CONTROL CHART FOR LARGEST AND SMALLEST VALUES

By John M. Howell

Los Angeles City College

1. Introduction. It may at times be desirable to use a control chart for largest and smallest values (L & S) in place of the conventional charts for averages and ranges ($\bar{X}$ & R). The chart for largest and smallest values has certain advantages: all information may be combined on one chart, computations are simple, and specifications may be placed on the chart. In this paper, constants for the use of this chart are developed and comparison is made with the average and range charts.

2. Constants for determining limits. Let L and S denote the largest and smallest values, respectively, in a sample of n pieces, and let $\bar{L}$ and $\bar{S}$ denote the averages of these values for k samples. Then $(\bar{L} + \bar{S})/2$ and $(\bar{L} - \bar{S})/d_2$ are unbiased estimates of the population mean and standard deviation, respectively, in the case of a random sample from a normal population. The value of the constant d_2 is given in [1] and repeated in table 1 for convenience. If we denote $(\bar{L} + \bar{S})/2$ by M and $(\bar{L} - \bar{S})$ by $\bar{R}$, control limits may be determined in terms of these statistics.

In conformance with usual control chart practice, we will set the upper control limit at $\bar{L} + 3\hat{\sigma}_L$ and the lower control limit at $\bar{S} - 3\hat{\sigma}_S$, where $\hat{\sigma}_L$ is an estimate of the standard deviation of the largest values in samples drawn from a normal population, and similarly for $\hat{\sigma}_S$. The results of Tippett [2] and Pearson [3] for $E(R)$ of samples from a normal population were used to determine expected values of L and S: $E(R) = d_2\sigma$. Here, R is the range of samples of size n: $R = L - S$. But since $E[(L + S)/2] = a$ for a symmetrical distribution, then $E(L) = a + d_2\sigma/2$ and $E(S) = a - d_2\sigma/2$, where a and σ are the mean and standard deviation of the normal population from which samples are drawn.

The probability element of the largest value [4] is given by:

$$n[F(L)]^{n-1}f(L)\,dL \quad \text{where} \quad f(x) = 1/\sqrt{2\pi}\,\sigma\, e^{-(x-a)^2/2\sigma^2} \quad \text{and} \quad F(x) = \int_{-\infty}^{x} f(y)\,dy.$$

Then $E(L^2) = n \int_{-\infty}^{\infty} L^2[F(L)]^{n-1}f(L)\,dL$. Integrals of this type, differing only by a constant factor have been evaluated by Hojo [5] and from his results d_4 was determined so that $\sigma_L = \sigma_S = d_4\sigma$. Values for d_4 for $n = 2, 5, 10$ are also given by Tippett [2]. "Three-sigma" control limits may then be given in the form: $M \pm A_3\bar{R}$, where $A_3 = 0.5 + 3d_4/d_2$. The expected value of the upper control limit will then be: $E(UCL) = a + A_4\sigma$, where $A_4 = (d_2/2) + 3d_4$. Values of these constants for various sample sizes are given in Table I.

In practice, it might be desired, in the case of control charts for individual measurements or for L and S, to have $E(UCL) = a + 3\sigma$, and the lower control limit symmetrically placed with respect to the central line. In this case, the formula for the limits would be: $M \pm 3\bar{R}/d_2$ or $M \pm \sqrt{n}A_2\bar{R}$, where $A_2 =$

$3/(d_2\sqrt{n})$ is given in [1]. Since the efficiency of M decreases rapidly with increasing sample size [6], it would probably be better to use $\bar{\bar{X}}$ in place of M for determining the central line for a control chart when the sample size is greater than five. $\bar{\bar{X}}$ is the "average of averages" as defined in [1].

The chart for largest and smallest values would then consist of a chart on which both the largest and smallest values are plotted, with the central line at M, and the limits as given above.

3. Comparison of charts for a particular case. A comparison of the L & S chart with the $\bar{X}$ chart for a particular case in which the sample size was three is given in Fig. 1. Measurements were the shear strength of spotweld coupons of aluminum in pounds. Since the range chart had no points above the "three-sigma" control limit and showed no other peculiarities, it has been omitted.

TABLE I

Constants for largest and smallest value chart

n	d_2	d_4	A_2	A_3	A_4	n
2	1.128	.825	1.880	2.72	3.03	2
3	1.693	.748	1.023	1.82	3.09	3
4	2.059	.709	.729	1.53	3.15	4
5	2.326	.670	.577	1.36	3.17	5
6	2.534	.648	.483	1.27	3.21	6
7	2.704	.627	.419	1.20	3.23	7
8	2.847	.614	.373	1.15	3.26	8
9	2.970	.600	.337	1.10	3.28	9
10	3.076	.588	.308	1.07	3.30	10

4. General comparison of charts. We assume a mean of zero and a standard deviation of unity as a "given standard," and then compute the probabilities when the true values are a and σ. The probability of a point being inside of "3-sigma" control limits on the range chart under these conditions is: $P_1 = \Pr(R < d_2D_4/\sigma)$, where D_4 is given in [1]. The probabilities for the range used here were found from the Pearson-Hartley tables [3]. The usual normality assumptions are made.

