CANADA

TOWARDS A COMMON BASIS FOR THE SAMPLING OF MATERIALS

J. VISMAN

DEPARTMENT OF MINES AND TECHNICAL SURVEYS, OTTAWA

FUELS AND MINING PRACTICE DIVISION

MINES BRANCH

RESEARCH REPORT

R 93

Price 75 cents

JULY 1962

ERRATA

p. 16 line 8, coefficient 0.0186 should read:		0.1086
line 3 from bottom Z = 0.31 " "		Z = 0.33

The same correction to be made on Fig. 5.

p. 27 Table 6, Column 4, $A = p(1-p)(a_1 - a_2)^2 d_1 d_2 / Dm$

should read $A = p(1-p)(a_1 - a_2)^2 d_1 d_2 / Dm^2$

p. 31, par. 4, lines 6, 7	A = 0.00288	should	read	A = 0.00180
and	B = 0.004186	"	"	B = 0.002616
p. 31, par. 5, lines 2, 3	N = 183	"	"	N = 115
	and 915 pounds	"	"	575 pounds.

equation $s^2 = A/W + B/N$ operates independently of the shape of the parent distribution of the variate. It applies generally for first-order estimates of the upper limit of the sampling variance (s^2). The degree of segregation (z) is described by $z^2 = B/A$.

A "sampling board" for small-scale experiments is introduced to demonstrate the above relationships for binominal distributions of the variate. Tests on this sampling board confirm that the above equations apply to systematic as well as random sampling conditions and can be used for assessing and predicting sampling precision for a large variety of materials.

---

* Head, Western Regional Laboratory (Edmonton), Fuels and Mining Practice Division, Mines Branch, Department of Mines and Technical Surveys, Ottawa, Canada.

Mines Branch Research Report R 93

TOWARDS A COMMON BASIS FOR THE SAMPLING OF MATERIALS

by

J. Visman*

SYNOPSIS

There is a need in many fields of investigation for a general method of estimating in advance the precision of samples drawn systematically from material consignments that are not random mixtures.

Materials and variates may vary over wide ranges and the circumstances under which the samples are collected can vary widely, but the causes of variation in sample value are limited. Two factors are inherent in the nature of the consignment, namely, random variation and "segregation". These can be determined as variance components from a specially designed test or estimated from previous information, if the material is known by composition and distribution.

The other factors influencing the precision of the sample are the number (N) of increments collected from all parts of the lot and the size (W) of the resultant gross sample, operating variables that can within certain limits be regulated at will by the sampler. The equation $s^2 = A/W + B/N$ operates independently of the shape of the parent distribution of the variate. It applies generally for first-order estimates of the upper limit of the sampling variance (s^2). The degree of segregation (z) is described by $z^2 = B/A$.

A "sampling board" for small-scale experiments is introduced to demonstrate the above relationships for binominal distributions of the variate. Tests on this sampling board confirm that the above equations apply to systematic as well as random sampling conditions and can be used for assessing and predicting sampling precision for a large variety of materials.

---

* Head, Western Regional Laboratory (Edmonton), Fuels and Mining Practice Division, Mines Branch, Department of Mines and Technical Surveys, Ottawa, Canada.

Direction des mines, Rapport de Recherches R 93

VERS UNE BASE COMMUNE POUR L'ÉCHANTILLONNAGE DES MATÉRIAUX

par

J. Visman*

RÉSUMÉ

Dans de nombreux domaines de recherches, on a besoin d'une méthode générale pour évaluer à l'avance la précision d'échantillons tirés systématiquement de lots de matériaux qui ne sont pas des mélanges au hasard.

Les matériaux et les variates peuvent varier considérablement et les circonstances dans lesquelles les échantillons sont pris peuvent varier beaucoup, mais les causes de variation de la valeur de l'échantillon sont limitées. Deux facteurs ne dépendent que de la nature du lot: la variation aléatoire et la "ségrégation". On peut les évaluer comme variances partielles, d'après une expérience spéciale ou bien d'après des connaissances antérieures si l'on connaît la composition et la distribution du matériau.

Les autres facteurs de la précision de l'échantillon sont: le nombre (N) des prélèvements obtenus de toutes les parties du lot, et la grandeur (W) de l'échantillon total; ces variables opératoires peuvent dans une certaine mesure être modifiées à volonté par l'échantillonneur. L'équation $s^2 = A/W + B/N$ est valable quelle que soit la forme de la distribution de la variate. Elle s'applique en général pour des estimations de premier ordre de la limite supérieure de la variance d'échantillonnage (s^2). Le degré de ségrégation est exprimé par $z^2 = B/A$.

L'auteur propose un "tableau d'échantillonnage" expérimental pour démontrer les rapports ci-dessus, dans le cas de distributions binomiales de la variate. Des expériences avec ce tableau confirment que ces équations s'appliquent aux conditions d'échantillonnage systématique aussi bien qu'à l'échantillonnage au hasard, et qu'on peut les utiliser pour évaluer et prédire la précision d'échantillonnage, pour un grand nombre de matériaux.

---

* Chargé de recherches principal, Laboratoire de la Région de l'Ouest de la Division des combustibles et du génie minier, Direction des mines, ministère des Mines et des Relevés techniques, Ottawa, Canada.

CONTENTS

FIGURES

TABLES

--

INTRODUCTION

There is today no unified theory for the systematic sampling of materials that are non-randomly distributed in space or in time (8). The existing theory deals essentially with random sampling. The methods derived from this theory vary with the type of material and the circumstances under which the sample is collected. Consequently, the literature on sampling is large and diversified, the methods having been adapted to the specific needs of each particular field.

There are, however, certain underlying principles common to all sampling experiments. To formulate these principles in terms of quantities easy to measure is the objective of this report. More specifically, the overall precision of sampling is expressed as a function of operating variables and constants for the purpose of estimating in advance the precision of samples collected from materials that are known by composition and distribution, or whose characteristics have been determined from a test.

Main Definitions

Every sampling operation consists essentially of either extracting one single sample from a given quantity of material or extracting, from different parts of the lot, a series of small portions or "increments" that are combined into one "gross sample". The latter method is known as "sampling by increments" and will be considered here. The former method can be regarded as a special case of incremental sampling in which the number of increments equals one.

The existing theory for sampling materials that are non-randomly distributed is known as "stratified sampling" or "representative (random) sampling (of stratified populations)". In this theory the precision of sampling is expressed as the sum of the variance "within-strata" and the variance "between-strata", the strata indicating parts of the material consignment whose mean values differ significantly from the overall mean value of the consignment. Sometimes, as in incremental sampling, these "strata" are imaginary, as they become identical to the portions represented by each individual increment. The "within" and "between" variance estimates are then a function of the size and the number of increments. It is common usage to identify the "between-strata" variance with the "trend variance" and the "within-strata" variance with the "random variance". It is clear, however, that with different size and number of increments the estimates of the between-strata variance and the within-strata variance will change. Therefore, these variance estimates cannot be regarded as constants and cannot be used, without certain corrections, for calculating the number and size of increments that would be required to evaluate in advance, a projected overall precision of sampling.

The meaning of "random sampling error" as used in this report goes back to a classical experiment where a number of black and white balls are mixed at random in a vase and a sample is withdrawn that consists of one or more balls. The random error is caused when the hand collecting the sample selects by chance a white ball instead of a black ball, or vice versa. The resulting variance is the "random variance", of which the "within-strata variance" used in representative sampling gives a biased estimate (depending on the size of the samples used) when dealing with materials that are non-randomly distributed. This random variance is determined by the average composition of the material (in this case the relative amount of black or white balls) and by the size of the sample only. The same definition of the random variance is adopted for variates with parent distributions that are not of the binomial type.

In this report, the term "random variance" is maintained in its original meaning; "trend variance" has been deleted because of its confusing nature. Instead, a new term "segregation variance" is introduced, denoting the variance caused solely by deviations resulting from the non-random distribution of a consignment. Its physical meaning is simple to explain. The deviation of any sample value from the true mean of the lot or consignment is the algebraic sum of its random error and a remaining error resulting from the fact that the variate is non-randomly distributed over the lot. The latter is called the segregation error and its variance the segregation variance. It will be shown that the segregation variance component of single samples is independent of sample size; it depends on the degree of segregation of the consignment only. It will further be shown that the maximum degree of segregation, as expressed by the variance of segregation, is directly related to the random variance. This relationship is utilized to estimate sampling precision.

