Designation: ~~E178 – 08~~ E178 – 16

An American National Standard

Standard Practice for Dealing With Outlying Observations[1]

This standard is issued under the fixed designation E178; the number immediately following the designation indicates the year of original adoption or, in the case of revision, the year of last revision. A number in parentheses indicates the year of last reapproval. A superscript epsilon (ε) indicates an editorial change since the last revision or reapproval.

1. Scope

1.1 This practice covers outlying observations in samples and how to test the statistical significance of ~~them. An outlying observation, or "outlier," is one that appears to deviate markedly from other members of the sample in which it occurs. In this connection, the following two alternatives are of interest:~~outliers.

~~1.1.1 An outlying observation may be merely an extreme manifestation of the random variability inherent in the data. If this is true, the value should be retained and processed in the same manner as the other observations in the sample.~~

~~1.1.2 On the other hand, an outlying observation may be the result of gross deviation from prescribed experimental procedure or an error in calculating or recording the numerical value. In such cases, it may be desirable to institute an investigation to ascertain the reason for the aberrant value. The observation may even actually be rejected as a result of the investigation, though not necessarily so. At any rate, in subsequent data analysis the outlier or outliers will be recognized as probably being from a different population than that of the other sample values.~~

1.2 ~~It is our purpose here to provide statistical rules that will lead the experimenter almost unerringly to look for causes of outliers when they really exist, and hence to decide whether alternative~~ The system of units for this standard ~~1.1.1 above, is not the more plausible hypothesis to accept, as compared to alternative~~ is not specified. Dimensional quantities ~~1.1.2,~~ in ~~order that the most appropriate action in further data analysis may be taken. The procedures covered herein apply primarily to the simplest kind of experimental data, that is, replicate measurements of some property of a given material, or observations in a supposedly single random sample. Nevertheless, the tests suggested do cover a wide enough range of cases in practice to have broad utility.~~the standard are presented only as illustrations of calculation methods. The examples are not binding on products or test methods treated.

1.3 *This standard does not purport to address all of the safety concerns, if any, associated with its use. It is the responsibility of the user of this standard to establish appropriate safety and health practices and determine the applicability of regulatory requirements prior to use.*

2. Referenced Documents

2.1 *ASTM Standards:*[2]

E456 Terminology Relating to Quality and Statistics

E2586 Practice for Calculating and Using Basic Statistics

3. Terminology

3.1 *Definitions:* The terminology defined in Terminology E456 applies to this standard unless modified herein.

3.1.1 *order statistic* $x_{(k)}$, *n*—value of the kth observed value in a sample after sorting by order of magnitude. (Practice E2586.)

[1] This practice is under the jurisdiction of ASTM Committee ~~D19~~E11 on ~~Water~~Quality and Statistics and is the direct responsibility of Subcommittee ~~D19.05~~E11.10 on ~~Inorganic Constituents in Water~~Sampling / Statistics.

Current edition approved June 1, 2016. Published ~~November 2008~~June 2016. Originally approved in 1961. Last previous edition approved in ~~2002~~2008 as ~~E178 – 02.~~E178 – 08. DOI: ~~10.1520/E0178-08.~~10.1520/E0178-16.

[2] For referenced ASTM standards, visit the ASTM website, www.astm.org, or contact ASTM Customer Service at service@astm.org. For *Annual Book of ASTM Standards* volume information, refer to the standard's Document Summary page on the ASTM website.

3.1.1.1 *Discussion—*

In this Practice, x_k is used to denote order statistics in place of $x_{(k)}$, to simplify the notation.

3.1.2 *outlier*—see **outlying observation**.

3.1.3 *outlying observation, n*—an observation extreme observation in either direction that appears to deviate markedly in value from other members of the sample in which it appears.

4. Significance and Use

4.1 When the experimenter is clearly aware that a gross deviation from prescribed experimental procedure has taken place, the resultant observation should be discarded, whether or not it agrees with the rest of the data and without recourse to statistical tests for outliers. If a reliable correction procedure, for example, for temperature, is available, the observation may sometimes be corrected and retained.

4.1 In many cases evidence for deviation from prescribed procedure will consist primarily of the discordant value itself. In such cases it is advisable to adopt a cautious attitude. Use of one of the criteria discussed below will sometimes permit a clear-cut decision to be made. In doubtful cases the experimenter's judgment will have considerable influence. When the experimenter cannot identify abnormal conditions, he should at least report the discordant values and indicate to what extent they have been used in the analysis of the data.An outlying observation, or "outlier," is an extreme one in either direction that appears to deviate markedly from other members of the sample in which it occurs.

4.2 Thus, for purposes of orientation relative to the over-all problem of experimentation, our position on the matter of screening samples for outlying observations is precisely the following: Statistical rules test the null hypothesis of no outliers against the alternative of one or more actual outliers. The procedures covered were developed primarily to apply to the simplest kind of experimental data, that is, replicate measurements of some property of a given material or observations in a supposedly random sample.

4.3.1 *Physical Reason Known or Discovered for Outlier(s):*
4.3.1.1 Reject observation(s).
4.3.1.2 Correct observation(s) on physical grounds.
4.3.1.3 Reject it (them) and possibly take additional observation(s).
4.3.2 *Physical Reason Unknown—Use Statistical Test:*
4.3.2.1 Reject observation(s).
4.3.2.2 Correct observation(s) statistically.
4.3.2.3 Reject it (them) and possibly take additional observation(s).
4.3.2.4 Employ truncated-sample theory for censored observations.

