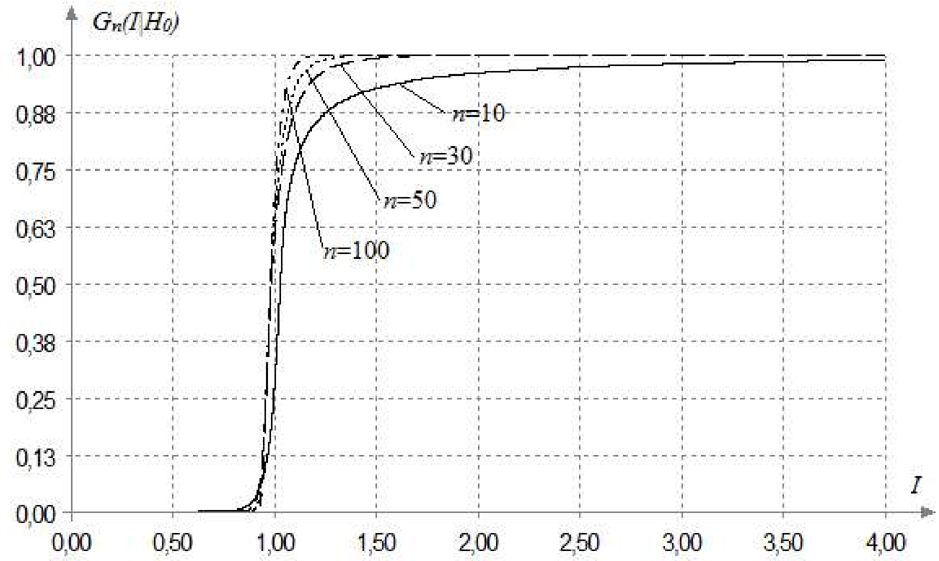

Рис. 1. Графики распределения статистики (3) критерия Мартинеса–Иглевича в зависимости от объема выборки при n =10, 30, 50, 100

Этот критерий считается более мощным относительно конкурирующих законов с длинными (тяжелыми) «хвостами». Критерий двусторонний, в табл. 1 приведены полученные критические значения статистики.

Таблица 1. Критические значения статистики критерия Мартинеса–Иглевича

n	90		95	
	0.05	0.95	0.25	0.975
10	0.926	1.781	0.899	2.644
20	0.918	1.305	0.898	1.470
30	0.924	1.204	0.91	1.3
40	0.928	1.157	0.917	1.228
50	0.931	1.128	0.922	1.181
100	0.935	1.074	0.933	1.102

3. Критерий Али–Черго–Ревиса

Али, Черго, Ревес предложили семейство статистик [7, 8] для проверки нормальности распределения случайных величин, которые основаны на взвешенных квадратах спейсингов, под которыми понимаются расстояния между соседними порядковыми статистиками.

Если $x_1,...,x_n$ – порядковые статистики наблюдаемого ряда случайных величин, то статистика критерия записывается в форме

$$M_n = \sum_{i=1}^{n} \left\{ \frac{x_n - \bar{x}}{s_n} - \Phi^{-1}\left(\frac{i}{n+1}\right) \right\}^2 \varphi\left[\Phi^{-1}\left(\frac{i}{n+1}\right)\right], \tag{4}$$

где $\Phi^{-1}(p) = u_p$ – p-квантиль стандартного нормального распределения; $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ – оценки математического ожидания и дисперсии.

Критерий правосторонний. В литературных источниках отсутствует указание на предельное распределение, но даны процентные точки. Уточненные процентные точки приведены в табл. 2.

Таблица 2. Критические значения статистики критерия Али–Черго–Ревиса

n	$1-\alpha$		
	0,90	0,95	0,99
5	0,268	0,308	0,403
10	0,333	0,390	0,524
15	0,354	0,414	0,569
20	0,358	0,422	0,567
30	0,362	0,430	0,585
40	0,361	0,427	0,581
50	0,361	0,431	0,582
60	0,362	0,426	0,591
70	0,360	0,427	0,586
80	0,360	0,428	0,595
90	0,359	0,432	0,592
100	0,357	0,423	0,587

На рис. 2 показано изменение распределения статистики (4) критерия Али–Черго–Ревиса в зависимости от объема выборки при справедливости гипотезы о нормальности.