Previous Work and Commentary

The method suggested here follows earlier work on the sampling of coal (2,3,4,17,18,19) and subsequent commentary, notably by R.C. Tomlinson, whose criticism is, briefly, that sampling theory applies only when the condition of randomness is fulfilled and that, even so, sample variances may be biased when samples are collected for determining a ratio (14,15).

A more practical course was followed by ASTM D5 subcommittee XXIII over the period 1954-58 when a comprehensive test program was carried out for testing an experimental method of forecasting sampling precision for coal (5,19,20). The results of this program showed that the theoretical objections against systematic sampling of segregated coals have been overrated.

As it is the objective of sampling theory to forecast the variance of a variate X (not X itself), less mathematical rigour is demanded than in a related field of statistics that deals with the prediction of certain events (e.g., the expected rainfall per annum, the fatality rate of air travel), in other words, with (X) itself.

It would appear from recent work, notably that of I.S.O./TC27 WG7-Sampling, that there is now a tendency towards a more quantitative evaluation of certain assumptions of sampling theory. For instance, one difficulty affecting the practical application of all sampling theory is how to deal with over-estimates of the sample variance. When a sample of given size is drawn from an infinite population, its theoretical variance is always larger than when a sample of the same size is drawn from a finite population with otherwise identical characteristics. In this report the above problem is dealt with as follows.

The fact that in practice all populations are finite does not necessarily invalidate the theoretical estimate of the variance, provided it is stipulated that it is an estimate of the maximum value that this variance will attain for an infinite population. The same problem is encountered when samples are drawn systematically or at random from a stratified population. Samples that straddle the boundaries between two strata contribute less to the sampling variance estimate than those that are drawn wholly from individual strata. Consequently, the latter variance estimate is always larger than the former.

It is suggested, firstly, that the above theoretical variance estimates are accepted with the qualification that they are estimates of the upper limit that the variance will attain under theoretical conditions (infinite population and stratified distribution of the variate).

Secondly, it is suggested that the meaning of the term "biased sample" is restricted to those sampling errors and deviations that are caused by the inclusion of components that are foreign to the population (contamination) and by the systematic exclusion of true components of the population (e.g., the exclusion of large particles by a faulty sampling device). A biased sample value could only be caused by such errors or deviations. All other samples are then to be regarded as representative of the population, regardless of the magnitude of their deviation from the true mean of the population.

It is held that, if the above qualifications and definitions are accepted, sampling theory can be of practical benefit in almost any field of human endeavour, especially in industry, without violating the basic mathematical concepts of statistical theory. The value of such a theory of sampling in regard to expertise, guarantee, and litigation can hardly be over-estimated.

Scope of Report

The objective of this report is to provide first-order estimates of the upper limit of the sampling variance in a general case where samples are collected from materials whose component parts are distributed throughout the population in a non-random pattern.

It is also the objective of this report to show that the overall accuracy of a given sampling procedure can be estimated in advance by this theory, for binomial as well as for non-binomial distributions of the variate.

It is claimed that costly and time-consuming experiments can be avoided. A versatile model population for small-scale experiments is introduced. It is of the binomial type. Several important questions can be answered with this model that apply to non-binomial population types as well, e.g.: Is there a difference between systematic sampling and random sampling? What is the effect of various degrees of segregation and patterns of distribution on the sample variance? What is the relationship between segregation and random variance? The results of this inductive study confirm the practical feasibility of applying statistical theory to the systematic sampling of segregated materials under conditions that can as a rule be fulfilled by a well-instructed, experienced sampler.

A duplicate sampling method with small and large samples (20) is also described. It is used for estimating the upper limit of the random variance component and segregation variance component ("sampling constants"), by first-order approximation, for materials whose composition and distribution are not known in advance. It will be shown that this method applies in principle for all materials and variates, including non-binomial parent distributions. In this duplicate sampling test, samples are drawn systematically from segregated consignments, one series of small samples and one series of relatively large samples. From it the two sampling constants (A, B) are found that can be used later on when either the same lot of material, or material consignments that are known to be similar to it, are to be sampled with a certain pre-assigned accuracy. The sampling constants are then used in an equation that provides estimates of the minimum number and size of increments required to attain the projected accuracy. Essential mathematical derivations are given in an Appendix.

ANALYSIS OF VARIABILITY

In this section the theory of sampling segregated binomial populations, originally developed for coal in 1947 (17,18), is shown to apply to other materials and to variates with parent distributions of any type. A model population is introduced to demonstrate the essential relationship and its general applicability. Variance values found from tests on this model are maximum estimates only, because the model represents the conditions that cause the largest possible variations. Sampling variances derived from the tests are accurate by first-order approximation only. Conditions other than those governing test results from the model will lead to variance estimates that are smaller, as for instance when the samples are very large or when the population is relatively small. Other conditions are discussed in the text. The above limitations do not seriously interfere with the requirements of industry regarding the testing and safeguarding of quality. It is believed that by limiting the sampling theory in this manner the broad objective of establishing a common basis for the sampling of materials is served as well as is practically possible.

Model Population

The model population of "black" and "non-black" items, as exemplified by a "sampling board" (Figure 1), is used for analyzing variability of samples drawn from segregated consignments.

This sampling board consists of a piece of 10" x 10" wire screen with 10 openings per linear inch and a supply of 5,000 lead pellets. The lead pellets can be used entirely, or in part, for making model populations that are segregated in different ways. The pellets can be distributed in any conceivable manner ranging from complete segregation to near-perfect random mixtures. The samples collected from this population are not removed but merely counted. A sample is taken by placing a square frame with its centre over the selected station and counting the number of pellets enclosed by it. The size of the samples can thus be varied and the number can be chosen at will. The samples can be collected either systematically at fixed stations marked off on the screen, or at random. In the latter case, a random sampling table is used for determining the co-ordinates.

The method of analysis consists essentially in collecting samples of different size from a given population and determining the relationship between sample variance and sample size.

It will be shown (Eq 3, p. 12) that the total variance of sampling (s^2) consists of a random variance component (s_p^2/w') that depends on the size (w) of the sample, and a segregation variance component (S_s^2) that is independent of sample size.

The results of experiments done with the sampling board are presented in the form of graphs showing the relationship between the variance of single samples and the sample size, the latter being determined by the number of screen openings in a square frame. In the tests reported here, three different sample sizes are used, namely:

$w_1 = 1$,

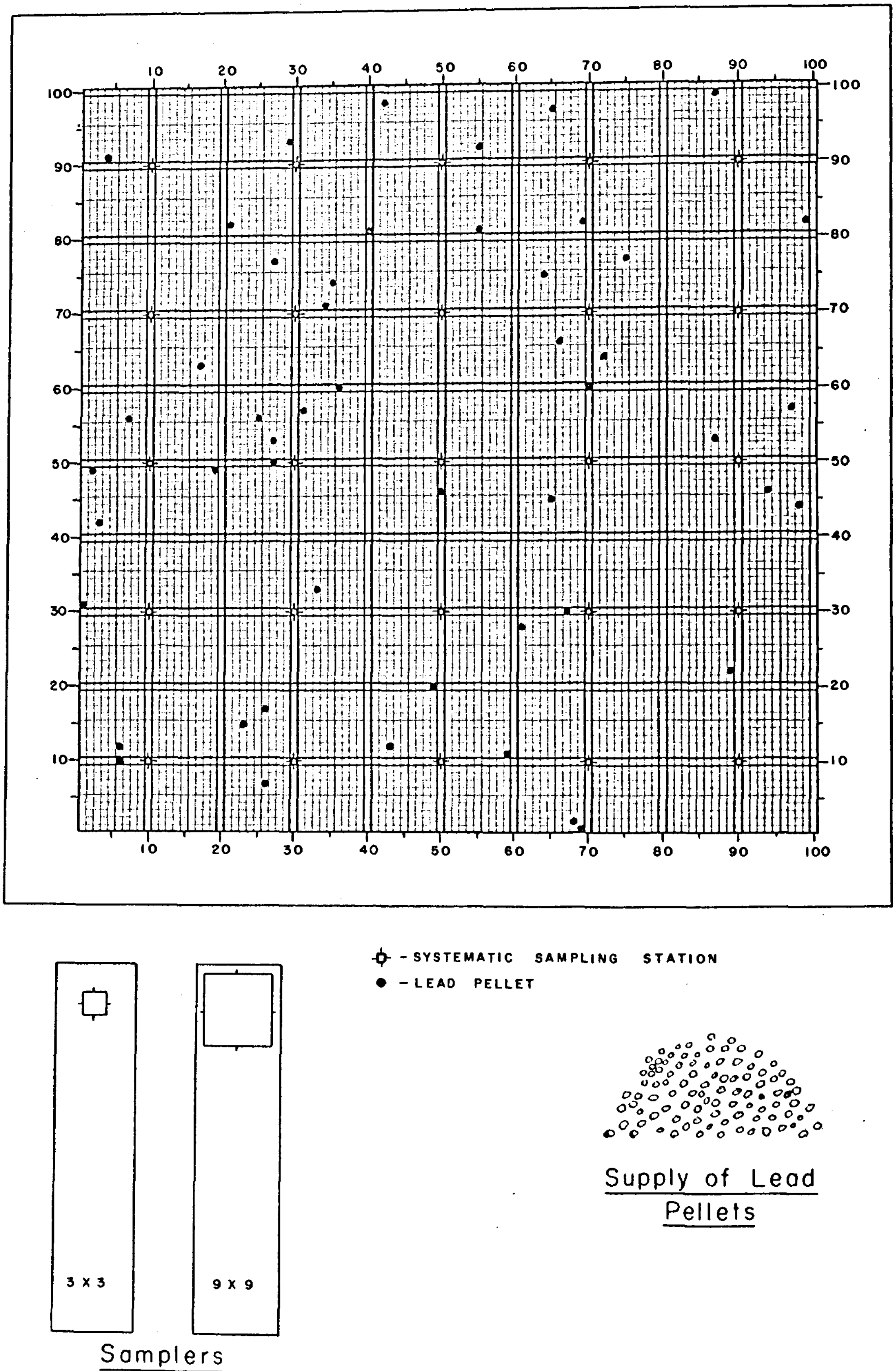
$w_2 = 9$ (located in the square of 3 x 3 openings),

and $w_3 = 81$ (9 x 9 openings).

The numbers of pellets (x) found within the square frames are marked down and the series thus obtained is used for calculating variance estimates. For the reader who is unfamiliar with statistics, it is noted that the variance is the square of the standard deviation (s), the "root mean square" of deviations. A simple formula for calculating

FIGURE I

SAMPLING BOARD



J.V./J.P. 13·6·60

this measure of dispersion for a series of observations is presented in Table 5 (p. 24) of this report, where $p = \frac{x}{w}$.

Relationship Between the Degree of Segregation and the Parent Frequency Distribution

Example 1

An example of complete segregation will be studied first by placing 2,500 beads in one corner of the sampling board (the lower-left corner as shown by the inset on Figure 2). This corresponds to a binomial population designated by p = 0.25. Samples collected from this mixture will be either 100% black or 100% white, except those that straddle the boundary between the black area and the white area. This latter restriction is of little consequence so long as the samples are small compared with the "patch" of 2,500 beads, as is shown on Table 1 (p. 9) where three series of systematic samples and three series of random samples are represented that have sizes 1, 9 and 81 respectively. Figure 2 illustrates that the six variance estimates found from these series do not deviate significantly from a straight horizontal line corresponding with the binomial variance $s^2 = p(1-p) = 0.1875$. The fiducial limits of the variance estimates correspond to variance ratios $F_{95} = 1.52$ (24 and ∞ deg. fr.) for variance estimates larger than 0.1875, and $F_{95} = 1.73$ (∞ and 24 deg. fr.) for variance estimates smaller than 0.1875. The result of this sampling experiment shows there is no significant difference between the samples drawn at random and the samples collected systematically. The same conclusion follows when the Chi-square test is used.

The experiments also show that, while the size-variance curve of a completely random mixture would be defined by a straight line sloping down at an angle of 45° on a double-log scale, the sample variance never exceeds the theoretical value of 0.1875 in the case of complete segregation and remains substantially constant over the entire interval.

Patterns showing partial segregation may take many forms that are impossible to deal with in every detail. The gradual transition of complete segregation into complete randomness can, however, be illustrated in an orderly fashion and the conclusions that can be drawn from it apply generally to any pattern of distribution.

To study the characteristics of partial segregation it will be assumed that mixing takes place in five equal steps, reducing the degree of segregation first from 1.0 to 0.8, then to 0.6, to 0.4, to 0.2, and finally to 0. When segregation is zero, the number of pellets within the black square should be 25% of the original number. The total reduction from 100% pellets to 25%, divided into five equal steps, is a reduction of 15% or 375 pellets for each step.

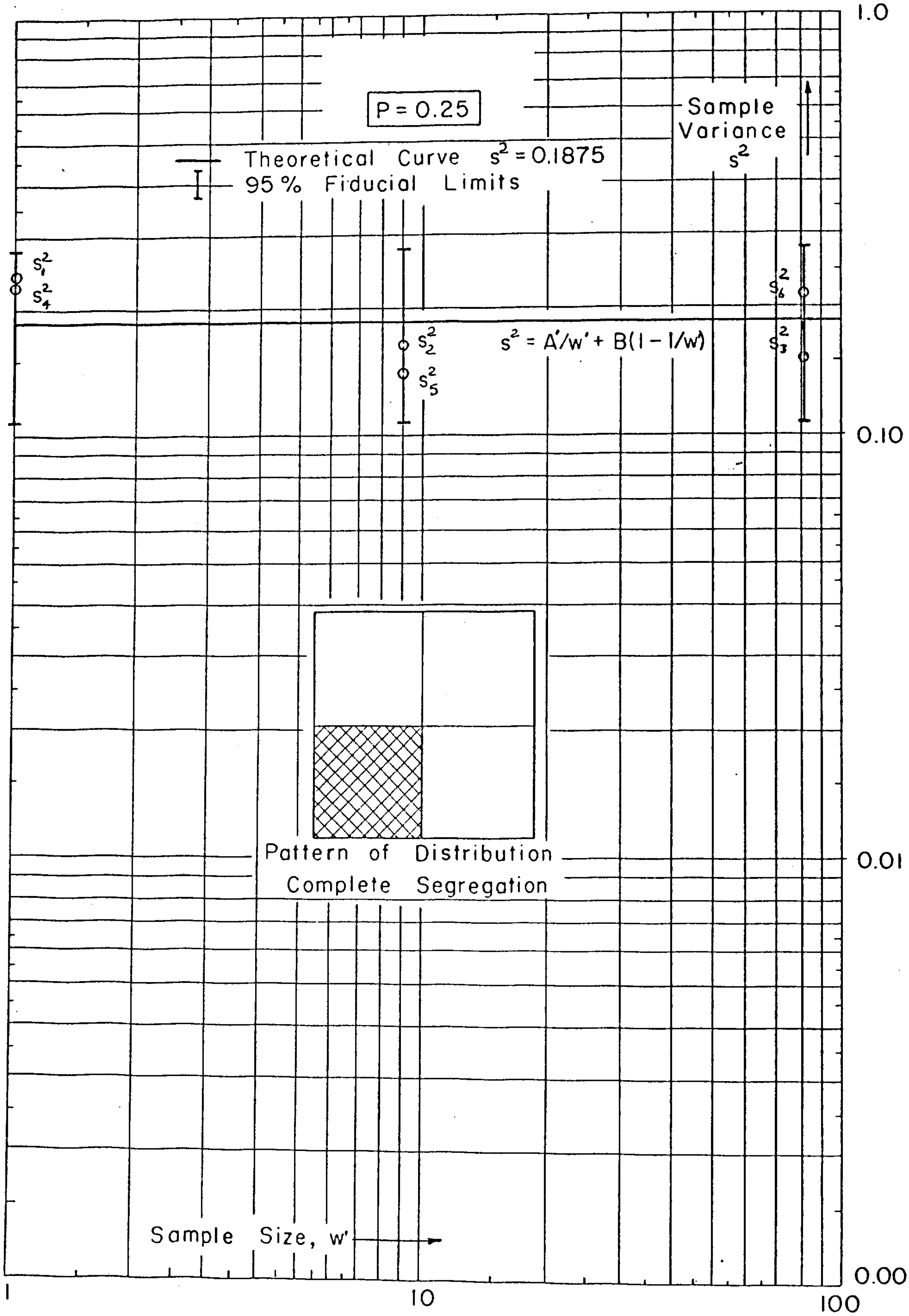


Figure 2. Size Variance Curve (Complete Segregation)

