Analysis Using Few Data, Part Two

Comparing the *W*-Ratio Test with an *XmR* Chart

Donald J. Wheeler

While an *XmR* chart is commonly used as a process behavior chart, it may also be used as a test of homogeneity for a finite number of values. This paper explains the difference in these two uses of the *XmR* chart and compares the homogeneity chart with the *W*-ratio test presented in my June column. In addition, this paper contains an expanded table of critical values for the *W*-ratio test.

PROCESS BEHAVIOR CHARTS

The *XmR* chart is generally thought of as a sequential procedure intended for use with a continuing stream of values. When an *XmR* chart is used in this way the result is what I call a process behavior chart. As each point is added to such a chart another act of analysis occurs. If the new point falls outside the limits we judge that the underlying process is likely to have changed. If the new point falls within the limits we judge that the underlying process is unlikely to have changed. With this emphasis upon making a decision each time we add a point to the chart, it is natural and appropriate to characterize how a process behavior chart works on a point by point basis. For example, we describe both the power of the chart and the risk of a false alarm in terms of each additional point added to the chart.

The power function for the *X* chart describes the likelihood of detecting a shift in location on the first point following the shift. I published the formulas for this power function in the October 1983 issue of *Journal of Quality Technology*. These formulas covered the use of all four of the Western Electric Zone Tests. Like most mathematical treatments, these power functions provide a reasonable approximation for what happens in practice. However, at that time I did not consider the effect of baseline length upon the power function. To remedy this oversight I have approximated the power function curves for different baseline lengths using a series of simulation studies. Figure 1 shows these power function curves.

The horizontal scale of Figure 1 shows the size of the shift in location in standard deviation units while the vertical scale shows the probability of detecting a shift. The left end point of each curve defines the risk of a false alarm when a new point is added to the chart. The remainder of the points on each curve show the probability of detecting a shift of a given size when a new point is added to the chart following that shift. The theoretical power function curve is shown in red. The remaining curves approximate the power function when the baseline length is 25, 10, and 6.

While the length of the baseline does affect the power function for a process behavior chart, these changes are not so large as to have an appreciable impact in practice.

One way to characterize a power function is to use the size of shift that will be detected half the time. This point is known as the detectable difference 50 value, DD_{50}. The DD_{50} values are expressed as multiples of the standard deviation of the original values, so that a DD_{50} value of 3.0 indicates that a shift equal to three standard deviations of *X* will be detected half the time on the first observation following that shift. Table 1 lists the risks of a false alarm and the DD_{50} values for different baseline lengths.

