DeweyBurdPubEm Resource

From: Burrows, Ronald
Sent: Wednesday, August 25, 2010 10:23 AM
To: DeweyBurdHrgFile Resource
Subject: ASTM Standard Dealing with Ouliers.pdf - Adobe Reader
Attachments: ASTM Standard Dealing with Ouliers.pdf

Hearing Identifier: Powertech_Uranium_Dewey_Burdock_LA_Public
Email Number: 118

Mail Envelope Properties (44CD2E65B0FF0E499CB32BC30CF781F0108E913D76)

Subject: ASTM Standard Dealing with Ouliers.pdf - Adobe Reader
Sent Date: 8/25/2010 10:23:28 AM
Received Date: 8/25/2010 10:24:09 AM
From: Burrows, Ronald

Created By: Ronald.Burrows@nrc.gov

Recipients:
"DeweyBurdHrgFile Resource" <DeweyBurdHrgFile.Resource@nrc.gov>
Tracking Status: None

Post Office: HQCLSTR01.nrc.gov

Files	**Size**	**Date & Time**
MESSAGE	3	8/25/2010 10:24:09 AM
ASTM Standard Dealing with Ouliers.pdf	179133	

Options
Priority: Standard
Return Notification: No
Reply Requested: No
Sensitivity: Normal
Expiration Date:
Recipients Received:

Designation: E 178 – 02

An American National Standard

Standard Practice for Dealing With Outlying Observations[1]

This standard is issued under the fixed designation E 178; 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:

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 1.1.1 above, is not the more plausible hypothesis to accept, as compared to alternative 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.

2. Referenced Documents

2.1 *ASTM Standards:*

E 456 Terminology Relating to Quality and Statistics[2]

3. Terminology

3.1 *Definitions:* The terminology defined in Terminology E 456 applies to this standard unless modified herein.

3.1.1 *outlier*—see **outlying observation**.

3.1.2 *outlying observation*, *n*—an observation 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.

4.2 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 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.

4.3 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:

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.

[1] This practice is under the jurisdiction of ASTM Committee E11 on Statistical Methods and is the direct responsibility of Subcommittee E11.30 on Data Analysis.
Current edition approved May 10, 2002. Published July 2002. Originally published as E 178 – 61 T. Last previous edition E 178 – 00.

[2] *Annual Book of ASTM Standards,* Vol 14.02.

4.4 The 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. Basis of Statistical Criteria for Outliers

5.1 There are a number of criteria for testing outliers. In all of these, 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 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.), then the observed value of the sample criterion used would 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.

NOTE 1—In this practice, we will usually illustrate the use of the 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.

5.2 It should be pointed out that almost all criteria for outliers are based on an assumed underlying normal (Gaussian) population or distribution. 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 experimenter is cautioned against interpreting the probabilities too literally.

5.3 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. Recommended Criteria for Single Samples

6.1 Let the sample of n observations be denoted in order of increasing magnitude by $x_1 \leq x_2 \leq x_3 \leq ... \leq x_n$. 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:

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

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:

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

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

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

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.

6.2 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 *only* in outliers that occur on the *high side*, we should always use the statistic $T_n = (x_n - \bar{x})/s$ and take as critical value the 0.05 point of Table 1. On the other hand, if we are interested *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 $T_n = (x_n - \bar{x})/s$ 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. 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.

6.2.1 *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. The doubtful observation is the high value, $x_{10} = 596$. Is the value of 596 significantly high? The mean is $\bar{x}$ = 575.2 and the estimated standard deviation is s = 8.70. We compute

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

From Table 1, for 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 % 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.

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

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
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
63	3.896	3.579	3.430	3.218	3.044	2.854
64	3.903	3.586	3.437	3.224	3.049	2.860
65	3.910	3.592	3.442	3.230	3.055	2.866
66	3.917	3.598	3.449	3.235	3.061	2.871
67	3.923	3.605	3.454	3.241	3.066	2.877
68	3.930	3.610	3.460	3.246	3.071	2.883
69	3.936	3.617	3.466	3.252	3.076	2.888
70	3.942	3.622	3.471	3.257	3.082	2.893
71	3.948	3.627	3.476	3.262	3.087	2.897
72	3.954	3.633	3.482	3.267	3.092	2.903
73	3.960	3.638	3.487	3.272	3.098	2.908
74	3.965	3.643	3.492	3.278	3.102	2.912
75	3.971	3.648	3.496	3.282	3.107	2.917
76	3.977	3.654	3.502	3.287	3.111	2.922
77	3.982	3.658	3.507	3.291	3.117	2.927
78	3.987	3.663	3.511	3.297	3.121	2.931
79	3.992	3.669	3.516	3.301	3.125	2.935
80	3.998	3.673	3.521	3.305	3.130	2.940
81	4.002	3.677	3.525	3.309	3.134	2.945
82	4.007	3.682	3.529	3.315	3.139	2.949
83	4.012	3.687	3.534	3.319	3.143	2.953
84	4.017	3.691	3.539	3.323	3.147	2.957
85	4.021	3.695	3.543	3.327	3.151	2.961
86	4.026	3.699	3.547	3.331	3.155	2.966
87	4.031	3.704	3.551	3.335	3.160	2.970
88	4.035	3.708	3.555	3.339	3.163	2.973
89	4.039	3.712	3.559	3.343	3.167	2.977
90	4.044	3.716	3.563	3.347	3.171	2.981
91	4.049	3.720	3.567	3.350	3.174	2.984
92	4.053	3.725	3.570	3.355	3.179	2.989
93	4.057	3.728	3.575	3.358	3.182	2.993
94	4.060	3.732	3.579	3.362	3.186	2.996
95	4.064	3.736	3.582	3.365	3.189	3.000
96	4.069	3.739	3.586	3.369	3.193	3.003
97	4.073	3.744	3.589	3.372	3.196	3.006
98	4.076	3.747	3.593	3.377	3.201	3.011
99	4.080	3.750	3.597	3.380	3.204	3.014
100	4.084	3.754	3.600	3.383	3.207	3.017
101	4.088	3.757	3.603	3.386	3.210	3.021
102	4.092	3.760	3.607	3.390	3.214	3.024
103	4.095	3.765	3.610	3.393	3.217	3.027
104	4.098	3.768	3.614	3.397	3.220	3.030
105	4.102	3.771	3.617	3.400	3.224	3.033
106	4.105	3.774	3.620	3.403	3.227	3.037
107	4.109	3.777	3.623	3.406	3.230	3.040
108	4.112	3.780	3.626	3.409	3.233	3.043
109	4.116	3.784	3.629	3.412	3.236	3.046
110	4.119	3.787	3.632	3.415	3.239	3.049
111	4.122	3.790	3.636	3.418	3.242	3.052
112	4.125	3.793	3.639	3.422	3.245	3.055
113	4.129	3.796	3.642	3.424	3.248	3.058
114	4.132	3.799	3.645	3.427	3.251	3.061

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
115	4.135	3.802	3.647	3.430	3.254	3.064
116	4.138	3.805	3.650	3.433	3.257	3.067
117	4.141	3.808	3.653	3.435	3.259	3.070
118	4.144	3.811	3.656	3.438	3.262	3.073
119	4.146	3.814	3.659	3.441	3.265	3.075
120	4.150	3.817	3.662	3.444	3.267	3.078
121	4.153	3.819	3.665	3.447	3.270	3.081
122	4.156	3.822	3.667	3.450	3.274	3.083
123	4.159	3.824	3.670	3.452	3.276	3.086
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

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

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

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

[A]Values of T are taken from Ref **(2)**. All values have been adjusted for division by n – 1 instead of n in calculating s.

6.3 An alternative system, the Dixon criteria, based entirely on ratios of differences between the observations is described in the literature **(1)**[3] and 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 % levels of significance.

6.3.1 *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 s. Table 2 indicates use of

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

Thus, for $n = 10$,

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

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, the 5 % critical value for $n = 10$. Under the Dixon criterion, we should therefore *not* consider this observation as an outlier at the 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

[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
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
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
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

[A] $x_1 \leq x_2 \leq ... \leq x_n$. (See Ref **(1)**, Appendix.)

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.

6.4 A test equivalent to T_n (or T_1) 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 **(5)**.

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

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:

$$w/s = (x_n - x_1)/s \tag{7}$$

where:

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

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.

