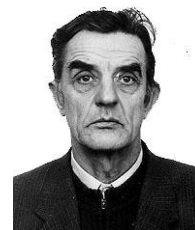


УДК 519.22/.25

М.В. Боярский

Боярский Михаил Владимирович родился в 1937 г., окончил в 1960 г. Ленинградскую лесотехническую академию, кандидат технических наук, доцент Марийского государственного университета. Имеет более 50 печатных работ в области исследования процессов резания древесины, ее физико-механических свойств и статистической обработки результатов исследований.



О ВЫЯВЛЕНИИ И ИСКЛЮЧЕНИИ АНОМАЛЬНЫХ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ В ПРОЦЕССАХ ДЕРЕВООБРАБОТКИ

Проанализированы применяемые критерии грубых ошибок, основанные на оценке нормированного отклонения подозрительных результатов от среднего арифметического. Показана их непригодность при малом числе измерений. Предложена новая методика оценки и рассчитаны критические значения нормированного отклонения при числе измерений от 3 до 10. Отмечено существенное влияние асимметрии выборки на критическое значение критерия.

Ключевые слова: статистика, промах, выявление, исключение.

Статистически достоверная обработка результатов научных исследований – трудоемкий процесс, предусматривающий наличие большого информационного массива данных, формирование которого представляется достаточно сложным.

При определении характеристик древесины, обработке результатов наблюдений и проведении исследований аномальные результаты оказывают наиболее существенное влияние на статистические характеристики контрольных выборок, особенно при малом числе испытаний, поэтому правильная оценка подозрительных результатов, их учет или исключение, является первоочередной задачей при обработке статистических рядов.

Для оценки подозрительных результатов в настоящее время имеется целый ряд критериев и рекомендаций, выбор которых позволяет формально обосновать диаметрально противоположное истолкование этих результатов, что существенно облегчает их субъективную (и столь же существенно затрудняет их объективную) оценку.

Насколько важна объективная оценка аномальности подозрительных результатов (при нынешнем стремлении к минимальному числу дублированных измерений) можно видеть из следующего примера.

Пусть при трех измерениях одного и того же объекта получены следующие значения: $y_1 = 1,0$; $y_2 = 4,5$ и $y_3 = 2,0$. С учетом всех трех измерений имеем $y_1 = 2,5$ и $S_1 = 1,803$, без учета y_2 получаем $y_2 = 1,5$ и $S_2 = 0,707$ (y_1 , y_2 и S_1 , S_2 – средние арифметические и средние квадратические отклонения для полной и усеченной (т.е. с учетом и без учета подозрительного результата) выборок).

Получается огромная разница: по среднему арифметическому – в 1,70 раза, по среднему квадратическому отклонению (и доверительному интервалу) – в 2,55 раза.

Рассмотрим статистические оценки аномальности подозрительного результата $y_2 = 4,5$ по разным критериям и рекомендациям. Наиболее многочисленными и научными являются критерии, основанные на оценке нормированного отклонения подозрительного результата от среднего арифметического:

$$t_p = |y_n - \bar{y}_1| / S < t_T, \quad (1)$$

где t_p – расчетное значение нормированного отклонения;

y_n – подозрительный результат;

t_T – табличное значение критерия аномальных результатов.

К этим критериям относятся [2]: критерий Райта – t_{TR} ; Стьюдента – t_{TS} ; Шарлье – $t_{Шр}$; Шовене – $t_{Шв}$; Греббса–Смирнова – $t_{ГС}$.

Табличные значения перечисленных критериев аномальных отклонений приведены в табл. 1.

Рассмотрим, как соотносятся расчетные значения нормированных отклонений подозрительного результата y_2 в нашем примере с табличными значениями различных критериев.

В примере для y_2 имеем

$$t_{p1} = |y_2 - \bar{y}_1| / S_1 = |4,5 - 2,5| / 1,803 = 1,109; \quad (2)$$

$$t_{p2} = |y_2 - \bar{y}_2| / S_2 = |4,5 - 1,5| / 0,707 = 4,24. \quad (3)$$

По критерию Райта $t_{p2} = 4,24 > 4,00$, т.е. y_2 следует признать аномальным результатом. По остальным критериям $t_{p1} = 1,109 < t_{TR} = 1,389 < t_{TS} = 1,406$, т.е. y_2 не является аномальным результатом.