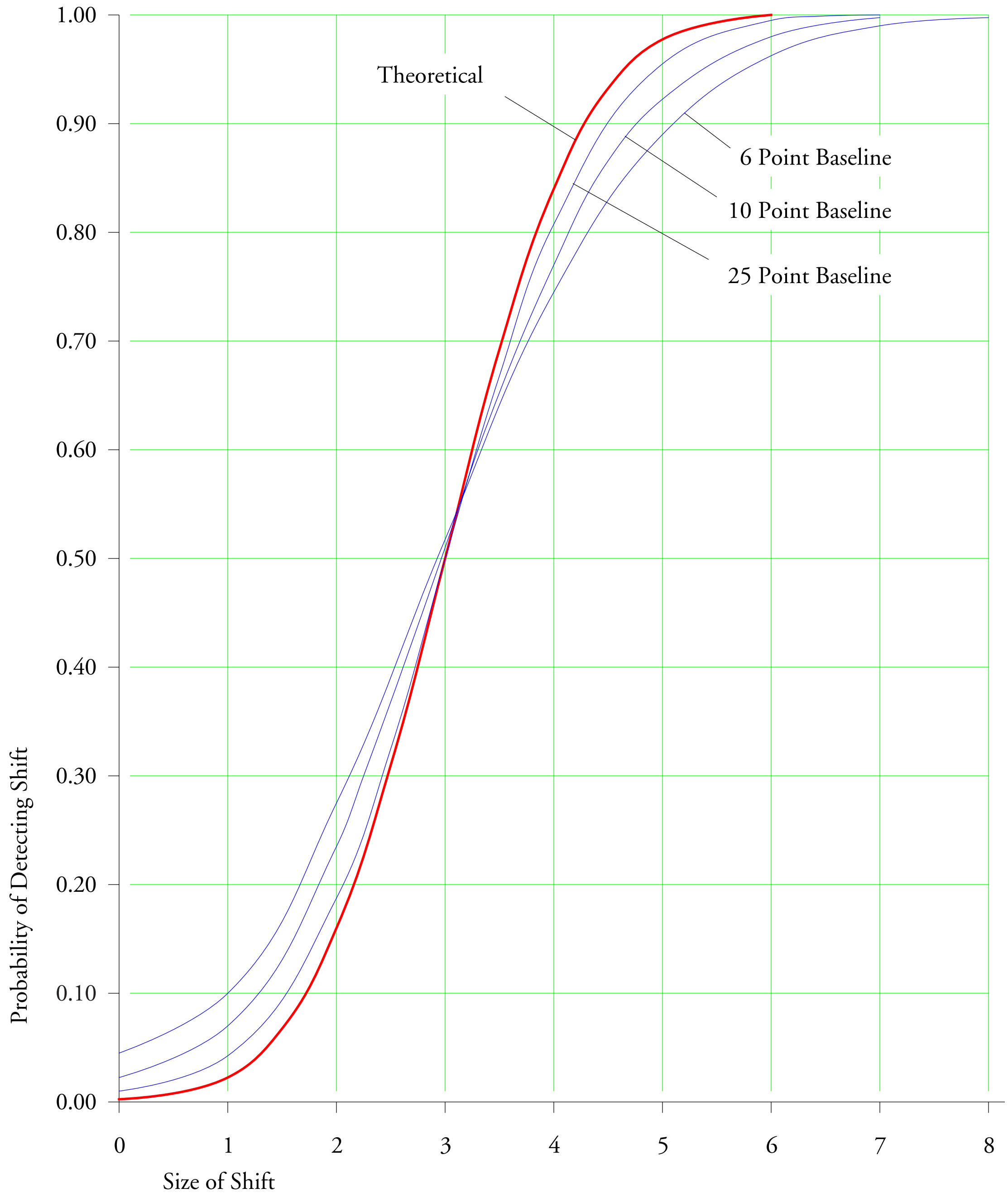

Figure 1: The Effect of Baseline Length Upon Power Function Curves for Process Behavior Charts

Figure 1 and Table 1 provide a theoretical answer to the question of how many data do we need to establish the limits for a process behavior chart. The DD_{50} values all stay reasonably close to the theoretical value of 3 sigma, suggesting that a short baseline length does not seriously affect the ability of the process behavior chart to detect signals that occur following the baseline. The only real risk to using short baselines is a slight increase in the risk of a false alarm as new data are added to the chart.

Table 1: False Alarm Risks, and DD_{50} Values for Process Behavior Charts

Baseline Length	False Alarm Risk	DD_{50}
6	4.4%	2.92
7	3.6%	2.94
8	3.1%	2.96
9	2.6%	2.96
10	2.3%	2.96
15	1.4%	2.99
20	1.0%	3.00
25	1.0%	3.00

As noted above, the description of both power and false alarm risks are given in terms of each new point added to a process behavior chart. This is correct and appropriate for a sequential procedure. However, using an *XmR* chart as a test of homogeneity is a different application of the technique, and it will require a different type of summary.

HOMOGENEITY CHARTS

When we use an *XmR* chart to assess the homogeneity of a finite set of values we are no longer performing a sequential procedure. We are essentially dumping all of the data onto the chart at once and performing a one-time analysis. To distinguish this use of an *XmR* chart from the sequential procedure above I will call this usage a homogeneity chart. While it is easy and appropriate to use an *XmR* chart in this way, the homogeneity chart is a one-time analysis. As such, we cannot use the power functions of Figure 1 to describe the properties of a homogeneity chart. Different power functions are required. These new power functions will have to consider how the one-time analysis of *k* values will detect signals within those *k* values.

The first difference between a process behavior chart and a homogeneity chart is in how we compute the risk of a false alarm (also known as the alpha-level). While the alpha levels in Figure 1 and Table 1 are given in terms of each additional point, here there will be no additional points. We are analyzing all *k* values at once, and we will need to consider the overall likelihood of *any* of the *k* values falling outside the limits by chance. This is generally referred to as the overall alpha level for the procedure. Bonferroni's inequality tells us that the overall alpha level for a homogeneity chart for *k* values will be less than or equal to *k* times the individual alpha level for each point. Thus, the overall alpha level will increase as the number of original data, *k*, increases.

