УДК 519.25

В. С. ПОПУКАЙЛО

Приднестровский государственный университет им. Т. Г. Шевченко, Тирасполь

ИССЛЕДОВАНИЕ КРИТЕРИЕВ ГРУБЫХ ОШИБОК ПРИМЕНИТЕЛЬНО К ВЫБОРКАМ МАЛОГО ОБЪЕМА

Рассмотрены различные критерии выявления грубых ошибок, используемые различными исследователями при обнаружении аномальных измерений в условиях ограниченной выборки. Проанализирована их возможность распознавать грубые ошибки, применительно к выборкам малого объема на примере информации, полученной при производстве кристаллов интегральных микросхем. Сделан вывод, что оптимальным критерием для поиска грубых ошибок в малых выборках является критерий Диксона и его модификации для различных законов распределения. Также хорошие результаты показали методы Ирвина, Львовского и Титьена-Мура. Рекомендовано для более достоверного определения грубой ошибки в выборке использовать как минимум два из указанных критериев.

Ключевые слова: *малая выборка, параметры распределения, грубые ошибки, критерии обнаружения выбросов.*

Введение

При решении задач статистического анализа проблема наличия в выборке аномальных измерений имеет чрезвычайно важное значение. Присутствие единственного аномального наблюдения может приводить к оценкам, которые совершенно не согласуются с выборочными данными [1]. Грубые ошибки возникают вследствие нарушения основных условий измерения или в результате ошибок экспериментатора. Они резко отличаются по величине от остальных значений, на чем и основаны критерии их исключения из рассмотрения. При обнаружении грубой ошибки результат следует отбросить, а само измерение, если возможно, повторить [2]. Однако необходимо быть уверенным, что найденная, существенно отличающаяся от средних значений, величина действительно является ошибкой при фиксировании случайной величины. Для проверки значимости подозрительных экспериментальных данных разработаны специальные статистические критерии. Рекомендуется проводить обязательный анализ значимости отклонения крайних значений выборки от остальных, так как если они являются выбросами, то их использование при оценке выборочных моментов и проверка различных статистических гипотез может привести к большим ошибкам [3].

1. Постановка задачи исследования

Установлено, что все выборки делятся на три диапазона: выборки малого объёма (n=3-20 элементов) [4], выборки среднего объёма (n=20-100 элементов) [5] и выборки большого объёма (свыше 100 элементов) [6]. Выборки каждого диапазона имеют свои особенности при обработке, которые позволяют уменьшить потери информации и тем самым повысить точность и достоверность рассчитываемых параметров.

Существует большое количество критериев, предназначенных для решения задачи отсева грубых ошибок из полученной экспериментальным путем информации[3,7]. Эффективность этих методов во многом зависит от объема исследуемой выборки. Анализ современной литературы показал, что существует сравнение различных критериев только для выборок среднего и большого объема[8]. Исходя из этого, актуальна задача определения наиболее эффективных критериев для малых выборок. Так, например, в современной промышленности существует большое количество отраслей производства, в которых, в силу технологических ограничений, получение достаточно большого количества информации по различным причинам невозможно. Так при производстве кристаллов интегральных микросхем для контроля электрических характеристик структур и качества проведения технологических операций используют специально изготавливаемые и размещаемые на подложке структуры: тестовые микросхемы и кристаллы. Основной принцип построения тестовых микросхем состоит в том, что она должна содержать все конструктивные элементы реальной микросхемы, а также должна быть изготовлена по тому же технологическому маршруту, кроме того, тестовая микросхема должна обеспечивать удобство контроля во время испытаний и оценку качества технологического процесса. Применение тестовых микросхем и кристаллов предоставляет возможность организовать эффективный технологический контроль производства интегральных микросхем и сократить трудоемкость при проведении испытаний на надежность больших интегральных схем.

Будем исследовать способность критериев распознать грубую ошибку при объеме выборки n=10.

Данный объем выбран, так как при производстве кристаллов интегральных микросхем из-за специфики топологии пластин на ней имеются от 5 до 10 тестовых ячеек, измерения в которых должны с некоторой вероятностью отражать поведение одноименных параметров 400-5000 рабочих кристаллов.

