ГОСУДАРСТВЕННЫЙ СТАНДАРТ
СОЮЗА ССР

ПРИКЛАДНАЯ СТАТИСТИКА

ПРАВИЛА ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

ГОСТ 11.002—73

Издание официальное

ГОСУДАРСТВЕННЫЙ КОМИТЕТ СТАНДАРТОВ
СОВЕТА МИНИСТРОВ СССР
Москва

РАЗРАБОТАН Всесоюзным научно-исследовательским институтом стандартизации (ВНИИС)

Директор д-р экон. наук **Гличев А. В.**
Научный консультант д-р техн. наук, проф. **Шор Я. Б.**
Научный руководитель канд. техн. наук **Бендерский А. М.**
Исполнители: канд. техн. наук **Лосицкий О. Г.**, **Филиппов Ю. Д.**, **Воробьева В. К.**

ВНЕСЕН Всесоюзным научно-исследовательским институтом стандартизации (ВНИИС)

Директор д-р экон. наук **Гличев А. В.**

ПОДГОТОВЛЕН К УТВЕРЖДЕНИЮ Техническим управлением Государственного комитета стандартов Совета Министров СССР

Начальник отдела общетехнических стандартов **Талашов В. А.**
Ст. инженер **Распевакина Н. Т.**

УТВЕРЖДЕН Государственным комитетом стандартов Совета Министров СССР от 29 сентября 1972 г. (протокол № 142)

Председатель отраслевой научно-технической комиссии зам. председателя Госстандарта СССР **Ткаченко В. В.**
Члены комиссии: **Лямин Б. Н.**, **Панфилов Е. А.**, **Шаронов Г. Н.**, **Гличев А. В.**, **Верченко В. Р.**, **Бурденков Г. К.**, **Киселев Б. Р.**

ВВЕДЕН В ДЕЙСТВИЕ Постановлением Государственного комитета стандартов Совета Министров СССР от 15 января 1973 г. № 84

УДК 658.562.012.7(083.74) Группа Т59

ГОСУДАРСТВЕННЫЙ СТАНДАРТ СОЮЗА ССР

Прикладная статистика
ПРАВИЛА ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

Applied statistic. Methods of testing of outlying observations

ГОСТ 11.002—73

Постановлением Государственного комитета стандартов Совета Министров СССР от 15 января 1973 г. № 84 срок введения установлен

с 01.01.1974 г.
до 01.01.1979 г.

Настоящий стандарт устанавливает правила оценки анормальности результатов наблюдений в условиях производства. Требования стандарта должны использоваться при обработке результатов наблюдений случайной величины, заведомо подчиняющейся нормальному закону распределения.

1. ОСНОВНЫЕ ПОЛОЖЕНИЯ

1.1. Анормальным называется результат наблюдения, резко отклоняющийся от группы результатов наблюдений, которые являются нормальными.

1.2. При оценке анормальных результатов наблюдений следует рассмотреть альтернативу:

резко отклоняющийся результат наблюдения получен в тех же условиях, что и остальная группа результатов наблюдений, принадлежит к той же генеральной совокупности, но вероятность получения его мала. В этом случае оцениваемый результат наблюдений не следует исключать и расчет выборочных характеристик для оценки параметров генеральной совокупности (средняя и дисперсия) должен производиться с учетом всех без исключения результатов наблюдений;

резко отклоняющийся результат наблюдения вполне мог бы быть следствием случайного нарушения нормальных условий или грубых ошибок при наблюдении или расчете. В этом случае указанный результат наблюдений не принадлежит той же генеральной совокупности, что и остальные, поэтому он может быть исключен из общей выборки. Тогда расчет выборочных характеристик

Издание официальное

для оценки параметров генеральной совокупности должен производиться по выборке, из которой исключен анормальный результат наблюдения.

1.3. Следует иметь в виду, что решение такой задачи в любом случае с определенной малой вероятностью может оказаться ошибочным:

а) принимается решение о том, что подозреваемый результат наблюдения нормален и он оставляется, в то время как в действительности он анормален и подлежит исключению;

б) принимается решение об исключении результата наблюдения как анормального, в то время как он нормален и его следует оставить.

Поэтому при принятии решения об исключении или сохранении резко отклоняющихся результатов наблюдений нужно проявлять большую осторожность и перед этим внимательно анализировать условия, в которых получился резко отклоняющийся результат наблюдения.

1.4. Когда определенно известно, что резкое отклонение одного из результатов наблюдений возникло при воздействии факторов, не свойственных условиям получения остальных результатов наблюдений, следует исправить этот результат пересчетом его для условий, при которых исключается несвойственный фактор. Если такой прием оказывается невозможным, то надо обратиться к методам статистической оценки.

Этими же методами необходимо пользоваться и в тех случаях, когда причина резкого отклонения одного из результатов наблюдений от остальных в условиях эксперимента остается незамеченной.

1.5. Принцип решения вопроса об анормальности заключается в том, что по результатам наблюдений рассчитывается определенная функция от случайной величины, для которой известно распределение вероятностей. Вычисленное по выборочным данным значение этой функции сравнивается с ее предельным значением, соответствующим заранее принятой малой вероятности. Если при этом выясняется, что вероятность подозреваемого в анормальности результата наблюдений меньше принятой, то выносится решение о том, что оцениваемый результат анормальный и подлежит исключению, в противном случае его считают нормальным и не исключают.

1.6. При выборе критерия оценки анормальности результатов наблюдений надо руководствоваться следующими принципами:

а) используемые в настоящем стандарте методы и критерии оценки предполагают нормальное распределение измеряемой величины. Поэтому предварительно экспериментатор должен тщательно оценить возможность принятия гипотезы нормального распределения;

б) надо всегда иметь в виду, что результаты обработки наблюдений будут тем точнее, чем больше информации будет использовано. Поэтому, когда известно генеральное среднеквадратическое отклонение или когда оно может быть найдено в результате обработки предшествующих опытов, следует пользоваться критериями, основанными на использовании известного среднеквадратического отклонения. И только когда оно неизвестно и нет возможности его получить на основании предшествующих опытов, следует пользоваться критериями, основанными на использовании выборочного среднеквадратического отклонения.

1.7. Критериям анормальности, приведенным в стандарте, равноценны критерии, приведенные в приложении 2.

1.8. Теоретическое обоснование критериев рассмотрено в справочном приложении 3.

2. КРИТЕРИИ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПРИ НЕИЗВЕСТНОМ ГЕНЕРАЛЬНОМ СРЕДНЕКВАДРАТИЧЕСКОМ ОТКЛОНЕНИИ σ

2.1. Для упорядоченной выборки результатов наблюдений случайной величины

$$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n$$

подсчитывают выборочное среднее

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

и выборочное среднеквадратическое отклонение

$$S = \left\{ \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2 \right\}^{1/2}. \tag{2}$$

Чтобы оценить принадлежность y_n или y_1 к данной нормальной совокупности и принять решение об исключении или оставлении $y_n (y_1)$ в составе выборки, находят отношение

$$U_n = \frac{y_n - \bar{y}}{S} \quad \text{или} \quad U_1 = \frac{\bar{y} - y_1}{S}. \tag{3}$$

Результат сравнивают с величиной β, взятой из табл. 1 для данного объема выборки n и принятого уровня значимости α.