The second difference in how we characterize a homogeneity chart has to do with the fact that any lack of homogeneity within the data will inflate the limits and desensitize the chart. This will have the effect of inflating the DD_{50} values for a homogeneity chart. When computing the power function for a homogeneity chart we use the simple case where one value out of the *k* values has been shifted by some amount.

The nature of the *XmR* chart also forces us to also consider where this shifted value occurs in the sequence of *k* values. If the shifted value is at either end of the chart it will be more easily detected than if it occurs within the sequence of values. However, this is not a license to change the order of the points. Because of the way the moving range works it is important to use the natural order of the data with an *XmR* chart. If you arrange the values on an *XmR* chart in numerical order you are guaranteed to have false alarms and the resulting chart will be useless. Therefore, since we cannot control where a signal may occur, we have to use the more conservative power functions and DD_{50} values for the internal points.

Moreover, since the arithmetic will prohibit an internal point from falling outside the limits of an *XmR* chart until $k = 8$, we do not consider homogeneity charts for fewer than 8 values. Figure 2 shows the power function curves for homogeneity charts having from 8 to 20 values. Table 2 lists the alpha levels and DD_{50} values for these curves.

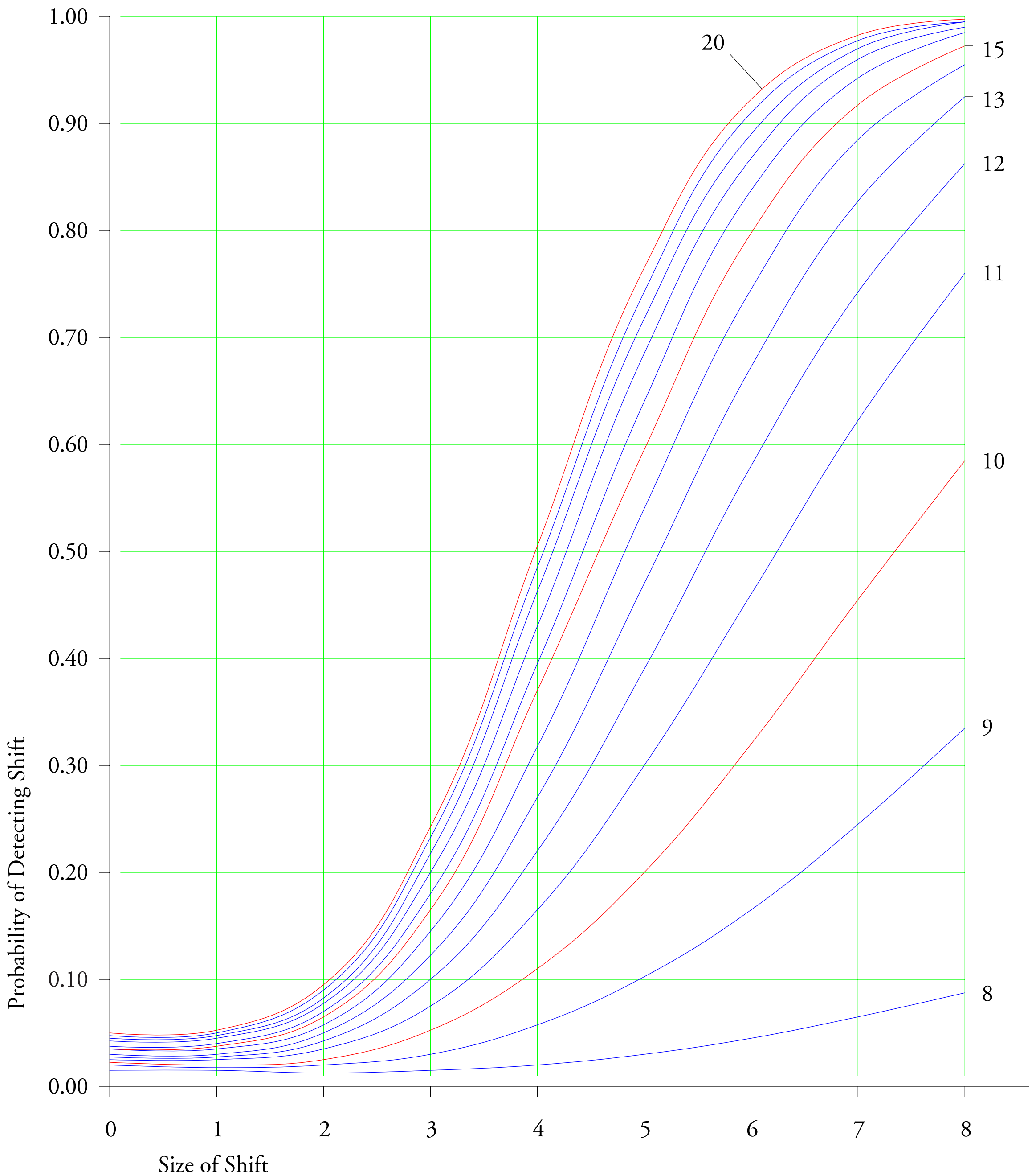

Figure 2: Power Functions for Homogeneity Charts

Even a glance will show the dramatic difference between the power functions in Figure 1 and those in Figure 2. It is this difference that makes it important to distinguish between these two different uses of an *XmR* chart.

THE W- RATIO TEST

The *W*-ratio test examines the homogeneity of a data set by looking for detectable gaps within the ordered set of values. When working with small data sets it is impractical to detect more than one such gap within the data. For this reason we shall only consider the power function of the *W*-ratio test when one value differs from the rest. Table 3 at the end of this paper gives the critical values and DD_{50} values for the *W*-ratio test for use with $k = 3$ to $k = 20$ values.