4.3 TheA statistical test may always be used to support a judgment that a physical reason does actually exist for an outlier, or the statistical criterion may be used routinely as a basis to initiate action to find a physical cause.

5. Procedure

5.1 In dealing with an outlier, the following alternatives should be considered:

5.1.1 An outlying observation might be the result of gross deviation from prescribed experimental procedure or an error in calculating or recording the numerical value. When the experimenter is clearly aware that a deviation from prescribed experimental procedure has taken place, the resultant observation should be discarded, whether or not it agrees with the rest of the data and without recourse to statistical tests for outliers. If a reliable correction procedure is available, the observation may sometimes be corrected and retained.

5.1.2 An outlying observation might be merely an extreme manifestation of the random variability inherent in the data. If this is true, the value should be retained and processed in the same manner as the other observations in the sample. Transformation of data or using methods of data analysis designed for a non-normal distribution might be appropriate.

5.1.3 Test units that give outlying observations might be of special interest. If this is true, once identified they should be segregated for more detailed study.

5.2 In many cases, evidence for deviation from prescribed procedure will consist primarily of the discordant value itself. In such cases it is advisable to adopt a cautious attitude. Use of one of the criteria discussed below will sometimes permit a clearcut decision to be made.

5.2.1 When the experimenter cannot identify abnormal conditions, he should report the discordant values and indicate to what extent they have been used in the analysis of the data.

5.3 Thus, as part of the over-all process of experimentation, the process of screening samples for outlying observations and acting on them is the following:

5.3.1 *Physical Reason Known or Discovered for Outlier(s):*
5.3.1.1 Reject observation(s) and possibly take additional observation(s).
5.3.1.2 Correct observation(s) on physical grounds.
5.3.2 *Physical Reason Unknown—Use Statistical Test:*
5.3.2.1 Reject observation(s) and possibly take additional observation(s).
5.3.2.2 Transform observation(s) to improve fit to a normal distribution.
5.3.2.3 Use estimation appropriate for non-normal distributions.

5.3.2.4 Segregate samples for further study.

6. Basis of Statistical Criteria for Outliers

6.1 ~~There are a number of criteria for testing outliers. In all of these,~~ In testing outliers, the doubtful observation is included in the calculation of the numerical value of a sample criterion (or statistic), which is then compared with a critical value based on the theory of random sampling to determine whether the doubtful observation is to be retained or rejected. The critical value is that value of the sample criterion which would be exceeded by chance with some specified (small) probability on the assumption that all the observations did indeed constitute a random sample from a common system of causes, a single parent population, distribution or universe. The specified small probability is called the "significance level" or "percentage point" and can be thought of as the risk of erroneously rejecting a good observation. ~~It becomes clear, therefore, that if there exists~~ If a real shift or change in the value of an observation ~~that~~ arises from nonrandom causes (human error, loss of calibration of instrument, change of measuring instrument, or even change of time of measurements, ~~etc.),~~ and so forth), then the observed value of the sample criterion used ~~would~~will exceed the "critical value" based on random-sampling theory. Tables of critical values are usually given for several different significance ~~levels, for example, 5 %, 1 %. For statistical tests of outlying observations, it is generally recommended that a low significance level, such as 1 %, be used and that significance levels greater than 5 % should not be common practice.~~levels. In particular for this Practice, significance levels 10, 5, and 1% are used.

NOTE 1—In this practice, we will usually illustrate the use of the ~~5 %~~5% significance level. Proper choice of level in probability depends on the particular problem and just what may be involved, along with the risk that one is willing to take in rejecting a good observation, that is, if the null-hypothesis stating "all observations in the sample come from the same normal population" may be assumed correct.

6.2 ~~It should be pointed out that almost~~ Almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population or distribution. The null hypothesis that we are testing in every case is that all observations in the sample come from the same normal population. In choosing an appropriate alternative hypothesis (one or more outliers, separated or bunched, on same side or different sides, and so forth) it is useful to plot the data as shown in the dot diagrams of the figures. When the data are not normally or approximately normally distributed, the probabilities associated with these tests will be different. ~~Until such time as criteria not sensitive to the normality assumption are developed, the~~ The experimenter is cautioned against interpreting the probabilities too literally.

6.3 Although our primary interest here is that of detecting outlying observations, ~~we remark that~~ some of the statistical criteria presented may also be used to test the hypothesis of normality or that the random sample taken ~~did~~ come from a normal or Gaussian population. The end result is for all practical purposes the same, that is, we really wish to know whether we ought to proceed as if we have in hand a sample of homogeneous normal observations.

6.4 One should distinguish between data to be used to estimate a central value from data to be used to assess variability. When the purpose is to estimate a standard deviation, it might be seriously underestimated by dropping too many "outlying" observations.