Анализу подлежали практически все известные на сегодняшний день методы нахождения аномального измерения:

1. Метод Ирвина.
2. Критерий Стьюдента.
3. Критерий наибольшего абсолютного отклонения.
4. Критерий максимального относительного отклонения.
5. Критерий Романовского.
6. Метод вариационного размаха.
7. Критерий 3 Сигм.
8. Критерий Райта.
9. Критерий Граббса.
10. Q-критерий(Диксона).
11. Критерий Львовского.
12. Критерий Шовине.
13. Критерий Дэвида.
14. Критерий Хоглина-Иглевича.
15. L-критерий(Критерий Титьена-Мура).
16. Критерий Смоляка-Титаренко.
17. Критерий Бродского-Бацаня-Власенко.
18. Критерий Кимбера.

Все вышеперечисленные критерии предлагаются различными исследователями для определения грубых ошибок в выборочных совокупностях сравнительно небольшого объема.

Целью данной работы является анализ существующих критериев обнаружения грубых ошибок и определение их эффективности применительно к выборкам малого объема.

2. Ход эксперимента

Изложение хода эксперимента будем вести на примере выборки, полученной при производстве кристаллов интегральных микросхем. В нашем распоряжении имеется выборка контрольных параметров объемом n=10.

x_i = 13,0; 13,2; 13,5; 13,7; 13,7 14,2; 14,3; 14,5; 14,6; 16,5.

Проведём исследование данной выборки на наличие грубых ошибок каждым из выбранных критериев.

1. Метод Ирвина.

Метод применяется для поиска грубых ошибок в выборках с известной дисперсией распределения.

Имеем: $\bar{x} = 14{,}12$; $s^2 = 0{,}99$; $s = 0{,}995$.

Вычисляем (для $x_i = 16{,}5$):

$$\tau = \frac{x_n - x_{n-1}}{s} = \frac{16{,}5 - 14{,}6}{0{,}995} = 1{,}909.$$

По таблице критических значений критерия Ирвина для n=10 и уровня значимости $\alpha = 0{,}95$ находим $\tau(0{,}95) = 1{,}46$.

Так как $\tau = 1{,}909 > \tau(0{,}95) = 1{,}46$, значение $x_{10} = 16{,}5$ можно считать грубой ошибкой.

2. Критерий Стьюдента.

Все разновидности критерия Стьюдента являются параметрическими и основаны на дополнительном предположении о нормальности выборки данных.

Используя полученные ранее значения среднего арифметического и дисперсии, вычисляем нормированное отклонение:

$$\tau = \frac{\max\left|x_i - \bar{x}\right|}{s} = \frac{16{,}5 - 14{,}12}{0{,}995} = 2{,}391.$$

Затем вычисляем критерий отбраковки:

$$\tau_{кр}(q, \nu) = \frac{t(q, \nu)\sqrt{n-1}}{\sqrt{(n-2) + [t(q, \nu)]^2}},$$

где $t(q,\nu)$ – критерий Стьюдента с q уровнем значимости и $\nu = n - 2$ числом степеней свободы.

Подозреваемое число следует оставить в выборке, если $\tau < \tau_{кр}(5\%,\nu)$; число можно оставить или выбросить по усмотрению исследователя, если $\tau(5\%,q) < \tau < \tau_{кр}(0{,}1\%,q)$; а если $\tau > \tau_{кр}(0{,}1\%,\nu)$, то число нужно обязательно исключить из выборки.

В рассматриваемом примере $\tau_{кр}(5\%,8) = 1{,}87$, а $\tau_{кр}(0{,}1\%,8) = 2{,}447$, таким образом решение о наличии или отсутствии грубого промаха в данной выборке предоставляется исследователю, однако учитывая тот факт, что рассчитанное значение очень близко к табличному с уровнем значимости q=0,1%, а при расчете остальных критериев мы принимаем уровень значимости равный 95%, то примем решение оставить максимальный элемент в выборке.

3. Критерий наибольшего абсолютного отклонения.

Для нахождения грубой ошибки вычисляем:

$$\tau = \frac{\max(x_i - \bar{x})}{s} = \frac{16{,}5 - 14{,}12}{0{,}995} = 2{,}391.$$

Из таблицы критических значений критерия наибольшего абсолютного отклонения выбираем значение для доверительной вероятности α =0,95 и объема выборки n=10. Так как табличное значение $\tau(0,95)=2,414>\tau=2,391$, то гипотеза о наличии выбросов не находит подтверждения.

4. Критерий максимального относительного отклонения.