TABLE 1

Complete Segregation (Figure 2): $p = 0.25$

	Systematic Samples						Random Samples											
Sample Size	1		9		81				1				9				81	
Sample No.	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	coordinates x_4		x_4	x_4^2	coordinates x_5		x_5	x_5^2	coordinates x_6		x_6	x_6^2
1							17	07	1	1	68	55			44	04	81	6,561
2							76	74			34	74			22	33	81	6,561
3							37	21	1	1	30	30	9	81	78	46		
4							13	19	1	1	13	77			84	09		
5							04	30	1	1	70	40			26	52	27	729
6							70	97			74	59			71	13		
7							33	77			57	29			91	58		
8							24	46	1	1	25	97			38	18	81	6,561
9							03	44	1	1	65	68			67	24		
10							54	80			76	60			54	76		
11	1	1	6	36	45	2,025	04	94			27	48	9	81	96	96		
12	1	1	6	36	45	2,025	43	77			42	55			57	46		
13	1	1	4	16	25	625	18	24	1	1	37	90			69	92		
14							66	21			86	65			36	42	81	6,561
15							79	90			53	72			10	45	81	6,561
16	1	1	9	81	81	6,561	12	99			00	66			77	10		
17	1	1	9	81	81	6,561	72	27			39	37	9	81	84	45		
18	1	1	6	36	45	2,025	07	72			68	32			57	65		
19							34	95			29	20	9	81	03	04	81	6,561
20							45	14	1	1	61	30			29	26	81	6,561
21	1	1	9	81	81	6,561	52	38			29	68			53	34	18	324
22	1	1	9	81	81	6,561	85	68			94	49			75	23		
23	1	1	6	36	45	2,025	66	88			98	69			91	20		
24							60	11			94	10			93	57		
25							44	80			24	82			30	27	81	6,561
Sum	9	9	64	484	529	34,969			8	8			36	324			693	53,541
s^2	0.2400		0.1647		0.1510				0.2267				0.1400				0.2180	

The following mental experiment can now be conducted: Three hundred and seventy-five (375) pellets are selected at random from the black square of 2,500 (Figure 2), and are redistributed randomly over the remaining three-quarters of the sampling board (the degree of segregation is reduced from 1 to 0.8).

A sample drawn from the black quarter of the sampling board will have an expected value

$$E(X)_{black} = (2500 - 375)/2500 = 0.85$$

Similarly, for samples drawn from the other three-quarters, we find expected sample values

$$E(X)_{white} = 375/7500 = 0.05$$

for each individual quarter.

The expected variance as calculated from these figures is, for a degree of segregation 0.8,

$$E(\text{variance}) = E\left[\left[X - E(X)\right]^2\right] = 0.1200.$$

The total variance for a degree of segregation of 0.8 is 0.64 times the total variance for the entirely segregated mixture.

By continuing the experiment for lower degrees of segregation the results presented in Table 2 are found, when collecting four samples (one from each quarter of the sampling board) for each individual test.

TABLE 2

Effect of Segregation on Total Variance

Degree of Segregation (z)	Deviation from Mean Grade p = 0.25 for Each Quarter	Total Expected Variance	
		$E(s^2)$	Fractional
1.0	0.75; 0.25; 0.25; 0.25	0.1875	1.00
0.8	0.60; 0.20; 0.20; 0.20	0.1200	0.64
0.6	0.45; 0.15; 0.15; 0.15	0.0675	0.36
0.4	0.30; 0.10; 0.10; 0.10	0.0300	0.16
0.2	0.15; 0.05; 0.05; 0.05	0.0075	0.04
0.0	0.00; 0.00; 0.00; 0.00	0.0000	0.00

This table shows that the degree of segregation (z) and the expected variance are related:

$$E(\text{variance}) = 0.1875\ z^2$$

A similar relationship holds for all ratios of "black" and "white" mixtures other than 2,500 out of 10,000.

The practical meaning of the expected variance is that it is the limit of the total variance as sample size increases. Therefore, the expected variance is identical with the segregation variance: $E(\text{variance}) = s_s^2$.

Furthermore, the variance for complete segregation appears to be identical with the parent variance, that is, the variance of single items which in this case follows from the binomial equation $s_p^2 = p(1 - p)$.

From the foregoing equations it follows that:

$$s_s = z\ s_p \quad \text{(Eq 1)}$$

Summarizing the conclusions from the above experiment, we have:

1. The segregation variance has a maximum value equal to that of the parent variance of the population.

2. The segregation variance is within the range of actual sampling practice, substantially independent of sample size. It never exceeds the parent variance.

3. The ratio between the segregation variance and the parent variance depends solely on the degree of segregation (z).

4. The total variance of samples consisting of one unit only, equals the parent variance (s_p^2) regardless of the degree of segregation.

We have conjectured, on the basis of experimental evidence, that the expected variance of sampling satisfies the following relationship:

$$E\ (s^2) = s_p^2/w' + E\ (s_s^2)\ (1 - 1/w') \quad \text{(Eq 2)}$$

where s_p^2 = parent variance; variance of single units;

$E\ (s_s^2)$ = expected value of the segregation variance;

w' = sample size, expressed in number of units.

For samples consisting of two units the total variance becomes, by first approximation,

$$s^2 = 1/2\ s_p^2 + 1/2\ s_s^2.$$

For samples consisting of ten or more units, Equation 2 can be written by first approximation as:

$$s^2 = s_p^2/w' + s_s^2 \quad \text{. . . . . (Eq 3)}$$

It is noted that the parent variance (s_p^2) is a constant which, according to the binomial equation, depends on the composition of the material only. It is designated as "sampling constant" A'.

The segregation variance (s_s^2) for one and the same material depends on the degree of segregation (z) only, in accordance with Equation 1. It is known from experience that, while (z) may range from zero to 1, the stability of the segregation variance under otherwise normal conditions of handling, storage and transportation is comparable to that of the parent variance. To illustrate this with figures, it is known that noticeable blending can be observed when a mixing device reduces the segregation variance of a product by a factor of 3 or more. Conversely, an increase of the segregation variance by a factor of 3 to 4 or more is equivalent to a distinct separating action. Therefore, while (s_s^2) may change, its value for a given material consignment will be constant within limits normal for variance estimate (F-ratio), unless the consignment is noticeably mixed or segregated. Segregation variance s_s^2 is designated "sampling constant" B.

The practical value of the "sampling constants" can be demonstrated by the following Examples 2 and 3:

Example 2

General Equation 2 was tested by distributing 2,500 lead pellets non-randomly over the sampling board. The samples of different sizes were collected systematically and at random as was done in the first example. The results are presented in Table 3 and Figure 3.