7. Recommended Criteria for Single Samples

~~6.1 Let the sample of n observations be denoted in order of increasing magnitude by $x_1 \le x_2 \le x_3 \le \ldots \le x_n$. Let x_n be the doubtful value, that is the largest value. The test criterion, T_n, recommended here for a single outlier is as follows:~~

$$T_n = (x_n - \bar{x})/s \quad (1)$$

~~where:~~

~~$\bar{x}$ = arithmetic average of all n values, and~~
~~s = estimate of the population standard deviation based on the sample data, calculated as follows:~~

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - n\cdot\bar{x}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2/n}{n-1}}$$

~~If x_1 rather than x_n is the doubtful value, the criterion is as follows:~~

$$T_1 = (\bar{x} - x_1)/s \quad (2)$$

~~The critical values for either case, for the 1 and 5 % levels of significance, are given in Table 1. Table 1 and the following tables give the "one-sided" significance levels. In the previous tentative recommended practice (1961), the tables listed values of significance levels double those in the present practice, since it was considered that the experimenter would test either the lowest or the highest observation (or both) for statistical significance. However, to be consistent with actual practice and in an attempt to avoid further misunderstanding, single-sided significance levels are tabulated here so that both viewpoints can be represented.~~

7.1 *Criterion for a Single Outlier*—~~The hypothesis that we are testing in every case is that all observations in the sample come from the same normal population. Let us adopt, for example, a significance level of 0.05. If we are interested~~ Let the sample of ~~*only*~~n ~~in outliers that occur on the~~ observations be denoted in order ~~high side, we should always use the statistic $T_n = (x_n - \bar{x})/s$~~

TABLE 1 Critical Values for *T* (One-Sided Test) When Standard Deviation is Calculated from the Same Sample[A]

Number of Observations, *n*	Upper 0.1 % Significance Level	Upper 0.5 % Significance Level	Upper 1 % Significance Level	Upper 2.5 % Significance Level	Upper 5 % Significance Level	Upper 10 % Significance Level
3	1.155	1.155	1.155	1.155	1.153	1.148
4	1.499	1.496	1.492	1.481	1.463	1.425
5	1.780	1.764	1.749	1.715	1.672	1.602
6	2.011	1.973	1.944	1.887	1.822	1.729
7	2.201	2.139	2.097	2.020	1.938	1.828
8	2.358	2.274	2.221	2.126	2.032	1.909
9	2.492	2.387	2.323	2.215	2.110	1.977
10	2.606	2.482	2.410	2.290	2.176	2.036
11	2.705	2.564	2.485	2.355	2.234	2.088
12	2.791	2.636	2.550	2.412	2.285	2.134
13	2.867	2.699	2.607	2.462	2.331	2.175
14	2.935	2.755	2.659	2.507	2.371	2.213
15	2.997	2.806	2.705	2.549	2.409	2.247
16	3.052	2.852	2.747	2.585	2.443	2.279
17	3.103	2.894	2.785	2.620	2.475	2.309
18	3.149	2.932	2.821	2.651	2.504	2.335
19	3.191	2.968	2.854	2.681	2.532	2.361
20	3.230	3.001	2.884	2.709	2.557	2.385
21	3.266	3.031	2.912	2.733	2.580	2.408
22	3.300	3.060	2.939	2.758	2.603	2.429
23	3.332	3.087	2.963	2.781	2.624	2.448
24	3.362	3.112	2.987	2.802	2.644	2.467
25	3.389	3.135	3.009	2.822	2.663	2.486
26	3.415	3.157	3.029	2.841	2.681	2.502
27	3.440	3.178	3.049	2.859	2.698	2.519
28	3.464	3.199	3.068	2.876	2.714	2.534
29	3.486	3.218	3.085	2.893	2.730	2.549
30	3.507	3.236	3.103	2.908	2.745	2.563
31	3.528	3.253	3.119	2.924	2.759	2.577
32	3.546	3.270	3.135	2.938	2.773	2.591
33	3.565	3.286	3.150	2.952	2.786	2.604
34	3.582	3.301	3.164	2.965	2.799	2.616
35	3.599	3.316	3.178	2.979	2.811	2.628
36	3.616	3.330	3.191	2.991	2.823	2.639
37	3.631	3.343	3.204	3.003	2.835	2.650
38	3.646	3.356	3.216	3.014	2.846	2.661
39	3.660	3.369	3.228	3.025	2.857	2.671
40	3.673	3.381	3.240	3.036	2.866	2.682
41	3.687	3.393	3.251	3.046	2.877	2.692
42	3.700	3.404	3.261	3.057	2.887	2.700
43	3.712	3.415	3.271	3.067	2.896	2.710
44	3.724	3.425	3.282	3.075	2.905	2.719
45	3.736	3.435	3.292	3.085	2.914	2.727
46	3.747	3.445	3.302	3.094	2.923	2.736
47	3.757	3.455	3.310	3.103	2.931	2.744
48	3.768	3.464	3.319	3.111	2.940	2.753
49	3.779	3.474	3.329	3.120	2.948	2.760
50	3.789	3.483	3.336	3.128	2.956	2.768
51	3.798	3.491	3.345	3.136	2.964	2.775
52	3.808	3.500	3.353	3.143	2.971	2.783
53	3.816	3.507	3.361	3.151	2.978	2.790
54	3.825	3.516	3.368	3.158	2.986	2.798
55	3.834	3.524	3.376	3.166	2.992	2.804
56	3.842	3.531	3.383	3.172	3.000	2.811
57	3.851	3.539	3.391	3.180	3.006	2.818
58	3.858	3.546	3.397	3.186	3.013	2.824
59	3.867	3.553	3.405	3.193	3.019	2.831
60	3.874	3.560	3.411	3.199	3.025	2.837
61	3.882	3.566	3.418	3.205	3.032	2.842
62	3.889	3.573	3.424	3.212	3.037	2.849