Для обнаружения грубой ошибки вычисляем:

$$\tau = \frac{\max(x_i - \bar{x})}{s} = \frac{16,5 - 14,12}{0,995} = 2,391.$$

Из таблицы критических значений критерия максимального относительного отклонения выбираем значение для доверительной вероятности α =0,95 и объема выборки n=10. Так как табличное значение $\tau(0,95)=2,29<\tau=2,391$, то гипотеза о признании значения $x_{10} = 16,5$ грубой ошибкой подтверждается.

5. Критерий Романовского.

Гипотеза о наличии грубых погрешностей в подозрительных результатах подтверждается, если выполняется неравенство:

$$\left|x_i - \bar{x}\right| \geq t_p \cdot s,$$

где t_p – квантиль распределения Стьюдента при заданной доверительной вероятности с числом степеней свободы $k = n - k_n$ (k_n - число подозрительных результатов наблюдений).

Точечные оценки распределения среднего значения и среднеквадратического отклонения вычисляются без учета подозрительных результатов наблюдений.

Отсюда получаем:

$$\bar{x} = 13,856;\ s = 0,328.$$

Проверим выполнение неравенства при t_p (0,95;9)=2,28:

|16,5-13,856|≥2,28·0,328, то есть 2,644>1,305.

Так как неравенство выполнилось, гипотеза о наличии грубой ошибки считается подтвержденной.

6. Метод вариационного размаха.

Является одним из самых простых методов исключения грубой погрешности измерений. Для его использования определяют размах вариационного ряда упорядоченной совокупности наблюдений: $R = 16,5 - 13,0 = 3,5$.

Затем производят проверку подозрительного значения из ряда, используя неравенство:

$$\bar{x} - z \cdot R < x_k < \bar{x} + z \cdot R,$$

где $\bar{x}$ – выборочное среднее арифметическое значение, вычисленное после исключения предполагаемого промаха; z – критериальное значение.

Нулевую гипотезу (об отсутствии грубого промаха) принимают, если указанное неравенство выполняется.

Проверим выполнение неравенства, используя полученные ранее значения среднего и размаха и табличное критериальное значение z=1,3:

13,856-1,3·3,5<16,5<3,856+1,3·3,5,

таким образом 9,306<16,5<18,406.

Неравенство выполняется, следовательно, проверяемое значение не является грубым промахом.

7. Критерий 3 сигм.

Сущность правила трех сигм: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднеквадратического отклонения:

$\left|x_k - \bar{x}\right| \geq 3 \cdot s$, где x_k - проверяемое значение.

Проверим выполнение неравенства:

|16,5-14,12|≥3·0,995, то есть 2,38≥2,98.

Так как неравенство не выполняется, то гипотеза о наличии грубой погрешности не находит подтверждения.

8. Критерий Райта.

Критерий Райта аналогичен предыдущему примеру и основан на том, что если остаточная погрешность больше четырех среднеквадратических отклонений, то результат является грубой погрешностью и должен быть исключен при дальнейшей обработке.

Очевидно, что результат проверки неравенства по критерию Райта также не подтвердит гипотезу о наличии погрешности в данной выборке.

9. Критерий Граббса.

Статистики критерия Граббса предусматривают возможность проверки на наличие в выборке максимальных, либо минимальных аномальных наблюдений.

Используя найденные ранее величины, вычисляем для максимального значения:

$$\tau = \frac{x_n - \bar{x}}{s} = \frac{16,5 - 14,6}{0,995} = 2,391.$$

По таблице критических значений статистик Граббса для n=10 и уровня значимости $\alpha = 0,95$ находим $\tau(0,95) = 2,441$.

Так как $\tau = 2,391 < \tau(0,95) = 2,441$, гипотеза о принятии $x_{10} = 16,5$ грубой ошибкой отклоняется.

10. Q-критерий (Диксона).

Используется для быстрого выявления выпадающих наблюдений в выборках небольшого объема по отношению размаха и подразмахов.

Вычисляем:

$$Q = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{16,5 - 14,6}{16,5 - 13,0} = 0,543.$$

Из таблицы критических значений статистик Диксона находим Q(0,95)=0,48.

Так как рассчитываемое значение больше табличного, то значение $x_{10} = 16{,}5$ считается грубой ошибкой.

11. Критерий Львовского.

Предлагается [9] для нахождения аномальных наблюдений при малом числе измерений и с этой целью вводит в метод максимального относительного отклонения уточняющий коэффициент.