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

[A] See Ref **(6)**, where:

$$w = x_n = x_1$$

$$x_1 \le x_2 \le \ldots \le x_n$$

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

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

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

From Table 3 for $n = 15$, we see that the value of $w/s = 4.374$ 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 *both* 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

$$T_1 = (\bar{x} - x_1)/s = [0.018 - (-1.40)]/0.551 = 2.574 \quad (10)$$

This value is greater than the critical value for the 5 % level, 2.409 from Table 1, so we reject − 1.40. Since we have decided that − 1.40 should be rejected, we use the remaining 14 observations and test the upper extreme 1.01, either with the criterion

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

or with Dixon's r_{22}. Omitting − 1.40 and renumbering the observations, we compute

$$\bar{x} = 1.67/14 = 0.119, \; s = 0.401, \quad (12)$$

and

$$T_{14} = (1.01 - 0.119)/0.401 = 2.22 \quad (13)$$

From Table 1, for $n = 14$, we find that a value as large as 2.22 would occur by chance more than 5 % of the time, so we should retain the value 1.01 in further calculations. We next calculate

$$\begin{aligned} r_{22} &= (x_{14} - x_{12})/x_{14} - x_3) \\ &= (1.01 - 0.48)/(1.01 + 0.24) \\ &= 0.53/1.25 \\ &= 0.424 \end{aligned} \quad (14)$$

From Table 2 for $n = 14$, we see that the 5 % critical value for r_{22} is 0.546. Since our calculated value (0.424) is less than the critical value, we also retain 1.01 by Dixon's test, and no further values would be tested in this sample.

NOTE 3—It should be noted that in repeated application of outlier tests to a sample, the overall significance level changes. If we apply k tests, an acceptable rule would be to use a significance level of α/k for each test so that the overall significance level will be approximately α.

6.8 For suspected observations on both the high and low sides in the sample, and to deal with the situation in which some of $k \geq 2$ suspected outliers are larger and some smaller than the remaining values in the sample, Tietjen and Moore **(7)** suggest the following statistic. Let the sample values be x_1, x_2, x_3, ... x_n and compute the sample mean, $\bar{x}$. Then compute the n absolute residuals

$$r_1 = |x_1 - \bar{x}|, r_2 = |x_2 - \bar{x}|, \ldots r_n = |x_n - \bar{x}| \quad (15)$$

Now relabel the original observations $x_1, x_2, \ldots, x_n$ as z's in such a manner that z_i is that x whose r_i is the ith largest absolute residual above. This now means that z_1 is that observation x which is closest to the mean and that z_n is the observation x which is farthest from the mean. The Tietjen-Moore statistic for testing the significance of the k largest residuals is then

$$E_k = [\sum_{i=1}^{n-k} (z_i - \bar{z}_k)^2 / \sum_{i=1}^{n} (z_i - \bar{z})^2] \quad (16)$$

where:

$$\bar{z}_k = \sum_{i=1}^{n-k} z_i/(n-k) \quad (17)$$

is the mean of the $(n - k)$ least extreme observations and z is the mean of the full sample.

6.8.1 Applying this test to the above data, we find that the total sum of squares of deviations for the entire sample is 4.24964. Omitting —1.40 and 1.01, the suspected two outliers, we find that the sum of squares of deviations for the reduced sample of 13 observations is 1.24089. Then $E_2 = 1.24089/4.24964 = 0.292$, and by using Table 4, we find that this observed E_2 is slightly smaller than the 5 % critical value of 0.317, so that the E_2 test would reject both of the observations, —1.40 and 1.01. We would probably take this latter recommendation, since the level of significance for the E_2 test is precisely 0.05 whereas that for the double application of a test for a single outlier cannot be guaranteed to be smaller than 1 — $(0.95)^2 = 0.0975$. The table of percentage points of E_k was computed by Monte Carlo methods on a high-speed electronic calculator.

6.9 We next turn to the case where we may have the two largest or the two smallest observations as probable outliers.

TABLE 4 1000 X Tietjen-Moore Critical Values (One-Sided Test) for E_k

k		n																							
	α	50	45	40	35	30	25	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
1[A]	0.01	748	728	704	669	624	571	499	484	459	440	422	404	374	337	311	274	235	197	156	110	68	29	4	...
	0.05	796	776	756	732	698	654	594	579	562	544	525	503	479	453	423	390	353	310	262	207	145	81	25	1
	0.10	820	802	784	762	730	692	638	624	610	593	576	556	534	510	482	451	415	374	326	270	203	127	49	3
2	0.01	636	607	574	533	482	418	339	323	306	290	263	238	207	181	159	134	101	78	50	28	12	2	...	...
	0.05	684	658	629	596	549	493	416	398	382	362	340	317	293	262	234	204	172	137	99	65	34	10	1	...
	0.10	708	684	657	624	582	528	460	442	424	406	384	360	337	309	278	250	214	175	137	94	56	22	2	...
3	0.01	550	518	480	435	386	320	236	219	206	188	166	146	123	103	83	64	44	26	14	6	1	...	...	...
	0.05	599	567	534	495	443	381	302	287	267	248	227	206	179	156	133	107	83	57	34	16	4	...	...	...
	0.10	622	593	562	523	475	417	338	322	304	284	263	240	216	189	162	138	108	80	53	27	9	...	...	...
4	0.01	482	446	408	364	308	245	170	156	141	122	107	90	72	56	42	30	18	9	4	...	...	...	...	...
	0.05	529	492	458	417	364	298	221	203	187	170	153	134	112	92	73	55	37	21	10	...	...	...	...	...
	0.10	552	522	486	443	391	331	252	234	217	198	182	160	138	116	94	73	52	32	16	...	...	...	...	...
5	0.01	424	386	347	299	250	188	121	108	94	79	68	54	42	31	20	12	6	...	...	...	...	...	...	...
	0.05	468	433	395	351	298	236	163	146	132	116	102	84	68	53	39	26	14	...	...	...	...	...	...	...
	0.10	492	459	422	379	325	264	188	172	156	140	122	105	86	68	52	36	22	...	...	...	...	...	...	...
6	0.01	376	336	298	252	204	146	86	74	62	52	40	32	22	14	8	...	...	...	...	...	...	...	...	...
	0.05	417	381	343	298	246	186	119	105	91	78	67	52	39	28	18	...	...	...	...	...	...	...	...	...
	0.10	440	406	367	324	270	210	138	124	110	95	82	67	52	38	26	...	...	...	...	...	...	...	...	...
7	0.01	334	294	258	211	166	110	58	50	41	32	24	18	12	...	...	...	...	...	...	...	...	...	...	...
	0.05	373	337	297	254	203	146	85	74	62	50	41	30	21	...	...	...	...	...	...	...	...	...	...	...
	0.10	396	360	320	276	224	168	102	89	76	64	53	40	29	...	...	...	...	...	...	...	...	...	...	...
8	0.01	297	258	220	177	132	87	40	32	26	18	14	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	334	299	259	214	166	114	59	50	41	32	24	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	355	320	278	236	186	132	72	62	51	42	32	...	...	...	...	...	...	...	...	...	...	...	...	...
9	0.01	264	228	190	149	108	66	26	20	14	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	299	263	223	181	137	89	41	33	26	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	319	284	243	202	154	103	51	42	34	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	0.01	235	200	164	124	87	50	17	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	268	233	195	154	112	68	28	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	287	252	212	172	126	80	35	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

[A] From Grubbs (1950, Table 1) for $n \leq 25$.

Here, we employ a test provided by Grubbs **(5, 10)** which is based on the ratio of the sample sum of squares when the two doubtful values are omitted to the sample sum of squares when the two doubtful values are included. If simplicity in calculation is the prime requirement, then the Dixon type of test (actually omitting one observation in the sample) might be used for this case. In illustrating the test procedure, we give the following Examples 4 and 5.

6.9.1 *Example 4*—In a comparison of strength of various plastic materials, one characteristic studied was the percentage elongation at break. Before comparison of the average elongation of the several materials, it was desirable to isolate for further study any pieces of a given material which gave very small elongation at breakage compared with the rest of the pieces in the sample. In this example, one might have primary interest only in outliers to the left of the mean for study, since very high readings indicate exceeding plasticity, a desirable characteristic.