TABLE 1 *Continued*

Number of Observations, *n*	Upper 0.1 % Significance Level	Upper 0.5 % Significance Level	Upper 1 % Significance Level	Upper 2.5 Significan Level
63	3.896	3.579	3.430	3.218
64	3.903	3.586	3.437	3.224
65	3.910	3.592	3.442	3.230
66	3.917	3.598	3.449	3.235
67	3.923	3.605	3.454	3.241
68	3.930	3.610	3.460	3.246
69	3.936	3.617	3.466	3.252
70	3.942	3.622	3.471	3.257
71	3.948	3.627	3.476	3.262
72	3.954	3.633	3.482	3.267
73	3.960	3.638	3.487	3.272
74	3.965	3.643	3.492	3.278
75	3.971	3.648	3.496	3.282
76	3.977	3.654	3.502	3.287
77	3.982	3.658	3.507	3.291
78	3.987	3.663	3.511	3.297
79	3.992	3.669	3.516	3.301
80	3.998	3.673	3.521	3.305
81	4.002	3.677	3.525	3.309
82	4.007	3.682	3.529	3.315
83	4.012	3.687	3.534	3.319
84	4.017	3.691	3.539	3.323
85	4.021	3.695	3.543	3.327
86	4.026	3.699	3.547	3.331
87	4.031	3.704	3.551	3.335
88	4.035	3.708	3.555	3.339
89	4.039	3.712	3.559	3.343
90	4.044	3.716	3.563	3.347
91	4.049	3.720	3.567	3.350
92	4.053	3.725	3.570	3.355
93	4.057	3.728	3.575	3.358
94	4.060	3.732	3.579	3.362
95	4.064	3.736	3.582	3.365
96	4.069	3.739	3.586	3.369
97	4.073	3.744	3.589	3.372
98	4.076	3.747	3.593	3.377
99	4.080	3.750	3.597	3.380
100	4.084	3.754	3.600	3.383
101	4.088	3.757	3.603	3.386
102	4.092	3.760	3.607	3.390
103	4.095	3.765	3.610	3.393
104	4.098	3.768	3.614	3.397
105	4.102	3.771	3.617	3.400
106	4.105	3.774	3.620	3.403
107	4.109	3.777	3.623	3.406
108	4.112	3.780	3.626	3.409
109	4.116	3.784	3.629	3.412
110	4.119	3.787	3.632	3.415
111	4.122	3.790	3.636	3.418
112	4.125	3.793	3.639	3.422
113	4.129	3.796	3.642	3.424
114	4.132	3.799	3.645	3.427
115	4.135	3.802	3.647	3.430
116	4.138	3.805	3.650	3.433
117	4.141	3.808	3.653	3.435
118	4.144	3.811	3.656	3.438
119	4.146	3.814	3.659	3.441
120	4.150	3.817	3.662	3.444
121	4.153	3.819	3.665	3.447
122	4.156	3.822	3.667	3.450
123	4.159	3.824	3.670	3.452

TABLE 1 *Continued*

Number of Observations, *n*	Upper 0.1 % Significance Level	Upper 0.5 % Significance Level	Upper 1 % Significance Level	Upper 2.5 % Significance Level	Upper 5 % Significance Level	Upper 10 % Significance Level
124	4.161	3.827	3.672	3.455	3.279	3.089
125	4.164	3.831	3.675	3.457	3.281	3.092
126	4.166	3.833	3.677	3.460	3.284	3.095
127	4.169	3.836	3.680	3.462	3.286	3.097
128	4.173	3.838	3.683	3.465	3.289	3.100
129	4.175	3.840	3.686	3.467	3.291	3.102
130	4.178	3.843	3.688	3.470	3.294	3.104
131	4.180	3.845	3.690	3.473	3.296	3.107
132	4.183	3.848	3.693	3.475	3.298	3.109
133	4.185	3.850	3.695	3.478	3.302	3.112
134	4.188	3.853	3.697	3.480	3.304	3.114
135	4.190	3.856	3.700	3.482	3.306	3.116
136	4.193	3.858	3.702	3.484	3.309	3.119
137	4.196	3.860	3.704	3.487	3.311	3.122
138	4.198	3.863	3.707	3.489	3.313	3.124
139	4.200	3.865	3.710	3.491	3.315	3.126
140	4.203	3.867	3.712	3.493	3.318	3.129
141	4.205	3.869	3.714	3.497	3.320	3.131
142	4.207	3.871	3.716	3.499	3.322	3.133
143	4.209	3.874	3.719	3.501	3.324	3.135
144	4.212	3.876	3.721	3.503	3.326	3.138
145	4.214	3.879	3.723	3.505	3.328	3.140
146	4.216	3.881	3.725	3.507	3.331	3.142
147	4.219	3.883	3.727	3.509	3.334	3.144