Вычисляем значение критерия по формуле:

$$\tau = \frac{x_n - \bar{x}}{\sqrt{\frac{n-1}{n}}s} = \frac{16{,}5-14{,}6}{0{,}949 \cdot 0{,}995} = 2{,}52.$$

Сравниваем его с табличным значением $\tau(0{,}95)$=2,29 и определяем, что значение $x_{10} = 16{,}5$ является грубой ошибкой.

12. Критерий Шовине.

Согласно критерию Шовине элемент выборки x_i объема n является выбросом, если вероятность его отклонения от среднего значения не более $1/(12 \cdot n)$.

Вычисляем значение критерия:

$$K = \frac{x_n - \bar{x}}{s} = 2{,}391.$$

Сравниваем его с табличным значением для объема выборки n=10.

$$K=2{,}391>K_{табл}=1{,}96$$

По результатам сравнения приходим к выводу, что крайнее значение в выборке является выбросом.

13. Критерий Дэвида.

Критерий является модификацией критерия Граббса.

Высчитываем:

$$\tau = \frac{x_n - \bar{x}}{s} = 2{,}391$$

и сравниваем с табличным значением:

$\tau = 2{,}391 < \tau(0{,}95) = 2{,}89$, из чего следует, что гипотеза о наличии выброса отклоняется.

14. Критерий Хоглина-Иглевича.

Критерий основывается на порядковых статистиках и признает наблюдение выбросом, если его значение находится вне интервала, ограниченного величинами:

$$(1+k)x_{[l]} - kx_{[n+1-l]} \text{ и } (1+k)x_{[n+1-l]} - kx_{[l]},$$

где x_i - i-ая порядковая статистика (т.е. i-ый по величине член выборки, упорядоченной по возрастанию).

Для выбора значения l используем формулу:

$$l = \frac{1}{2}\left[\frac{n+3}{2}\right], \text{ где } [...] \text{ целая часть числа.}$$

Для n=10 и $\alpha = 0{,}95$, найдем табличное значение k=2,4.

Таким образом получаем:

$$(1+k)x_{[l]} - kx_{[n+1-l]} = x_3 - 2{,}4 \cdot (x_8 - x_3) = 11{,}1;$$

$$(1+k)x_{[n+1-l]} - kx_{[l]} = x_8 + 2{,}4 \cdot (x_8 - x_3) = 16{,}9.$$

Все значения данной выборки попадают в рассчитанный диапазон, что позволяет с вероятностью 0,95 утверждать отсутствие выбросов в предложенной выборке.

15. L-критерий (Критерий Титьена-Мура)

Критерий Титьена-Мура является обобщением критерия Граббса на случай выявления нескольких выбросов в выборке.

Для выделения k наибольших выбросов используется статистика:

$$L_k = \frac{\sum_{i=1}^{n-k}(x_i - \overline{x_k})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}, \text{ где } \overline{x_k} = \frac{1}{n-k}\sum_{i=1}^{n-k} x_i .$$

Рассчитав по формуле значение L_k получим:

$$L_k = \frac{13{,}856}{14{,}12} = 0{,}331 .$$

Сравним полученный результат с табличным значением критерия:

$$L_k = 0{,}331 < L_k(0{,}95) = 0{,}418.$$

Так как рассчитываемое значение меньше табличного, следует признать наличие выброса значимым с достоверностью 0,95.

16. Критерий Смоляка-Титаренко.

Данный критерий предложен для обнаружения грубых ошибок в ограниченных выборках, извлекаемых из генеральных совокупностей, распределенных по экспоненциальному закону.

Критерий основан на отношении величины наибольшего члена к выборочному среднему.

Вычисляем (для x_i=16,5):

$$C_1 = \frac{x_n}{\bar{x}} = \frac{16{,}5}{14{,}12} = 1{,}169.$$

При $\alpha = 0{,}95$ и n=10 находим табличное значение $C_1(0{,}95)$=4,449.

Так как C_1 не превышает соответствующего критического значения, то наибольшее значение не может быть признано выбросом.

17. Критерий Бродского-Быцаня Власенко.

Критерий является аналогом критерия Диксона для случая экспоненциального распределения. Рассчитаем статистику критерия проверки на выброс для наибольшего значения выборки:

$$z = \frac{x_n - x_{n-1}}{x_n - x_1} = \frac{16{,}5-14{,}6}{16{,}5-13{,}0} = 0{,}543.$$

Вычисляем:

$$P = (n-1)!(1-z)^{n-2}\prod_{j=1}^{n-2}\frac{1}{1+j(1-z)} = 1{,}538.$$

Так как рассчитанная величина P=1,538 значительно больше уровня значимости $\alpha = 0,95$, то можно утверждать, что наибольшее значение не является выбросом.

