acid and glycerophosphoric acid. The compound concerned is evidently a glyceroinositophosphatidic acid, which is perhaps identical with the one previously found in soybean phosphatides (13).

The third weight-curve peak, around transfer No. 250, is like the last-mentioned peak, reflected both in the phosphorus and in the glycerol curve but neither in the nitrogen nor in the meso-inositol curve. As the molar ratio phosphorus:fatty acid is about 1:2 and since glycerophosphoric acid is the only phosphoric acid ester occurring in an acid hydrolysate, it would be reasonable to assume that the compound concerned was an ordinary glycerophosphatidic acid. The position of the peak does not however correspond to the position it should have if this were true, and electrometric titration shows it to be a monobasic acid with an equivalent weight of about 2,700. After being washed with acid for a short period, it may be titrated as a dibasic acid with the same equivalent weight. These facts suggest the presence of a polyglycerophosphatidic acid.

The fourth weight-curve peak occurs at about tube No. 40. It is reflected in the glycerol, phosphorus, and nitrogen curves. The molar ratio between these compounds is very nearly equal to 1. Moreover this peak coincides with a peak of the fatty acid curve. The molar ratio of fatty acid:phosphorus is about 2:1. Paper chromatographic separation of the substances released by acid hydrolysis (6 N HCl for 24 hrs.) shows that the only substances present are choline, ethanolamine, and serine. Therefore it is a reasonable assumption that what we are dealing with is a mixture of phosphatidylcholine, phosphatidylethanolamine, and phosphatidylserine.

Summary

The object of the present work has been to study those soybean phosphatides which cannot be extracted by means of nonpolar solvents but only by means of mixtures of nonpolar and polar solvents, for instance hexane and ethanol. These phosphatides were fractionated by the countercurrent distribution technique, and the following groups of substances were found: a) carbohydrates with d-inositol in the form of the methyl ether called pinitol; b) a number of nitrogen-containing substances, the nature of which is not as yet fully elucidated but which is perhaps merely decomposition products of proteins; c) a glyceroinositophosphatidic acid which contains equimolar quantities of glycerophosphoric acid and inositolmonophosphoric acid and phosphorus and fatty acids at a ratio of about 1 to 2; d) a high-molecular phosphatidic acid; and e) a mixture of the three glycerophosphatides: phosphatidylcholine, -ethanolamine, and -serine.

REFERENCES

1. Buchanan, J. G., Dekker, C. A., and Long, A. G., J. Chem. Soc., 3162–3167 (1950).
2. Carter, H. E., Celmer, W. D., Lands, W. E. M., Mueller, K. L., and Tomizawa, H. H., J. Biol. Chem., *206*, 613–623 (1954).
3. Carter, H. E., Celmer, W. D., Galanos, D. S., Gigg, R. H., Lands, W. E. M., Law, J. H., Mueller, K. L., Nakayama, T., Tomizawa, H. H., and Weber, E., J. Am. Oil Chemists' Soc., *35*, 335 (1958).
4. Cole, P. G., Lathe, G. H., and Ruthven, C. R. J., Biochem. J., *54*, 449–457 (1953).
5. Fleury, P., Courtois, J. E., and Malangeau, P., Bull. soc. chim. biol., *35*, 537–540 (1953).
6. Folch, J., Proc. Fed. Am. Soc. Exptl. Biol., *6* (No. 1), Part II, 252 (1947).
7. Hawthorne, J. N., and Chargaff, E., J. Biol. Chem., *206*, 27–37 (1954).
8. Hill, U. T., Ind. Eng. Chem., Anal. Ed., *18*, 317–319 (1946).
9. Lovern, J. A., Nature, *169*, 969 (1952).
10. Mitchell, J., "Organic Analysis," Vol. 1, Interscience Publishers Inc., New York, 1953.
11. Nelson, N., Chem. Abstr., *38*, 4634 (1944); J. Biol. Chem., *153*, 375–380 (1944).
12. Nielsen, K., "Studies on the Nonhydratable Soybean Phosphatides," København and London, 1956.
13. Okuhara, E., and Nakayama, T., J. Biol. Chem., *215*, 295–302 (1955).
14. Plouvier, V., Compt. rend., *230*, 125–126 (1950).
15. Scholfield, C. R., Dutton, H. J., Tanner, F. W., and Cowan, J. C., J. Am. Oil Chemists' Soc., *25*, 368–372 (1948).
16. Scholfield, C. R., and Dutton, H. J., J. Biol. Chem., *208*, 461–469 (1954).
17. Somogyi, M., J. Biol. Chem., *160*, 61–68 (1945).
18. Toyama, Y., and Toyama, T., Chem. Abst. *48*, 11086 (1954); Research Rept., Nagoya Ind. Sci. Research Inst., No. *4*, 56–61 (1951).
19. van Handel, E., "The Chemistry of Phosphoaminolipids," Amsterdam, 1954.