The probability of a point being inside of "3-sigma" control limits on the average chart under the same conditions is:

$$P_2 = \int_{\sqrt{n}/\sigma\,((-3/\sqrt{n})-a)}^{\sqrt{n}/\sigma\,((3/\sqrt{n})-a)} \varphi(t)\,dt \quad \text{where} \quad \varphi(t) = \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2}.$$

Since Daly [7] has shown that the average and range of samples from a normal

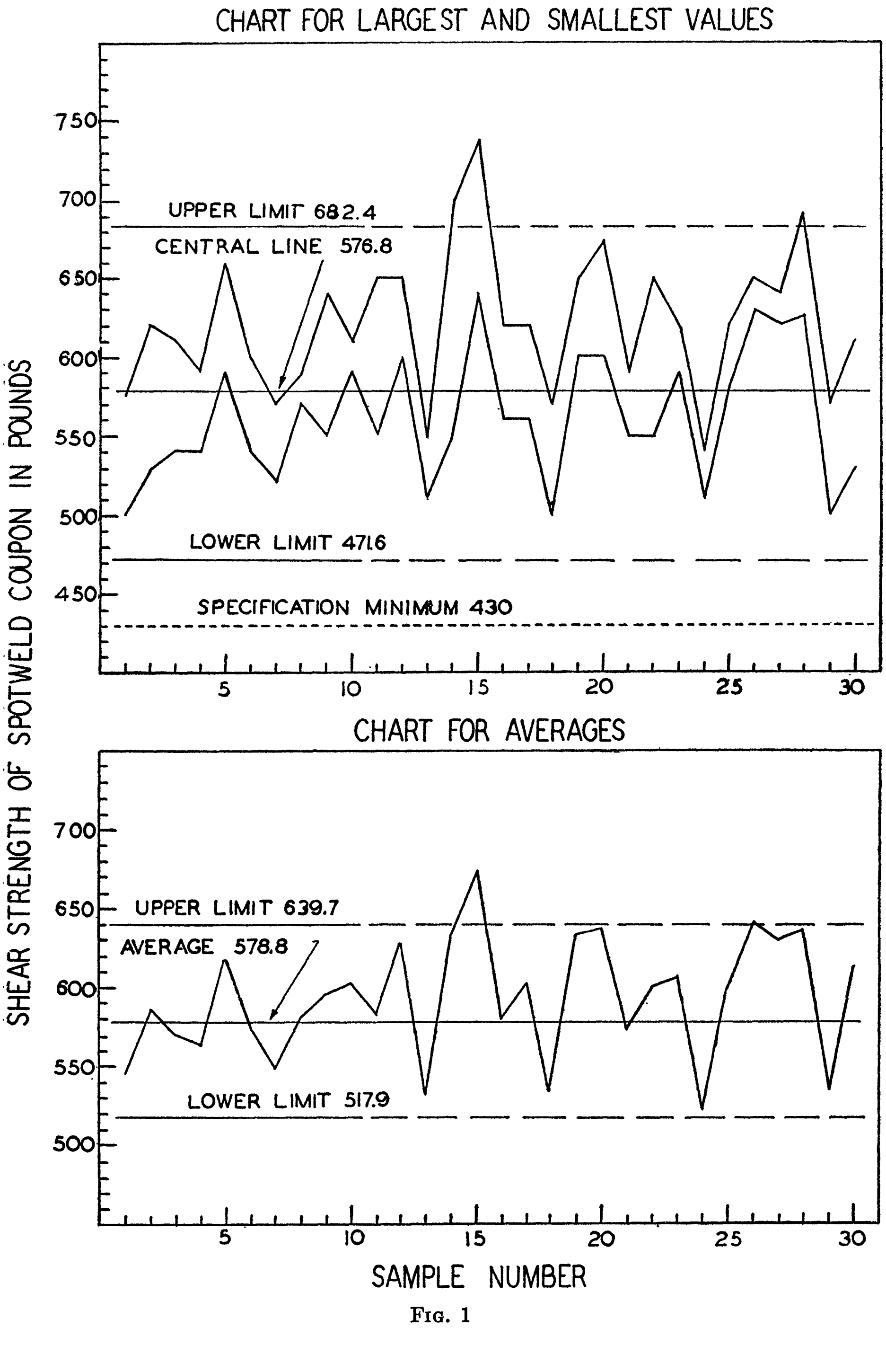
CHART FOR LARGEST AND SMALLEST VALUES
UPPER LIMIT 682.4
CENTRAL LINE 576.8
LOWER LIMIT 471.6
SPECIFICATION MINIMUM 430
750
700
650
600
550
500
450
5
10
15
20
25
30
SHEAR STRENGTH OF SPOTWELD COUPON IN POUNDS
CHART FOR AVERAGES
UPPER LIMIT 639.7
AVERAGE 578.8
LOWER LIMIT 517.9
700
650
600
550
500
SAMPLE NUMBER