6.9.1.1 Ten measurements of percentage elongation at break made on material No. 23 follow: 3.73, 3.59, 3.94, 4.13, 3.04, 2.22, 3.23, 4.05, 6.11, and 2.02. Arranged in ascending order of magnitude, these measurements are: 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94,. 4.05, 6.11, 4.13. The questionable readings are the two lowest, 2.02 and 2.22. We can test these two low readings simultaneously by using the following criterion of Table 5:

$$S_{1,2}{}^2/S^2 \tag{18}$$

For the above measurements:

$$\begin{aligned} S^2 &= \sum_{i=1} (x_i - \bar{x})^2 \\ &= [n\Sigma x_i{}^2 - (\Sigma x_i)^2]/n \\ &= [10(121.3594) - (34.06)^2]/10 \\ &= 5.351, \end{aligned} \tag{19}$$

and

$$\begin{aligned} S_{1,2}{}^2 &= \sum_{i=3}^{n} (x_i - \bar{x}_{1,2})^2 \\ &= [(n-2)\sum_{i=3}^{n} x_i{}^2 - (\sum_{i=3}^{n} x_i)^2]/(n-2) \\ &= [8(112.3506) - (29.82)^2]/8 \\ &= 9.5724/8 \\ &= 1.197 \end{aligned} \tag{20}$$

$$[\text{where } \bar{x}_{1,2} = \sum_{i=3}^{n} x_i/(n-2)] \tag{21}$$

We find:

$$S_{1,2}{}^2/S^2 = 1.197/5.351 = 0.224 \tag{22}$$

From Table 5 for n = 10, the 5 % significance level for $S_{1,2}{}^2/S^2$ is 0.2305. Since the calculated value is less than the critical value, we should conclude that both 2.02 and 2.22 are outliers. In a situation such as the one described in this example, where the outliers are to be isolated for further analysis, a significance level as high as 5 % or perhaps even 10 % would probably be used in order to get a reasonable size of sample for additional study. The problem may really be one of economics, and we use probability as a sensible basis for action.

6.9.2 *Example 5*—The following ranges (horizontal distances in yards from gun muzzle to point of impact of a projectile) were obtained in firings from a weapon at a constant angle of elevation and at the same weight of charge of propellant powder.

Distances in Yards

4782	4420
4838	4803
4765	4730
4549	4833

6.9.2.1 It is desired to make a judgment on whether the projectiles exhibit uniformity in ballistic behavior or if some of the ranges are inconsistent with the others. The doubtful values are the two smallest ranges, 4420 and 4549. For testing these two suspected outliers, the statistic $S_{1,2}^2/S^2$ of Table 5 is probably the best to use.

NOTE 4—Kudo **(11)** indicates that if the two outliers are due to a shift in location or level, as compared to the scale σ, then the optimum sample criterion for testing should be of the type:

$$\min (2\bar{x} - x_i - x_j)/s = (2\bar{x} - x_1 - x_2)/s \text{ in our Example 5.}$$

6.9.2.2 The distances arranged in increasing order of magnitude are:

4420	4782
4549	4803
4730	4833
4765	4838

TABLE 5 Critical Values for $S^2_{n-1,n}/S^2$, or $S^2_{1,2}/S^2$ for Simultaneously Testing the Two Largest or Two Smallest Observations[A]

Number of Observations, *n*	Lower 0.1 % Significance Level	Lower 0.5 % Significance Level	Lower 1 % Significance Level	Lower 2.5 % Significance Level	Lower 5 % Significance Level	Lower 10 % Significance Level
4	0.0000	0.0000	0.0000	0.0002	0.0008	0.0031
5	0.0003	0.0018	0.0035	0.0090	0.0183	0.0376
6	0.0039	0.0116	0.0186	0.0349	0.0564	0.0920
7	0.0135	0.0308	0.0440	0.0708	0.1020	0.1479
8	0.0290	0.0563	0.0750	0.1101	0.1478	0.1994
9	0.0489	0.0851	0.1082	0.1492	0.1909	0.2454
10	0.0714	0.1150	0.1414	0.1864	0.2305	0.2863
11	0.0953	0.1448	0.1736	0.2213	0.2667	0.3227
12	0.1198	0.1738	0.2043	0.2537	0.2996	0.3552
13	0.1441	0.2016	0.2333	0.2836	0.3295	0.3843
14	0.1680	0.2280	0.2605	0.3112	0.3568	0.4106
15	0.1912	0.2530	0.2859	0.3367	0.3818	0.4345
16	0.2136	0.2767	0.3098	0.3603	0.4048	0.4562
17	0.2350	0.2990	0.3321	0.3822	0.4259	0.4761
18	0.2556	0.3200	0.3530	0.4025	0.4455	0.4944
19	0.2752	0.3398	0.3725	0.4214	0.4636	0.5113
20	0.2939	0.3585	0.3909	0.4391	0.4804	0.5270
21	0.3118	0.3761	0.4082	0.4556	0.4961	0.5415
22	0.3288	0.3927	0.4245	0.4711	0.5107	0.5550
23	0.3450	0.4085	0.4398	0.4857	0.5244	0.5677
24	0.3605	0.4234	0.4543	0.4994	0.5373	0.5795
25	0.3752	0.4376	0.4680	0.5123	0.5495	0.5906
26	0.3893	0.4510	0.4810	0.5245	0.5609	0.6011
27	0.4027	0.4638	0.4933	0.5360	0.5717	0.6110
28	0.4156	0.4759	0.5050	0.5470	0.5819	0.6203
29	0.4279	0.4875	0.5162	0.5574	0.5916	0.6292
30	0.4397	0.4985	0.5268	0.5672	0.6008	0.6375
31	0.4510	0.5091	0.5369	0.5766	0.6095	0.6455
32	0.4618	0.5192	0.5465	0.5856	0.6178	0.6530
33	0.4722	0.5288	0.5557	0.5941	0.6257	0.6602
34	0.4821	0.5381	0.5646	0.6023	0.6333	0.6671
35	0.4917	0.5469	0.5730	0.6101	0.6405	0.6737
36	0.5009	0.5554	0.5811	0.6175	0.6474	0.6800
37	0.5098	0.5636	0.5889	0.6247	0.6541	0.6860
38	0.5184	0.5714	0.5963	0.6316	0.6604	0.6917
39	0.5266	0.5789	0.6035	0.6382	0.6665	0.6972
40	0.5345	0.5862	0.6104	0.6445	0.6724	0.7025
41	0.5422	0.5932	0.6170	0.6506	0.6780	0.7076
42	0.5496	0.5999	0.6234	0.6565	0.6834	0.7125
43	0.5568	0.6064	0.6296	0.6621	0.6886	0.7172
44	0.5637	0.6127	0.6355	0.6676	0.6936	0.7218
45	0.5704	0.6188	0.6412	0.6728	0.6985	0.7261
46	0.5768	0.6246	0.6468	0.6779	0.7032	0.7304
47	0.5831	0.6303	0.6521	0.6828	0.7077	0.7345
48	0.5892	0.6358	0.6573	0.6876	0.7120	0.7384
49	0.5951	0.6411	0.6623	0.6921	0.7163	0.7422
50	0.6008	0.6462	0.6672	0.6966	0.7203	0.7459
51	0.6063	0.6512	0.6719	0.7009	0.7243	0.7495
52	0.6117	0.6560	0.6765	0.7051	0.7281	0.7529
53	0.6169	0.6607	0.6809	0.7091	0.7319	0.7563
54	0.6220	0.6653	0.6852	0.7130	0.7355	0.7595
55	0.6269	0.6697	0.6894	0.7168	0.7390	0.7627
56	0.6317	0.6740	0.6934	0.7205	0.7424	0.7658
57	0.6364	0.6782	0.6974	0.7241	0.7456	0.7687
58	0.6410	0.6823	0.7012	0.7276	0.7489	0.7716
59	0.6454	0.6862	0.7049	0.7310	0.7520	0.7744
60	0.6497	0.6901	0.7086	0.7343	0.7550	0.7772

