О критериях одиночных выбросов

About single outliers tests

Заляжных В.В.
Канд. техн. наук, доцент кафедры стандартизации, метрологии и сертификации Северного (Арктического) федерального университета имени М.В. Ломоносова
e-mail: zalvladimir@yandex.ru

Zaliazhnykh V.V.
Candidate of Technical Sciences, Associate Professor, Department of «Standardization, Metrology and Certification», Northern (Arctic) Federal University named after M.V. Lomonosov
e-mail: zalvladimir@yandex.ru

Аннотация
Выборки, получаемые в различных исследованиях, часто принадлежат к нормальному распределению. При этом иногда в них встречаются значения, существенно отличающиеся от основной группы значений. Рассмотрены наиболее популярные статистические критерии для проверки на выброс единственного такого значения. Показано, что в этом случае статистически обоснованы критерии Граббса, Диксона и Ирвина. Не рекомендуются критерии Шовене и Шарлье ввиду слишком низких или слишком высоких уровней значимости табличных значений. Показано также, что критерий Романовского эквивалентен одностороннему критерию Граббса, и поэтому не имеет самостоятельного значения. Процентные точки критерия Романовского, приводимые в информационных источниках, ошибочны. Рассчитаны действительные значения его процентных точек.
Ключевые слова: выброс, критерий Граббса, критерий Романовского, критерий Шовене, критерий Шарлье, критерий Диксона, критерий Ирвина.

Abstract
The samples obtained in various studies often belong to a normal distribution. At the same time, sometimes they contain values that differ significantly from the main group of values. The most popular statistical tests for checking for an outlier of a single such value is considered. It do shown that in this case the tests of Grubbs, Dixon and Irwin are statistically justified. The Chauvenet and Charlie tests do not recommended due to too low or too high levels of significance of tabular values. It do also shown that the Romanovsky's test is equivalent to the one-sides Grubbs test, and so do not independent. The percentage points of the Romanovsky's test given in information sources are erroneous. The actual values of its percentage points are calculated.
Keywords: outlier, Grubbs`s test, Romanovsky's test, Chauvenet`s test, Charlie's test, Dixon’s test, Irvin’s test.

1. Введение

Числовые данные, получаемые в различных процессах, иногда содержат значения, заметно отличающиеся по величине от основной группы значений. Их появление может иметь случайный характер, например, из-за резкого изменения условий измерений, сбоев средств измерений, ошибок при передаче данных, ошибок оператора и др. В этом случае они являются выбросами, поэтому необходимо их выявление и отбраковка. В данной статье рассмотрены статистические критерии для проверки на единственный выброс, применимые к выборкам, принадлежащим к нормальному распределению, и наиболее часто встречающиеся в научных публикациях и учебных материалах.

По количеству значений, проверяемых на выбросы, бывают критерии для проверки одного значения (на одиночный выброс) – Граббса, Романовского, Шовене, Шарлье, и для проверки до двух или нескольких значений (множественное тестирование) – Диксона, Ирвина, Титьена-Мура, Роснера. Для проверки на одиночный выброс применимы критерии обеих групп. При этом критерии Титьена-Мура и Роснера при тестировании на одиночный выброс равнозначны соответственно одностороннему и двустороннему критерию Граббса, и в этом случае самостоятельного значения не имеют.

При проверке на выбросы нулевая гипотеза состоит в том, что проверяемые значения принадлежат тому же распределению, что и вся выборка, т.е. не являются выбросами. Альтернативная гипотеза состоит в том, что проверяемые значения принадлежат другому распределению, и потому являются выбросами.

Критерии Граббса, Романовского, Диксона, Ирвина основаны на сравнении расчётного значения статистики критерия с процентной точкой (табличным значением), соответствующей некоторому уровню значимости α. В то же время критерии Шарлье и Шовене основаны на других статистических закономерностях, в них не используются уровни значимости и процентные точки. Уровень значимости – это вероятность отклонения верной нулевой гипотезы.