Если $U_n \geqslant \beta$, то подозреваемый в анормальности результат наблюдения анормален и может быть исключен, в противном случае его считают нормальным и не исключают.

Оценка результата наблюдения y_1 производится аналогично.

Примеры применения критериев оценки анормальности результатов наблюдений при неизвестном среднеквадратическом отклонении приведены в приложении 1 (примеры 1 и 2).

Таблица 1

Предельные значения β для случая неизвестного генерального среднеквадратического отклонения σ

$$\alpha = \text{Вер}\left(\frac{y_n - \bar{y}}{S} > \beta\right) \text{ и}$$

$$\alpha^* = \text{Вер}\left(\max\left|\frac{y_n - \bar{y}}{S}\right| > \beta\right)$$

Объем выборки n	Предельное значение β при уровне значимости α			
	0,100	0,075	0,050	0,025
3	1,15	1,15	1,15	1,15
4	1,42	1,44	1,46	1,48
5	1,60	1,64	1,67	1,72
6	1,73	1,77	1,82	1,89
7	1,83	1,88	1,94	2,02
8	1,91	1,96	2,03	2,13
9	1,98	2,04	2,11	2,21
10	2,03	2,10	2,18	2,29
11	2,09	2,14	2,23	2,36
12	2,13	2,20	2,29	2,41
13	2,17	2,24	2,33	2,47
14	2,21	2,28	2,37	2,50
15	2,25	2,32	2,41	2,55
16	2,28	2,35	2,44	2,58
17	2,31	2,38	2,48	2,62
18	2,34	2,41	2,50	2,66
19	2,36	2,44	2,53	2,68
20	2,38	2,46	2,56	2,71
	Предельное значение β при уровне значимости α^*			
	0,200	0,150	0,100	0,050

3. КРИТЕРИЙ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПРИ ИЗВЕСТНОМ ГЕНЕРАЛЬНОМ СРЕДНЕКВАДРАТИЧЕСКОМ ОТКЛОНЕНИИ σ И НЕИЗВЕСТНОМ ГЕНЕРАЛЬНОМ СРЕДНЕМ a

3.1. В практике часто встречается, когда из большого числа предшествующих опытов известно генеральное среднеквадратическое отклонение результатов наблюдений, но по некоторым причинам генеральное среднее зависит от ряда условий и неизвестно.

В случае, когда генеральное среднеквадратическое отклонение известно, следует пользоваться критериями, приведенными в этом разделе. Пользование другими критериями приводит к потере информации и снижению эффективности оценки.

3.2. Для упорядоченной выборки результатов наблюдений

$$y_1 \leqslant y_2 \leqslant \cdot \cdot \cdot \leqslant y_n$$

подсчитывают выборочное среднее

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

и вычисляют отношение

$$t_n = \frac{y_n - \bar{y}}{\sigma}, \qquad (4)$$

которое сравнивается с величиной β, взятой из табл. 2 для данного объема выборки n и принятой вероятности α.

Таблица 2

Предельные значения β для случая известного генерального среднеквадратического отклонения σ и неизвестного среднего a

$$\alpha = \text{Вер}\left(\frac{y_n - \bar{y}}{\sigma} \geqslant \beta\right),$$

$$\alpha = \text{Вер}\left(\frac{\bar{y} - y_1}{\sigma} \geqslant \beta\right) \text{ и}$$

$$\alpha^* = \text{Вер}\left(\max\left|\frac{y_k - \bar{y}}{\sigma}\right| \geqslant \beta\right) \; k=1,n$$

Объем выборки n	Предельное значение β при уровне значимости α			
	0,100	0,050	0,010	0,005
3	1,497	1,738	2,215	2,396
4	1,696	1,941	2,431	2,618
5	1,835	2,080	2,574	2,764
6	1,939	2,184	2,679	2,870
7	2,022	2,267	2,761	2,952
8	2,091	2,334	2,828	3,019
9	2,150	2,392	2,884	3,074
10	2,200	2,441	2,931	3,122
11	2,245	2,484	2,973	3,163
12	2,284	2,523	3,010	3,199
13	2,320	2,557	3,043	3,232
14	2,352	2,589	3,072	3,261
15	2,382	2,617	3,099	3,287
16	2,409	2,644	3,124	3,312
17	2,434	2,668	3,147	3,334
18	2,458	2,691	3,168	3,355
19	2,480	2,712	3,188	3,375
20	2,500	2,732	3,207	3,393
21	2,519	2,750	3,224	3,409
22	2,538	2,768	3,240	3,425
23	2,555	2,784	3,255	3,439
24	2,571	2,800	3,269	3,453
	Предельное значение β при уровне значимости α^*			
	0,200	0,100	0,020	0,010

Если $t_n \gg \beta$, то результат наблюдения y_n анормален и может быть исключен, в противном случае его считают нормальным и не исключают. Оценка анормальности результата наблюдения y_1 производится аналогично.

Пример применения критерия оценки анормальности результатов наблюдений при известном генеральном среднеквадратическом отклонении и неизвестном генеральном среднем приведен в приложении 1 (пример 3).

4. КРИТЕРИЙ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПРИ ИЗВЕСТНОМ ГЕНЕРАЛЬНОМ СРЕДНЕКВАДРАТИЧЕСКОМ ОТКЛОНЕНИИ σ И ИЗВЕСТНОМ ГЕНЕРАЛЬНОМ СРЕДНЕМ a

4.1. В практике такой случай встречается, когда надо проверить, что генеральная совокупность, из которой извлечена выборка, имеет среднее a при условии, когда известно генеральное среднеквадратическое отклонение σ.

4.2. Для оценки анормальности результата наблюдений y_n, где

$$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n,$$

подсчитывают величину

$$V_n = \frac{y_n - a}{\sigma}. \qquad (5)$$

Величину V_n сравнивают с величиной β, взятой из табл. 3 для принятой вероятности.

Если $V_n \gg \beta$, то y_n можно считать анормальным и исключить из выборки, в противном случае его считают нормальным и не исключают.

Оценка анормальности результата наблюдения y_1 производится аналогично.

Пример применения критерия анормальности при известном генеральном среднеквадратическом отклонении и генеральном среднем приведен в приложении 1 (пример 4).

Таблица 3

Предельные значения β для случая известного генерального среднеквадратического отклонения σ и известного генерального среднего a

$$\alpha = \text{Вер}\left(\frac{y_n - a}{\sigma} \geqslant \beta\right) \text{ или}$$

$$\alpha = \text{Вер}\left(\frac{a - y_1}{\sigma} \geqslant \beta\right)$$

Объем выборки n	Предельное значение β при уровне значимости α				
	0,100	0,050	0,0010	0,005	0,001
1	1,282	1,645	2,326	2,576	3,090
2	1,632	1,955	2,575	2,807	3,290
3	1,818	2,121	2,712	2,935	3,403

Продолжение

Объем выборки n	Предельное значение β при уровне значимости α				
	0,100	0,050	0,0010	0,005	0,001
4	1,943	2,234	2,806	3,023	3,481
5	2,036	2,319	2,877	3,090	3,540
6	2,111	2,386	2,934	3,143	3,588
7	2,172	2,442	2,981	3,188	3,628
8	2,224	2,490	3,022	3,227	3,662
9	2,269	2,531	3,057	3,260	3,692
10	2,309	2,568	3,089	3,290	3,719
15	2,457	2,705	3,207	3,402	3,820
20	2,559	2,799	3,289	3,480	3,890
25	2,635	2,870	3,351	3,539	3,944
30	2,696	2,928	3,402	3,587	3,988
40	2,792	3,015	3,480	3,662	4,054
50	2,860	3,082	3,541	3,716	4,108
100	3,076	3,285	3,723	3,892	4,263
250	3,339	3,534	3,946	4,108	4,465
500	3,528	3,703	4,108	4,263	4,607

5. КРИТЕРИЙ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПО МОДУЛЮ ИХ ОТКЛОНЕНИЯ ОТ СРЕДНЕГО

5.1. В некоторых случаях при обработке результатов наблюдений требуется оценивать анормальность, исходя из того, что его отклонение от среднего с заданной вероятностью α * по крайней мере будет больше заданной величины $\beta\sigma$ или меньше величины $-\beta\sigma$ при известном σ и соответственно βS и $-\beta S$ при неизвестном σ.