TABLE 5 *Continued*

Number of Observations, *n*	Lower 0.1 % Significance Level	Lower 0.5 % Significance Level	Lower 1 % Significance Level	Lower 2.5 % Significance Level	Lower 5 % Significance Level	Lower 10 % Significance Level
61	0.6539	0.6938	0.7121	0.7375	0.7580	0.7798
62	0.6580	0.6975	0.7155	0.7406	0.7608	0.7824
63	0.6620	0.7010	0.7189	0.7437	0.7636	0.7850
64	0.6658	0.7045	0.7221	0.7467	0.7664	0.7874
65	0.6696	0.7079	0.7253	0.7496	0.7690	0.7898
66	0.6733	0.7112	0.7284	0.7524	0.7716	0.7921
67	0.6770	0.7144	0.7314	0.7551	0.7741	0.7944
68	0.6805	0.7175	0.7344	0.7578	0.7766	0.7966
69	0.6839	0.7206	0.7373	0.7604	0.7790	0.7988
70	0.6873	0.7236	0.7401	0.7630	0.7813	0.8009
71	0.6906	0.7265	0.7429	0.7655	0.7836	0.8030
72	0.6938	0.7294	0.7455	0.7679	0.7859	0.8050
73	0.6970	0.7322	0.7482	0.7703	0.7881	0.8070
74	0.7000	0.7349	0.7507	0.7727	0.7902	0.8089
75	0.7031	0.7376	0.7532	0.7749	0.7923	0.8108
76	0.7060	0.7402	0.7557	0.7772	0.7944	0.8127
77	0.7089	0.7427	0.7581	0.7794	0.7964	0.8145
78	0.7117	0.7453	0.7605	0.7815	0.7983	0.8162
79	0.7145	0.7477	0.7628	0.7836	0.8002	0.8180
80	0.7172	0.7501	0.7650	0.7856	0.8021	0.8197
81	0.7199	0.7525	0.7672	0.7876	0.8040	0.8213
82	0.7225	0.7548	0.7694	0.7896	0.8058	0.8230
83	0.7250	0.7570	0.7715	0.7915	0.8075	0.8245
84	0.7275	0.7592	0.7736	0.7934	0.8093	0.8261
85	0.7300	0.7614	0.7756	0.7953	0.8109	0.8276
86	0.7324	0.7635	0.7776	0.7971	0.8126	0.8291
87	0.7348	0.7656	0.7796	0.7989	0.8142	0.8306
88	0.7371	0.7677	0.7815	0.8006	0.8158	0.8321
89	0.7394	0.7697	0.7834	0.8023	0.8174	0.8335
90	0.7416	0.7717	0.7853	0.8040	0.8190	0.8349
91	0.7438	0.7736	0.7871	0.8057	0.8205	0.8362
92	0.7459	0.7755	0.7889	0.8073	0.8220	0.8376
93	0.7481	0.7774	0.7906	0.8089	0.8234	0.8389
94	0.7501	0.7792	0.7923	0.8104	0.8248	0.8402
95	0.7522	0.7810	0.7940	0.8120	0.8263	0.8414
96	0.7542	0.7828	0.7957	0.8135	0.8276	0.8427
97	0.7562	0.7845	0.7973	0.8149	0.8290	0.8439
98	0.7581	0.7862	0.7989	0.8164	0.8303	0.8451
99	0.7600	0.7879	0.8005	0.8178	0.8316	0.8463
100	0.7619	0.7896	0.8020	0.8192	0.8329	0.8475
101	0.7637	0.7912	0.8036	0.8206	0.8342	0.8486
102	0.7655	0.7928	0.8051	0.8220	0.8354	0.8497
103	0.7673	0.7944	0.8065	0.8233	0.8367	0.8508
104	0.7691	0.7959	0.8080	0.8246	0.8379	0.8519
105	0.7708	0.7974	0.8094	0.8259	0.8391	0.8530
106	0.7725	0.7989	0.8108	0.8272	0.8402	0.8541
107	0.7742	0.8004	0.8122	0.8284	0.8414	0.8551
108	0.7758	0.8018	0.8136	0.8297	0.8425	0.8563
109	0.7774	0.8033	0.8149	0.8309	0.8436	0.8571
110	0.7790	0.8047	0.8162	0.8321	0.8447	0.8581
111	0.7806	0.8061	0.8175	0.8333	0.8458	0.8591
112	0.7821	0.8074	0.8188	0.8344	0.8469	0.8600
113	0.7837	0.8088	0.8200	0.8356	0.8479	0.8610
114	0.7852	0.8101	0.8213	0.8367	0.8489	0.8619
115	0.7866	0.8114	0.8225	0.8378	0.8500	0.8628
116	0.7881	0.8127	0.8237	0.8389	0.8510	0.8637
117	0.7895	0.8139	0.8249	0.8400	0.8519	0.8646
118	0.7909	0.8152	0.8261	0.8410	0.8529	0.8655
119	0.7923	0.8164	0.8272	0.8421	0.8539	0.8664
120	0.7937	0.8176	0.8284	0.8431	0.8548	0.8672
121	0.7951	0.8188	0.8295	0.8441	0.8557	0.8681

TABLE 5 *Continued*

Number of Observations, *n*	Lower 0.1 % Significance Level	Lower 0.5 % Significance Level	Lower 1 % Significance Level	Lower 2.5 % Significance Level	Lower 5 % Significance Level	Lower 10 % Significance Level
122	0.7964	0.8200	0.8306	0.8451	0.8567	0.8689
123	0.7977	0.8211	0.8317	0.8461	0.8576	0.8697
124	0.7990	0.8223	0.8327	0.8471	0.8585	0.8705
125	0.8003	0.8234	0.8338	0.8480	0.8593	0.8713
126	0.8016	0.8245	0.8348	0.8490	0.8602	0.8721
127	0.8028	0.8256	0.8359	0.8499	0.8611	0.8729
128	0.8041	0.8267	0.8369	0.8508	0.8619	0.8737
129	0.8053	0.8278	0.8379	0.8517	0.8627	0.8744
130	0.8065	0.8288	0.8389	0.8526	0.8636	0.8752
131	0.8077	0.8299	0.8398	0.8535	0.8644	0.8759
132	0.8088	0.8309	0.8408	0.8544	0.8652	0.8766
133	0.8100	0.8319	0.8418	0.8553	0.8660	0.8773
134	0.8111	0.8329	0.8427	0.8561	0.8668	0.8780
135	0.8122	0.8339	0.8436	0.8570	0.8675	0.8787
136	0.8134	0.8349	0.8445	0.8578	0.8683	0.8794
137	0.8145	0.8358	0.8454	0.8586	0.8690	0.8801
138	0.8155	0.8368	0.8463	0.8594	0.8698	0.8808
139	0.8166	0.8377	0.8472	0.8602	0.8705	0.8814
140	0.8176	0.8387	0.8481	0.8610	0.8712	0.8821
141	0.8187	0.8396	0.8489	0.8618	0.8720	0.8827
142	0.8197	0.8405	0.8498	0.8625	0.8727	0.8834
143	0.8207	0.8414	0.8506	0.8633	0.8734	0.8840
144	0.8218	0.8423	0.8515	0.8641	0.8741	0.8846
145	0.8227	0.8431	0.8523	0.8648	0.8747	0.8853
146	0.8237	0.8440	0.8531	0.8655	0.8754	0.8859
147	0.8247	0.8449	0.8539	0.8663	0.8761	0.8865
148	0.8256	0.8457	0.8547	0.8670	0.8767	0.8871
149	0.8266	0.8465	0.8555	0.8677	0.8774	0.8877

$$x_1 \leq x_2 \leq \ldots \leq x_n$$

$$S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$$

$$S^2_{1,2} = \sum_{i=1}^{n} (x_i - \bar{x}_{1,2})^2$$

$$\bar{x}_{1,2} = \frac{1}{n-2} \sum_{i=3}^{n} x_i$$

$$S^2_{n-1,n} = \sum_{i=1}^{n-2} (x_i - \bar{x}_n - 1, n)^2$$

$$\bar{x}_{n-1,n} = \frac{1}{n-2} \sum_{i=1}^{n-2} x_i$$

[A]These significance levels are taken from Table 11, Ref **(2)**. An observed ratio less than the appropriate critical ratio in this table calls for rejection of the null hypothesis.