Таким образом, выбирая вид критерия, можно получить желаемую оценку подозрительного результата. Какова объективная оценка?

Таблица 1

Табличные значения критериев аномальных результатов

Число измерений n	t_{TS} при $q, \%$			$t_{Шр}$	$t_{Шв}$	$t_{ГС}$ при $q, \%$		
	10	5	1			10	5	1
3	1,640	2,350	4,540	–	1,389	1,406	1,412	1,414
4	1,530	2,130	3,750	1,150	1,530	1,645	1,689	1,710
5	1,480	2,015	3,360	1,280	1,645	1,791	1,869	1,917
6	1,440	1,943	3,140	1,389	1,730	1,894	1,996	2,067
7	1,415	1,895	3,000	1,465	1,800	1,974	2,093	2,182
8	1,397	1,860	2,900	1,530	1,860	2,041	2,172	2,273
9	1,383	1,833	2,820	1,592	1,910	2,097	2,238	2,349
10	1,372	1,812	2,760	1,645	1,960	2,146	2,294	2,540

Примечание. Табличный критерий Райта $t_{TR} = 4$.

Таблица 2

Зависимость расчетной величины нормированного отклонения t_p подозрительного измерения от среднего арифметического при $n = 3$

Показатели	Значения показателей для разных y_2								
	3	4	5	6	8	10	15	30	∞
\bar{y}_l	2,0	2,333	2,667	3,000	3,667	4,333	6,000	11,000	–
S_l	1,0	15,280	2,082	2,646	3,786	4,933	7,810	16,462	–
$y_2 - y_l$	1,0	1,667	2,333	3,000	4,333	5,666	9,000	19,000	–
t_{p1}	1,0	1,091	1,121	1,134	1,145	1,149	1,152	1,154	1,1547

Для этого рассмотрим зависимость t_{p1} от величины отклонения y_2 от y_1 при ранее принятых значениях $y_1 = 1$ и $y_3 = 2$ (табл. 2).

Из табл. 2 видно, что расчетные значения нормированного отклонения по статистикам полной выборки при $n = 3$ не могут быть меньше 1,0 и больше 1,1547.

Сравнивая эти значения t_{p1} со значениями критериев в табл. 1, можно сделать вывод, что ни один из критериев не пригоден для выявления аномальных отклонений при $n = 3$, так как все они больше максимально возможного значения t_{p1} .

Для приближенной оценки аномальных отклонений можно предложить расчетное отношение отклонения подозрительного результата от границы основной группы измерений к размаху основной группы по критерию Стьюдента:

$$t_s = (y_2 - y_3) / (y_3 - y_1). \quad (4)$$

По величине t_s для основных процентных точек определим абсолютную величину y_2 и основные статистики:

$$y_2 = y_3 + t_s(y_3 - y_1); \quad (5)$$

$$\bar{y} = (y_1 + y_2 + y_3) / 3; \quad (6)$$

$$S = \sqrt{\sum_{i=1}^3 (y_i - \bar{y})^2 / (n-1)}; \quad (7)$$

$$t_p = (y_2 - y_1) / S. \quad (8)$$

Результаты расчетов приведены в табл. 3.

Таблица 3

Значения нормированной величины t_p аномальных отклонений от среднего арифметического T_c при $n = 3$

Показатели	Значения показателей для разных уровней значимости q , %							
	10,00	5,00	2,50	1,00	0,50	0,25	0,10	0,05
T_c	1,638	2,353	3,182	4,541	5,841	7,453	10,214	12,924
t_p	1,070	1,105	1,124	1,138	1,144	1,148	1,151	1,152

Таблица 4

**Значения нормированной величины t_p аномальных отклонений
от среднего арифметического для основных уровней значимости q**

q , %	Значение t_p при разных n							
	3	4	5	6	7	8	9	10
10	1,070	1,392	1,661	1,896	2,106	2,300	2,477	2,644
5	1,105	1,433	1,706	1,946	1,161	2,357	2,539	2,709
1	1,138	1,473	1,753	1,959	2,216	2,416	2,601	2,775
0	1,1547	1,500	1,789	2,040	2,270	2,475	2,667	2,848

Аналогично можно получить табличные значения t_p при других значениях n для основных уровней значимости $q = 10, 5$ и 1% (табл. 4).