$T_n = (x_n - \bar{x})/s$

$$\sqrt{\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - n\cdot\bar{x}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}x_i^2 - \left(\sum_{i=1}^{n}x_i\right)^2/n}{n-1}}$$

$T_1 = [(\bar{x} - x_1)/s]\; x_1 \leq x_2 \leq \ldots \leq x_n$

TABLE 1 Critical Values for *T* (One-Sided Test) When Standard Deviation is Calculated from the Same Sample[A]

Number of Observations, *n*	Upper 10% Significance Level	Upper 5% Significance Level	Upper 1% Significance Level
3	1.1484	1.1531	1.1546
4	1.4250	1.4625	1.4925
5	1.602	1.672	1.749
6	1.729	1.822	1.944
7	1.828	1.938	2.097
8	1.909	2.032	2.221
9	1.977	2.110	2.323
10	2.036	2.176	2.410
11	2.088	2.234	2.485
12	2.134	2.285	2.550
13	2.175	2.331	2.607
14	2.213	2.371	2.659
15	2.247	2.409	2.705
16	2.279	2.443	2.747
17	2.309	2.475	2.785
18	2.335	2.504	2.821
19	2.361	2.532	2.854
20	2.385	2.557	2.884
21	2.408	2.580	2.912
22	2.429	2.603	2.939
23	2.448	2.624	2.963
24	2.467	2.644	2.987
25	2.486	2.663	3.009
26	2.502	2.681	3.029
27	2.519	2.698	3.049
28	2.534	2.714	3.068
29	2.549	2.730	3.085
30	2.563	2.745	3.103
35	2.628	2.811	3.178
40	2.682	2.866	3.240
45	2.727	2.914	3.292
50	2.768	2.956	3.336

[A] Values of T are taken from RefGrubbs (1)., Table 1)[3]. All values have been adjusted for division by $n - 1$ instead of n in calculating s. Use Ref. (1) for higher sample sizes up to n = 147.

and take as critical value the 0.05 point of Table 1. On the other hand, if we are interested increasing magnitude by *only in outliers occurring on the low side, we would always use the statistic $T_1 = (\bar{x} - x_1)/s$ and again take as a critical value the 0.05 point of* Table 1. *Suppose, however, that we are interested in outliers occurring on either side, but do not believe that outliers can occur on both sides simultaneously. We might, for example, believe that at some time during the experiment something possibly happened to cause an extraneous variation on the high side or on the low side, but that it was very unlikely that two or more such events could have occurred, one being an extraneous variation on the high side and the other an extraneous variation on the low side. With this point of view we should use the statistic $T_n = (x_{n1} - \leq \bar{x}\, x_2)/ \leq s x_3$* or the statistic ≤ ... ≤ $T x_{n,T} = (\bar{x} -$ Let the largest value, $x_{n,T})/$ sbe whichever is larger. If in this instance we use the 0.05the doubtful value, that is the largest value. The test criterion, *point*T_n, offor Table 1 as our a single outlier is as follows:

$$T_n = (x_n - \bar{x})/s \qquad (1)$$

critical

where:

$\bar{x}$ = arithmetic average of all n values, and
s = estimate of the population standard deviation based on the sample data, calculated as follows:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2 / n}{n-1}}$$

value, the true significance level would be twice 0.05 or 0.10. If we wish a significance level of 0.05 and not 0.10, we must in this case use as a critical value the 0.025 point of Table 1. Similar considerations apply to the other tests given below.

If x_1 rather than x_n is the doubtful value, the criterion is as follows:

$$T_1 = (\bar{x} - x_1)/s \qquad (2)$$

The critical values for either case, for the 1, 5, and 10% levels of significance, are given in Table 1.

7.1.1 The test criterion T_n can be equated to the Student's t test statistic for equality of means between a population with one observation x_n and another with the remaining observations x_1, ... , x_{n-1}, and the critical value of T_n for significance level α can be approximated using the α/n percentage point of Student's t with n-2 degrees of freedom. The approximation is exact for small enough values of α, depending on n, and otherwise a slight overestimate unless both α and n are large.

$$T_n(\alpha) \leq \frac{t_{\alpha/n,n-2}}{\sqrt{1 + \frac{n t^2_{\alpha/n,n-2} - 1}{(n-1)^2}}}$$

7.1.2 To test outliers on the *high side*, use the statistic $T_n = (x_n - \bar{x})/s$ and take as critical value the 0.05 point of Table 1. To test outliers on the *low side*, use the statistic $T_1 = (\bar{x} - x_1)/s$ and again take as a critical value the 0.05 point of Table 1. If we are interested in outliers occurring on *either side*, use the statistic $T_n = (x_n - \bar{x})/s$ or the statistic $T_1 = (\bar{x} - x_1)/s$ whichever is larger. If in this instance we use the 0.05 point of Table 1 as our critical value, the true significance level would be twice 0.05 or 0.10. Similar considerations apply to the other tests given below.