Обычно проверку на выбросы проводят при уровне значимости 0,01 или 0,05. В ГОСТ Р ИСО 5725-2-2002 «Точность (правильность и прецизионность) методов и результатов испытаний», часть 2, рекомендуется значения выборки, определяемые как выбросы при уровне значимости 0,05, идентифицировать как квазивыбросы, при уровне 0,01 – как статистические выбросы. Если они не могут быть объяснены техническими ошибками и заменены правильными значениями, рекомендуется квазивыбросы сохранять в наборе данных, а статистические выбросы исключать. Уровень значимости в сочетании с соответствующей процентной точкой критерия представляется наиболее объективной мерой при проверке на выбросы.

2. Критерии Граббса

Рассмотрим вариационный ряд выборки из нормального распределения: x_1, x_2, …, x_i, …, x_n. По одностороннему критерию Граббса [1-5] (встречаются также названия «критерий Н.В. Смиронова» и «критерий Смирнова-Граббса») для проверки на выброс значения выборки рассчитывают статистику

$$G_1 = \frac{|\,\overline{x} - x_c\,|}{s}, \tag{1}$$

где $\overline{x}$ – среднее арифметическое выборки, x_c – проверяемое (сомнительное) значение, т.е. x_1 или x_n, $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$ – выборочное среднеквадратическое отклонение (СКО).

Если G_1 больше табличного значения одностороннего критерия Граббса, x_c идентифицируют как выброс.

По двустороннему критерию Граббса [6] (встречается также название «критерий наибольшего абсолютного отклонения») проверяют на выброс крайнее значение выборки, максимально удалённое от $\overline{x}$. Статистика критерия:

$$G_2 = \max\left(\frac{x_n - \overline{x}}{s}, \frac{\overline{x} - x_i}{s}\right)$$

Если G_2 больше табличного значения двустороннего критерия Граббса, x_c идентифицируют как выброс.

Критерии Граббса рассмотрены во множестве публикаций, широко применяются на практике, их процентные точки получены как аналитически, так и методом Монте-Карло. Можно рассматривать эти критерии как статистически обоснованные.

3. Критерий Романовского

Критерий Романовского [7] основан на методе проверке гипотезы о равенстве средних двух выборок. Если при проверке на выброс сомнительное значение выборки x_c выделить в отдельную выборку, то по этому методу получим статистику критерия:

$$B = \frac{|\overline{x}^* - x_c|}{s^*},$$

где $\overline{x}^*$– среднее арифметическое выборки без учёта x_c, s^* − выборочное среднеквадратическое отклонение без учёта x_c. Если B больше табличного значения критерия Романовского, x_c идентифицируют как выброс.

Табличные значения критерия Романовского $B_{\text{табл}}$ в соответствии с [7] рассчитывают по выражению

$$B_{\text{табл}} = t_{n\text{-}1,\alpha}\sqrt{\frac{n+1}{n}},$$

где $t_{n\text{-}1,\alpha}$ – квантиль распределения Стьюдента при числе степеней свободы n-1 и уровне значимости α. Иногда для упрощения принимают $B_{\text{табл}} = t_{n\text{-}1,\alpha}$.

Некоторые процентные точки критерия Романовского по [7] приведены в табл. 1.

Таблица 1

Процентные точки критерия Романовского по [7]

α	n = 3	n = 10	n = 20	n = 40	n = 70	n = 100
0,01	11,460	3,408	2,932	2,742	2,668	2,640
0,05	4,968	2,273	2,145	2,048	2,009	1,994

Расчёт процентных точек с использованием метода проверки равенства средних двух выборок выглядит, однако, сомнительно, поскольку при выделении в отдельную выборку проверяемого значения, минимального или максимального, она не будет случайной и независимой.