По рисунку 2 видно, что с ростом числа n распределения статистик сходятся к некоторому предельному распределению.

4. Критерий Чена-Шапиро

Чен и Шапиро ввели критерий [9, 10], основанный на нормализованных расстояниях. Его статистика определяется следующим образом

$$QH = \frac{1}{(n-1)s}\sum_{i=1}^{n-1}\frac{x_{(i+1)} - x_{(i)}}{H_{i+1} - H_i}, \tag{5}$$

где $H_i = \Phi^{-1}\{(i-3/8)/(n+1/4)\}$.

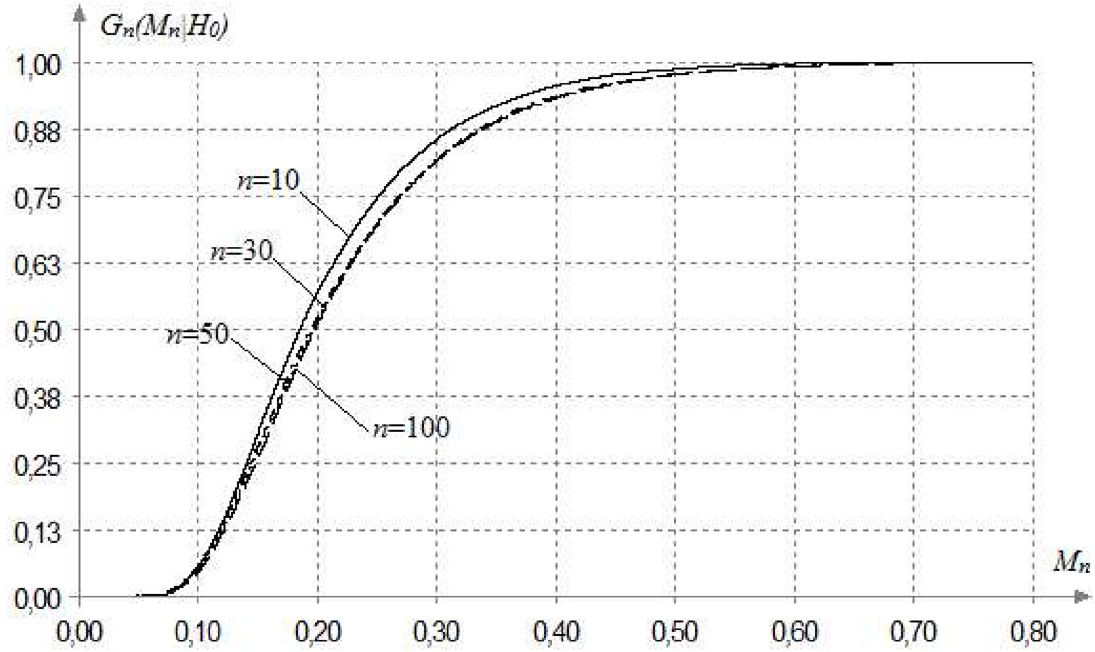

Рис. 2. Графики распределения статистики (4) критерия Али–Черго–Ревиса в зависимости от объема выборки при n=10, 30, 50, 100

Критерий левосторонний: гипотеза о нормальности отвергается при малых значениях статистики.

В табл. 3 представлены процентные точки, полученные в данной работе, для статистики

$$QH^* = \sqrt{n}(1-QH). \tag{6}$$

Таблица 3. Критические значения статистики (6) критерия Чена-Шапиро

n	α			
	0.10	0.05	0.025	0.01
10	0.019	0.066	0.114	0.174
15	0.018	0.059	0.098	0.159
20	0.013	0.050	0.088	0.136
30	0.008	0.040	0.072	0.120
40	0.005	0.032	0.062	0.099
50	0.002	0.029	0.054	0.09
60	0.0005	0.025	0.048	0.084

На рис. 3 показана зависимость распределения статистики (5) критерия Чен-Шапиро от объема выборки при справедливости проверяемой гипотезы о нормальности.