Сравнивая данные табл. 4 с табл. 1, получаем, что критерий Шарлье при $n \geq 4$ допускает отсеивание случайных отклонений в качестве аномальных, причем с большой разницей в величине нормированных отклонений: при $n = 4$ имеем $t_{тШр} = 1,150$ вместо 1,392, а при $n = 10$ имеем $t_{тШр} = 1,645$ вместо 2,644 при $q = 10\%$

По критерию Шовене при $n = 3$ и 4 , наоборот, аномальные отклонения выявляться не будут, так как табличное значение критерия больше максимально возможной величины t_p .

По критерию Греббса–Смирнова такая ситуация наблюдается до $n \leq 5$, и только при $n > 5$ возможно выявление аномальных отклонений. Зато при $n > 6$ по критериям Шовене и Греббса–Смирнова возможно повышенное выявление случайных отклонений в качестве аномальных (по сравнению с табл. 4).

Для выявления аномальных результатов по данным табл. 4 можно использовать рекомендации из работы [1]:

1) если $t_p < t_t$ при $q \geq 10\%$, то подозрительный результат следует рассматривать как случайное отклонение, т.е. не считать его аномальным;

2) если $t_p < t_t$ при $10\% > q > 5\%$, то подозрительный результат следует считать скорее аномальным отклонением, чем случайным;

3) если $t_p > t_t$ при $q \leq 5\%$, то подозрительный результат, как правило, является аномальным отклонением.

В дополнение к изложенному, можно отметить, что табличные значения критерия аномальных отклонений имеют однозначную величину только при $n = 3$, а при $n \geq 4$ они зависят не только от q , но и от распределения результатов внутри основной группы.

Так, для $n = 4$ и $q = 10\%$ получены следующие величины t_t в зависимости от распределения измерений внутри основной группы (табл. 5).

Из табл. 5 видно, что отклонения от среднего значения $t_p = 1,367$ могут достигать $\pm 7\%$ (при $n = 4$).

Аналогично получены значения для $n = 5$ (от 1,661 до 1,900 при $q = 10\%$) и для $n = 6$ (от 1,868 до 2,181 при $q = 10\%$).

Таблица 5

Значения нормированной величины t_p аномальных отклонений от среднего значения и показателя асимметрии A для $q = 10\%$ и $n = 4$ в зависимости от распределения результатов внутри основной группы при $y_4 = 3,533$

Число измерений для y			t_p	A
$y_1=1,0$	$y_2=1,5$	$y_3=2,0$		
2	0	1	1,265	0,2134
1	1	1	1,392	0,4543
1	0	2	1,445	0,5359

При этом выявлена довольно четкая зависимость t_p от показателя асимметрии всей выборки (с учетом подозрительного результата).

Например, для $n = 6$ и $q = 10\%$ получено уравнение

$$t_p = 1,11 + 0,628A \quad (9)$$

со средней ошибкой аппроксимации менее 0,4 %.

В отношении отсеивания аномальных результатов можно дать следующие рекомендации:

1) аномальные результаты многократных измерений одного и того же объекта надо отсеивать как промахи в измерениях;

2) аномальные результаты измерений нескольких однородных объектов следует тщательно изучать в качестве основных носителей информации об источниках брака или, наоборот, об условиях получения аномальных положительных результатов.

Таким образом, объективное выявление и отсеивание аномальных результатов представляется весьма ответственной и достаточно сложной операцией, требующей от исследователя осмысленного (а не формального) применения имеющихся критериев.

СПИСОК ЛИТЕРАТУРЫ

1. Брюханов В.А., Маркова Е.В. Проблемы внедрения концепции неопределенности результатов измерений в практику обработки и представления измерительной информации // Материалы 5-ой Всероссийской научно-техн. конф. «Состояние и проблемы технических измерений». – М.: МГТУ, 1998 – С. 6–7.

2. Рейх Н.Н., Тупиченков А.А., Цейтлин В.Г. Метрологическое обеспечение производства. – М.: Изд-во стандартов, 1987. – 248 с.

Марийский государственный технический университет

Поступила 18.06.02

M.V. Boyarsky

Detection and Elimination of Abnormal Observation Results in Woodworking Processes

The criteria of gross mistakes based on estimation of normalized deviation of suspicious results from the mean value have been analyzed. Their unsuitability at low number of measurements is shown. New technique of estimation is offered and critical values of normalized deviation at the number of measurements from 3 to 10 are calculated. The sufficient influence of sample dissymetry on criterion critical value is registered.