В [8] приведена статистика одностороннего критерия Граббса для максимального значения выборки в виде

$$G_1 = \frac{n-1}{n} \times \frac{x_n - \overline{x}^*}{\sqrt{\frac{n-2}{n-1}s^{*2} + \frac{(x_n - \overline{x}^*)^2}{n}}}$$

Преобразуя это выражение, получим

$$B = \frac{n}{n-1}\sqrt{\frac{n-2}{(n-1)\left(\frac{1}{G_1^2} - \frac{n}{(n-1)^2}\right)}} \qquad (2)$$

Из (2) видно, что критерий Романовского эквивалентен одностороннему критерию Граббса и, таким образом, не имеет самостоятельного значения. Подставляя в (2) значения процентных точек для G_1, можно получить соответствующие действительные значения процентных точек критерия Романовского. Некоторые из них приведены в табл. 2.

Таблица 2

Действительные процентные точки критерия Романовского

α	$n = 3$	$n = 10$	$n = 20$	$n = 40$	$n = 70$	$n = 100$

0,01	117,093	4,746	4,024	3,855	3,854	3,884
0,05	23,391	3,537	3,280	3,277	3,346	3,407

Расчёт $B_{табл}$ методом Монте-Карло даёт те же значения (отличающиеся в пределах погрешности метода), что и в табл. 2. Приведённые в табл. 2 действительные значения процентных точек критерия Романовского значительно отличаются от имеющихся в литературных источниках (табл. 1), что указывает на ошибочность литературных данных.

4. Критерий Шовене

По критерию Шовене [9] сомнительное значение является выбросом, если

$$\frac{|\bar{x} - x_c|}{s} > K,$$

где K – квантиль функции стандартного нормального распределения $F(K)$, равной 1-1/(4n), т.е. $K = F^{-1}[1-1/(4n)]$.

Статистика критерия Шовене соответствует статистике (1) одностороннего критерия Граббса, но при неизвестных уровнях значимости. Найденные методом Монте-Карло при моделировании 1 млн выборок для каждого n, уровни значимости приведены в табл. 3.

Таблица 3

Уровни значимости для критерия Шовене

n	3	4	5	6	10	30	50	100	500	1000	5000
α	0	0	0,068	0,095	0,138	0,187	0,190	0,201	0,214	0,216	0,218

Из табл. 3 видно, что при n = 3 и n = 4 уровни значимости для критерия Шовене слишком малы, практически равны нулю. При n = 5 и более уровни значимости слишком велики. В пределе

$$\lim_{n\to\infty} \alpha = \lim_{n\to\infty} \left(1 - \frac{1}{4n}\right)^n = \exp\left(-\frac{1}{4}\right) \approx 0,221$$

Таким образом, критерий Шовене не может быть рекомендован для практического применения.

5. Критерий Шарлье

Критерий Шарлье подобен критерию Шовене. По критерию Шарлье [10] сомнительное значение является выбросом, если

$$|\bar{x} - x_c| > sK_ш,$$

где $К_ш$ − квантиль функции стандартного нормального распределения находится из соотношения $F(K_ш)$ = 1-1/(2n).

Статистика критерия Шарлье также соответствует статистике (1) одностороннего критерия Граббса при иных значениях процентных точек и неизвестных уровнях значимости. Уровни значимости, полученные методом Монте-Карло, при моделировании 1 млн выборок при каждом n, приведены в табл. 4.

Таблица 4

Уровни значимости для критерия Шарлье

n	3	4	5	6	10	30	50	100	500	1000	5000
α	0,552	0,467	0,434	0,417	0,398	0,384	0,384	0,385	0,390	0,390	0,391

Как видно из табл. 4, уровни значимости для табличных значений критерия Шарлье недопустимо велики. В пределе

$$\lim_{n\to\infty} \alpha = \lim_{n\to\infty} \left(1 - \frac{1}{2n}\right)^n = \exp\left(-\frac{1}{2}\right) \approx 0,393$$

Таким образом, критерий Шарлье также не может быть рекомендован для практического применения.