The large DD_{50} values for small k found in Table 3 emphasize the absurdity of trying to use a small alpha level with small data sets. Inferences based on small data sets are necessarily full of uncertainty. In order to have any meaningful sensitivity we must accept a larger than normal risk of a false alarm. For this reason I recommend using an alpha level of 20% with $k = 3$ and $k = 4$. With $k = 5$, 6, or 7 I recommend a minimum alpha level of 15%. With $k = 8$ or 9 I recommend an alpha level of at least 10%. For larger values of k I recommend an alpha level of at least 5%. In Table 2 I list the DD_{50} values for these recommended *W*-ratio tests.

Table 2: Comparisons of the Homogeneity Chart and the *W*-Ratio Test

Number	*XmR* Homogeneity Chart		*W*-Ratio Test	
of Data	Alpha Level	DD_{50}	Alpha Level	DD_{50}
3	—	—	20%	7.8
4	—	—	20%	5.2
5	—	—	15%	5.1
6	—	—	15%	4.7
7	—	—	15%	4.5
8	1.6%	18.0	10%	4.8
9	2.0%	9.7	10%	4.7
10	2.2%	7.3	5%	5.2
11	2.5%	6.2	5%	5.1
12	2.7%	5.6	5%	5.0
13	3.0%	5.1	5%	5.0
14	3.4%	4.8	5%	4.9
15	3.5%	4.6	5%	4.9
16	3.8%	4.4	5%	4.9
17	4.2%	4.3	5%	4.9
18	4.5%	4.1	5%	4.9
19	4.8%	4.0	5%	4.8
20	4.9%	4.0	5%	4.8

Table 2 provides a direct comparison between the alpha levels and the DD_{50} values for the homogeneity chart and the *W*-ratio test. When we have 13 or more values, the homogeneity chart will have comparable or better sensitivity with a comparable alpha-level. With fewer data the *W*-ratio test will provide greater sensitivity in return for a higher risk of a false alarm.

EXAMPLE ONE

Fifteen determinations of the charge of an electron by R. A. Millikan are given by the following values. The first question of analysis is whether or not these 15 values are homogeneous.

4805, 4801, 4808, 4809, 4791, 4788, 4783, 4740, 4792, 4810, 4799, 4797, 4790, 4806, 4790

The average is 4793.9 and the average moving range is 14.5. When we place these 15 values in their time order on an *XmR* chart we get the *X* chart shown in Figure 3. This homogeneity chart shows that the observation of 4740 is detectably different from the rest of the observations. The overall risk of a false alarm for this homogeneity chart is 3.5%.