Two variance estimates, s_1^2 and s_3^2, obtained from the systematic samples were used to evaluate the sampling constants by using Equation 2, which can now be written as:

$$s^2 = A'/w' + B(1 - 1/w') \quad \text{. . . . . (Eq 4)}$$

The following values were found for the sampling constants, using Equations 8 and 9 on page 23:

A' = 0.1824

B = 0.00761

TABLE 3

Partial Segregation (Figure 3): p = 0.25

	Systematic Samples						Random Samples											
Sample Size	1		9		81		1				9				81			
Sample No.	x_1	x_1^2	x_2	x_2^2	x_3	x_3^2	coordinates x_4		x_4	x_4^2	coordinates x_5		x_5	x_5^2	coordinates x_6		x_6	x_6^2
1			2	4	21	441	03	36			46	33			60	16	17	289
2			2	4	20	400	47	96	1	1	98	26			11	22	18	324
3			2	4	14	196	43	47	1	1	63	16	2	4	14	77	20	400
4			2	4	16	256	73	36	1	1	71	80	1	1	10	94	7	49
5			2	4	15	225	86	61			62	45	4	16	95	39	28	784
6			1	1	15	225	97	42			42	27	6	36	24	84	26	676
7			1	1	30	900	74	81			53	07	1	1	51	42	23	529
8			3	9	35	1,225	24	14			32	36	1	1	79	17	35	1,225
9					25	625	67	57	1	1	37	07	5	25	89	53	20	400
10			4	16	19	361	62	20			32	51	2	4	73	31	20	400
11			1	1	19	361	16	56			32	13			88	63	20	400
12			2	4	21	441	76	50	1	1	90	55			97	01	40	36
13	1	1	1	1	20	400	62	26			79	38	1	1	54	63	27	729
14			1	1	22	484	27	71	1	1	78	58	2	4	14	78	20	400
15	1	1	5	25	28	784	66	07			53	59	4	16	10	59	8	64
16	1	1	4	16	26	676	12	96			05	57			88	33	16	256
17	1	1	6	36	27	729	56	96			03	12			26	21	22	585
18			1	1	21	441	85	68			72	10	5	25	49	12	23	529
19	1	1	3	9	22	484	99	27			93	14	4	16	81	34	11	121
20					15	225	26	31			15	21	3	9	76	29	29	841
21			1	1	11	121	55	38			31	06	1	1	23	57	21	441
22			1	1	20	400	59	54			62	18	2	4	83	60	21	441
23			2	4	19	361	56	82			43	44	1	1	01	86	45	36
24	1	1	2	4	38	1,444	35	46	1	1	09	32	4	16	30	32	22	484
25			8	64	46	2,116	64	22			90	53	3	9	30	44	17	289
Sum	6	6	57	215	565	14,321			7	7			52	190			483	10,627
s^2	0.1900		0.0438		0.00986				0.2100				0.04210				0.00823	

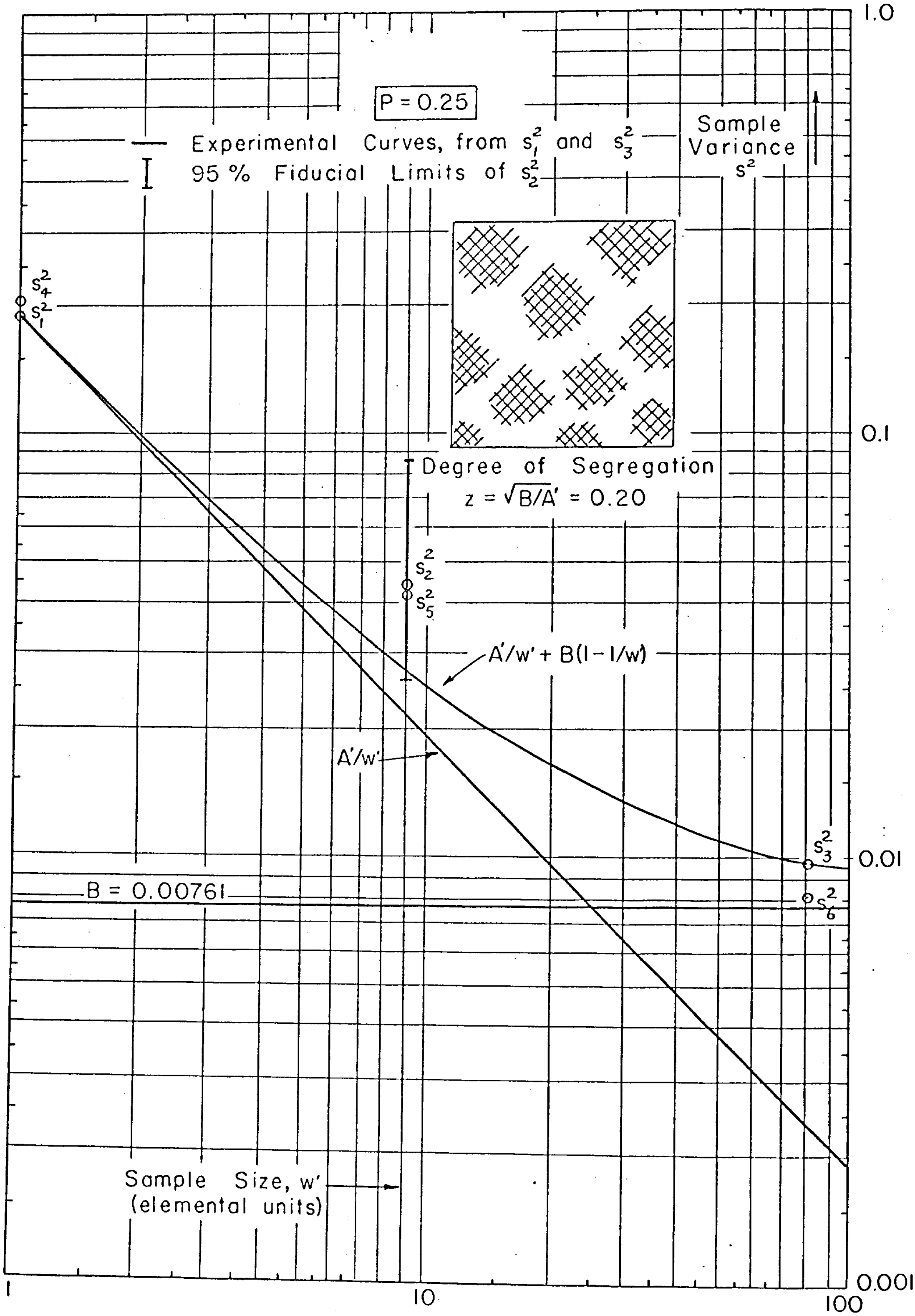


Figure 3. Size Variance Curve (Partial Segregation)

J.V./J.P. 20·4·60

From these values the size-variance curve representing Equation 4 is found; it is illustrated in Figure 3. This size-variance curve is approximately the algebraic sum of a straight line, A'/w', sloping down at 45 degrees from point ($w' = 1$; $s^2 = 0.1824$) and a straight horizontal line, $B = 0.00761$; the former represents the random variance component, the latter the segregation variance component. The degree of segregation is found from the equivalent of Equation 1:

$$z = \sqrt{B/A'} = 0.20$$

Here, the Chi-square test provides spot-checks for the goodness of fit of Equation 4, using experimental variance estimates s_2^2, s_4^2, s_5^2 and s_6^2.

The size-variance curve calculated from s_1^2 and s_3^2 falls within the confidence interval defined by $\frac{n-1}{\text{Chi-square}} \cdot s_1^2$ of each one of the above four variance estimates for probability levels P = 0.025 and 0.975. For example, the confidence interval of the variance estimate s_2^2, which was found from 25 (systematic) samples, is 0.026 - 0.088 at the 95 per cent level. The calculated variance (Equation 4) falls within this range at 0.027. As s_2^2 shows the largest difference of all, the Chi-square test confirms the statistical identity of the calculated variance (Equation 4) and all four experimental variance estimates at the 95% level. On Figure 3, the confidence interval is shown for experimental variance s_2^2 only. It is noted that similar results were found when using the F-test. The Chi-square test was preferred, it being the more rigorous one of the two tests. Frequency distributions of samples with size larger than 1 unit ($w' = 1$) will generally show deviations from the binomial distribution when the material is segregated. When the samples contain only a small number of units, as they necessarily do in the experiments performed with the sampling board, these departures from the theoretical binomial frequency distribution cannot always be proved significant. When, however, the number of units contained in the sample becomes very large, such as in molecular binomial mixtures (fluids, pulps, etc.), the difference between the frequency curve of sample values as found from a test and the frequency curve of the sample values observed in the same material consignment when randomly mixed, will be generally significant, the more so when the degree of segregation is high. In fact, the frequency distribution of large samples from segregated mixtures can take on any shape, independently of the shape of the parent distribution, but the variance of such large samples is directly related to the variance of the frequency distribution of the single units. The theory presented here utilizes this relationship and is demonstrated for variates that can be expressed by parameters having a binomial parent distribution. It will be shown later on in the report that the same concept applies to parent distributions of different type, including normal, poissonian, and irregular parent distributions (see under "Non-binomial variates", p. 32).