The value of S^2 is 158 592. Omission of the two shortest ranges, 4420 and 4549, and recalculation, gives $S_{1,2}{}^2$ equal to 8590.8. Thus,

$$S_{1,2}{}^2/S^2 = 8590.8/158\ 592 = 0.054 \tag{23}$$

which is significant at the 0.01 level (See Table 5). It is thus highly unlikely that the two shortest ranges (occurring actually from excessive yaw) could have come from the same population as that represented by the other six ranges. It should be noted that the critical values in Table 5 for the 1 % level of significance are smaller than those for the 5 % level. So for this particular test, the calculated value is significant if it is *less* than the chosen critical value.

6.10 By Monte Carlo methods using an electronic calculator, Tietjen and Moore **(7)** have recently extended the tables of percentage points for the two highest or the two lowest observations to $k > 2$ highest or lowest sample values. Their results are given in Table 6 where

$$L_k = \sum_{i=1}^{n-k} (x_i - \bar{x}_k)^2 / \sum_{i=1}^{n} (x_i - \bar{x})^2 \quad \text{and} \quad \bar{x}_k = \sum_{i=1}^{n-k} x_i/(n-k).$$

Note that their L_2 equals our $S_{n,n-1}{}^2/S^2$. *For k = 1, their critical values agreed with exact values calculated by Grubbs (1950). This new table may be used to advantage in many practical problems of interest.*

6.11 If simplicity in calculation is very important, or if a large number of samples must be examined individually for outliers, the questionable observations may be tested with the application of Dixon's criteria. Disregarding the lowest range, 4420, we test if the next lowest range, 4549, is outlying. With $n = 7$, we see from Table 2 that $r_{1\,0}$ is the appropriate statistic. Renumbering the ranges as x_i to x_7, beginning with 4549, we find:

TABLE 6 1000 X Tietjen-Moore Critical Values (One-Sided Test) for L_k

k	n																							
	α	50	45	40	35	30	25	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3
1[A]	0.01	768	745	722	690	650	607	539	522	504	485	463	440	414	386	355	321	283	241	195	145	93	44	10	...
	0.025	796	776	756	732	699	654	594	579	562	544	525	503	479	453	423	390	353	310	262	207	145	81	25	1
	0.05	820	802	784	762	730	692	638	624	610	593	576	556	534	510	482	451	415	374	326	270	203	127	49	3
	0.10	840	826	812	792	766	732	685	673	660	646	631	613	594	573	548	520	488	450	405	350	283	199	98	11
2[B]	0.01	667	641	610	573	527	468	391	373	353	332	310	286	261	233	204	174	141	108	75	44	19	4	...	...
	0.025	697	667	644	610	567	512	439	421	403	382	360	337	311	284	254	221	186	149	110	71	35	9	...	...
	0.05	720	698	673	641	601	550	480	464	446	426	405	382	357	330	300	267	230	191	148	102	56	18	1	...
	0.10	746	726	702	674	637	591	527	511	494	476	456	435	411	384	355	323	286	245	199	148	92	38	3	...
3	0.01	592	558	522	484	434	377	300	272	260	237	219	194	172	147	120	98	70	48	28	10	2	...	...	...
	0.025	622	592	561	527	479	417	341	321	299	282	261	239	214	184	162	129	100	73	45	21	5	...	...	...
	0.05	646	618	588	554	506	450	377	354	337	322	300	276	250	224	196	162	129	99	64	32	10	...	...	...
	0.10	673	648	622	586	523	489	420	398	384	364	342	322	298	270	240	208	170	134	95	56	20	...	...	...
4	0.01	531	498	460	418	369	308	231	211	192	171	151	132	113	94	70	52	32	18	8	...	...	...	...	...
	0.025	559	529	491	455	408	342	265	243	226	208	185	167	145	122	96	74	52	30	13	...	...	...	...	...
	0.05	588	556	523	482	434	374	299	277	259	240	219	197	174	150	125	98	70	45	22	...	...	...	...	...
	0.10	614	586	554	516	472	412	339	316	302	282	260	236	212	186	159	128	98	68	38	...	...	...	...	...
5	0.01	483	444	408	364	312	246	175	154	140	126	108	90	72	56	38	26	12	...	...	...	...	...	...	...
	0.025	510	473	433	398	352	282	209	189	171	151	135	113	95	77	57	40	23	...	...	...	...	...	...	...
	0.05	535	502	468	424	376	312	238	217	200	181	159	140	122	98	76	54	34	...	...	...	...	...	...	...
	0.10	562	533	499	458	411	350	273	251	236	216	194	172	150	126	103	74	51	...	...	...	...	...	...	...
6	0.01	438	399	364	321	268	204	136	118	104	91	72	57	46	33	19	...	...	...	...	...	...	...	...	...
	0.025	466	430	387	348	302	233	165	145	129	117	96	78	63	47	31	...	...	...	...	...	...	...	...	...
	0.05	490	456	421	376	327	262	188	168	154	136	115	97	79	60	42	...	...	...	...	...	...	...	...	...
	0.10	518	488	451	410	359	296	220	199	184	165	144	124	104	82	62	...	...	...	...	...	...	...	...	...
7	0.01	400	361	324	282	229	168	104	88	76	64	49	37	27	...	...	...	...	...	...	...	...	...	...	...
	0.025	428	391	348	308	261	192	128	108	95	82	65	51	38	...	...	...	...	...	...	...	...	...	...	...
	0.05	450	417	378	334	283	222	150	130	116	100	82	66	50	...	...	...	...	...	...	...	...	...	...	...
	0.10	477	447	408	365	316	251	176	158	142	125	104	86	68	...	...	...	...	...		...	...	...	...	...
8	0.01	368	328	292	250	196	144	78	64	53	44	30	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.025	392	356	314	274	226	159	98	80	68	58	45	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	414	382	342	297	245	184	115	99	86	72	55	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	442	410	372	328	276	213	140	124	108	92	73	...	...	...	...	...	...	...	...	...	...	...	...	...
9	0.01	336	296	262	220	166	112	58	46	36	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.025	363	325	283	242	193	132	73	59	48	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	383	350	310	264	212	154	88	74	62	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	410	378	338	294	240	180	110	94	80	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
10	0.01	308	270	234	194	142	92	42	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.025	334	295	257	213	165	108	54	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.05	356	320	280	235	183	126	66	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	0.10	380	348	307	262	210	152	85	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

[A]From Grubbs (1950, Table I) for $n \leq 25$.
[B]From Grubbs (1972, Table II).

$$r_{10} = (x_2 - x_1)/(x_7 - x_1)$$

$$= (4730 - 4549)/(4838 - 4549)$$
$$= 181/289$$
$$= 0.626 \tag{24}$$

which is only a little less than the 1 % critical value, 0.637, for $n = 7$. So, if the test is being conducted at any significance level greater than a 1 % level, we would conclude that 4549 is an outlier. Since the lowest of the original set of ranges, 4420, is even more outlying than the one we have just tested, it can be classified as an outlier without further testing. We note here, however, that this test did not use all of the sample observations.

6.12 *Rejection of Several Outliers*— So far we have discussed procedures for detecting one or two outliers in the same sample, but these techniques are not generally recommended for repeated rejection, since if several outliers are present in the sample the detection of one or two spurious values may be "masked" by the presence of other anomalous observations. Outlying observations occur due to a shift in level (or mean), or a change in scale (that is, change in variance of the observations), or both. Ferguson **(3,4)** has studied the power of the various rejection rules relative to changes in level or scale. For several outliers and repeated rejection of observations, Ferguson points out that the sample coefficient of skewness,

$$\sqrt{b_1} = \sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3/(n-1)^{3/2}s^3 \tag{25}$$

$$= \sqrt{n} \sum_{i=1}^{n} (x_i - \bar{x})^3/[\Sigma(x_i - \bar{x})^2]^{3/2}$$

should be used for "one-sided" tests (change in level of several observations in the same direction), and the sample coefficient of kurtosis,

$$b_2 = n \sum_{i=1}^{n} (x_i - \bar{x})^4/(n-1)^2 s^4 \tag{26}$$

$$= n \sum_{i=1}^{n} (x_i - \bar{x})^4/[\Sigma(x_i - \bar{x})^2]^2$$

is recommended for "two-sided" tests (change in level to higher and lower values) and also for changes in scale (variance) (see Note 5). In applying the above tests, the $\sqrt{b_1}$ or the b_2, or both, are computed and if their observed values exceed those for significance levels given in Table 7 and Table 8, then the observation farthest from the mean is rejected and the same procedure repeated until no further sample values are judged as outliers. (As is well-known $\sqrt{b_1}$ and b_2 are also used as tests of normality.)