5.2. В случае неизвестного генерального среднеквадратического отклонения максимальным по модулю отклонением может быть $U_n^* = U_n$, если $U_n > (-U_1)$ или $U_n^* = U_1$, если $U_n \leqslant (-U_1)$.

Величина U_n^* сравнивается с величиной β, взятой из табл. 1 для данного объема выборки n и вероятности α^*.

Если $U_n^* > \beta$, то можно результат наблюдения y_k, соответствующий U_n^*, считать анормальным и исключить, в противном случае его считают нормальным и не исключают.

5.3. При известном генеральном среднеквадратическом отклонении и неизвестном среднем максимальном по модулю отклонениями могут быть $t_n^* = t_n$ и $V_n^* = V_n$, если соответственно $t_n > (-t_1)$ и $V_n > (-V_1)$ или $t_n^* = t_1$ и $V_n^* = V_1$, если соответственно $t_n < (-t_1)$ и $V_n < (-V)_1$.

Величины t_n^* и V_n^* сравнивают с величинами β, взятыми соответственно для t_n^* из табл. 3 и для V_n^* из табл. 4 для данного объема выборки n и вероятности α^*. Если величина t_n^* или V_n^* окажется больше β, то соответствующий результат наблюдения y_k можно считать анормальным и исключить, в противном случае его считают нормальным и не исключают.

Таблица 4

Предельные значения β для случая известного генерального среднеквадратического отклонения σ и известного генерального среднего a

$$\alpha^* = \text{Вер}\left(\max\left|\frac{y_k - a}{\sigma}\right| \geqslant \beta\right) \quad k = 1, n.$$

Объем выборки n	Предельное значение β при уровне значимости α^*							
	0,500	0,200	0,100	0,050	0,020	0,010	0,002	0,001
1	0,674	1,281	1,646	1,963	2,327	2,577	3,089	3,292
2	1,052	1,619	1,949	2,239	2,577	2,806	3,292	3,480
3	1,261	1,801	2,111	2,388	2,711	2,934	3,399	3,588
4	1,410	1,929	2,226	2,489	2,806	3,022	3,480	3,662
5	1,518	2,017	2,313	2,570	2,873	3,089	3,541	3,716
6	1,605	2,091	2,381	2,624	2,934	3,143	3,588	3,764
7	1,673	2,152	2,435	2,648	2,981	3,190	3,629	3,804
8	1,733	2,206	2,482	2,725	3,022	3,224	3,662	3,838
9	1,787	2,246	2,522	2,765	3,065	3,258	3,689	3,868
10	1,835	2,286	2,556	2,799	3,089	3,292	3,716	3,892
15	2,003	2,435	2,698	2,927	3,204	3,399	3,818	3,986
20	2,118	2,543	2,792	3,015	3,285	3,480	3,892	4,054
25	2,206	2,617	2,867	3,082	3,352	3,541	3,946	4,108
30	2,273	2,678	2,921	3,157	3,399	3,399	3,986	4,148
40	2,381	2,772	3,008	3,224	3,480	3,662	4,054	4,214
50	2,462	2,846	3,076	3,285	3,534	3,716	4,108	4,263
100	2,698	3,055	3,278	3,474	3,728	3,892	4,263	4,418
250	2,995	3,325	3,528	3,710	3,939	4,108	4,465	4,607
500	3,197	3,514	3,703	3,878	4,108	4,263	4,607	4,755

5.4. Следует иметь в виду, что $\alpha^* \cong 2\alpha$ и поэтому для оценки анормальности результатов наблюдений по модулю отклонения от среднего в табл. 1—2 внизу даны числовые значения для замены α на 2α.

Примеры применения критериев оценки анормальности результатов наблюдений по модулю отклонения от среднего приведены в приложении 1 (примеры 5 и 6).

6. ОБРАБОТКА РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ ПО МНОГИМ ВЫБОРКАМ (СРАВНЕНИЕ ЧАСТОСТИ ПОЯВЛЕНИЯ ПОДОЗРИТЕЛЬНЫХ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ С ВЕРОЯТНОСТЯМИ α И α^*)

6.1. В некоторых случаях при обработке результатов наблюдений ставят задачу оценить анормальность результатов при условии, что имеется N выборок объемом n каждая и в m из них содержится по одному результату, вероятность которого в выборке как нормального равна α. Для решения этой задачи следует убедиться в том, что выборки независимые. Затем вычисляют вероятность того, что по крайней мере в m выборках из N будут результаты наблюдений, вероятность которых как нормальных равна α. Эту вероятность вычисляют по формуле

$$R_{N,m} = \sum_{i=m}^{N} \binom{N}{i} \alpha^i (1-\alpha)^{N-i}$$

(приближенные значения $R_{N,m}$ приведены в табл. 5). Если она окажется малой величиной, то подозреваемые результаты можно считать анормальными и исключить, в противном случае их считают нормальными и не исключают.

Таблица 5

Значения функции $R_{N,m} = \sum_{i=m}^{N} \binom{N}{i} \alpha^i (1-\alpha)^{N-i}$ при $N\alpha$

m	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	2,0	3,0
0	0,0952	0,1813	0,2592	0,3297	0,3035	0,4512	0,5034	0,5507	0,5934	0,6321	0,8647	0,9502
1	0,0047	0,0175	0,0369	0,0616	0,0902	0,0902	0,1558	0,1912	0,2275	0,2642	0,5940	0,8009
2	0,0002	0,0011	0,0036	0,0118	0,0144	0,0021	0,0341	0,0774	0,0629	0,0903	0,3233	0,5768
3	0,0001	0,0001	0,0003	0,0079	0,0018	0,0024	0,0058	0,0091	0,0115	0,0190	0,1429	0,3528
4	0,0000	0,0001	0,0001	0,0001	0,0001	0,0004	0,0008	0,0014	0,0023	0,0037	0,0527	0,1847
5	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0002	0,0003	0,0005	0,0166	0,0839
6	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0001	0,0001	0,0101	0,0335
7	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0001	0,0001	0,0001	0,0001	0,0045	0,0119
8	—	—	—	—	—	—	0,0000	0,0000	0,0000	0,0001	0,0002	0,0038
9	—	—	—	—	—	—	—	—	—	0,0000	0,0001	0,0011
10	—	—	—	—	—	—	—	—	—	—	0,0001	0,0003
11	—	—	—	—	—	—	—	—	—	—	0,0001	0,0001
12	—	—	—	—	—	—	—	—	—	—	0,0000	0,0001
13	—	—	—	—	—	—	—	—	—	—	—	0,0001
14	—	—	—	—	—	—	—	—	—	—	—	0,0001
15	—	—	—	—	—	—	—	—	—	—	—	0,0000

ПРИЛОЖЕНИЕ 1 к ГОСТ 11.002—73
Рекомендуемое

ПРИМЕРЫ ПРИМЕНЕНИЯ КРИТЕРИЕВ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ

Пример 1. Определяют твердость образцов металла стальным шариком (по Бринеллю). При этом получено пять значений твердости: НВ 180, НВ 182, НВ 183, НВ 184, НВ 196.