18. Критерий Кимбера.

Кимбером была предложена последовательная процедура для выявления нескольких выбросов в выборке из экспоненциального распределения.

Для проверки гипотезы вычисляем значение критерия для последних двух элементов выборки:

$$S_1 = \frac{x_{10}}{\sum_{i=1}^{10} x_i} = \frac{16,5}{141,2} = 0,117;$$

$$S_2 = \frac{x_9}{\sum_{i=1}^{9} x_i} = \frac{14,6}{124,7} = 0,117.$$

Используя табличные значения для $\alpha = 0,95$, n=10, убеждаемся, что $S_1=0,117<S_1(0,95)=0,48$ и $S_2=0,117<S_1(0,95)=0,43$.

Следовательно, на основании полученных данных, можно утверждать, что верхних выбросов в выборке не наблюдается.

Таблица 1

Результаты выявления грубой ошибки

Наименование критерия	Выявление ошибки
Метод Ирвина	Выявлена
Критерий Стьюдента	На усмотрение
Критерий наибольшего абсолютного отклонения	Не выявлена
Критерий максимального относительного отклонения	Выявлена
Критерий Романовского	Выявлена
Метод вариационного размаха	Не выявлена
Критерий 3 Сигм	Не выявлена
Критерий Райта	Не выявлена
Критерий Граббса	Не выявлена
Q-критерий(Диксона)	Выявлена
Критерий Львовского	Выявлена
Критерий Шовине	Выявлена
Критерий Дэвида	Не выявлена
Критерий Хоглина-Иглевича	Не выявлена
L-критерий (Критерий Титьена-Мура)	Выявлена
Критерий Смоляка-Титаренко	Не выявлена
Критерий Бродского-Быцаня-Власенко	Не выявлена
Критерий Кимбера	Не выявлена

Таким образом, рассчитав для одной и той же малой выборки 18 методов нахождения грубой ошибки, применяемых различными исследователями при решении задачи поиска аномального наблюдения при небольшом числе испытаний, получили результаты, приведённые в таблице 1.

3. Выводы и предложения

Подобным образом было рассмотрено несколько десятков выборок, что позволило сделать следующие выводы:

1. Такие критерии как: наибольшего абсолютного отклонения, вариационного размаха, 3 сигм, Райта, Граббса, Дэвида, Хоглина-Иглевича, Смоляка-Титаренко, Бродского-Быцаня-Власенко и Кимбера не могут достоверно определить наличие грубой ошибки в выборке малого объема.
2. Критерии Стьюдента и максимального относительного отклонения довольно точно определяют вероятность грубой ошибки в выборке малого объема, но оставляют решение об исключении из выборки на усмотрение исследователя. При применении этих методов рекомендуется проверить выборку дополнительным критерием.
3. Критерии Романовского и Шовине хорошо распознают наличие грубой ошибки в выборке малого объема, однако, в некоторых случаях выдают за выброс неаномальный результат.
4. Лучше всего в определении грубых ошибок в малых выборках показали себя критерии Ирвина, Львовского, Диксона и Титьена-Мура. Учитывая различные подходы для определения грубой ошибки, лежащие в основе этих методов, для получения достоверного результата рекомендуется проверять выборку как минимум двумя из данных критериев.

В работе были рассмотрены критерии, предназначенные для выборок с нормальным и экспоненциальным законами распределения. Следует отметить, что критерии Граббса и Диксона могут быть трансформированы для обнаружения выбросов в выборках, имеющих распределение Вейбулла [3].

Таким образом, оптимальным критерием для поиска грубых промахов в выборках малого объема следует считать Критерий Диксона, который в случае известного закона распределения также может быть трансформирован, как для совокупностей, распределенных экспоненциально, так и по закону Вейбулла.

Так как закон распределения малой выборки не может быть определен с достаточной достоверностью, следует продолжить работу по уточнению методов определения закона распределения для малых выборок, а также с критериями обнаружения грубых ошибок для других законов распределения, чтобы определить какой из них обладает наилучшей результативностью в заданных условиях.

Заключение

Рассмотрена задача поиска грубых ошибок в выборках малого объема.