NOTE 5—In the above equations for $\sqrt{b_1}$ and b_2, s is defined as used in this standard:

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

6.12.1 The significance levels in Table 7 and Table 8 for sample sizes of 5, 10, 15, and 20 (and 25 for b_2) were obtained by Ferguson on an IBM 704 computer using a sampling experiment or "Monte Carlo'' procedure. The significance levels for the other sample sizes are from Pearson, E. S. "Table of Percentage Points of $\sqrt{b_1}$ and b_2 in Normal Samples; a Rounding Off," *Biometrika*, Vol 52, 1965, pp. 282–285.

6.12.2 The $\sqrt{b_1}$ and b_2 statistics have the optimum property of being "locally" best against one-sided and two-sided alternatives, respectively. The $\sqrt{b_1}$ test is good for up to 50 % spurious observations in the sample for the one-sided case, and the b_2 test is optimum in the two-sided alternatives case for up to 21 % "contamination" of sample values. For only one or two outliers the sample statistics of the previous paragraphs are recommended, and Ferguson **(3)** discusses in detail their optimum properties of *pointing out* one or two outliers.

6.12.2.1 Instead of the more complicated $\sqrt{b_1}$ and b_2 statistics, one can use Table 4 and Table 6 **(7)** for sample sizes and percentage points given.

7. Recommended Criterion Using Independent Standard Deviation

7.1 Suppose that an independent estimate of the standard deviation is available from previous data. This estimate may be from a single sample of previous similar data or may be the result of combining estimates from several such previous sets of data. In any event, each estimate is said to have degrees of freedom equal to one less than the sample size that it is based on. The proper combined estimate is a weighted average of the several values of s^2, the weights being proportional to the respective degrees of freedom. The total degrees of freedom in the combined estimate is then the sum of the individual degrees of freedom. When one uses an independent estimate of the standard deviation, s_v, the test criterion recommended here for an outlier is as follows:

TABLE 7 Significance Levels (One-Sided Test) for $\sqrt{b_1}$

Significance Level, percent	n									
	5[A]	10[A]	15[A]	20[A]	25	30	35	40	50	60
1	1.34	1.31	1.20	1.11	1.06	0.98	0.92	0.87	0.79	0.72
5	1.05	0.92	0.84	0.79	0.71	0.66	0.62	0.59	0.53	0.49

[A]These values were obtained by Ferguson, using a Monte Carlo procedure.

TABLE 8 Significance Levels (One-Sided Test) for b_2

Significance Level, percent	n							
	5[A]	10[A]	15[A]	20[A]	25[A]	50	75	100
1	3.11	4.83	5.08	5.23	5.00	4.88	4.59	4.39
5	2.89	3.85	4.07	4.15	4.00	3.99	3.87	3.77

[A]These values were obtained by Ferguson, using a Monte Carlo procedure. For $n = 25$; Ferguson's Monte Carlo values of b_2 agree with Pearson's computed values.

$$T'_1 = (\bar{x} - x_1)/s \qquad (27)$$

or:

$$T'_n = (x_n - \bar{x})/s_v \qquad (28)$$

where:

v = total number of degrees of freedom.

7.2 The critical values for T'_1 and T'_n for the 5 % and 1 % significance levels are due to David **(12)** and are given in Table 9. In Table 9 the subscript v = df indicates the total number of degrees of freedom associated with the independent estimate of standard deviation σ and n indicates the number of observations in the sample under study. We illustrate with an example on interlaboratory testing.

7.3 *Example 6—Interlaboratory Testing*—In an analysis of interlaboratory test procedures, data representing normalities of sodium hydroxide solutions were determined by twelve different laboratories. In all the standardizations, a 0.1 *N* sodium hydroxide solution was prepared by the Standard Methods Committee using carbon-dioxide-free distilled water. Potassium acid phthalate (P.A.P.), obtained from the National Institute of Standards and Technology, was used as the test standard.

7.3.1 Test data by the twelve laboratories are given in Table 10. The P.A.P. readings have been coded to simplify the calculations. The variances between the three readings within all laboratories were found to be homogeneous. A one-way classification in the analysis of variance was first analyzed to determine if the variation in laboratory results (averages) was statistically significant. This variation was significant and indicated a need for action, so tests for outliers were then applied to isolate the particular laboratories whose results gave rise to the significant variation.

7.3.2 Table 11 shows that the variation between laboratories is highly significant. To test if this (very significant) variation is due to one (or perhaps two) laboratories that obtained "outlying" results (that is, perhaps showing nonstandard technique), we can test the laboratory averages for outliers. From the analysis of variance, we have an estimate of the variance of an individual reading as 0.008793, based on 24 degrees of freedom. The estimated standard deviation of an individual measurement is $\sqrt{0.008793} = 0.094$ and the estimated standard deviation of the average of three readings is therefore $0.094/\sqrt{3} = 0.054$.

7.3.3 Since the estimate of within-laboratory variation is independent of any difference between laboratories, we can use the statistic T'_1 of 7.1 to test for outliers. An examination of the deviations of the laboratory averages from the grand average indicates that Laboratory 10 obtained an average reading much lower than the grand average, and that Laboratory 12 obtained a high average compared to the over-all average. To first test if Laboratory 10 is an outlier, we compute

TABLE 9 Critical Values (One-Sided Test) for T' When Standard Deviation s_v is Independent of Present Sample[A]

$$T' = \frac{x_n - \bar{x}}{s_v}, or \frac{\bar{x} - x_1}{s_v}$$

v = d.f.	n								
	3	4	5	6	7	8	9	10	12
					1 percentage point				
10	2.78	3.10	3.32	3.48	3.62	3.73	3.82	3.90	4.04
11	2.72	3.02	3.24	3.39	3.52	3.63	3.72	3.79	3.93
12	2.67	2.96	3.17	3.32	3.45	3.55	3.64	3.71	3.84
13	2.63	2.92	3.12	3.27	3.38	3.48	3.57	3.64	3.76
14	2.60	2.88	3.07	3.22	3.33	3.43	3.51	3.58	3.70
15	2.57	2.84	3.03	3.17	3.29	3.38	3.46	3.53	3.65
16	2.54	2.81	3.00	3.14	3.25	3.34	3.42	3.49	3.60
17	2.52	2.79	2.97	3.11	3.22	3.31	3.38	3.45	3.56
18	2.50	2.77	2.95	3.08	3.19	3.28	3.35	3.42	3.53
19	2.49	2.75	2.93	3.06	3.16	3.25	3.33	3.39	3.50
20	2.47	2.73	2.91	3.04	3.14	3.23	3.30	3.37	3.47
24	2.42	2.68	2.84	2.97	3.07	3.16	3.23	3.29	3.38
30	2.38	2.62	2.79	2.91	3.01	3.08	3.15	3.21	3.30
40	2.34	2.57	2.73	2.85	2.94	3.02	3.08	3.13	3.22
60	2.29	2.52	2.68	2.79	2.88	2.95	3.01	3.06	3.15
120	2.25	2.48	2.62	2.73	2.82	2.89	2.95	3.00	3.08
∞	2.22	2.43	2.57	2.68	2.76	2.83	2.88	2.93	3.01
					5 percentage points				
10	2.01	2.27	2.46	2.60	2.72	2.81	2.89	2.96	3.08
11	1.98	2.24	2.42	2.56	2.67	2.76	2.84	2.91	3.03
12	1.96	2.21	2.39	2.52	2.63	2.72	2.80	2.87	2.98
13	1.94	2.19	2.36	2.50	2.60	2.69	2.76	2.83	2.94
14	1.93	2.17	2.34	2.47	2.57	2.66	2.74	2.80	2.91
15	1.91	2.15	2.32	2.45	2.55	2.64	2.71	2.77	2.88
16	1.90	2.14	2.31	2.43	2.53	2.62	2.69	2.75	2.86
17	1.89	2.13	2.29	2.42	2.52	2.60	2.67	2.73	2.84
18	1.88	2.11	2.28	2.40	2.50	2.58	2.65	2.71	2.82
19	1.87	2.11	2.27	2.39	2.49	2.57	2.64	2.70	2.80
20	1.87	2.10	2.26	2.38	2.47	2.56	2.63	2.68	2.78
24	1.84	2.07	2.23	2.34	2.44	2.52	2.58	2.64	2.74
30	1.82	2.04	2.20	2.31	2.40	2.48	2.54	2.60	2.69
40	1.80	2.02	2.17	2.28	2.37	2.44	2.50	2.56	2.65
60	1.78	1.99	2.14	2.25	2.33	2.41	2.47	2.52	2.61
120	1.76	1.96	2.11	2.22	2.30	2.37	2.43	2.48	2.57
∞	1.74	1.94	2.08	2.18	2.27	2.33	2.39	2.44	2.52

[A]The percentage points are reproduced from Ref **(12)**.

$$T' = (1.871 - 0.745)/0.054 = 20.9 \quad (29)$$

7.3.4 This value of T' is obviously significant at a very low level of probability ($P << 0.01$—Refer to Table 9 with $n = 12$ and $v = 24$ degrees of freedom). We conclude, therefore, that the test methods of Laboratory 10 should be investigated.