Требуется оценить результат НВ 196 при заданном $\alpha = 0{,}050$.

Для критерия $U_n = \frac{y_n - \bar{y}}{S}$ вычисляют

$$\bar{y} = \frac{180 + 182 + 183 + 184 + 196}{5} = 185.$$

$$S = \sqrt{\frac{1}{4}\{(180 - 185)^2 + (182 - 185)^2 + (183 - 185)^2 + (184 - 185)^2 + (196 - 185)^2\}} = 6{,}3,$$

откуда

$$U_n = \frac{196 - 185}{6{,}3} = 1{,}75.$$

В табл. 1 настоящего стандарта для $n=5$ и $\alpha = 0{,}050$ находят $\beta = 1{,}67$.

Тогда $U_n > \beta$ и результат НВ 196 можно считать анормальным и исключить. Более того, такой вывод можно сделать с большей вероятностью, так как при $n=5$ и $\alpha = 0{,}025$ $\beta = 1{,}72$.

Необходимость тщательного подхода к оценке анормальности можно показать на другом примере, в котором выборочные результаты наблюдений мало отличаются от приведенного.

Пример 2. Была получена твердость пяти деталей: НВ 178, НВ 180, НВ 184, НВ 186, НВ 197.

Вычисляют

$$\bar{y} = \frac{178 + 180 + 184 + 186 + 197}{5} = 185;$$

$$S = \sqrt{\frac{1}{4}\{(178 - 185)^2 + (180 - 185)^2 + (184 - 185)^2 + (186 - 185)^2 + (197 - 185)^2\}} = 7{,}4,$$

откуда

$$U_n = \frac{197 - 185}{4} = 1{,}62.$$

В этом случае $U_n < \beta$ (1,67) для $n=5$ и $\alpha = 0{,}050$ и, следовательно, для исключения результата НВ 197 нет оснований.

Пример 3. Испытывают пробегом выборку 10 шин. Получены следующие результаты пробега до полного износа в км: 65000, 66100, 65700, 65800, 66500, 67000, 64700, 65000, 64000, 60200. Генеральное среднеквадратическое отклонение $\sigma = 970$ км.

Требуется оценить результат 60200 км при заданном $\alpha = 0{,}005$

Вычисляют

$$\bar{y}=\frac{65,0+66,1+65,7+65,8+66,5+67,7+64,7+65,0+64,0+60,2}{10}=65,0;$$

$$t_1=\frac{60,2-65,0}{0,97}=-4,948.$$

В табл. 2 настоящего стандарта для $n=10$ и $\alpha=0,005$ находят $\beta=3,122$. Поскольку $t_1<-\beta$, то для заданных условий можно считать результат $y_1=60200$ км анормальным и исключить.

Пример 4. Оценка анормальности при известном генеральном среднеквадратическом отклонении и известном генеральном среднем (см. 4.1).

Для проверки стабильности технологической операции обточки вала в течение заданного времени была извлечена выборка 12 валов диаметрами (мм): 40,00; 40,02; 39,99; 39,98; 40,00; 40,03; 39,99; 39,98; 40,01; 40,08; 40,04; 39,97.

Генеральное среднеквадратическое отклонение $\sigma=0,024$, генеральное среднее $a=40,00$, требуется оценить результат 40,08.

Вычисляют

$$V_n=\frac{40,08-40,00}{0,024}=3,33.$$

Путем линейного интерполирования в табл. 3 для $n=12$ и $\alpha=0,005$ находят $\beta=3,346$. Отсюда вероятность того, что результат 40,08 принадлежит данной нормальной совокупности мала ($\alpha=0,005$), поэтому его можно считать анормальным и исключить.

Пример 5. При определении плотности электролита в относительных единицах получены следующие результаты X: 215; 210; 210; 201; 217; 215; 215; 214; 209; 217; 228.

Подозревают в анормальности результаты 201 и 228. Требуется оценить их по модулю отклонения от среднего.

Вычисляют

$$\bar{X}=\frac{215+210+210+201+217+215+215+214+209+217+228}{11}=212,9;$$

$$S=\sqrt{\frac{1}{10}\{(2,1)^2+(-2,9)^2+(-2,9)^2+(-11,9)^2+4,1^2+2,1^2+2,1^2+1,1^2+(-3,9)^2+4,1^2+15,1^2\}}=6,7,$$

откуда

$$U_n^*=\left|\frac{228-212,9}{6,7}\right|=2,25.$$

В табл. 1 настоящего стандарта для $n=11$ значениям $\beta^*=2,36$ и $\beta=2,23$ соответствуют вероятности $\alpha^*=0,025$ и $\alpha^*=0,050$. После интерполирования их с вероятностью $\alpha^*=0,092$ приходят к выводу, что результаты можно считать нормальными и для исключения их нет оснований. Оценку $X=201$ производят аналогично.

Пример 6. Микроизмерения штриха на масштабе дали следующие результаты:

3,68;	5,08;	1,81;	4,43;	3,11;
2,95;	4,65;	3,43;	4,76;	6,35;
3,27;	3,26;	2,75;	3,78;	4,08;
2,48;	4,15;	4,49;	4,51;	4,84.

Точность этих измерений характеризуется известной $\sigma=1,00$. Требуется оценить анормальность результатов 1,81 и 6,35.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = 3,893;$$

$$t_n = \max\left|\frac{X_i - \overline{X}}{\sigma}\right| = \frac{6,35 - 3,89}{1} = 2,46.$$

Из табл. 2 настоящего стандарта видно, что при $n=20$ и нормальных условиях это значение имеет вероятность $\alpha > 0,20$, поэтому считать его анормальным нет оснований. Оценку результата 1,81 производят аналогично.

Пример 7. При съемке местности для нанесения рельефа на топографическую карту было получено 100 выборок по восьми измерений в каждой. В трех выборках было по одному измерению y_n такому, что

$$U_n = \frac{y_n - \overline{y}}{S} \gg 2,13.$$

Требуется оценить анормальность этих результатов.

Из табл. 1 настоящего стандарта следует, что при $n=8$ и $U_n = 2,13$ $\alpha = 0,025$. При этой вероятности вероятность того, что по крайней мере в шести выборках из 100 будут иметь место указанные результаты равна:

$$R_{100,6} = \sum_{i=6}^{100} \binom{100}{i} 0,025^i (1-0,025)^{100-i}.$$

Имея в виду, что для $N=100$ и $\alpha=0,025$ биномиальное распределение хорошо аппроксимируется распределением Пуассона с параметром $\lambda = N\alpha = 100 \cdot 0,025 = 2,5$, это выражение можно заменить:

$$R_{100,6} \cong e^{-2,5}\sum_{i=6}^{100}\frac{(2,5)^i}{i!} = 1 - e^{-2,5}\sum_{i=0}^{6}\frac{(2,5)^i}{i!}.$$

По таблице распределения Пуассона определяют для $\lambda = 2,5$ и $m=6$

$$e^{-2,5}\sum_{i!}^{6}\frac{(2,5)^i}{i!} = 0,9805 \text{ и}$$

$$R_{100,6} = 0,0195.$$

Вследствие малой вероятности $R_{100,6}$ есть основания считать результаты y_n анормальными и исключить их.