Проработаны 18 критериев обнаружения аномальных значений в выборочной совокупности малого объёма (n=10).

Выбраны критерии наиболее достоверно обнаруживающие грубую ошибку в выборках малого объема.

Ввиду того, что из-за ограниченного количества элементов невозможно точно определить вид закона распределения для малой выборки предлагается рассмотреть также критерии, предназначенные для других законов распределения [7], а также нечувствительные к закону распределения критерии [10] и проанализировать их возможности при обнаружении грубых промахов в малых выборках.

Литература

1. Лемешко, Б. Ю. Робастные методы оценивания и отбраковка аномальных измерений [Текст] / Б. Ю. Лемешко // Заводская лаборатория. – 1997. – Т. 63, №5. – С. 43-49.

2. Кокунин, В. А. Статистическая обработка данных при малом числе опытов [Текст] / В. А. Кокунин // Укр. биохим. журнал. – 1975. – № 47(6). – С. 776-790.

3. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников [Текст] / А. И. Кобзарь. – М. : ФИЗМАТЛИТ, 2012. – 816 с.

4. Столяренко, Ю. А. Контроль кристаллов интегральных схем на основе статистического моделирования методом точечных распределений [Текст] : дис. ... канд. техн. наук : 05.27.01 / Столяренко Юлия Александровна. – М. : ГУП НПЦ «СПУРТ», 2006. – 191 с.

5. Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans [Text] / B. Efron. – Philadelphia, Pa.: SIAM, 1982. – 135 p.

6. Митропольский, А. К. Техника статистических вычислений [Текст] / А. К. Митропольский. – М. : Наука, 1971. – 576 с.

7. Barnett, V. Outliers in statistical data [Text] / V. Barnet, T. Lewis. – Wiley, 1994. – 584 p.

8. Seo, S. A review and comparison of methods for detecting outliers in univariate data sets [Electronic resource] / S. Seo // University of Pittsburgh. – 2006. – Access: http://d-scholarship.pitt.edu/ 7948/1/Seo.pdf. – 20.05.2015.

9. Львовский, Е. Н. Статистические методы построения эмпирических формул [Текст] : учеб. пособие для вузов / Е. Н. Львовский. – М. : Высшая школа, 1988. – 239 с.

10. Candelon, B. A distribution-free test for outliers [Text] / B. Candelon, N. Metiu // Discussion Paper. Deutsche Bundesbank. – 2013. – № 2. – 10 p.

Поступила в редакцию 21.07.2015, рассмотрена на редколлегии 11.09.2015

ДОСЛІДЖЕННЯ КРИТЕРІЇВ ГРУБИХ ПОМИЛОК СТОСОВНО ДО ВИБІРКИ МАЛОГО ОБСЯГУ

В. С. Попукайло

Розглянуто різні критерії виявлення грубих помилок, що використовуються різними дослідниками при виявленні аномальних вимірювань в умовах обмеженої вибірки. Проаналізовано їх можливість розпізнавати грубі помилки, стосовно до вибірок малого обсягу на прикладі інформації, отриманої при виробництві кристалів інтегральних мікросхем. Зроблено висновок, що оптимальним критерієм для пошуку грубих помилок в малих вибірках є критерій Діксона і його модифікації для різних законів розподілу. Також хороші результати показали методи Ірвіна, Львівського і Тітьена-Мура. Рекомендовано для більш достовірного визначення грубої помилки у вибірці використовувати як мінімум два із зазначених критеріїв.

Key words: мала вибірка, параметри розподілу, грубі помилки, критерії виявлення викидів.

THE OUTLIER CRITERIA RESEARCH IN RELATION TO SMALL VOLUME SAMPLES

V. S. Popykaylo

This article describes almost all currently known identifying outliers criteria used for abnormal measurements detection in a limited volume sample by various researchers. The criteria ability to recognize outliers was analyzed in relation to small volume samples on an information example obtained in the production of crystals of integrated chips. The conclusion is that Dickson criterion and its modifications for various distribution laws is optimum criterion for outliers search in small volume samples. Also, the methods of Irvin, Lvovsky and Tietjen-Moore showed good results in outliers determination. For more reliable outlier determination in the sample it is recommended to use at least two of these criteria.

Key words: small volume samples, outlier criterion, distribution parameters.

Попукайло Владимир Сергеевич – аспирант каф. ИТиАУПП, Приднестровский государственный университет им. Т. Г. Шевченко, Тирасполь, Молдова.