7.3.5 Excluding Laboratory 10, we compute a new grand average of 1.973 and test if the results of Laboratory 12 are outlying. We have

$$T' = (2.327 - 1.973)/0.054 = 6.56 \quad (30)$$

and this value of T' is significant at $P<< 0.01$ (Refer to Table 7 with $n = 11$ and $v = 24$ degrees of freedom). We conclude that the procedures of Laboratory 12 should also be investigated.

7.3.6 To verify that the remaining laboratories did indeed obtain homogeneous results, we might repeat the analysis of variance omitting Laboratories 10 and 12. The calculations give the results shown in Table 12.

7.3.6.1 For this analysis, the variation between laboratories is not significant at the 5 % level and we conclude that all the laboratories except No. 10 and No. 12 exhibit the same capability in testing procedure.

7.3.6.2 In conclusion, there should be a systematic investigation of test methods for Laboratories No. 10 and No. 12 to determine why their test procedures are apparently different from the other ten laboratories. (In this type of problem, the tables of Greenhouse, Halperin, and Cornfield **(13)** could also be used for testing outlying laboratory averages.)

7.3.7 *Cautionary Remarks*—In the use of the tests for outliers as given above, our interest was to direct the statistical tests of significance toward picking out those laboratories which have different levels of measurement than the others. Thus, we have assumed that there should not exist any component of variance among the laboratory true means of measurement. On the other hand, it is well known that in

TABLE 10 Standardization of Sodium Hydroxide Solutions as Determined by Plant Laboratories

Standard used: Potassium Acid Phthalate (P.A.P.)

Laboratory	(P.A.P. − 0.096000 $\times 10^3$)	Sums	Averages	Deviation of Average from Grand Average
1	1.893			
	1.972			
	1.876	5.741	1.914	+ 0.043
2	2.046			
	1.851			
	1.949	5.846	1.949	+ 0.078
3	1.874			
	1.792			
	1.829	5.495	1.832	−0.039
4	1.861			
	1.998			
	1.983	5.842	1.947	+ 0.076
5	1.922			
	1.881			
	1.850	5.653	1.884	+ 0.013
6	2.082			
	1.958			
	2.029	6.069	2.023	+ 0.152
7	1.992			
	1.980			
	2.066	6.038	2.013	+ 0.142
8	2.050			
	2.181			
	1.903	6.134	2.045	+ 0.174
9	1.831			
	1.883			
	1.855	5.569	1.856	−0.015
10	0.735			
	0.722			
	0.777	2.234	0.745	−1.126
11	2.064			
	1.794			
	1.891	5.749	1.916	+ 0.045
12	2.475			
	2.403			
	2.102	6.980	2.327	+ 0.456
Grand sum		67.350		
Grand average			1.871	

TABLE 11 Analysis of Variance

Source of Variation	Degrees of Freedom (d.f.)	Sum of Squares (SS)	Mean Square (MS)	F-ratio
Between laboratories	11	4.70180	0.4274	F = v 48.61
Within laboratories	24	0.21103	0.008793	(highly significant)
Total	35	4.91283		

TABLE 12 Analysis of Variance (Omitting Labs 10 and 12)

Source of Variation	d.f.	SS	MS	F-ratio
Between laboratories	9	0.13889	0.015430	F = v 2.36
Within laboratories	20	0.13107	0.00655	F0.05(9, 20) = v 2.40 F0.01(9, 20) = v 3.45
Total	29	0.26996		

practically all interlaboratory tests one does indeed find a nonzero component of variance among the laboratory levels. Often the variance among the laboratory means may be several times that within individual laboratories. Thus, if we knew the size of the actual component of variance among laboratories we must live with—or guard against—then the observed F ratio could be multiplied by the within variance of a sample mean and divided by this quantity plus the among laboratory variance, in order to adjust the F test to detect the undesirable deviations of those laboratories which departed in average level from measurements of the common or acceptable level of the closely agreeing laboratories. Also, a somewhat similar adjustment, if desired, could be applied to the tests for isolated outliers. In our particular example, however, we desired to detect those particular laboratories which departed in average level from that of the closely agreeing laboratories. In fact, this should be the aim of many interlaboratory testing programs, if we are to seek high precision and accuracy of measurement.

8. Recommended Criteria for Known Standard Deviation

8.1 Frequently the population standard deviation σ may be known accurately. In such cases, Table 13 may be used for single outliers and we illustrate with the following example:

8.2 *Example 7* (σ *known*)—Passage of the Echo I (Balloon) Satellite was recorded on star-plates when it was visible. Photographs were made by means of a camera with shutter automatically timed to obtain a series of points for the Echo path. Since the stars were also photographed at the same times as the Satellite, all the pictures show star-trails and are thus called "star-plates."

8.2.1 The *x*- and *y*-coordinate of each point on the Echo path are read from a photograph, using a stereo-comparator. To eliminate bias of the reader, the photograph is placed in one position and the coordinates are read; then the photograph is rotated 180 deg and the coordinates reread. The average of the two readings is taken as the final reading. Before any further calculations are made, the readings must be "screened" for

TABLE 13 Critical Values (One-Sided Test) of $T'_{1\infty}$ and $T'_{n\infty}$ When the Population Standard Deviation σ is Known[A]

Number of Observations, *n*	5 Percent Significance Level	1 Percent Significance Level	0.5 Percent Significance Level
2	1.39	1.82	1.99
3	1.74	2.22	2.40
4	1.94	2.43	2.62
5	2.08	2.57	2.76
6	2.18	2.68	2.87
7	2.27	2.76	2.95
8	2.33	2.83	3.02
9	2.39	2.88	3.07
10	2.44	2.93	3.12
11	2.48	2.97	3.16
12	2.52	3.01	3.20
13	2.56	3.04	3.23
14	2.59	3.07	3.26
15	2.62	3.10	3.29
16	2.64	3.12	3.31
17	2.67	3.15	3.33
18	2.69	3.17	3.36
19	2.71	3.19	3.38
20	2.73	3.21	3.39
21	2.75	3.22	3.41
22	2.77	3.24	3.42
23	2.78	3.26	3.44
24	2.80	3.27	3.45
25	2.81	3.28	3.46

$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$

$T'_1 = (\bar{x} - x_1)/\sigma;\ T'_n = (x_n - \bar{x})/\sigma$

[A] This table is taken from Ref **(13)**.

gross reading or tabulation errors. This is done by examining the difference in the readings taken at the two positions of the photograph.

8.2.2 Table 14 records a sample of six readings made at the two positions and the differences in these readings. On the third reading, the differences are rather large. Has the operator made an error in placing the cross hair on the point?

8.2.3 For this example, an independent estimate of σ is available since extensive tests on the stereo-comparator have shown that the standard deviation in reader's error is about 4 µm. The determination of this standard error was based on such a large sample that we can assume σ = 4 µm. The standard deviation of the difference in two readings is therefore

$$\sqrt{4^2 + 4^2} = \sqrt{32} \text{ or } 5.7 \ \mu\text{m} \tag{31}$$

8.2.4 For the six readings above, the mean difference in the x-coordinates is $\bar{\Delta} x = 3.5$ and the mean difference in the y-coordinates is $\bar{\Delta} y = 1.8$. For the questionable third reading, we have

$$T'_x = (24 - 3.5)/5.7 = 3.60 \tag{32}$$

$$T'_y = (22 - 1.8)/5.7 = 3.54 \tag{33}$$

From Table 13 we see that for $n = 6$, values of $T'_{n\infty}$ as large as the calculated values would occur by chance less than 1 % of the time so that a significant reading error seems to have been made on the third point.