Подобный вывод был бы неправильным, если бы в двух выборках из 100 встретились результаты наблюдений, вероятность которых как нормальных $\alpha = 0,025$.

В этом случае вероятность

$$R_{100,6} = 1 - e^{-2,5}\sum_{i=0}^{2}\frac{(2,5)^i}{i!} = 1 - 0,5499 = 0,4501$$

достаточно велика и для исключения указанных результатов нет оснований.

ПРИЛОЖЕНИЕ 2 к ГОСТ 11.002—73
Справочное

КРИТЕРИИ ОЦЕНКИ АНОРМАЛЬНОСТИ РЕЗУЛЬТАТОВ НАБЛЮДЕНИЙ, РАВНОЦЕННЫЕ ПРИВЕДЕННЫМ В СТАНДАРТЕ

1. Наряду с критериями анормальности, приведенными в стандарте, имеются равноценные им. Применение этих критериев правомерно так же, как приведенных в разд. 2 и 3 настоящего стандарта.

2. Вместо критерия анормальности

$$U_n = \frac{y_n - \overline{y}}{S}$$

может быть использован критерий

$$U'_n = \frac{y_n - \overline{y}_{n-1}}{S},$$

в котором

$$\overline{y}_{n-1} = \frac{1}{n-1}\sum_{j=1}^{n-1} y_j;$$

$$S_{n-1} = \left\{\frac{1}{n-2}\sum_{j=1}^{n-1}(y_j - \overline{y}_{n-1})^2\right\}^{1/2},$$

т. е. $\overline{y}_{n-1}$ и $\overline{S}_{n-1}$ подсчитаны по всем результатам наблюдений, за исключением результата, подозреваемого в анормальности.

В этом случае величину $\overline{U}'_n$ следует сравнивать с величиной β, взятой из табл. 1, и принимать решение тем же путем, что и в случае с U_n.

Оценка результата y_1 с помощью критерия

$$U'_1 = \frac{\overline{y}_1 - y_1}{S_1},$$

в котором $\overline{y}_1$ и S_1 подсчитаны по всем результатам наблюдений, за исключением подозреваемого в анормальности y_1, производится аналогично.

Доказательства равноценности критериев U_n и U'_n (U_1 и U'_1) приведены в приложении 3.

3. Вместо критерия оценки анормальности t_n можно использовать критерий анормальности

$$t'_n = \frac{y_n - \overline{y}_{n-1}}{\sigma},$$

в котором

$$\overline{y}_{n-1} = \frac{1}{n-1}\sum_{j=1}^{n-1} y_j,$$

т. е. выборочное среднее для выборки без подозреваемого в анормальности результата наблюдения y_n.

В этом случае полученное значение t'_n следует сравнить с величиной β, взятой из табл. 2, и действовать точно в соответствии с требованиями разд. 3 настоящего стандарта.

Оценка анормальности результата наблюдения y_1 с помощью критерия

$$t_1' = \frac{\bar{y}_1 - y_1}{\sigma},$$

в котором

$$\bar{y}_1 = \frac{1}{n-1}\sum_{j=2}^{n} y_j$$

производится аналогично.

Следует иметь в виду, что оба критерия t_n и t'_n (t_1 и t_1') равноценны.

4. Примеры применения равноценных критериев

4.1. В примере 1 приложения 1 установлено, что вероятность результата НВ 196 как нормального не более $\alpha = 0{,}025$ и поэтому его можно считать анормальным и исключить.

Такой же вывод получается при применении критерия U_n'. В этом случае для оценки твердости металла следует принять

$$\bar{y}_{n-1} = \frac{180+182+183+184}{4} = 182{,}2,$$

$$S_{n-1} = \sqrt{\frac{1}{3}\{(180-182{,}2)^2+(182-182{,}2)^2+(183-182{,}2)^2 + (184-182{,}2)^2\}} = 1{,}67,$$

откуда

$$U_n' = \frac{196{,}0 - 182{,}2}{1{,}67} = 8{,}25,$$

что больше $\beta = 6{,}509$ при $n=5$ и $\alpha = 0{,}025$ (см. табл. 1).

4.2. В примере 2 приложения 1 установлено, что для исключения результата НВ 197 нет оснований.

Точно такая же оценка получается при использовании критерия U_n' по табл. 1.

Выше уже подсчитаны

$$\bar{y}_{n-1} = 182{,}2 \text{ и } S_{n-1} = 1{,}67,$$

тогда

$$U_n' = \frac{196 - 182{,}2}{1{,}67} = 8{,}25,$$

что больше $\beta = 5{,}086$ для $n=5$ и $\alpha = 0{,}050$ (см. табл. 1).

Для второй рассмотренной выборки получается

$$\bar{y}_{n-1} = \frac{178+180+184+186}{4} = 182,$$

$$S_{n-1} = \sqrt{\frac{1}{3}\{(178-182)^2+(180-182)^2+(184-182)^2+(186-182)^2\}} = 3{,}65,$$

откуда

$$U_n' = \frac{197 - 182}{3{,}65} = 4{,}1;$$

что меньше $\beta = 5{,}086$. Как и следовало ожидать, выводы при использовании критериев U_n и U'_n получились одинаковые.

4.3. В примере 3 приложения 1 установлено, что результат $y_1 = 60200$ км можно считать анормальным и исключить его.

К такому же выводу приходят при использовании критерия t_1'

$$\bar{y}_1 = \frac{65{,}0+66{,}1+65{,}7+65{,}8+66{,}5+67{,}0+64{,}7+65{,}0+64{,}0}{9} = 65{,}7$$

$$\text{и } t_1' = \frac{60{,}2 - 65{,}7}{0{,}97} = 5{,}650.$$

Это меньше, чем $-\beta = -3{,}467$.