Example 3

A test similar to the ones above was done with 1,000 lead pellets that were distributed as evenly as possible over the sampling board. The curve representing Equation 4 was based on variance estimates s_1^2 and s_3^2 (see Figure 4). All the other values which were determined independently appear to check, within the limits of chance variation, with the curve

$$s^2 = 0.0186/w' + 0.00137\ (1 - 1/w').$$

The degree of segregation found from $z = \sqrt{B/A'} = 0.11$.

The three examples discussed here confirm the correctness of the general Equation 4 for a range of conditions varying between complete segregation and near-random dispersion of the variate.

In sampling practice the use of samples consisting of only a few units is common in such fields as microscopic analysis of particle mixtures and sampling for defectives. In many cases, however, the samples collected consist, of necessity, of a very large number of units that cannot be counted. Consequently, sample size is expressed in some unit of measurement (1 gram, 1 pound, etc.); each unit of measurement may contain thousands or millions of elementary units of the binomial. As a result, the size-variance curve of such samples will be generally determined by the segregation variance component only. In other words, the actual range of sample sizes lies somewhere within the less steep section of the size-variance curve.

For this type of material it would be impractical to use the parent variance for sampling constant A', because the number (w') of binomial units is too large to be counted. Instead, sampling constant A' can be determined for one unit of measurement. It is then necessary to indicate to what unit of measurement this sampling constant does refer.

Practical Units and Proximate Equation

To illustrate the use of practical units and their relationship to the general equation, the results of another test are presented in Figure 5. One thousand lead pellets were distributed with a high degree of segregation (see inset Figure 5) and the sampling constants calculated from variance estimates s_1^2 and s_3^2 as before:

$$s^2 = 0.09923/w' + 0.01078\ (1 - 1/w')$$

degree of segregation $z = 0.31$.

The other variance estimates (obtained from random samples as well as systematic samples) correspond within the 95% fiducial

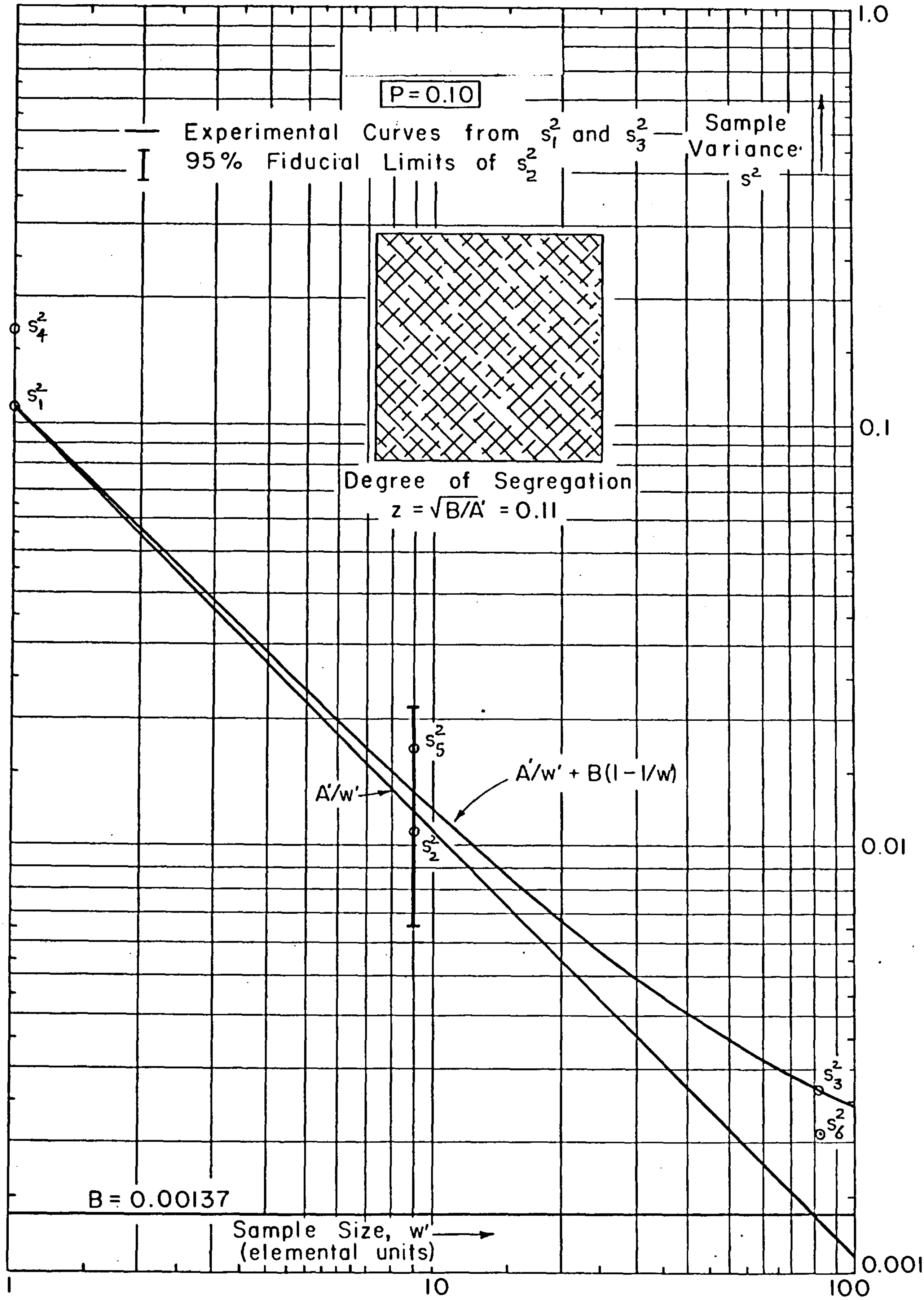


Figure 4. Size Variance Curve (Minor Segregation)