7.1.3 *Example 1*—As an illustration of the use of T_{n} and Table 1, consider the following ten observations on breaking strength (in pounds) of 0.104-in. hard-drawn copper wire: 568, 570, 570, 570, 572, 572, 572, 578, 584, 596. See Fig. 1. The doubtful observation is the high value, x_{10} = 596. = 596. Is the value of 596 significantly high? The mean is $\bar{x}$ = 575.2 = 575.2 and the estimated standard deviation is s = 8.70. = 8.70. We compute

$$T_{10} = (596 - 575.2)/8.70 = 2.39 \qquad (3)$$

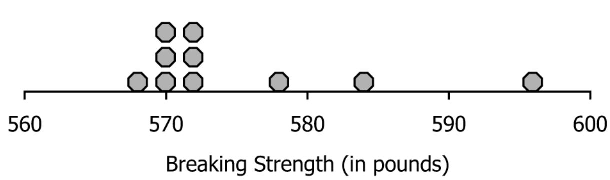

FIG. 1 Ten Observations of Breaking Strength from Example 1 in 6.2.1

From Table 1, for ~~$n = 10$,~~ $n = 10$, note that a T_{10} as large as 2.39 would occur by chance with probability less than 0.05. In fact, so large a value would occur by chance not much more often than ~~1 %~~1% of the time. Thus, the weight of the evidence is against the doubtful value having come from the same population as the others (assuming the population is normally distributed). Investigation of the doubtful value is therefore indicated.

7.2 *Dixon Criteria for a Single Outlier*—An alternative system, the Dixon ~~criteria, based entirely on ratios of differences between the observations is described in the literature~~ criteria (**2**)[3] ~~and~~, based entirely on ratios of differences between the observations may be used in cases where it is desirable to avoid calculation of s or where quick judgment is called for. For the Dixon test, the sample criterion or statistic changes with sample size. Table 2 gives the appropriate statistic to calculate and also gives the critical values of the statistic for the 1, 5, and ~~10 %~~10% levels of significance. In most situations, the Dixon criteria is less powerful at detecting an outlier than the criterion given in section 7.1.

7.2.1 *Example 2*—As an illustration of the use of Dixon's test, consider again the observations on breaking strength given in Example ~~1, and suppose that a large number of such samples had to be screened quickly for outliers and it was judged too time-consuming to compute~~ 1. ~~s.~~Table 2 indicates use of

~~$$r_{11} = (x_n - x_{n-1})/(x_n - x_2) \quad (4)$$~~

$$r_{11} = (x_n - x_{n-1})/(x_n - x_2) \quad (4)$$

Thus, for $n = 10$,

$$r_{11} = (x_{10} - x_9)/(x_{10} - x_2) \quad (5)$$

[3] The boldface numbers in parentheses refer to the list of references at the end of this practice.

TABLE 2 Dixon Criteria for Testing of Extreme Observation (Single Sample)[A]

n	Criterion	Significance Level (One-Sided Test) 10 percent	5 percent	1 percent
~~3~~	~~$r_{10} = (x_2 - x_1)/(x_n - x_1)$ if smallest value is suspected;~~	~~0.886~~	~~0.941~~	~~0.988~~
3	$r_{10} = (x_2 - x_1)/(x_n - x_1)$ if smallest value is suspected;	0.886	0.941	0.988
~~4~~	~~$= (x_n - x_{n-1})/(x_n - x_1)$ if largest value is suspected~~	~~0.679~~	~~0.765~~	~~0.889~~
4	$= (x_n - x_{n-1})/(x_n - x_1)$ if largest value is suspected	0.679	0.765	0.889
5		0.557	0.642	0.780
6		0.482	0.560	0.698
7		0.434	0.507	0.637
~~8~~	~~$r_{11} = (x_2 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;~~	~~0.479~~	~~0.554~~	~~0.683~~
~~9~~	~~$= (x_n - x_{n-1})/(x_n - x_2)$ if largest value is suspected.~~	~~0.441~~	~~0.512~~	~~0.635~~
~~10~~		~~0.409~~	~~0.477~~	~~0.597~~
8	$r_{11} = (x_2 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;	0.479	0.554	0.683
9	$= (x_n - x_{n-1})/(x_n - x_2)$ if largest value is suspected.	0.441	0.512	0.635
10		0.409	0.477	0.597
~~11~~	~~$r_{21} = (x_3 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;~~	~~0.517~~	~~0.576~~	~~0.679~~
~~12~~	~~$= (x_n - x_{n-2})/(x_n - x_2)$ if largest value is suspected.~~	~~0.490~~	~~0.546~~	~~0.642~~
~~13~~		~~0.467~~	~~0.521~~	~~0.615~~
11	$r_{21} = (x_3 - x_1)/(x_{n-1} - x_1)$ if smallest value is suspected;	0.517	0.576	0.679
12	$= (x_n - x_{n-2})/(x_n - x_2)$ if largest value is suspected.	0.490	0.546	0.642
13		0.467	0.521	0.615
14	$r_{22} = (x_3 - x_1)/(x_{n-2} - x_1)$ if smallest value is suspected;	0.492	0.546	0.641
~~15~~	~~$= (x_n - x_{n-2})/(x_n - x_3)$ if largest value is suspected.~~	~~0.472~~	~~0.525~~	~~0.616~~
15	$= (x_n - x_{n-2})/(x_n - x_3)$ if largest value is suspected.	0.472	0.525	0.616
16		0.454	0.507	0.595
17		0.438	0.490	0.577
18		0.424	0.475	0.561
19		0.412	0.462	0.547
20		0.401	0.450	0.535
21		0.391	0.440	0.524
22		0.382	0.430	0.514
23		0.374	0.421	0.505
24		0.367	0.413	0.497
25		0.360	0.406	0.489
26		0.354	0.399	0.486
27		0.348	0.393	0.475
28		0.342	0.387	0.469
29		0.337	0.381	0.463
30		0.332	0.376	0.457
35		0.311	0.354	0.431
40		0.295	0.337	0.412
45		0.283	0.323	0.397
50		0.272	0.312	0.384