Таблица 1

Предельные значения β для случая неизвестного генерального среднеквадратического отклонения σ

$$\alpha = \text{Вер}\left(\frac{y_n - \bar{y}_{n-1}}{S_{n-1}} > \beta\right) \quad \text{или}$$

$$\alpha = \text{Вер}\left(\frac{\bar{y}_{n-1} - y_1}{S_{n-1}} > \beta\right)$$

Объем выборки n	Предельное значение β при уровне значимости α			
	0,100	0,050	0,025	0,010
3	11,332	21,786	70,822	70,822
4	4,958	7,217	10,117	15,918
5	3,899	5,086	6,509	8,971
6	3,494	4,337	5,308	6,880
7	3,285	3,973	4,735	5,868
8	3,177	3,764	4,398	5,324
9	3,103	3,629	4,182	4,978
10	3,048	3,534	4,040	4,742
11	3,022	3,474	3,939	4,580
12	2,995	3,426	3,858	4,452
13	2,981	3,386	3,791	4,350
14	2,968	3,359	3,750	4,276
15	2,961	3,332	3,710	4,209
16	2,961	3,318	3,676	4,162
17	2,954	3,305	3,656	4,114
18	2,954	3,298	3,629	4,081
19	2,954	3,285	3,615	4,047
20	2,954	3,278	3,602	4,027
21	2,954	3,271	3,588	4,000
22	2,954	3,271	3,582	3,979
23	2,961	3,265	3,568	3,966
24	2,961	3,265	3,561	3,953
25	2,961	3,265	3,555	3,939

Таблица 2

Предельные значения β для случая известного генерального среднеквадратического отклонения σ и неизвестного среднего c

$$\alpha = \text{Вер}\left(\frac{y_n - \bar{y}_{n-1}}{\sigma} \geqslant \beta\right) \text{ и}$$

$$\alpha^* = \text{Вер}\left(\max\left|\frac{y_k - \bar{y}_{k-1}}{\sigma}\right| \geqslant \beta\right)\ k=1,n$$

Объем выборки n	Предельное значение β при уровне значимости α				
	0,100	0,050	0,010	0,005	0,001
3	2,246	2,604	3,325	3,595	4,168
4	2,260	2,590	3,244	3,487	4,027
5	2,293	2,597	3,217	3,453	3,959
6	2,327	2,624	3,217	3,447	3,932
7	2,361	2,644	3,224	3,447	3,926
8	2,388	2,664	3,231	3,447	3,919
9	2,421	2,691	3,244	3,460	3,919
10	2,442	2,711	3,258	3,467	3,926
11	2,469	2,732	3,271	3,480	3,932
12	2,489	2,752	3,285	3,487	3,939
13	2,516	2,772	3,298	3,501	3,946
14	2,529	2,876	3,305	3,514	3,953
15	2,550	2,806	3,318	3,521	3,959
16	2,570	2,819	3,332	3,534	3,966
17	2,583	2,833	3,345	3,541	3,973
18	2,604	2,846	3,352	3,555	3,979
19	2,617	2,860	3,366	3,561	3,986
20	2,630	2,873	3,372	3,568	3,993
21	2,644	2,887	3,386	3,582	4,000
22	2,657	2,900	3,392	3,588	3,006
23	2,671	2,097	3,399	3,595	4,013
24	2,684	2,921	3,413	3,602	4,020
25	2,691	2,934	3,420	3,609	4,027

Предельное значение β при уровне значимости α^*

0,200	0,100	0,020	0,010	0,002

ПРИЛОЖЕНИЕ 3 к ГОСТ 11.002—73
Справочное

ТЕОРЕТИЧЕСКОЕ ОБОСНОВАНИЕ ФОРМУЛ И ТАБЛИЦ, ИСПОЛЬЗУЕМЫХ В СТАНДАРТЕ

1. Впервые правильное решение задачи об оценке анормальных результатов наблюдений было дано Н. В. Смирновым в 1941 г. В 1950 г. Ф. Е. Груббс повторил результат Н. В. Смирнова без ссылки на него [3]. Поэтому в литературе часто неправомерно критерий Н. В. Смирнова приписывается Ф. Е. Груббсу. Существовавшие до 1941 г. критерии оценки анормальности, в том числе известный критерий Шовенэ, оказались либо пригодными только в некоторых частных случаях, либо вообще научно необоснованными и непригодными для использования. Подробный анализ и критика указанных критериев приведены в [2].

В 1952 г. А. М. Бендерский распространил критерий Н. В. Смирнова на случай оценки модуля максимального отклонения от среднего в ряду наблюдений [4]. Обзор дальнейших работ по оценке анормальности наблюдений приведен в [5].

В последнее время появились работы [8—15] по вопросу принятия решений об оставлении или исключении подозреваемых в анормальности результатов наблюдений в зависимости от дополнительных данных, выражающих функцию риска ошибочных решений. Выводы, сделанные в этих работах, в стандарт не включены ввиду их дискуссионного характера.

2. Вывод критерия Н. В. Смирнова (см. таблицы в [1, 2, 3, 6]).

Рассматривается выборка независимых случайных величин из генеральной нормальной совокупности (0, σ) с известной дисперсией σ^2

$$X_1, \ldots, X_n \tag{1}$$

и получающаяся из нее упорядоченная выборка

$$y_1 \leqslant \ . \ . \ . \leqslant y_n, \text{ где } X_i = y_j. \tag{2}$$

Требуется определить распределение вероятности величины

$$U_n = \frac{y_n - \overline{y}}{S}; \quad \overline{y} = \frac{1}{n} \sum_{j=1}^{n} y_j = \frac{1}{n} \sum_{i=1}^{n} X_i;$$

$$S = \left\{ \frac{1}{n} \sum_{j=1}^{n} (y_j - \overline{y}) \right\}^{1/2*}. \tag{3}$$

При указанных выше условиях величина

$$U_k = \frac{X_k - \overline{X}}{S} \quad (k = 1, 2, \ . \ . \ . \ , n) \tag{4}$$

---

* В связи с тем, что при $S^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$ получаются более компактные выражения для границ интервалов и рекуррентного соотношения при теоретическом рассмотрении, принято такое выражение вместо

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2.$$

имеет распределение с плотностью вероятностей

$$\varphi(U)=\frac{1}{\sqrt{\pi(n-1)}}\,\frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)}\left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}}. \tag{5}$$

Вероятность того, что значение U_n будет заключено в пределах U, $U+dU$, равна вероятности того, что одно из отношений U_k будет в этом интервале $[\varphi(U)d\bar{U}]$, умноженное на условную вероятность того, что остальные $n-1$ значений U_i будут меньше U_k и умноженное на число способов, которыми может быть получено U_n.

$$\text{Вер}\{U\leqslant U_n\leqslant U+dU\}=nP_{n-1}(U)\varphi(U)dU. \tag{6}$$

Для краткости можно записать:

$$\text{Вер}\{U_1,\ \ldots\ ,U_{n-1}<U\}=P_{n-1}(U). \tag{7}$$

В [2] подробно показано, что $U_n\leqslant\sqrt{n-1}$ и в интервалах

$$\begin{aligned}
&\sqrt{\frac{n-2}{2}},\quad \sqrt{n-1}\quad &&\text{— вероятность } P_{n-1}(U)=1\\
&\sqrt{\frac{n-3}{3}},\quad \sqrt{\frac{n-2}{2}}\quad &&\text{— вероятность } P_{n-2}(U)=1\\
&\sqrt{\frac{n-4}{4}},\quad \sqrt{\frac{n-3}{3}}\quad &&\text{— вероятность } P_{n-3}(U)=1
\end{aligned} \tag{8}$$

В [2] показано, что существует рекуррентное соотношение

$$\delta(U)=\sqrt{\frac{n}{n-1}}\quad\frac{U}{\sqrt{1-\frac{1}{n-1}U^2}}, \tag{9}$$

при котором в интервале

$$\sqrt{\frac{n-3}{3}}<\beta\leqslant\sqrt{\frac{n-2}{2}} \tag{10}$$

вероятность $P_{n-1}(U)$ того, что $U_{n-1}<U$ при условии $U_n=U$ равна вероятности того, что

$$U'_{n-1}\leqslant\delta(U), \tag{11}$$

где

$$U'_{n-1}=\frac{y_{n-1}-\bar{y}_{n-1}}{S_{n-1}}, \tag{12}$$

тогда

$$\bar{y}_{n-1}=\frac{1}{n-1}\sum_{j=1}^{n-1}y_j;$$

$$S_{n-1}=\left\{\frac{1}{n-2}\sum_{j=1}^{n-1}(y_j-\bar{y}_{n-1})^2\right\}^{1/2},$$