J.V./J.P. 20·4·60

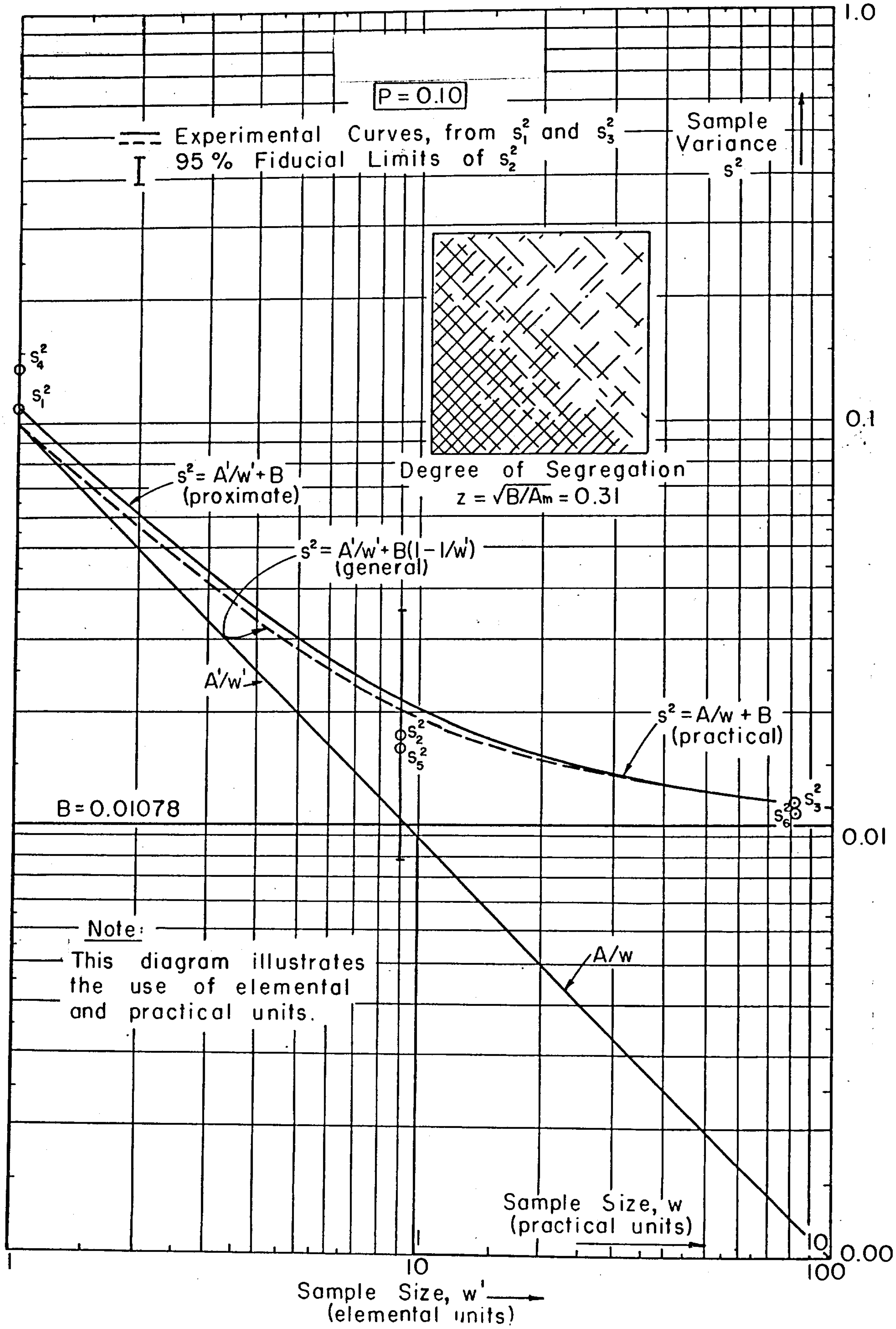


Figure 5. Size Variance Curve (Practical Units)

J.V./J.P. 30·1·61

limits with this curve as before. It will be assumed for the sake of convenience that the size of samples is expressed in a practical unit of measurement equal to ten elementary units. The general Equation 4 now changes to:

$$s^2 = A/w + B\,(1 - 1/10w) \ldots\ldots\ldots\ldots\ldots \text{(Eq } 4^a\text{)}$$

where A = variance of samples of 1 unit of measurement,

w = sample size expressed in same unit of measurement, and

A/w = random variance component.

It is noted that the numerical value of the random variance component does not change by this transformation, as shown in Figure 5. The only difference is that $A = 1/10\ A'$.

It is also noted that the segregation variance B is independent of the unit of measurement.

In those cases where samples have to be expressed in some unit of measurement that is many times the size of an elemental binomial unit, the upper part of the size variance curve as shown in Figure 5 is not used. Consequently, the general Equation 4 can be replaced by:

$$s^2 = A'/w' + B$$

or, when using practical units of measurement,

$$s^2 = A/w + B \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(Eq 5)}$$

The curve corresponding to this equation is also shown in Figure 5. The discrepancy between the general curve and the practical curve turns out to be negligible for a first approximation of the total variance estimate. The same conclusion holds for higher degrees of segregation. Equation 5 will be used from here on, unless otherwise indicated.

Equation 1 for the degree of segregation (z) likewise changes, when practical units of measurement are used, to:

$$z = \sqrt{B/Am} \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(Eq 6)}$$

where m = number of elemental units per unit of measurement.

Equation 6 will appear to be useful as (z) can often be estimated from available data on the average composition and distribution of a material consignment. Examples 4 and 5 (pp. 24 and 25) illustrate the application of Equation 6.

It is noted that the product (Am) is dimensionless and can be estimated from any other unit for which the value of (A) is known.

In view of the above tests, it can be concluded that the variance of single samples drawn systematically or at random from segregated materials consignments can be expressed as a function of two constants determined by the composition of the material and the degree of segregation of the consignment, and by the size of the sample.

When single samples are combined, as is done in incremental sampling, the total variance of a gross sample consisting of (N) increments has a maximum value equal to 1/N times the total variance of the single samples. Theoretically, this maximum value will be attained only when the "patches" caused by segregation of the consignment are themselves distributed at random. In actual practice this condition may not prevail and the total variance as formulated for gross samples consisting of (N) increments,

$$s^2 = A/Nw + B/N,$$

is, in fact, an estimate of the upper limit of the gross sample variance. The estimate of the total variance obtained from this equation is therefore a safe estimate; the same equation can be written as follows:

$$s^2 = A/W + B/N \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(Eq 7)}$$

where $W = Nw$ = the gross sample size.

This equation, originally introduced for the sampling of broken coal (17,18), is suggested as a general expression of variability, for gross samples drawn from material consignments that are not perfect mixtures.

<u>Comparison With Existing Theory</u>

The theory of sampling that is presently being applied when assessing the precision of incremental sampling of segregated material consignments is a modified random sampling method known as "representative sampling", as has already been mentioned in the Introduction. When applying this method it is necessary to determine the number of increments and their distribution over individual "strata" in such a manner that all strata are represented in the gross sample in direct proportion to the individual size and variability of the strata as expressed by the within-stratum standard deviation (16). The increments are drawn in a random manner.

The advantage that can be claimed for the representative sampling method is that the precision of the gross sample is not affected by any "trend", that is, by variations between strata.

The requirements of "proportional representation" may, on the one hand, cause some complications when the strata differ in size and in variability, as is often the case in census surveying.

It is then necessary to evaluate the size and the standard deviation for each individual stratum. On the other hand, the theory of representative sampling can be simplified in many instances, such as in bulk sampling, by choosing imaginary "strata" of equal size and finding an average estimate for the within-stratum standard deviation or from previous knowledge regarding those same variations in a similar material.

This simplified method of representative sampling is generally applied to the systematic sampling of bulk materials as well as to "discrete populations", under which can be classified a great variety of mass-produced articles. Manufactured goods generally show considerably less variability than do raw materials, and quality control systems for such goods can be handled by representative sampling theory without much trouble.

For those categories of materials where the variability is very pronounced, special techniques have been developed based upon the theory of representative sampling.

Hansen, Hurwitz and Madow, in a recent publication on census sampling (9), list no less than ten different sampling techniques, including simple random sampling, cluster sampling, systematic sampling, stratified simple random sampling, simple one-and two-stage cluster sampling, stratified single and multi-stage cluster sampling, multi-stage sampling with large primary sampling units, double sampling, sampling for time series, and purposive sampling. This book, which deals exclusively with finite populations, is indicative of the complexity of present sampling theory, even in limited fields such as census surveying.

Binomial Sampling Theory

Application of the binomial theory to segregated materials has been studied by W.M. Bertholf for broken coal (2,3,4,5) and by H.J. de Wijs for ores in place (21).

In the "trend variance" theory suggested by Bertholf a formula identical to Equation 7 is used. The true nature of the "unit increment variance" (random variance) is left in doubt, because two different methods are used to determine this variance. In the publication first mentioned (2), Bertholf defines the "unit increment variance" as the variance "within sets", as distinct from the variance "between sets". Thus, like the "intra-class" and "inter-class" components used in representative sampling, the "unit increment variance" and the "trend" variance proposed by Bertholf are not independent of the size and number of samples from which they are derived. In a contemporaneous paper (3), however, the same author defines the "unit increment variance" (random component) correctly as $s_1^2 = \frac{4}{g} pq\overline{w}$. This is an approximation of the binomial variance for single, average coal particles. The two definitions are not identical.

The method introduced by de Wijs (21) is a very significant application of the binomial theory to the sampling of solid ores. Briefly, this theory deals with the analysis of a series of samples representing equal masses of the ore body. The variance of the sample mean is determined from the mean value of the samples and from the differences between adjacent samples. A coefficient (d) is introduced for expressing the "dispersion of grade" (page 367 of the article), which, like (z), varies from 0 to 1 and is identical with the latter, except for the manner in which it is determined. The author quotes the following values for (d):

TABLE 4

Segregation of Ores in Place (after H.J. de Wijs)

TYPE OF ORE	Dispersion of grade expressed by (d)			
	conspicuously regular	"no comment"	fairly irregular	extremely irregular
Hydrothermal fissure veins Cu, Pb, Zn, Sn	< 0.15	0.15–0.25	0.25–0.35	> 0.35
Hydrothermal deposits of Au, Pt, Ag		0.35–0.45		
Ta, Nb or Be in pegmatites	more irregular than gold, etc.			
Stratified deposits of Fe, Mn				> 0.20

A more recent publication on a graphical approximation of the mean grade of ores, based on the binomial distribution by M. Bruté de Rémur (7), is of interest to note, as well as the work of R.M. Becker and Scott W. Hazen (1) on the binomial distribution of ore grade.