[A] $x_1 \leq x_2 \leq \ldots \leq x_{n:n}$. ~~(See Ref~~ Original Table in Dixon (**2**)~~),~~, Appendix). ~~Appendix.)~~Critical values updated by calculations by Bohrer (**3**) and Verma-Ruiz (**4**).

For the measurements of breaking strength above,

$$r_{11} = (596 - 584)/(596 - 570) = 0.462 \quad (6)$$

For the measurements of breaking strength above,

$$r_{11} = (596 - 584)/(596 - 570) = 0.462 \quad (6)$$

which is a little less than 0.477,0.478, the 5 %5% critical value for $n = 10. = 10.$ Under the Dixon criterion, we should therefore not consider this observation as an outlier at the 5 %5% level of significance. These results illustrate how borderline cases may be accepted under one test but rejected under another. It should be remembered, however, that the T-statistic discussed above is the best one to use for the single-outlier case, and final statistical judgment should be based on it. See Ferguson (3, 4).

6.3.2 Further examination of the sample observations on breaking strength of hand-drawn copper wire indicates that none of the other values need testing.

NOTE 2—With experience we may usually just look at the sample values to observe if an outlier is present. However, strictly speaking the statistical test should be applied to all samples to guarantee the significance levels used. Concerning "multiple" tests on a single sample, we comment on this below.

7.3 *Recursive Testing for Multiple Outliers in Univariate Samples*—A test equivalent to For testing multiple outliers in a sample, recursive application of a test for a single outlier may be used. In recursive testing, a test for an outlier, Tx_{n1} (oror $Tx_{n,T}$) based on the sample sum of squared deviations from the mean for all the observations and the sum of squared deviations omitting the "outlier" is given by Grubbs is first conducted. If this is found to be significant, then the test is repeated, omitting the outlier found, to test the point on the opposite side of the sample, or an additional point on the same side. The performance of most tests for single outliers is affected by masking, where the probability of detecting an outlier using a test for a single outlier is reduced when there are two or more outliers. Therefore, the recommended procedure is to use a (criterion5). designed to test for multiple outliers, using recursive testing to investigate after the initial criterion is significant.

6.5 The next type of problem to consider is the case where we have the possibility of two outlying observations, the least and the greatest observation in a sample. (The problem of testing the two highest or the two lowest observations is considered below.) In testing the least and the greatest observations simultaneously as probable outliers in a sample, we use the ratio of sample range to sample standard deviation test of David, Hartley, and Pearson (**6**). The significance levels for this sample criterion are given in Table 3. Alternatively, the largest residuals test of Tietjen and Moore (**7**) could be used. An example in astronomy follows.

6.5.1 *Example 3*—There is one rather famous set of observations that a number of writers on the subject of outlying observations have referred to in applying their various tests for "outliers." This classic set consists of a sample of 15 observations of the vertical semidiameters of Venus made by Lieutenant Herndon in 1846 (**8**). In the reduction of the observations, Prof. Pierce assumed two unknown quantities and found the following residuals which have been arranged in ascending order of magnitude:

−1.40 in.	−0.24	−0.05	0.18	0.48
−0.44	−0.22	−0.06	0.20	0.63
−0.30	−0.13	−0.10	0.39	1.01

See Fig. 2.

The deviations − 1.40 and 1.01 appear to be outliers. Here the suspected observations lie at each end of the sample. Much less work has been accomplished for the case of outliers at both ends of the sample than for the case of one or more outliers at only one end of the sample. This is not necessarily because the "one-sided" case occurs more frequently in practice but because "two-sided'' tests are much more difficult to deal with. For a high and a low outlier in a single sample, we give two procedures below, the first being a combination of tests, and the second a single test of Tietjen and Moore (**7**) which may have nearly optimum properties. For optimum procedures when there is an independent estimate at hand, s^2 or σ^2, see (**9**).

6.6 For the observations on the semi-diameter of Venus given above, all the information on the measurement error is contained in the sample of 15 residuals. In cases like this, where no independent estimate of variance is available (that is, we still have the single sample case), a useful statistic is the ratio of the range of the observations to the sample standard deviation:

$$w/s = (x_n - x_1)/s \quad (7)$$

where:

$$s = \sqrt{\sum[(x_i - \bar{x})^2/(n-1)]} \quad (8)$$

If x_n is about as far above the mean, $\bar{x}$, as x_1 is below $\bar{x}$, and if w/s exceeds some chosen critical value, then one would conclude that *both* the doubtful values are outliers. If, however, x_1 and x_n are displaced from the mean by different amounts, some further test would have to be made to decide whether to reject as outlying only the lowest value or only the highest value or both the lowest and highest values.