[Received June 15, 1959]

An Interlaboratory Study of Test Methods

A. C. ROHLOFF and R. J. HOULE,[1] Lever Brothers Company, Edgewater, New Jersey

AN INTERLABORATORY STUDY of the accuracy and precision of test methods is one of the activities of the Smalley subcommittee on glycerin of the American Oil Chemists' Society. In this paper the statistical methods used to analyze the results of the 1957–58 study are described. The scoring system used to select the two laboratories awarded certificates of merit is also explained.

In addition to fulfilling Smalley Committee objectives, an interlaboratory test study could shed light on the following questions:

Do any of the laboratories have a constant error for the test?
What degree of variation can be expected when the test is used
a) by the same analyst on the same day?
b) over a period of several months within the same laboratory?
c) in different laboratories over a period of several months?
Can the variation of the test be considered the same
a) from month to month within the same laboratory?
b) from month to month within different laboratories?

[1] Past chairman of the A.O.C.S. Smalley Subcommittee for Glycerin.

To the individual participating laboratory, the first question is probably the most important. For companies using the services of referee laboratories the degree of laboratory-to-laboratory variation for the test method is also an important consideration.

Discussion and Calculations

To realize the full potential information in an interlaboratory study it is necessary to have the results reported in the same uniform way by all participating laboratories. It is desirable to have each laboratory run and report the same number of determinations per sample. If laboratories run from one to 10 determinations per sample and report only the average value, not only is a great deal of information lost but the results are almost impossible to interpret statistically.

Twenty-five laboratories participated in the glycerin subcommittee program. Five samples were distributed at monthly intervals. The series included