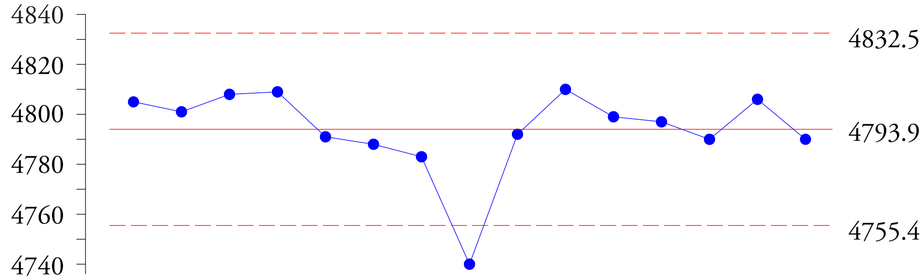

Figure 3: Homogeneity Chart for 15 Determinations of the Charge of an Electron

To use the *W*-ratio Test with Millikan's data we arrange the 15 values in numerical order:

4740 4783 4788 4790 4790 4791 4792 4797 4799 4801 4805 4806 4808 4809 4810

Find the 14 differences:

43 5 2 0 1 1 5 2 2 4 1 2 1 1

And divide these differences by the total span of the original fifteen values (110 – 40) = 70 to get the *W*-ratios:

0.614, 0.071, 0.029, 0.00, 0.014, 0.014, 0.071, 0.029, 0.029, 0.057, 0.014, 0.029, 0.014, 0.014

These *W*-ratios are compared to their respective critical values for an alpha-level of 0.05. The first *W*-ratio of 0.614 exceeds the W_1 critical value of 0.489, which indicates a detectable difference. None of the remaining *W*-ratios exceed their respective critical values. Thus, the value of 4740 is detectably different from the others and should be treated as a discrepant value.

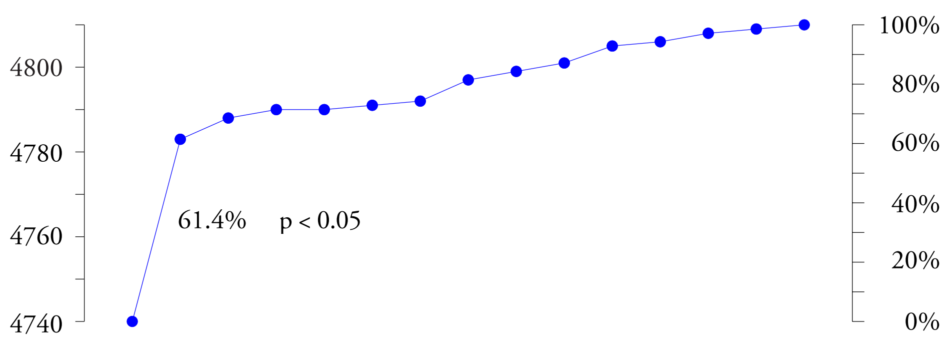

Figure 4: W-Ratio Test for 15 Determinations of the Charge of an Electron

EXAMPLE TWO

Fifteen successive determinations of the weight of a 10 gram standard at the Bureau of Standards resulted in the following values. The values recorded are the number of micrograms in excess of 9.999000 grams.

563, 582, 585, 596, 599, 599, 593, 588, 625, 591, 594, 602, 594, 597, 596

The homogeneity chart for these fifteen determinations of the weight of the standard has an average of 593.6 micrograms and an average moving range of 10.07 micrograms. The homogeneity chart is shown in Figure 5, where we find two weighings to be detectably different from the rest. The alpha-level for this homogeneity chart is 3.5%.

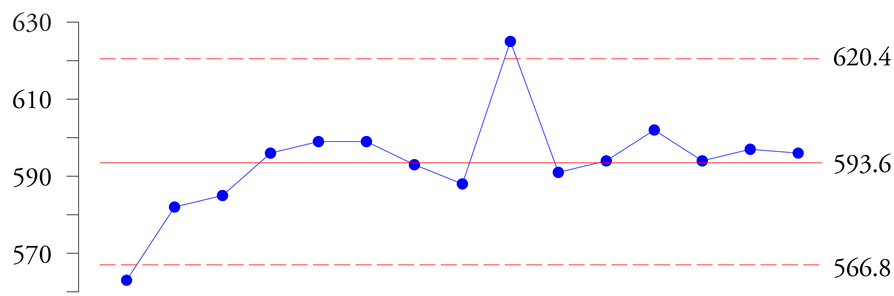

Figure 5: Homogeneity Chart for 15 Determinations of the Weight of a Standard

To use the *W*-ratio test we arrange these 15 values in numerical order:

563, 582, 585, 588, 591, 593, 594, 594, 596, 596, 597 599, 599, 602, 625

We compute the 14 differences:

19, 3, 3, 3, 2, 1, 0, 2, 0, 1, 2, 0, 3, 23

We divide by the total span of 62 to get the *W*-ratios:

0.306, 0.048, 0.048, 0.048, 0.032, 0.016, 0.000, 0.032, 0.000, 0.016, 0.032, 0.000, 0.048, 0.371

These *W*-ratios are compared with their respective critical values and none are found to exceed their critical values regardless of what alpha level is used. Here the fact that there was an inconsistent reading on both the high and low sides fooled the computations of the *W*-ratio by inflating the total span of the data.

If we restrict our analysis to the first 14 values in the ordered set above we get a total span of 39 which results in the following *W*-ratios:

0.487, 0.077, 0.077, 0.077, 0.051, 0.026, 0.000, 0.051, 0.000, 0.026, 0.051, 0.000, 0.077

And the first *W*-ratio exceeds the 10% W_1 critical value of 0.465. Thus, the reading of 563 may be said to be detectably different from the other 13 values examined in this analysis.

If we then restrict our analysis to the last 14 values in the ordered set above we get a total span of 43 which results in the following *W*-ratios:

0.070, 0.070, 0.070, 0.047, 0.023, 0.000, 0.047, 0.000, 0.023, 0.047, 0.000, 0.070, 0.535

And the last *W*-ratio exceeds the 5% W_{13} critical value of 0.500. Thus, the reading of 625 is detectably different from the other 13 values examined in this analysis. Thus, with two separate analyses we find the two extreme weighings to be detectably different from the others.

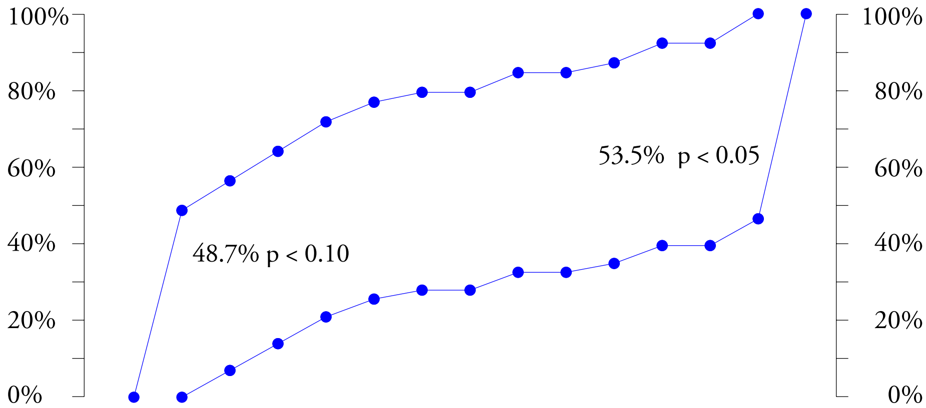

Figure 6: W-Ratio Tests for 15 Determinations of the Weight of a Standard

SUMMARY

While both the *W*-ratio test and the homogeneity chart examine the data for evidence of a lack of homogeneity, they do so in somewhat different ways. The *W*-ratio test can be used with fewer data. With more than thirteen data the homogeneity chart will provide greater sensitivity with a comparable overall alpha level. With twelve or fewer data the *W*-ratio test is recommended.

Table 3: Critical Values for the *W*-Ratio Test

k	alpha	DD_{50}	W_1	W_2
3	1%	155	0.994	0.994
	5%	30.9	0.970	0.970
	10%	15.5	0.941	0.941
	15%	10.4	0.913	0.913
	20%	**7.8**	**0.885**	**0.885**

k	alpha	DD_{50}	W_1	W_2	W_3
4	1%	23.5	0.934	0.919	0.934
	5%	10.6	0.860	0.828	0.860
	10%	7.4	0.806	0.765	0.806
	15%	6.0	0.766	0.719	0.766
	20%	**5.2**	**0.733**	**0.681**	**0.733**

k	alpha	DD_{50}	W_1	W_2	W_3	W_4
5	1%	13.1	0.859	0.819	0.819	0.859
	5%	7.5	0.766	0.707	0.707	0.766
	10%	5.9	0.710	0.641	0.641	0.710
	15%	**5.1**	**0.672**	**0.597**	**0.597**	**0.672**
	20%	4.5	0.642	0.563	0.563	0.642

k	alpha	DD_{50}	W_1	W_2	W_3	W_4	W_5
6	1%	9.8	0.793	0.732	0.716	0.732	0.793
	5%	6.4	0.698	0.620	0.598	0.620	0.698
	10%	5.3	0.646	0.560	0.536	0.560	0.646
	15%	**4.7**	**0.612**	**0.522**	**0.496**	**0.522**	**0.612**
	20%	4.3	0.585	0.492	0.467	0.492	0.585