6. Критерии для множественного тестирования

По критерию Диксона [9,11] x_n является выбросом, если

$$\frac{x_n - x_{n-1}}{x_n - x_1} > r_{10}$$

Выбросом является x_1, если

$$\frac{x_2 - x_1}{x_n - x_1} > r_{10}$$

Здесь r_{10} – процентная точка.

Критерий Диксона широко применяется на практике, его процентные точки получены как аналитически, так и методом Монте-Карло. Можно рассматривать этот критерий как статистически обоснованный.

По критерию Ирвина x_n является выбросом, если

$$\frac{x_n - x_{n-1}}{s} > \lambda_{\alpha,n}$$

Выбросом является x_1, если

$$\frac{x_2 - x_1}{s} > \lambda_{\alpha,n}$$

Процентные точки $\lambda_{\alpha,n}$ для критерия Ирвина изначально были рассчитаны для случая известного генерального среднеквадратического отклонения [12], однако обычно применялись при замене его выборочным среднеквадратическим отклонением s, что приводит к некоторой неточности, особенно при небольших объёмах выборки. В [13] приведены процентные точки критерия Ирвина для случая выборочного среднеквадратического отклонения. Критерий Ирвина применим для оценки на одиночный выброс при использовании процентных точек по [13].

7. Заключение

Из рассмотренных критериев проверки на единственный выброс при нормальном распределении приемлемы критерии Граббса, Диксона, Ирвина. Критерии Шовене и Шарлье не рекомендуются для практического применения. Критерий Романовского эквивалентен одностороннему критерию Граббса, и поэтому самостоятельного значения не имеет.

Литература

1. Pearson E.S., Chandra Sekar C. The efficiency of statistical tools and a criterion for the rejection of outlying observations // Biometrika. – 1936. – Vol. 28. – P. 308–320.

2. *Смирнов Н.В.* Оценка максимального члена в ряду наблюдений // Доклады АН СССР. – 1941. – Т. 33. – № 5. – С. 346–349.

3. Grubbs F. E. Sample criteria for testing outlying observations// Annals of Mathematical Statistics. – 1950. – Vol. 21, № 1. – P. 27–58.

4. Grubbs F. E. Procedures for Detecting Outlying Observations in Samples // Technometrics. – 1969. – Vol. 11, № 1. – P.1–21.

5. Grubbs F. E., Beck G. Extension of sample sizes and percentage points for significance tests of outlying observations // Technometrics. – 1972. – Vol. 14, №. 4. – P. 847–854.

6. Quesenberry, C.P., David, H.A. Some tests for outliers // Biometrika. – 1961. – Vol. 48, P. 379–390.

7. Взаимозаменяемость и технические измерения в машиностроении : монография / Б. С. Балакшин, С.С. Волосов, А.Н. Журавлёв. – Москва: Машиностроение, 1972. – 615 с.

8. *Ширяева Л.К.* О нулевом и альтернативном распределении статистики критерия наибольшего по абсолютной величине нормированного отклонения // Изв. вузов. Математика. – 2014. – № 10. – С. 62–78.

9. *Кобзарь А.И.* Прикладная математическая статистика. – 2е изд., испр. и доп. – Москва: Физматлит, 2012. – 816 с.

10. *Схиртладзе А.Г., Радкевич Я.М.* Метрология, стандартизация и технические измерения: учебник. – Старый Оскол: ТНТ, 2010. – 420 с.

11. Dixon, W.J. Ratios involving extreme values // Annals of Mathematical Statistics. – Vol. 22, № 1. – P. 68–78.

12. Irvin J.O. On a criterion for the rejection of outlying observation // Biometrika. – 1925. – V. 17. – P. 238–250.

13. *Заляжных В.В.* Расширение области применения критерия Ирвина при обнаружении аномальных измерений // Вестник СибГУТИ. – 2020. – № 2 – . С. 95–104.