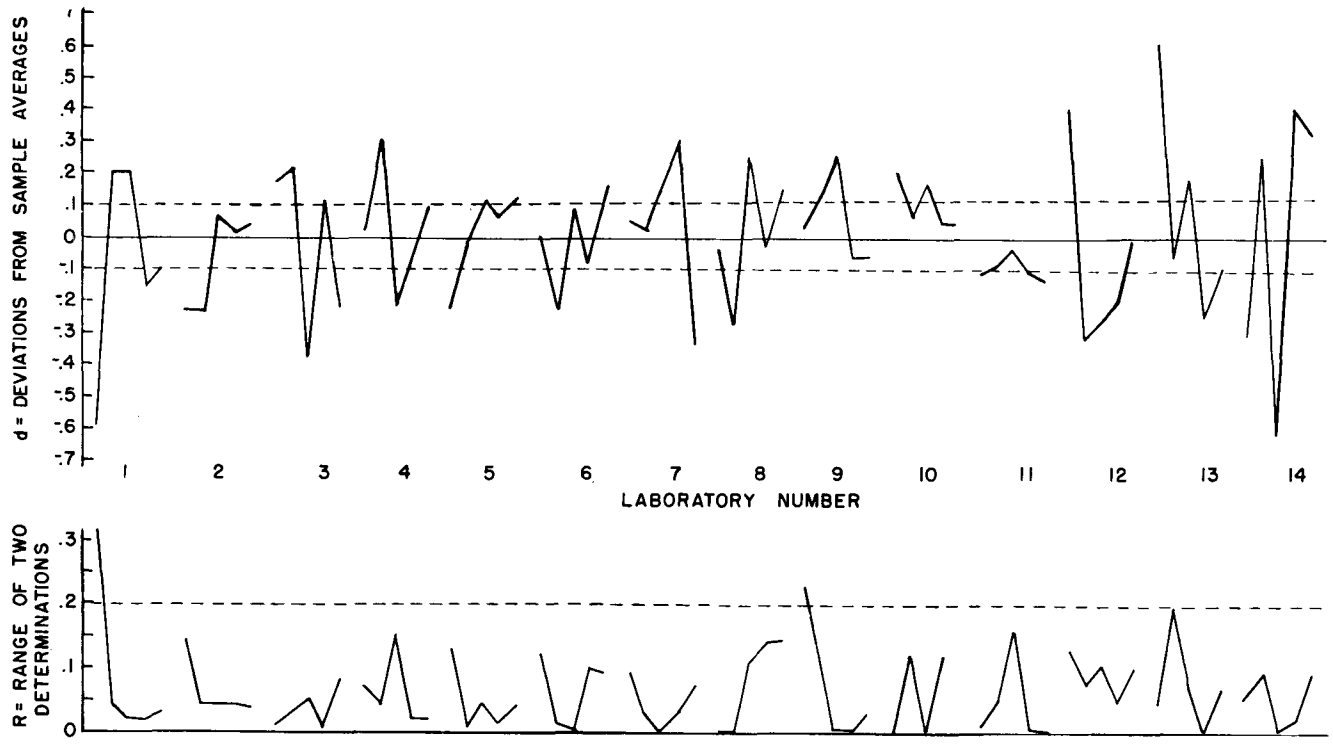

FIG. 1. Control charts for glycerol by sodium periodate.

soap lye crude, saponification crude, USP, and synthetic glycerin samples.

As the sodium periodate method for determining glycerol is of great interest, the numerical examples are based solely on the results of this test as reported by 14 of the laboratories participating in the 1957–58 study.

When results are received from the participants, the question invariably arises as to which, if any, of the values should be discarded as some of the values are often quite obviously out-of-line. Therefore it is desirable to have an objective method for discarding such values. The Dixon (1) criteria for rejecting them, using the 99% confidence level, was consequently adopted. The calculations for the first sample are illustrated in the Appendix. Using these criteria, none of the results reported by the 14 laboratories were discarded.

After rejecting such values the arithmetic average of the remaining values is computed for each sample, and, in the absence of other information, this average is taken as the true value. The deviation of the average of each laboratory from the over-all sample average (d) is next computed. This procedure is followed for each of the five samples, as shown in Table I. These deviations are also plotted in control chart form, as in the upper portion of Figure 1.

The variation of the test when run in duplicate on the same day is now estimated. If the average of two values for a given laboratory was discarded when testing for out-of-line values, these two values are not included in the calculation of test variation. The absolute difference between duplicate determinations (R) is found and plotted for each sample as illustrated by the lower part of Figure 1. We notice that the range for the first sample analyzed by Laboratory 1 is considerably larger when compared with the other ranges. Before calculating the average for a sample, outlying values were discarded; similarly we want to discard atypical values before obtaining an estimate of the variation of the test.

To do this we computed the average range ($\bar{R}$) and followed the statistical quality control methods to find the three-sigma limits for the range chart. From the computation given in the Appendix, the first estimate of the upper control limit for the range chart is 0.21%. The chances are 3 in 1,000 that a point will lie above this limit when there is no real change in the test variation. Two of the values do exceed this upper limit; these values are discarded as atypical, $\bar{R}$ and the limits are recomputed. Other than the discarded values, all points now lie within the revised limits shown on the range chart in Figure 1.

Based on the average range of $\bar{R} = .058\%$, the three-sigma limits for the chart of average deviations are $\pm .109\%$. These limits are shown as dashed lines in the upper part of Figure 1, where it is seen that none of the laboratories have deviations falling within these limits for all five samples.

It is typical of many test methods that duplicate determinations run nearly at the same time agree more closely than determinations made at longer time-intervals; hence it is not surprising to find a number of points outside the control limits based on duplicate determinations. It is for this reason that we may wish to have and, in fact, may need another estimate of the variation of the test.

The total variance, σ^2, of the observed results (other than that caused by differences between samples) can be partitioned into three components: variation due to differences between duplicates, $\sigma^{2/d}$; variation due to the interaction between laboratories and samples, $\sigma^{2/ls}$; and variation due to differences between laboratories, $\sigma^{2/l}$. Algebraically we have $\sigma^2=\sigma^{2/d}+\sigma^{2/ls}+\sigma^{2/l}$.