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6
7	*1%*	8.2	0.740	0.666	0.636	0.636	0.666	0.740
	5%	5.8	0.649	0.559	0.525	0.525	0.559	0.649
	10%	5.0	0.601	0.505	0.469	0.469	0.505	0.601
	15%	**4.5**	**0.570**	**0.470**	**0.434**	**0.434**	**0.470**	**0.570**
	20%	4.1	0.545	0.444	0.408	0.408	0.444	0.545

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7
8	*1%*	7.4	0.698	0.612	0.574	0.566	0.574	0.612	0.698
	5%	5.5	0.613	0.515	0.473	0.462	0.473	0.515	0.613
	10%	**4.8**	**0.567**	**0.465**	**0.423**	**0.411**	**0.423**	**0.465**	**0.567**
	15%	4.4	0.537	0.434	0.392	0.380	0.392	0.434	0.537
	20%	4.0	0.515	0.410	0.368	0.357	0.368	0.410	0.515

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
9	*1%*	7.1	0.669	0.573	0.530	0.512	0.512	0.530	0.573	0.669
	5%	5.3	0.584	0.482	0.434	0.414	0.414	0.434	0.482	0.584
	10%	**4.7**	**0.541**	**0.436**	**0.389**	**0.369**	**0.369**	**0.389**	**0.436**	**0.541**
	15%	4.3	0.514	0.406	0.360	0.341	0.341	0.360	0.406	0.514
	20%	4.0	0.493	0.385	0.339	0.321	0.321	0.339	0.385	0.493

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9
10	*1%*	6.6	0.636	0.542	0.492	0.469	0.444	0.469	0.492	0.542	0.636
	5%	5.2	0.560	0.455	0.406	0.381	0.374	0.381	0.406	0.455	0.560
	10%	4.6	0.519	0.412	0.363	0.339	0.332	0.339	0.363	0.412	0.519
	15%	4.3	0.493	0.386	0.337	0.313	0.307	0.313	0.337	0.386	0.493
	20%	4.0	0.473	0.366	0.318	0.295	0.288	0.295	0.318	0.366	0.473

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5
11	*1%*	6.4	0.616	0.514	0.461	0.435	0.419
	5%	5.1	0.541	0.433	0.381	0.354	0.342
	10%	4.6	0.503	0.393	0.342	0.316	0.304
	15%	4.2	0.478	0.368	0.318	0.293	0.281
	20%	4.0	0.459	0.350	0.300	0.276	0.265
			W_{10}	W_9	W_8	W_7	W_6

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6
12	*1%*	6.2	0.596	0.492	0.439	0.408	0.392	0.389
	5%	5.0	0.525	0.417	0.363	0.333	0.317	0.315
	10%	4.5	0.488	0.379	0.326	0.298	0.283	0.280
	15%	4.2	0.464	0.355	0.303	0.276	0.261	0.258
	20%	4.0	0.447	0.337	0.287	0.260	0.247	0.243
			W_{11}	W_{10}	W_9	W_8	W_7	W_6

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6
13	*1%*	6.1	0.581	0.476	0.419	0.386	0.369	0.362
	5%	5.0	0.512	0.401	0.346	0.316	0.299	0.292
	10%	4.5	0.475	0.365	0.312	0.283	0.266	0.259
	15%	4.2	0.453	0.342	0.291	0.262	0.246	0.239
	20%	4.0	0.436	0.327	0.275	0.247	0.232	0.226
			W_{12}	W_{11}	W_{10}	W_9	W_8	W_7

k	*alpha*	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7
14	*1%*	6.0	0.568	0.461	0.404	0.370	0.351	0.340	0.337
	5%	4.9	0.500	0.388	0.333	0.301	0.283	0.273	0.270
	10%	4.5	0.465	0.354	0.300	0.270	0.252	0.244	0.240
	15%	4.2	0.443	0.332	0.280	0.251	0.233	0.225	0.222
	20%	3.9	0.426	0.317	0.265	0.237	0.220	0.212	0.210
			W_{13}	W_{12}	W_{11}	W_{10}	W_9	W_8	W_7