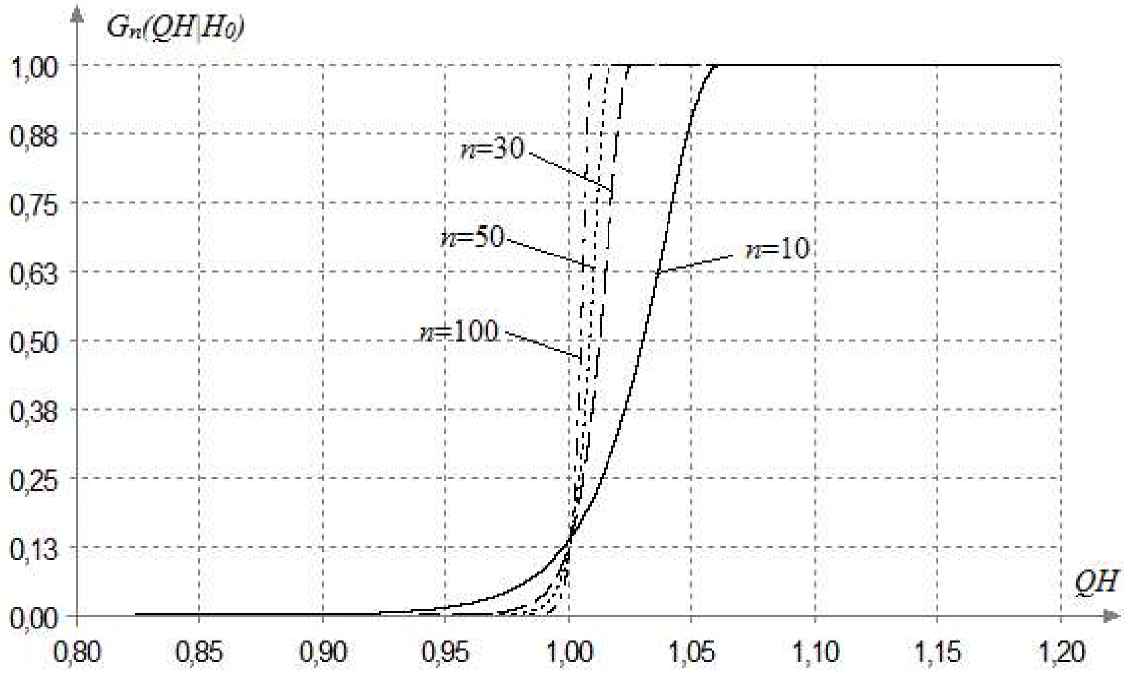

Рис. 3. Графики распределения статистики (5) критерия Чен-Шапиро в зависимости от объема выборки при n =10, 30, 50, 100

5. Сравнительный анализ мощности

При сравнительном анализе мощности критериев рассматриваются следующие гипотезы: в качестве конкурирующей гипотезы H_1 рассмотрено распределение семейства

$$f(x) = \frac{\theta_2}{2\theta_1 \Gamma(1/\theta_2)} \exp\left\{ -\frac{|x-\theta_0|^{\theta_2}}{\theta_1} \right\} \tag{7}$$

с параметром формы $\theta_2 = 1$ (распределение Лапласа), параметрами масштаба $\theta_1 = 1$ и сдвига $\theta_0 = 0$; конкурирующей гипотезе H_2 соответствует логистическое распределение с плотностью

$$f(x) = \frac{\pi}{\theta_1\sqrt{3}} \exp\left\{ -\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}} \right\} / \left[1 + \exp\left\{ -\frac{\pi(x-\theta_0)}{\theta_1\sqrt{3}} \right\} \right]^2 \tag{8}$$

с параметрами $\theta_1 = 1$ и $\theta_0 = 0$; в качестве конкурирующей гипотезы H_3 рассмотрено распределение семейства (7) с параметром формы $\theta_2 = 4$, параметром масштаба $\theta_1 = 1$ и параметром сдвига $\theta_0 = 0$.