For an illustrative example, any variation within laboratories, over and above that caused by differ-

TABLE I

Duplicate Determinations Reported for Percentage of Glycerol by Sodium Periodate with Averages ($\bar{x}$), Differences Between Duplicate (R), and Deviations from Sample Averages (d)

		Laboratory number												
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Soap lye crude glycerin	81.29	81.41	81.87	81.77	81.43	81.66	81.80	81.66	81.84	81.92	81.60	82.18	82.31	81.43
	80.97	81.55	81.88	81.70	81.56	81.78	81.71	81.66	81.61	81.92	81.61	82.05	82.35	81.37
$\bar{\bar{x}} = 81.71$	$\bar{x} = 81.13$	81.48	81.88	81.74	81.50	81.72	81.76	81.66	81.72	81.92	81.60	82.12	82.33	81.40
	R = .32[a]	.14	.01	.07	.13	.12	.09	.00	.23[a]	.00	.01	.13	.04	.06
	d = −.58	−.23	.17	.03	−.21	.01	.05	−.05	.01	.21	−.11	.41	.62	−.31
Soap lye crude glycerin	79.77	79.33	79.78	79.88	79.55	79.37	79.63	79.31	79.77	79.72	79.53	79.31	79.62	79.89
	79.81	79.37	79.81	79.92	79.56	79.35	79.60	79.31	79.66	79.60	79.48	79.23	79.43	79.80
$\bar{\bar{x}} = 79.59$	$\bar{x} = 79.79$	79.35	79.80	79.90	79.56	79.36	79.62	79.31	79.72	79.66	79.50	79.27	79.52	79.84
	R = .04	.04	.03	.04	.01	.02	.03	.00	.11	.12	.05	.08	.19	.09
	d = .20	−.24	.21	.31	−.03	−.23	.03	−.28	.13	.07	−.09	−.32	−.07	.25
Saponification crude glycerin	76.85	76.68	76.23	76.34	76.77	76.75	76.81	76.84	76.89	76.83	76.52	76.43	76.89	76.02
	76.83	76.72	76.28	76.49	76.72	76.75	76.81	76.95	76.89	76.83	76.68	76.32	76.82	76.02
$\bar{\bar{x}} = 76.64$	$\bar{x} = 76.84$	76.70	76.26	76.42	76.74	76.75	76.81	76.90	76.89	76.83	76.60	76.38	76.86	76.02
	R = .02	.04	.05	.15	.05	.00	.00	.11	.00	.00	.16	.11	.07	.00
	d = .20	.06	−.38	−.22	.10	.11	.17	.26	.25	.19	−.04	−.26	.22	−.62
USP glycerin	98.43	98.59	98.69	98.52	98.67	98.46	98.89	98.49	98.53	98.57	98.49	98.42	98.34	98.99
	98.45	98.63	98.70	98.54	98.65	98.56	98.92	98.63	98.53	98.69	98.48	98.37	98.34	99.01
$\bar{\bar{x}} = 98.59$	$\bar{x} = 98.44$	98.61	98.70	98.53	98.66	98.51	98.90	98.56	98.53	98.63	98.48	98.40	98.34	99.00
	R = .02	.04	.01	.02	.02	.10	.03	.14	.00	.12	.01	.05	.00	.02
	d = −.15	.02	.11	−.06	.07	−.08	.31	−.03	−.06	.04	−.11	−.19	−.25	.41
Synthetic USP glycerin	99.39	99.52	99.24	99.60	99.64	99.62	99.20	99.57	99.43	99.59	99.36	99.45	99.36	99.80
	99.42	99.55	99.32	99.58	99.60	99.71	99.13	99.71	99.46	99.48	99.36	99.55	99.43	99.91
$\bar{\bar{x}} = 99.50$	$\bar{x} = 99.40$	99.54	99.28	99.59	99.62	99.66	99.16	99.64	99.44	99.54	99.36	99.50	99.40	99.86
	R = .03	.03	.08	.02	.04	.09	.07	.14	.03	.11	.00	.10	.07	.11
	d = −.10	.04	−.22	.09	.12	.16	−.34	.14	−.06	.04	−.14	.00	−.10	.36

[a] Not included in the final estimate of $\bar{R}$.

ences between duplicates, will make the second component, $\sigma^{2/1s}$, larger. The computational procedure for obtaining estimates of these components of variance is given by Anderson and Bancroft (3). The results of these computations are shown in Table II.