Fig. 1

TABLE II

n	a	σ	P_1	P_2	P_1P_2	P_3	N_1	N_2
3	0	1.0	.994	.997	.991	.991	510	510
		1.2	.973	.988	.961	.963	116	122
		1.5	.901	.955	.860	.868	31	33
		2.0	.721	.866	.624	.645	10	11
3	0.5	1.0	.994	.983	.977	.980	198	228
		1.2	.973	.935	.935	.939	69	74
		1.5	.901	.917	.826	.834	25	27
		2.0	.721	.830	.598	.694	9	13
3	1.0	1.0	.994	.898	.893	.931	41	65
		1.2	.973	.855	.832	.860	25	31
		1.5	.901	.802	.723	.740	15	17
		2.0	.721	.746	.538	.550	8	8
3	2.0	1.0	.994	.323	.321	.590	5	9
		1.2	.973	.352	.342	.510	5	7
		1.5	.901	.378	.341	.414	5	6
		2.0	.721	.408	.294	.321	4	5
5	0	1.0	.995	.997	.992	.992	570	570
		1.2	.969	.988	.957	.957	105	105
		1.5	.855	.955	.817	.878	23	36
		2.0	.588	.866	.509	.545	7	8
5	0.5	1.0	.995	.970	.965	.980	130	227
		1.2	.969	.942	.913	.927	51	62
		1.5	.855	.891	.762	.791	17	20
		2.0	.588	.805	.473	.505	7	7
5	1.0	1.0	.995	.776	.722	.923	15	58
		1.2	.969	.736	.713	.828	14	25
		1.5	.855	.695	.594	.661	9	12
		2.0	.588	.648	.381	.426	5	6
5	2.0	1.0	.995	.071	.071	.512	2	7
		1.2	.969	.110	.107	.402	3	6
		1.5	.855	.164	.140	.286	3	4
		2.0	.588	.230	.135	.185	3	3

population are independent, the probability that a sample is within control limits on both charts is the product of the probabilities: P_1P_2. Thus the probability that a sample be outside of control limits on either chart is $1 - P_1P_2$.

The probability of the largest and smallest values both lying in the interval from $-c$ to c is: $P_3 = \Pr(-c < S, L < c) = \left[\int_{(-c-a)/\sigma}^{(c-a)/\sigma} \varphi(t)\, dt\right]^n$. Values of this expression with lower limit $-\infty$ are given in table XXI of [8] for sample of sizes 3, 5, and 10. For the purpose of comparing the charts, we choose c so that the probabilities of Type 1 errors are equal, that is: $1 - P_1P_2 = 1 - P_3$ or $P_1P_2 = P_3$ when the mean is zero and the standard deviation unity. Substituting in this equation and solving, we find: $F(c) = 0.5 + 0.5\,(.9973P_1)^{1/n}$, where $F(x) = \int_{-\infty}^{x} \varphi(t)\, dt$. For $n = 3$, $c = 2.99$ and for $n = 5$, $c = 3.15$.

Comparing P_1P_2 with P_3 when the true values are a and σ will then show the relative power of the $\bar{X}$ & R charts and the L & S chart for detecting lack of control.

Finally the charts are compared by finding the number (N_1 for the $\bar{X}$ & R charts and N_3 for the L & S chart) of samples which will detect lack of control with a .99 probability under the conditions given above. This is done by finding the smallest integer which satisfies the following inequalities: $(P_1P_2)^{N_1} < .01$ and $P_3^{N_3} < .01$. As may be seen from table II, under most conditions, the L & S chart is nearly as good as the $\bar{X}$ & R charts for detecting lack of control.

REFERENCES

[1] AMERICAN STANDARDS ASSOCIATION, *Control Chart Method of Controlling Quality during Production*, Z1.3—1942.

[2] L. H. C. TIPPETT, "On the extreme individuals and the range of samples taken from a normal population," *Biometrika*, Vol. 18 (1925), pp. 364–387.

[3] E. S. PEARSON, "The probability integral of the range in samples of n observations from a normal population," *Biometrika*, Vol. 32 (1942), pp. 301–308.

[4] S. S. WILKS, *Mathematical Statistics*, Princeton University Press, 1943, p. 91.

[5] HOJO, "Distribution of median from a normal population," *Biometrika*, Vol. 23 (1931), p. 315.

[6] W. A. SHEWHART, *Economic Control of Quality of Manufactured Product*, D. Van Nostrand Co., 1931, p. 282.

[7] J. F. DALY, "On the use of the sample range in an analogue of Student's t-test," *Annals of Math. Stat.*, Vol. 17 (1946), pp. 71–74.

[8] KARL PEARSON, *Tables for Statisticians and Biometricians*, Cambridge University Press, 1914.

SUFFICIENCY, TRUNCATION AND SELECTION[1]

BY JOHN W. TUKEY

Princeton University

1. Summary. The fact that the mean and variance were sufficient statistics for a univariate normal distribution truncated at a fixed point was known to

[1] Prepared in connection with work sponsored by the Office of Naval Research.