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7
15	**1%**	5.9	0.553	0.444	0.389	0.351	0.334	0.319	0.311
	5%	4.9	0.489	0.378	0.323	0.290	0.271	0.258	0.253
	10%	4.5	0.455	0.344	0.290	0.260	0.241	0.230	0.225
	15%	4.2	0.433	0.324	0.271	0.241	0.223	0.213	0.208
	20%	4.0	0.417	0.309	0.257	0.228	0.211	0.201	0.196
			W_{14}	W_{13}	W_{12}	W_{11}	W_{10}	W_9	W_8

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
16	**1%**	5.8	0.545	0.438	0.379	0.343	0.320	0.308	0.297	0.293
	5%	4.9	0.480	0.368	0.313	0.279	0.258	0.246	0.239	0.238
	10%	4.4	0.447	0.337	0.282	0.251	0.231	0.219	0.213	0.211
	15%	4.2	0.426	0.316	0.263	0.233	0.214	0.203	0.197	0.195
	20%	4.0	0.411	0.302	0.249	0.221	0.203	0.192	0.186	0.184
			W_{15}	W_{14}	W_{13}	W_{12}	W_{11}	W_{10}	W_9	W_8

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
17	**1%**	5.7	0.532	0.422	0.366	0.328	0.307	0.291	0.281	0.277
	5%	4.9	0.471	0.359	0.303	0.270	0.247	0.235	0.227	0.224
	10%	4.4	0.439	0.328	0.274	0.243	0.222	0.210	0.202	0.200
	15%	4.2	0.419	0.309	0.257	0.226	0.207	0.195	0.188	0.185
	20%	4.0	0.405	0.295	0.244	0.214	0.195	0.184	0.177	0.174
			W_{16}	W_{15}	W_{14}	W_{13}	W_{12}	W_{11}	W_{10}	W_9

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9
18	**1%**	5.7	0.527	0.417	0.361	0.323	0.298	0.281	0.273	0.266	0.266
	5%	4.9	0.463	0.352	0.296	0.261	0.241	0.226	0.217	0.213	0.212
	10%	4.4	0.432	0.322	0.268	0.236	0.215	0.203	0.194	0.190	0.188
	15%	4.2	0.413	0.304	0.251	0.219	0.200	0.188	0.180	0.176	0.175
	20%	4.0	0.399	0.290	0.239	0.208	0.189	0.178	0.170	0.166	0.165
			W_{17}	W_{16}	W_{15}	W_{14}	W_{13}	W_{12}	W_{11}	W_{10}	W_9

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9
19	**1%**	5.7	0.516	0.405	0.349	0.311	0.286	0.271	0.258	0.254	0.251
	5%	4.8	0.457	0.345	0.288	0.255	0.234	0.220	0.209	0.204	0.201
	10%	4.4	0.426	0.316	0.262	0.229	0.209	0.196	0.187	0.182	0.179
	15%	4.2	0.408	0.299	0.246	0.214	0.195	0.182	0.173	0.168	0.166
	20%	4.0	0.394	0.285	0.234	0.203	0.184	0.172	0.164	0.159	0.156
			W_{18}	W_{17}	W_{16}	W_{15}	W_{14}	W_{13}	W_{12}	W_{11}	W_{10}

k	**alpha**	DD_{50}	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}
20	**1%**	5.7	0.509	0.396	0.336	0.301	0.276	0.260	0.248	0.240	0.236	0.236
	5%	4.8	0.452	0.339	0.282	0.248	0.227	0.211	0.202	0.194	0.192	0.189
	10%	4.4	0.421	0.311	0.257	0.224	0.204	0.190	0.180	0.174	0.171	0.169
	15%	4.2	0.403	0.294	0.241	0.210	0.190	0.177	0.168	0.162	0.159	0.157
	20%	4.0	0.389	0.281	0.229	0.199	0.180	0.167	0.159	0.153	0.150	0.149
			W_{19}	W_{18}	W_{17}	W_{16}	W_{15}	W_{14}	W_{13}	W_{12}	W_{11}	W_{10}