The largest variance component is $\sigma^{2/1s} = 0.0606$. If some laboratories show significantly better accuracy for USP samples while other laboratories show better accuracy for soap lye crude samples, we expect $\sigma^{2/1s}$ to be large. However $\sigma^{2/1s}$ will also be large if not all the within-laboratory variation is reflected in $\sigma^{2/d}$. Assuming the latter is true, we obtain a new estimate of within-laboratory variance by pooling $\sigma^{2/d}$ and $\sigma^{2/1s}$: $\sigma^{2/w} = \sigma^{2/d} + \sigma^{2/1s} = 0.0645$ ($\sigma_w = 0.25\%$).

The variance of the average of duplicate determinations is equal to $\sigma^{2/d}/2 + \sigma^{2/1s} = 0.0625$ (a standard deviation of 0.25%). Three-sigma limits, $\pm 0.75\%$ based on this standard deviation, include all the values of d, plotted in the upper part of Figure 1.

To what extent $\sigma^{2/1s}$ does in fact include within-laboratory variation can be determined by a specially designed interlaboratory series. In part, such a design would call for the same sample to be reanalyzed within each laboratory after at least a month. Until such a program is completed, we believe that the value of 0.25% can be taken as a reasonable estimate of the standard deviation for the precision of the test.

TABLE II

Components of Variance for the Percentage of Glycerol Analysis

Source of variation	Degrees of freedom	Sums of squares	Mean squares	Expected mean squares
Samples	4	13458.3892	.....	.....
Laboratories	13	0.5259	0.0404	$\sigma^{2/d} + 10\ \sigma^{2/1}$
Samples × laboratories	52	6.5036	0.1251	$\sigma^{2/d} + 2\ \sigma^{2/1s}$
Between duplicates	70	0.2730	0.0039	$\sigma^{2/d}$
Total	139	13465.6917		

$\sigma^{2/d} = 0.0039$ $\quad \sigma_d = 0.06\%$

$\sigma^{2/1s} = (0.1251 - 0.0039)/2 = 0.0606$ $\quad \sigma_{1s} = 0.25\%$

$\sigma^{2/1} = (0.0404 - 0.0039)/10 = 0.0036$ $\quad \sigma_1 = 0.06\%$

Scoring System

As an incentive for participation in the study and also as a means of emphasizing the importance of controlling test variability, recognition is given to the three most accurate and precise laboratories. To select these laboratories an objective scoring system is needed. The scoring system has been devised so that: ties in the final standings will be avoided; improvement even among the better laboratories will be reflected in the scoring; each of the test methods is weighted according to relative importance in the industry; a laboratory does not reach the neighborhood of the maximum possible scores until the month-to-month variability of the test method is near that evidenced by duplicate determinations run on the same day.

The score depends on the magnitude of the mean deviations from the over-all sample average (d). To

TABLE III

The Maximum Possible Score for Each Test

All crudes	
Percentage of glycerol	45
Percentage of ash	15
Percentage of total alkalinity	10
Percentage of sodium chloride	10
Percentage of total residue at 175°C.	20
Synthetic USP and USP glycerin	
Percentage of glycerol	45
Specific gravity	45
Percentage of Karl Fischer moisture	10

TABLE IV

Percentage of Maximum Score to Be Assigned for Different Values of D

D	Percentage of maximum score
0.00 to 0.49	100
0.50 to 0.99	90
1.00 to 1.49	80
1.50 to 1.99	70
2.00 to 2.49	60
2.50 to 2.99	50
3.00 to 3.99	30
4.00 to 4.99	10
More than 5.00	0

TABLE V

Scores for Each Laboratory and Sample for Percentage of Glycerol by Sodium Periodate[a]

Sample	Laboratory number													
	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Soap lye crude glycerin	0	0	4.5	40.5	0	45.0	36.0	36.0	40.5	0	13.5	0	0	0
Soap lye crude glycerin	0	0	0	0	40.5	0	40.5	0	13.5	31.5	22.5	0	31.5	0
Saponification crude glycerin	0	31.5	0	0	22.5	13.5	4.5	0	0	0	36.0	0	0	0
USP glycerin	4.5	40.5	13.5	31.5	31.5	27.0	0	40.5	31.5	36.0	13.5	0	0	0
Synthetic USP glycerin	22.5	36.0	0	22.5	13.5	4.5	0	13.5	31.5	36.0	13.5	45.0	22.5	0
Total score	27.0	108.0	18.0	94.5	108.0	90.0	81.0	90.0	117.0	103.5	99.0	45.0	54.0	0
Rank order	10	2	11	5	2	6	7	6	1	3	4	9	8	12