т. е.

$$P_{n-1}(U)=1-(n-1)\int_{\delta(U)}^{\sqrt{n-2}}\varphi(U')dU'. \tag{13}$$

Следовательно, в интервале

$$\sqrt{\frac{n-2}{2}} \leqslant \beta \leqslant \sqrt{n-1}$$

$$\text{Вер}\{U_n \geqslant \beta\} = \frac{n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_{\beta}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU, \qquad (14)$$

в интервале

$$\sqrt{\frac{n-3}{3}} \leqslant \beta \leqslant \sqrt{\frac{n-2}{2}}.$$

$$\text{Вер}\{U_n \geqslant \beta\} = \frac{n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \cdot \int_{\beta}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU -$$

$$- \frac{n}{\pi} \sqrt{\frac{n-1}{n-2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} \int_{\beta}^{\sqrt{\frac{n-2}{2}}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU \int_{\delta(U)}^{\sqrt{n-2}} \left(1-\frac{\Theta^2}{n-2}\right)^{\frac{n-5}{2}} d\Theta. \qquad (15)$$

Аналогично получают формулы для следующих интервалов. На практике для расчета таблиц значений β, соответствующих установленным малым вероятностям, оказывается достаточно приведенных формул.

3. Две формы критериев оценки анормальности результатов наблюдений при неизвестном σ [2].

Простыми алгебраическими преобразованиями доказывают, что

$$U_n' = \frac{\sqrt{\frac{n}{n-1}} U_n}{\sqrt{1-\frac{1}{n-1}U_n^2}} \qquad (16)$$

(при $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \overline{X})^2$ эта зависимость имеет вид

$$U_n' = \sqrt{\frac{n(n-2)}{(n-1)^2}} \frac{U_n}{\sqrt{1-\frac{nU_n^2}{(n-1)(n-2)}}}\ )$$

и таким образом, критерии U_n и U'_n равноценны.

Для расчета таблиц предельных значений, соответствующих заданным малым вероятностям, используется то, что неравенство $U_n \geqslant \beta$ эквивалентно неравенству

$$U_n' \geqslant \sqrt{\frac{n}{n-1}} \frac{\beta}{\sqrt{1-\frac{\beta^2}{n-1}}}.$$

Следовательно, после подстановки формулы (16) в формулу (14) в интервале $\sqrt{\frac{n-2}{2}} < \beta \leqslant \sqrt{n-1}$ получается

$$\text{Вер}\{U_n' \geqslant V\} = \frac{n}{\sqrt{\pi(n-2)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_V^\infty \left(1+\frac{U'}{n-2}\right)^{\frac{-n+1}{2}} dU', \qquad (17)$$

где

$$V = \sqrt{\frac{n}{n-1}} \frac{\beta}{\sqrt{1-\frac{\beta^2}{n-1}}}, \qquad \sqrt{n-2} \leqslant V \leqslant \infty. \qquad (18)$$

Для следующих интервалов формулы получают аналогично.

4. Вывод критерия оценки анормальности результатов наблюдений при известном σ и неизвестном a [2] (см. таблицы в [2, 3]).

Вероятность получения упорядоченной выборки

$$y_1 \leqslant y_2 \leqslant \ldots \leqslant y_n \qquad (19)$$

из нормальной генеральной совокупности с генеральной средней 0 и генеральным среднеквадратическим отклонением 1, за счет смещения начала отсчета и изменения масштаба всегда можно добиться, чтобы $a=0$, $\sigma=1$ равна

$$\frac{n!}{(\sqrt{2\pi})^n} e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2} dy_1 dy_2 \ldots dy_n. \qquad (20)$$

Множитель $n!$ в формуле (20) поставлен вследствие того, что одни и те же значения выборочных y_j могут появляться в различном порядке, а $n!$ дает число возможных перестановок.

Величины y_j коррелированы, поэтому путем ортогонального преобразования получают

$$\left.\begin{aligned}
\eta_1 &= \frac{1}{\sqrt{1\cdot 2}}(y_2 - y_1),\\
\eta_2 &= \frac{1}{\sqrt{2\cdot 3}}[2y_3 - (y_2+y_1)],\\
&\ldots\ldots\ldots\ldots\\
\eta_i &= \frac{1}{\sqrt{i(i+1)}}[iy_{i+1} - (y_i + \ldots + y_1)]\\
&\ldots\ldots\ldots\ldots\\
\eta_{n-1} &= \frac{1}{\sqrt{(n-1)n}}[(n-1)y_n - (y_{n-1} + \ldots + y_1)]\\
\eta_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i = \sqrt{n}\cdot \bar{y}.
\end{aligned}\right\} \qquad (21)$$

Величины y преобразуются в независимые величины. С учетом свойств использованного ортогонального преобразования формула (21) преобразуется в

$$\frac{n}{(\sqrt{2\pi})^n} e^{\frac{n\bar{y}^2}{2} - \frac{1}{2}\sum_{i=1}^{n-1}\frac{\eta_i}{i(i+1)}} d\bar{y}\, d\eta_1 d\eta_2 \ldots d\eta_{n-1}. \qquad (22)$$

Интегрирование формулы (22) по $\bar{y}$ в пределах от $-\infty$ до ∞ дает после несложных преобразований выражение плотности распределения вероятности максимального отклонения от среднего $t_n = y_n - \bar{y}$

$$\Psi(t_n) = \frac{n\sqrt{n}}{(2\pi)^{n-1}} e^{-\frac{nt_n}{2(n-1)}} \int_0^{nt_n} e^{-\frac{\eta_n - 2^2}{2(n-2)(n-1)}} d\eta_{n-2} \times$$

$$\times \int_0^{\eta_{n-2}} e^{-\frac{\eta_{n-3}}{2(n-3)(n-2)}} d\eta_{n-3} \ldots \int_0^{\eta_2} e^{-\frac{\eta_1^2}{4}} d\eta_1. \qquad (23)$$

Следовательно, вероятность

$$\text{Вер}\{t_n \geqslant \beta\} = 1 - \int_0^{\beta} \Psi(t_n)dt_n. \quad (24)$$

По формуле (24) составлены таблицы [3], использованные в настоящем стандарте.

Если используется критерий $t'_n = \frac{y_n - \overline{y}_{n-1}}{\sigma}$, то его предельное значение β' (табл. 4) получают по формуле

$$\beta' = \frac{n}{n-1}\beta[2]$$

5. Замечание об оценке анормальности результатов наблюдений при известных a и σ [2].

Неравенство $V_n = \frac{y_n - a}{\sigma} \geqslant \beta$ оценивается по вероятности выхода за границу β хотя бы одного отношения V, т. е.

$$\text{Вер}\{V \geqslant \beta\} = 1 - \left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\beta} e^{-\frac{X^2}{2}} dX\right]^n. \quad (25)$$

6. Вывод критерия оценки анормальности результатов наблюдений по модулю отклонений от среднего [4] (см. таблицы в [10]).

Рассматривается упорядоченная выборка

$$0 \leqslant U_1^* \leqslant \ . \ . \ . \ \leqslant U_n^*, \quad (26)$$

в которой $U_j^* = |U_i|$.

