Testing for Randomness

The Mean Successive Difference Test

BY DEAN V. NEUBAUER

Q: How can I tell if my data are random — or not?

A. Many of us get data sets with the assumption that the process was in control, i.e., the data are random about a target; or suspicion that the process was out of control, i.e., the data show some indication of nonrandomness. An obvious solution is to simply plot the data on a control chart and determine whether or not the process was indeed in control. Of course, you must first know how to construct the chart and its control limits. E2587, Standard Practice for Use of Control Charts in Statistical Process Control, or ASTM's *Manual on Presentation of Data and Control Chart Analysis*, 8th Ed., will provide you with the information you need to solve the problem graphically.

But suppose that you don't have a copy of E2587, Manual 7 or a computer program such as Minitab available to help you plot the data and check for process stability (randomness). In this case, you can use a simple statistic called the mean successive difference test. Let's suppose that we have n data values, x_1, x_2, ..., x_n. Successive differences are just the differences $x_2 - x_1$, $x_3 - x_2$, ... $x_{n-1} - x_n$. (In the field of quality control, the absolute value of these differences are just the moving ranges of $n = 2$ typically used to construct the control limits for the individual control chart.) When we square these differences and average them, that gives us the mean successive difference. In order to construct a test statistic, we will divide the MSD by the sample variance. The result is a ratio of the sum of squares of the MSD and usual sum of squares of the sample variance. The formula looks like this:

$$M = \frac{MSD}{s^2} = \frac{\frac{1}{n-1}\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2}{\frac{1}{n-1}\sum_{i=1}^{n}(x_1 - \bar{x})^2} = \frac{\sum_{i=1}^{n-1}(x_{i+1} - x_i)^2}{\sum_{i=1}^{n}(x_1 - \bar{x})^2}$$

We can use M to detect nonrandomness in any sequence of observations, typically referred to as serial correlation. If the data come from an in-control process, then the average value for M is 2. If the observations fluctuate excessively, e.g., saw-tooth pattern, then M will be large. Conversely, if the observations exhibit a long-term cycle, then M will be small. So, how can you determine whether M is too large or too small statistically? Table 1 shows lower and upper critical values for M for a variety of sample sizes and levels of significance (α), e.g., α = 0.05 level of significance is 95 percent confidence. (Note that the upper critical values are simply obtained from $U_M = 4 - L_M$.)

The mean successive difference test as described by Bennett and Franklin can address two aspects of process capability studies.[1] First, the test can identify the type of nonrandomness present (too many or too few cycles). Second, given that there may be cycles or trends, the mean successive difference provides an estimate of the process variance about the cycle or trend line. In other words, the MSD implies the value of the process variance if the cycle or trend can be eliminated.

EXAMPLE

Bennett and Franklin provide a dataset of plant yields that we can use to test for serial correlation. Data for $n = 25$ consecutive weeks were: 81.02, 80.08, 80.05, 79.70, 79.13, 77.09, 80.09, 79.40, 80.56, 80.97, 80.17, 81.35, 79.64, 80.82, 81.26, 80.75, 80.74, 81.59, 80.14, 80.75, 81.01, 79.09, 78.73, 78.45 and 79.56. The sum of squares of the 24 successive differences is 31.6772, and the usual sum of squares associated with the sample variance is 25.1343. Thus, $M = 31.6772/25.1343 = 1.26$. Table 1 shows that M is between the critical values of the significance levels of 0.05 ($L_M = 1.367$) and 0.01 ($L_M = 1.128$). Since the value of M is less than 2, we can conclude that there is a slow

nonrandom fluctuation in plant yield values over this 25-week period with at least 95 percent confidence. Figure 1 shows an individual control chart from Minitab where the control limits are based on the MSD. Here we can see the slow cyclical pattern in the data suggested by M.

Figure 1 — Individual Control Chart for Plant Yield Data
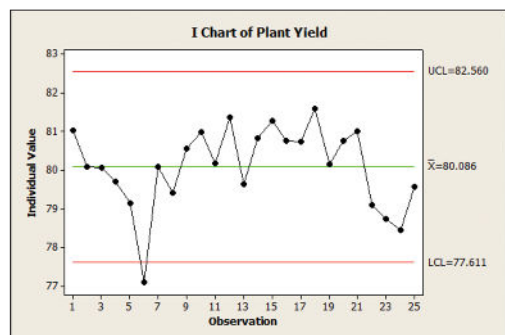
REFERENCE

1. Bennett, C.A. and Franklin, N.L., *Statistical Analysis in Chemistry and the Chemical Industry*, John Wiley and Sons, New York, 1954, p. 679.

DEAN V. NEUBAUER, *Corning Inc., Corning, N.Y., coordinates the DataPoints column; an ASTM International fellow, he is chairman of Committee E11 on Quality and Statistics and chairman of E11.90.03 on Publications.*

snonline

Get more tips for ASTM standards development at www.astm.org/sn-tips.

Find other DataPoints articles at www.astm.org/standardization-news/datapoints.

Table 1 — Critical Values for the Mean Successive Difference Test

	Lower Critical Values, L_M			Upper Critical Values, U_M		
n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
10	1.251	1.062	0.752	2.749	2.938	3.248
11	1.280	1.096	0.792	2.720	2.904	3.208
12	1.306	1.128	0.828	2.694	2.872	3.172
13	1.329	1.156	0.862	2.671	2.844	3.138
14	1.351	1.182	0.893	2.649	2.818	3.107
15	1.370	1.205	0.922	2.630	2.795	3.078
16	1.388	1.227	0.949	2.612	2.773	3.051
17	1.405	1.247	0.974	2.595	2.753	3.026
18	1.420	1.266	0.998	2.580	2.734	3.002
19	1.434	1.283	1.020	2.566	2.717	2.980
20	1.447	1.300	1.041	2.553	2.700	2.959
25	1.502	1.367	1.128	2.498	2.633	2.872
30	1.543	1.418	1.195	2.457	2.582	2.805
40	1.602	1.492	1.293	2.398	2.508	2.707
50	1.642	1.544	1.363	2.358	2.456	2.637
60	1.673	1.582	1.415	2.327	2.418	2.585
70	1.697	1.612	1.457	2.303	2.388	2.543
80	1.716	1.636	1.490	2.284	2.364	2.51
90	1.732	1.657	1.518	2.268	2.343	2.482
100	1.745	1.674	1.542	2.255	2.326	2.458
200	1.819	1.768	1.674	2.181	2.232	2.326
300	1.852	1.811	1.733	2.148	2.189	2.267
400	1.872	1.836	1.768	2.128	2.164	2.232
500	1.886	1.853	1.793	2.114	2.147	2.207
600	1.895	1.866	1.811	2.105	2.134	2.189
800	1.909	1.884	1.836	2.091	2.116	2.164
1000	1.919	1.896	1.853	2.081	2.104	2.147
	2.000	2.000	2.000	2.000	2.000	2.000