Мощность критериев сравнивалась также с критериями Гири и Хегази–Грина T_1, T_2 [Книги]. Оценки мощности представлены в табл. 4-6. В таблицах критерии расположены по возрастанию мощности.

Среди сравниваемых критериев относительно конкурирующей гипотезы H_1 наиболее мощным оказался критерий Мартинеса–Иглевича. Однако при объемах выборок (n=40, 50) мощность критерия Гири и Хегази–Грина T_2 превышают мощность критерия Мартинеса–Иглевича. В данном случае критерий Чена–Шапиро показал наименьшую мощность среди рассмотренных критериев, что означает, что он хуже всего распознает различия между гипотезами H_0 и H_1.

Таблица 4. Мощность критериев относительно конкурирующей гипотезы H_1

Критерий	Объем выборки n	α				
		0.15	0.1	0.05	0.025	0.01
Чена–Шапиро	10	0.278	0.222	0.149	0.102	0.063
	20	0.394	0.334	0.254	0.194	0.136
	30	0.487	0.426	0.336	0.267	0.192
	40	0.562	0.45	0.417	0.339	0.261
	50	0.639	0.576	0.482	0.407	0.321
Али–Черго–Ревиса	10	0.23	0.23	0.152	0.103	0.059
	20	0.429	0.358	0.266	0.199	0.135
	30	0.537	0.466	0.364	0.282	0.195
	40	0.629	0.559	0.456	0.361	0.271
	50	0.713	0.644	0.534	0.451	0.343
Хегази–Грина T_1	10	0.307	0.242	0.161	0.106	0.060
	30	0.569	0.496	0.390	0.302	0.213
	50	0.736	0.674	0.572	0.478	0.368
Гири	10	0,267	0,204	0,130	0,082	0,043
	20	0,442	0,371	0,274	0,201	0,132
	30	0,583	0,513	0,408	0,321	0,229
	40	0,694	0,630	0,528	0,435	0,330
	50	0,776	0,722	0,628	0,535	0,425
Хегази–Грина T_2	10	0.347	0.277	0.185	0.123	0.071
	30	0.632	0.559	0.447	0.350	0.251
	50	0.782	0.723	0.622	0.525	410
Мартинеса–Иглевича	10	0.294	0.223	0.136	0.085	0.039
	20	0.461	0.389	0.290	0.217	0.141
	30	0.58	0.511	0.406	0.312	0.226
	40	0.675	0.609	0.503	0.417	0.328
	50	0.748	0.695	0.603	0.518	0.412

Таблица 5. Мощность критериев относительно конкурирующей гипотезы H_2

Критерий	Объем выборки n	α				
		0.15	0.1	0.05	0.025	0.01
Али–Черго–Ревиса	10	0.192	0.138	0.078	0.046	0.022
	20	0.233	0.172	0.105	0.065	0.034
	30	0.259	0.194	0.124	0.079	0.044
	40	0.282	0.218	0.14	0.089	0.055
	50	0.31	0.241	0.155	0.106	0.061
Чена–Шапиро	10	0.191	0.137	0.081	0.047	0.025
	20	0.234	0.179	0.114	0.076	0.045
	30	0.258	0.203	0.142	0.097	0.057
	40	0.285	0.224	0.161	0.109	0.072
	50	0.308	0.253	0.181	0.132	0.087
Хегази–Грина T_1	10	0.197	0.141	0.081	0.047	0.022
	20	0.277	0.212	0.136	0.087	0.048
	50	0.336	0.267	0.179	0.120	0.070

Продолжение таблицы 5

Критерий	Объем выборки n	α				
		0.15	0.1	0.05	0.025	0.01
Гири	10	0.182	0.127	0.069	0.038	0.017
	20	0.227	0.169	0.103	0.063	0.033
	30	0.267	0.205	0.132	0.086	0.049
	40	0.305	0.240	0.161	0.108	0.064
	50	0.338	0.272	0.188	0.129	0.079
Мартинеса–Иглевича	10	0.190	0.132	0.071	0.037	0.014
	20	0.250	0.187	0.116	0.070	0.035
	30	0.299	0.233	0.15	0.093	0.053
	40	0.345	0.273	0.181	0.12	0.073
	50	0.377	0.311	0.22	0.158	0.097
Хегази–Грина T_2	10	0.219	0.160	0.093	0.055	0.027
	30	0.346	0.277	0.189	0.128	0.077
	50	0.433	0.359	0.260	0.187	0.121