8.3 A great number of points are read and automatically tabulated on star-plates. Here we have chosen a very small sample of these points. In actual practice, the tabulations would probably be scanned quickly for very large errors such as tabulator errors; then some rule-of-thumb such as ±3 standard deviations of reader's error might be used to scan for outliers due to operator error (Note 6). In other words, the data are probably too extensive to allow repeated use of precise tests like those described above (especially for varying sample size), but this example does illustrate the case where σ is assumed known. If gross disagreement is found in the two readings of a coordinate, then the reading could be omitted or reread before further computations are made.

NOTE 6—Note that the values of Table 13 vary between about 1.4σ and 3.5σ.

9. Additional Comments

9.1 In the above, we have covered only that part of screening samples to detect outliers statistically. However, a large area remains after the decision has been reached that outliers are present in data. Once some of the sample observations are branded as "outliers," then a thorough investigation should be initiated to determine the cause. In particular, one should look for gross errors, personal errors, errors of measurement, errors in calibration, etc. If reasons are found for aberrant observations, then one should act accordingly and perhaps scrutinize also the other observations. Finally, if one reaches the point that some observations are to be discarded or treated in a special manner based solely on statistical judgment, then it must be decided what action should be taken in the further analysis of the data. We do not propose to cover this problem here, since in many cases it will depend greatly on the particular case in hand. However, we do remark that there could be the outright rejection of aberrant observations once and for all on physical grounds (and preferably not on statistical grounds generally) and only the remaining observations would be used in further analyses or in estimation problems. On the other hand, some may want to replace aberrant values with newly taken observations and others may want to "Winsorize" the outliers, that is, replace them with the next closest values in the sample. Also with outliers in a sample, some may wish to use the median instead of the mean, and so on. Finally, we remark that perhaps a fair or appropriate practice might be that of using truncated-sample theory **(11)** for cases of samples where we have "censored" or rejected some of the observations. We cannot go further into these problems here. For additional reading on outliers, see Refs **(12,14,15,16,17,18,19)**.

9.2 A sample test criterion for non-normality, and hence possibly for outliers, not covered above is the Wilk-Shapiro W statistic for a sample of size n given by

$$W = \frac{\left[\sum_{i=1}^{[n/2]} a_{n-i+1}(x_{n-i+1} - x_i)\right]^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{34}$$

where:

$x_1 \leq x_2 \leq x_3 \leq \ldots \leq x_n$,

$\bar{x} = \sum_{i=1}^{n} x_i / n$,

$[n/2]$ is the greatest integer in $n/2$, and the coefficients a_{n-i+1} are the order statistics for n = 2(1)50 given in Ref **(20)**.

The Wilk-Shapiro W statistic has been found to be quite sensitive to departures from normality and generally may compare most favorably with the $\sqrt{b_1}$ and b_2 tests discussed above. In addition, therefore, the W statistic may also be used as a test for outliers, or otherwise as a general test for heterogeneity of sample values. Our significance tests given above have been selected and recommended since they specifically point out particular suspected outliers in the sample. We therefore are inclined to favor the above tests for specific outliers in samples for the case where they will be used routinely, for example, by engineers.

10. Keywords

10.1 dixon test; gross deviation; Grubbs test; outlier

TABLE 14 Measurements, µm

x-Coordinate			*y*-Coordinate		
Position 1	Position 1 + 180 deg	Δx	Position 1	Position 1 + 180 deg	Δy
−53011	−53004	− 7	70263	70258	+ 5
−38112	−38103	− 9	−39739	−39729	− 6
−2804	−2828	+ 24	81162	81140	+ 22
18473	18467	+ 6	41477	41485	− 8
25507	25497	+ 10	1082	1076	+ 6
87736	87739	− 3	−7442	−7434	− 8

REFERENCES

(**1**) Dixon, W. J., "Processing Data for Outliers,'' *Biometrics*, BIOMA March 1953, Vol 9, No. 1, pp. 74–89.

(**2**) Grubbs, F. E., and Beck, G.," Extension of Sample Sizes and Percentage Points for Significance Tests of Outlying Observations,'' *Technometrics*, TCMTA, Vol 14, No. 4, November 1972, pp. 847–854.

(**3**) Ferguson, T. S., *On the Rejection of Outliers*, Fourth Berkeley Symposium on Mathematical Statistics and Probability, edited by Jerzy Neyman, University of California Press, Berkeley and Los Angeles, Calif., 1961.

(**4**) Ferguson, T. S., "Rules for Rejection of Outliers,'' *Revue Inst. Int. de Stat.*, RINSA Vol 29, Issue 3, 1961, pp. 29–43.

(**5**) Grubbs, F. E., "Sample Criteria for Testing Outlying Observations,'' *Annals of Mathematical Statistics*, AASTA Vol 21, March 1950, pp. 27–58.

(**6**) David, H. A., Hartley, H. O., and Pearson, E. S., "The Distribution of the Ratio in a Single Sample of Range to Standard Deviation, *Biometrika*, BIOKA Vol 41, 1954, pp. 482–493.

(**7**) Tietjen, G. L., and Moore, R. H., "Some Grubbs-Type Statistics for the Detection of Several Outliers,'' *Technometrics*, TCMTA, Vol 14, No. 3, August 1972, pp. 583–597.

(**8**) Chauvenet, W., *A. Manual of Spherical and Practical Astronomy*,Vol 2, Fifth Edition.

(**9**) David, H. A., and Quesenberry, C. P., "Some Tests for Outliers,'' Technical Report No. 47, OOR(ARO) Project No. 1166, Virginia Polytechnic Inst., Blacksburg, Va.

(**10**) Grubbs, F. E., "Procedures for Detecting Outlying Observations in Samples,'' *Technometrics*, TCMTA, Vol 11, No. 4, February 1969, pp. 1–21.

(**11**) Kudo, A., "On the Testing of Outlying Observations," *Sankhyā*, *The Indian Journal of Statistics*, SNKYA Vol 17, Part 1, June 1956, pp. 67–76.

(**12**) David, H. A., "Revised Upper Percentage Points of the Extreme Studentized Deviate from the Sample Mean,'' *Biometrika*, BIOKA Vol 43, 1956, pp. 449–451.

(**13**) Greenhouse, S. W., Halperin, M., and Cornfield, J.," Tables of Percentage Points for the Studentized Maximum Absolute Deviation in Normal Samples,'' *Journal of the American Statistical Association*, JSTNA Vol 50, No. 269, 1955, pp. 185–195.

(**14**) Anscombe, F. J., "Rejection of Outliers,'' *Technometrics*, TCMTA Vol 2, No. 2, 1960, pp. 123–147.

(**15**) Chew, Victor," Tests for the Rejection of Outlying Observations," RCA Systems Analysis Technical Memorandum No. 64-7, 31 Dec. 1964, Patrick Air Force Base, Fla.

(**16**) Kruskal, W. H., "Some Remarks on Wild Observations," *Technometrics*, TCMTA Vol 2, No. 1, 1960, pp. 1–3.

(**17**) Proschan, F., "Testing Suspected Observations," *Industrial Quality Control*, IQCOA Vol XIII, No. 7, January 1957, pp. 14–19.

(**18**) Sarhan, A. E., and Greenberg, B. G., Editors, *Contributions to Order Statistics*, John Wiley and Sons, Inc., New York 1962.

(**19**) Thompson, W. R. "On a Criterion for the Rejection of Observa-tions and the Distribution of the Ratio of the Deviation to the Sample Standard Deviation,'' *The Annals of Mathematical Statistics*, AASTA Vol 6, 1935, pp. 214–219.

(**20**) Shapiro, S. S., and Wilk, M. B., "An Analysis of Variance Test for Non-Normality (Complete Samples)," *Biometrika*, BIOKA, Vol 52, 1965, pp. 591–611.