7.4 For this example the mean of the deviations is $\bar{x} = 0.018$, $s = 0.551$, and

$$w/s = [1.01 - (-1.40)]/0.551 = 2.41/0.551 = 4.374 \quad (9)$$

Criterion for Two Outliers on Opposite Sides of a Sample—

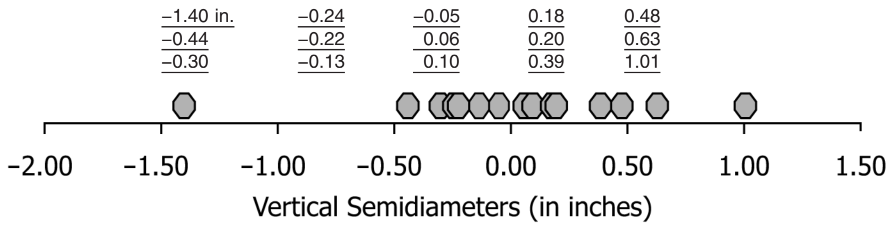

FIG. 2 Fifteen Residuals from the Semidiameters of Venus from Example 3 ~~in 6.5.1~~

~~From~~In ~~Table 3~~ ~~for~~ testing ~~n = 15, we see that the value of~~ the least ~~w/s = 4.374 falls between the critical values for the 1 and 5 % levels, so if the test were being run at the 5 % level of significance, we would conclude that this sample contains one or more outliers. The lowest measurement, − 1.40 in., is 1.418 below the sample mean, and the highest measurement, 1.01 in., is 0.992 above the mean. Since these extremes are not symmetric about the mean, either~~ and the greatest observations simultaneously as probable outliers in a sample, use the ratio of sample range to sample standard deviation test of David, Hartley, and Pearson ~~**both**~~(**5**). ~~extremes are outliers, or else only − 1.40 is an outlier. That − 1.40 is an outlier can be verified by use of the T_1 statistic. We have~~

~~$$T_1 = (\bar{x} - x_1)/s = (0.018 - (-1.40))/0.551 = 2.574 \quad (7)$$~~

$$w/s = (x_n - x_1)/s \quad (7)$$

~~This value is greater than the critical value for the 5 % level, 2.409 from~~The significance levels for this sample criterion are given ~~Table 1,~~in Table 3~~so we reject − 1.40. Since we have decided that − 1.40 should be rejected, we use the remaining 14 observations and test the~~. Alternatively, the largest residuals test of Tietjen and Moore (7.5 ~~upper extreme 1.01, either with the criterion~~) could be used.

TABLE 3 Critical Values (One-Sided Test) for *w/s* (Ratio of Range to Sample Standard Deviation)[A]

Number of Observations, *n*	5 Percent Significance Level	1 Percent Significance Level	0.5 Percent Significance Level
3	2.00	2.00	2.00
4	2.43	2.44	2.45
5	2.75	2.80	2.81
6	3.01	3.10	3.12
7	3.22	3.34	3.37
8	3.40	3.54	3.58
9	3.55	3.72	3.77
10	3.68	3.88	3.94
11	3.80	4.01	4.08
12	3.91	4.13	4.21
13	4.00	4.24	4.32
14	4.09	4.34	4.43
15	4.17	4.43	4.53
16	4.24	4.51	4.62
17	4.31	4.59	4.69
18	4.38	4.66	4.77
19	4.43	4.73	4.84
20	4.49	4.79	4.91
30	4.89	5.25	5.39
40	5.15	5.54	5.69
50	5.35	5.77	5.91
60	5.50	5.93	6.09
80	5.73	6.18	6.35
100	5.90	6.36	6.54
150	6.18	6.64	6.84
200	6.38	6.85	7.03
500	6.94	7.42	7.60
1000	7.33	7.80	7.99

TABLE 3 Critical Values[A] (One-Sided Test) for *w/s* (Ratio of Range to Sample Standard Deviation)

Number of Observations, *n*	10 Percent Significance Level	5 Percent Significance Level	1 Percent Significance Level
3	1.9973	1.9993	2.0000
4	2.409	2.429	2.445
5	2.712	2.755	2.803
6	2.949	3.012	3.095
7	3.143	3.222	3.338
8	3.308	3.399	3.543
9	3.449	3.552	3.720
10	3.574	3.685	3.875
11	3.684	3.803	4.011
12	3.782	3.909	4.133
13	3.871	4.005	4.244
14	3.952	4.092	4.344
15	4.025	4.171	4.435
16	4.093	4.244	4.519
17	4.156	4.311	4.597
18	4.214	4.374	4.669
19	4.269	4.433	4.736
20	4.320	4.487	4.799
21	4.368	4.539	4.858
22	4.413	4.587	4.913
23	4.456	4.633	4.965
24	4.497	4.676	5.015
25	4.535	4.717	5.061
26	4.572	4.756	5.106
27	4.607	4.793	5.148
28	4.641	4.829	5.188
29	4.673	4.863	5.226
30	4.704	4.895	5.263
35	4.841	5.040	5.426
40	4.957	5.162	5.561
45	5.057	5.265	5.674
50	5.144	5.356	5.773

$$T_n = (x_n - \bar{x})/s \quad (11)$$