Таблица 6. Мощность критериев относительно конкурирующей гипотезы H_3

Критерий	Объем выборки n	α				
		0.15	0.1	0.05	0.025	0.01
Хегази–Грина T_2	10	0.110	0.067	0.028	0.012	0.004
	30	0.084	0.044	0.014	0.004	0.001
	50	0.113	0.061	0.020	0.006	0.001
Хегази–Грина T_1	10	0.148	0.095	0.043	0.019	0.007
	30	0.209	0.140	0.068	0.032	0.011
	50	0.302	0.218	0.118	0.060	0.023
Мартинеса–Иглевича	10	0.151	0.103	0.058	0.031	0.015
	20	0.170	0.117	0.066	0.038	0.019
	30	0.215	0.155	0.087	0.051	0.025
	40	0.269	0.192	0.109	0.063	0.031
	50	0.355	0.258	0.143	0.077	0.038
Али–Черго–Ревиса	10	0.162	0.106	0.051	0.024	0.009
	20	0.208	0.141	0.074	0.039	0.015
	30	0.264	0.184	0.099	0.053	0.02
	40	0.315	0.233	0.134	0.071	0.033
	50	0.373	0.282	0.167	0.100	0.046
Чена–Шапиро	10	0.167	0.104	0.047	0.020	0.007
	20	0.224	0.149	0.071	0.031	0.009
	30	0.29	0.205	0.103	0.047	0.013
	40	0.358	0.256	0.144	0.071	0.026
	50	0.430	0.33	0.186	0.099	0.039
Гири	10	0.174	1.120	0.064	0.034	0.014
	20	0.247	0.181	0.105	0.061	0.0.029
	30	0.329	0.253	0.158	0.097	0.050
	40	0.409	0.326	0.215	0.138	0.075
	50	0.480	0.394	0.273	0.182	0.103

Среди рассмотренных критериев относительно конкурирующей гипотезы H_2 наиболее мощным оказался критерий Хегази–Грина T_2.

В то же время, относительно H_3 Критерии Хегази–Грина имеют наименьшую мощность и соответственно хуже всего различают гипотезы.

6. Заключение

Анализ мощности критериев Мартинеса–Иглевича, Али–Черго–Ревиса и Чена–Шапиро показал, что, в принципе, если они и уступают в мощности наиболее предпочтительным критериям нормальности, то это не настолько значительно. То есть критерии могут быть рекомендованы к использованию.

В качестве общего недостатка этих критериев можно назвать сильную зависимость распределений статистик от объемов выборок и неизвестность аналитических распределений этих статистик. По этой причине на практике для принятия решения о результатах проверки гипотезы приходится рассматривать таблицы процентных точек.

В то же время, для критерия Али–Черго–Ревиса можно численными методами, опираясь на статистическое моделирование, подобрать модель для предельного распределения этой статистики, которой можно пользоваться при объёмах выборок больше 30.

Принятие решения о результатах проверки гипотезы H_0 на основании достигнутого уровня значимости (**p**-value) всегда более обосновано, чем в результате сравнения полученного значения статистики с заданным критическим значением. В последнем случае остаётся не ясным, насколько далеко на самом деле истинное распределение, которому принадлежит анализируемая выборка (и которое в действительности всегда остается неизвестным), от нормального закона.

К сожалению, распределения многих специальных критериев проверки нормальности существенно зависят от объемов выборок, в связи с чем при формировании решения о результатах проверки гипотезы H_0 (отклонять – не отклонять) опираются на таблицы процентных точек. К тому же, реальные объемы выборок, которые приходится анализировать, как правило, достаточно ограничены. А при малых n действительные распределения статистик критериев могут существенно отличаться от предельных или приближенных асимптотических распределений этих же статистик. Поэтому использование последних вместо действительных распределений статистик при малых n приводит к ошибкам при вычислении достигнутого уровня значимости.