Последние m членов выборки (см. формулу 26) могут получаться как i первых и k последних членов выборки [см. формулу (3)]

Доказывается, что в интервале

$$\sqrt{\frac{n}{2}} < \beta^* \leqslant \sqrt{n-1} \qquad P(U^*_{n-1} \leqslant \beta^*) = 1, \quad (27)$$

в интервале

$$\sqrt{\frac{n-3}{3-8/n}} < \beta^* \leqslant \sqrt{\frac{n}{2}} \qquad P(U^*_{n-2} \leqslant \beta^*) = 1. \quad (28)$$

Далее следует, что в интервале

$$\sqrt{\frac{n-2}{2}} < \beta^* \leqslant \sqrt{\frac{n}{2}} \quad (29)$$

$P(U^*_{n-1} \leqslant \beta^*) \neq 1$ только в том случае, если $i=k=1$,

в интервале

$$\sqrt{\frac{n-3}{3-8/n}} < \beta^* = \sqrt{\frac{n-2}{2}} \quad (30)$$

$$P(U^*_{n-1} \leqslant \beta^*) \neq 1,$$

если:

а) $i=k=1$;

б) $i=0$, $k=2$;

в) $i=2$; $k=0$, причем условия *а*, *б* и *в* несовместимы в интервале [см. формулу (30)].

Аналогичным путем получают следующие значения для границ интервалов:

$$\sqrt{\frac{n-k-i}{k+i-4ki/n}}, \text{ если } k \neq 1; \quad \sqrt{\frac{n}{k+1}}, \text{ если } k = i \quad (31)$$

наконец, для вычисления вероятности $P(U^*_{n-1} \ll \beta)$ при условии, что $U^*_n \gg \beta^*$ существуют соотношения

$$\gamma(U) = \frac{n-2}{\sqrt{(n-1)n}} \frac{U}{\sqrt{1-\frac{U^2}{n-1}}}, \text{ если } i \neq 0,\ k \neq 0 \qquad (32)$$

$$\delta(U) = \sqrt{\frac{n}{n-1}} \cdot \frac{U}{\sqrt{1-\frac{n^2}{n-1}}}, \text{ если } i=0 \text{ или } k=0. \qquad (33)$$

Для функции Вер$\{U^*_n > \beta^*\}$ получаются выражения: в интервале [см. формулу (27)]

$$\text{Вер}\{U^*_n \gg \beta^*\} = \frac{2n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_{\beta^*}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU; \qquad (34)$$

в интервале [см. формулу (29)]

$$\text{Вер}\{U^*_n \gg \beta^*\} = \frac{2n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_{\beta^*}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU -$$

$$-\frac{2n}{\pi}\sqrt{\frac{n-1}{n-2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} \int_{\beta^*}^{\sqrt{\frac{n}{2}}} \left(1-\frac{k^2}{n-1}\right)^{\frac{n-4}{2}} dU \int_{\gamma(U)}^{\sqrt{n-2}} \left(1-\frac{\xi^2}{n-2}\right)^{\frac{n-5}{2}} d\xi. \qquad (35)$$

в интервале [см. формулу (30)]

$$\text{Вер}\{U^*_n \gg \beta^*\} = \frac{2n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_{\beta^*}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU -$$

$$-\frac{2n}{\pi}\sqrt{\frac{n-1}{n-2}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} \left\{ \int_{\beta^*}^{\sqrt{\frac{n}{2}}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU \int_{\gamma(U)}^{\sqrt{n-2}} \left(1-\frac{\xi^2}{n-2}\right)^{\frac{n-5}{2}} d\xi + \right.$$

$$\left. + \int_{\beta^*}^{\sqrt{(n-2):2}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU \int_{\delta(U)}^{\sqrt{n-2}} \left(1-\frac{\eta^2}{n-2}\right)^{\frac{n-5}{2}} d\eta \right\}. \qquad (36)$$

Л. Н. Большев показал [7], что для обработки опытных данных можно пользоваться приближенным выражением для распределения модуля максимального отклонения от среднего в ряду наблюдений

$$\text{Вер}\{U^*_n \gg \beta^*\} = \frac{2n}{\sqrt{\pi(n-1)}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \int_{\beta^*}^{\sqrt{n-1}} \left(1-\frac{U^2}{n-1}\right)^{\frac{n-4}{2}} dU \qquad (37)$$

и таблицами, приведенными в [1, 3], удвоив при этом уровни значимости.

Вывод приближенного выражения для $Вер\{t^*_{\nu} \gg \beta^*\}$ приведен в [2] и также основан на пренебрежительно малой величине вероятности одновременного выхода двух результатов наблюдений за установленные границы по сравнению с вероятностью невыхода и вероятностью выхода одного результата наблюдения за эти границы.

ЛИТЕРАТУРА

1. Смирнов Н. В. Оценка максимального члена в ряду наблюдений. Доклады АН СССР, 1941, т. 33, № 5, с. 346—349.
2. Шор Я. Б., Бендерский А. М. Методы оценки анормальности результатов измерений. М., ААН, 1953.
3. Grubbs F. E. Sample Criteria for Testing Outlying Observations «Annals of Math Statistics». 1950, 21, № 1.
4. Бендерский А. М. О распределении модуля максимального отклонения от среднего в ряду наблюдений. Доклады АН СССР, 1952, т. XXXV, № 1, с. 5—8.
5. Микешина Н. Г. Выявление и исключение анормальных значений «Заводская лаборатория» 1966, т. XXII, № 3, с. 310—318.
6. Большев Л. Н., Смирнов Н. В. Таблицы математической статистики. М., «Наука», 1968.
7. Большев Л. Н. О критериях исключения резко выделяющихся наблюдений. Труды Института прикладной математики Тбилисского государственного университета, т. II, 1969, с. 159—177.
8. Ansombe F. Y. Rejection of outliers «Technometrics», 1960, 2, pp. 123—147.
9. B. De Finetli The Baysian approach to the rejection of outliers Proc. Fourth Berkeley Symp. Math Statist Prob. 1961, 1, pp, 199—210.
10. Ferguson F. E. On the rejection on of outliers. Proc. Fourth Berkeley Symp. Math. Statist Prob. 1961, 1, pp. 253—287.
11. Halperin M. Greenhouse, Corufield y-and Lalokar J. «Tables percentage point for maximum absolute deviate in normal Samples» J. Amer. Statist Assos». 1955, 50, pp. 185—195.
12. Karlin S. and Truax D. Slippage problems «Annals of Math. Statist». 31, pp. 296—324.
13. Kudo A. On the testing of outlying observations «Sanhya» . 1953—1957, 17, pp. 67—76.
14. Qneseuberru C. P. and David H. A. Some tests for outliers «Biometrika», 1961, 48, pp. 379—387.
15. Tykey Y. W. The future of data annalysis «Annals of Math, Statist». 1962, v. 33, pp. 1—67.

Редактор *И. И. Топильская*
Технический редактор *С. Ю. Миронова*
Корректор *О. Я. Гавриленко*

Сдано в набор 05. 03. 73 г. Подп. в печ. 22. 05. 73 г. 1,5 п. л. Тир. 12000

Издательство стандартов, Москва, Д-22, Новопресненский пер., 3
Калужская типография стандартов, ул. Московская, 256. Зак. 454