[a] Based on $S = 0.051\%$ and $S_1 = S/\sqrt{2} = 0.036\%$.

score a laboratory on a given test we compute $D = [d]/S_1$, where S_1 is the standard deviation of averages of two duplicate determinations run on the same day. In Table III we show the maximum possible score for each test method in the study. Table IV gives the percentage of the total possible score awarded for various values of D. Table V shows the total scores for the 14 laboratories.

Although there will be ties between laboratories for a given test method, these ties will almost always disappear when the scores for the test methods are pooled.

For laboratories interested in improving and controlling the precision and accuracy of test methods, statistical quality control techniques provide a useful tool. One approach is to prepare a large number of aliquots from the same sample, include these aliquots periodically with routine samples, and use statistical quality control charts to spot when a shift occurs in the values obtained. When the control chart indicates a shift in the average, an effort can then be made to locate and eliminate the source of difficulty. If all analysts within a laboratory report values consistently too high or low, this may not be detected by statistical control charts within the laboratory. This type of bias will however probably be detected by a large-scale interlaboratory comparison such as the A.O.C.S. glycerin sample series.

Conclusions

Our aim has been to construct a simple method of scoring consistent with the objectives of the study and to present the results in a manner that does not require too much statistical sophistication. However, if realistic estimates of the precision of test methods are to be obtained, the interlaboratory tests must be more thoughtfully designed and the statistical analysis will tend to be somewhat more complicated as in the components of variance analysis we have presented.

Acknowledgments

The authors wish to express their grateful thanks to the other members of the subcommittee and all participating laboratories whose cooperation made this study possible.

REFERENCES

1. Dixon, W. J., Biometrics, *9*, 74–89 (1953).
2. "A.S.T.M. Manual on Quality Control of Materials," American Society for Testing Materials, Philadelphia, 1951.
3. Anderson, R. L., and Bancroft, T. A., "Statistical Theory in Research," McGraw-Hill Book Company Inc., New York, 1952.

[Received October 20, 1958]

APPENDIX

Discarding Outlying Values. For the first soap lye crude sample the averages of the two values reported by each laboratory are in rank order: 81.13, 81.40, 81.48, 81.50, 81.60, 81.66, 81.72, 81.72, 81.74, 81.76, 81.88, 81.92, 82.12, 82.33. To determine if 82.33 should be discarded, following Dixon (1), we compute:

$$r_{22} = \frac{X_{(n)} - X_{(n-2)}}{X_{(n)} - X_{(3)}} = \frac{82.33 - 81.92}{82.33 - 81.48} = 0.48 < 0.64$$

where r_{ij} for r_{22} at .99 confidence level = 0.64
$X_{(n)}$ is the largest value
$X_{(n-2)}$ is the second from the largest value
$X_{(3)}$ is the value third in rank

Since 0.48 is less than 0.64, the 0.99 confidence point given by Dixon, we do not discard 82.33. To test if 81.13 is an outlying value, the averages are ranked from smallest to largest and the formula is used as before with $X_{(n)}$ now representing the smallest value.

Limits for Control Charts. The constants used in calculating control chart limits and further details may be found in the "A.S.T.M. Manual on Quality Control of Materials" (2).

The Range Chart. The sum of the values of R for all laboratories and samples is 4.46.

$\bar{R} = 4.46/70 = 0.064\%$
Lower limit $= D_3\bar{R} = 0.000\ (0.064) = 0$
Upper limit $= D_4\bar{R} = 3.267\ (0.064) = 0.21\%$

The revised limits are found after discarding two values, 0.32 and 0.23:

$\bar{R} = (4.46 - 0.23 - 0.32)/68 = 0.058\%$
Lower limit = 0
Upper limit $= 3.267\ (0.058) = 0.19\%$

The estimate of a single test standard deviation is:

$s = \bar{R}/d_2 = 0.058/1.128 = 0.051\%$

The Chart for Mean Deviations from Sample Averages (d). The limits are based on $\bar{R} = 0.058\%$:

$\pm A_2\bar{R} = \pm 1.88\ (0.058) = \pm 0.109\%$.