Каким может быть решение для повышения качества статистических выводов? Для объективности анализа достигнутый уровень значимости может вычисляться на основании моделирования распределений статистик применяемых критериев при тех объемах выборок n, которые соответствуют объёмам выборок анализируемых экспериментов. При таком подходе необходимое для проверки гипотезы эмпирическое распределение статистики соответствующего критерия строится в результате статистического моделирования с точностью, зависящей от числа экспериментов N в методе Монте-Карло. Затем по эмпирическому распределению $G_N(S_n|H_0)$ статистики и значению статистики S^*, вычисленному по анализируемой выборке, можно найти оценку достигнутого уровня значимости **p**-value.

Такое статистическое моделирование может проводиться в интерактивном режиме в ходе осуществляемого статистического анализа, а затем его результаты могут использоваться при формировании вывода по итогам проверки гипотезы. Такой подход эффективно используется при использовании различных критериев проверки статистических гипотез, в том числе, при использовании критериев проверки нормальности и непараметрических критериев согласия при проверке сложных гипотез [2, 11].

Литература

1. *Лемешко, Б.Ю.* Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход: Монография / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов, Е.В. Чимитова. – Новосибирск: Изд-во НГТУ, 2011. 888 с.
2. *Лемешко, Б.Ю.* Критерии проверки отклонения распределения от нормального закона. Руководство по применению: Монография. / Лемешко Б.Ю. – М.: ИНФРА-М, 2015. 160 с.
3. *Лемешко Б.Ю., Рогожников А.П.* Исследование особенностей и мощности некоторых критериев нормальности // Метрология. 2009. № 4. С. 3-24.
4. *Лемешко Б.Ю., Лемешко С.Б.* Сравнительный анализ критериев проверки отклонения распределения от нормального закона // Метрология. 2005. № 2. С. 3-24.
5. *Лемешко Б.Ю.* Критерии согласия типа хи-квадрат при проверке нормальности // Измерительная техника. 2015. № 6. – С.3-9.
6. *Martinez, J.* A test for departure from normality based on a biweight estimator of scales / J. *Matrinez, B. Iglewitcz* // Biometrica. 1981. V. 68, № 1. P. 331-333.
7. *Кобзарь, А.И.* Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. М.: ФИЗМАТЛИТ, 2006. 816 с.
8. *Aly, E.E.* On some goodness-of-fit tests for the normal, logistic and extreme-value distributions / E.E. Aly, M.A. Shayib // Common. Stat.-Theory Meth. 1992. V.21, №5. P. 1297-1308.
9. *Brezinski, M.* The Chen-Shapiro test for normality / M. Brezinski // The Stata Journal. 2012. V.12, №3. P. 368-374.
10. *Chen, L.* An alternative test for normality based on normalized spacings / L. Chen, S.S. Shapiro // Journal of Statistical and Simulation. 2012. V.53. P. 269-288.
11. *Лемешко Б.Ю.* Непараметрические критерии согласия. Руководство по применению. М.: ИНФРА-М, 2014. 163 с.

Лемешко Борис Юрьевич

д.т.н., профессор, г.н.с. кафедры теоретической и прикладной информатики Новосибирского государственного технического университета (630073, Новосибирск, просп. К. Маркса, 20), e-mail: `lemeshko@ami.nstu.ru`.

Белоцерковец Валерия Николаевна

магистрант факультета прикладной математики и информатики Новосибирского государственного технического университета (630073, Новосибирск, просп. К. Маркса, 20), e-mail: `belocerkovecz.v@mail.ru`.

Properties and power of some tests of a rejection from the normal law

B. Yu. Lemeshko, V. N. Belocerkovecz

Novosibirsk State Technical University

Several tests are considered, they designed to test the hypothesis of sampling accessories normal distribution. Advantages and disadvantages are studied, power are estimated for the competing hypotheses.

Keywords: normality test, Martinez–Iglewitcz test, Aly–Csorgo–Revesz test, Chen–Shapiro test.