While the emphasis in the report presented here is on the design of sampling experiments for the purpose of predicting sample precision, it is of interest to mention a simple and effective method for checking the precision actually obtained after the sampling experiment has been completed. This is the duplicate sampling method, introduced by R.L. Brown (6) and R.C. Tomlinson (14), that has been incorporated in the new British specification BS-1017 - Sampling of Coal and Coke. In this method, alternate increments are collected in one bin, the other ones in a second bin; the precision obtained is estimated from the difference between the mean values of the two samples. Other materials may require different methods for the a posteriori determination of sampling accuracy; the discussion of such techniques falls outside the scope of this report, the main objective being the evaluation of the precision of a sampling experiment in advance. This can be done by determining the sampling constants (A,B) from a test, if the material is unknown, or from available data if the material is known by composition and distribution.

MATERIALS OF UNKNOWN COMPOSITION

Sampling constants (A,B) and the degree of segregation (z) for materials of unknown composition can be determined with the duplicate sampling method, using small and large samples (20). This test requires the collection of two series of single samples from which an estimate of the total variance (s^2) is found. For the first series relatively small samples (w_1) are chosen, to ensure that the first term, A/w, in Equation 5 contributes more to the total variance than the second term. The estimate (s_1^2) therefore largely reflects random sampling component (A/w). The second series of samples are of relatively large size (w_2); consequently the variance found from this series is caused mainly by the segregation component (B). The following equations 8 and 9 provide maximum estimates, by first-order approximation, of sampling constants A and B (see also p. 4).

$$A = w_1 w_2 (s_1^2 - s_2^2)/(w_2 - w_1) \quad \text{.....................(Eq 8)}$$

$$B = s_2^2 - A/w_2 \quad \text{.....................................(Eq 9)}$$

The error of reduction and analysis of individual samples has been ignored in these equations; the inflation caused in the estimates of (A, B) is generally of no consequence. The sample sizes (w_1, w_2) should generally be the smallest and largest sizes practically possible.

The degree of segregation (z) is expressed by Equation 6. In many materials that are mass-produced the degree of segregation (z) does not change very much, although the pattern of distribution may vary; and it is thus possible to estimate B without a test when (A) and (z) are known.

A condensed schedule of the calculations required for determining sampling constants (A, B) and the degree of segregation (z) is presented in Table 5.

TABLE 5

Calculation of (A,B) and (z) for Materials of Unknown Composition

Sample No.	Small Samples		Large Samples		Calculations
1 . . . n (see note)	p_1 . . .	p_1^2 . . .	p_2 . . .	p_2^2 . . .	Determine the variance for each series, (s_1^2) and (s_2^2), with the equation: $s^2 = \frac{\text{sum } p^2 - (\text{sum } p)^2/n}{n - 1}$ Determine (A,B) from equations 8 and 9 . Find (z) from equation 6 .
	sum p_1	sum p_1^2	sum p_2	sum p_2^2	NOTE: It is recommended to collect a minimum of 25 to 30 samples for each series.
Average size of samples	w_1		w_2		

Example 4

An untreated stove coal ($1\frac{1}{2}$ x $2\frac{3}{8}$ in.) was sampled by collecting 35 increments with an average weight of 185 grams, and a second series of 35 samples with an average weight of 6,539 grams each. These samples were analyzed for ash content. The variance for the small samples (calculated from fractional ash content) was $s_1^2 =$ 0.0234; the variance for the large samples was $s_2^2 = 0.00219$. Sampling constants found from Equations 8 and 9 are:

A = 4.04 for samples of 1 gram

B = 0.00157.

The weight of the gross sample and the number of increments can be found, for any pre-assigned accuracy, from Equation 7:

$$s^2 = 4.04/W + 0.00157/N$$

For instance, a sampling precision of 1% ash would be obtained 19 times out of 20 when collecting 128 increments with a total weight of 320 kilograms. The average particle weight of the coal was found to be 29.6 grams. Consequently, the number of particles per gram of sample is m = 1/29.6. The degree of segregation, as calculated from Equation 6, is found to be z = 0.11.

Example 5

The results of a general election were used in the following duplicate sampling test: the variance s_1^2 of the individual political adherence to a certain party (X) was compared with the variance of the average political adherence to the same party in the ridings. The average number of votes per riding was $w_2 = 15,430$, while $w_1 = 1$. The variance s_1^2 was found to be 0.27; variance s_2^2 appeared to be 0.0045. The resulting variance formula is:

$$s^2 = 0.27/W + 0.0045/N$$

The number of investigators required for probing the political opinion of the same population at some future date, and the number of interviews to be made by each investigator, can be estimated in advance with this equation. For instance, public opinion regarding the same party (X) could be determined to the nearest 1.5% by about 320 pollsters who would each interview 20 persons. The degree of segregation (z) for this population, with regard to its political adherence to party (X), follows from Equation 6 for m = 1; it follows that z = 0.13.

The following example demonstrates the application of Equations 5, 6 and 7 for materials that are characterized by a variate (X) but that do not consist of mixtures of identical units.

Example 6

Mixtures of particles of unequal size that are sampled for size analysis can be regarded as binomial mixtures by defining variate (X) as a particle size interval within two given size limits. The material consignment can then be regarded as to consist of two fractions (X) and (non-X), as before. The precision of the weight percentage of particles (X) found from a sample is determined by Equations 5 and 6. Estimates of the sampling constants A and B can be found from a duplicate sampling test as demonstrated above by collecting two series of samples, one series consisting of relatively small samples and the second series of relatively large samples.

The substance to be sampled may occur in the form of broken aggregate, solids in suspension, or droplets in an emulsion. When a material occurring in one of these forms is sampled, the chance error as expressed by the binomial variance is now caused by the accidental interchange of units of differing size and depends therefore on the size and relative abundance of the units.

When the particles are small and the number of particles per unit of weight is large, the value of the sampling constant A for samples of unit weight will generally be small in comparison with that of sampling constant B. The effect of segregation prevails over random variation; the frequency distribution of (X) will generally show an irregular form, depending on the pattern of segregation and the number of particles contained in each sample used for the determination of (X).

Solid Aggregates

When the material consignment consists of a solid aggregate, random errors caused by the accidental interchange of units (X) and (non-X) are automatically precluded because no movement of these units relative to one another is possible. While this does not exclude all random variation, most of the variations are caused by segregation when the elemental units that are the carriers of the variate are very small in comparison with the sample.

In materials of this type the variability of (X) is often of the binomial kind, as, for instance, when sampling ore in place for its metal content. The ore consists of a mixture of molecular units (X) and other constituents (non-X). All variability originates from this binomial mixture, but substantially in the form of segregation. The sampling constant (B) for molecular units can be calculated with the binomial equation or measured directly.

The practical value of the binomial theory lies in its application to materials of known composition and distribution, as will be demonstrated in the next section.

MATERIALS OF KNOWN COMPOSITION AND DISTRIBUTION

When the main characteristics and distribution of a material consignment are known, its sampling constants can often be determined without a test. Sampling precision as expressed by the total variance of sampling can be determined from Equations 5, 6 and 7 for binomial variates when the average value of the variate and the degree of segregation (z) of the consignment are known.

Binomial Variates

The sampling constant (A) is calculated from the binomial equation, which takes different forms depending on the type of material and variate. The sampling constant (B) is calculated from (A), the degree of segregation (z), and the ratio (m) denoting the number of units of the material contained in the unit of measurement used for expressing variate (X).

The "materials" are subdivided into three main classes (see Table 6). The first class deals with materials consisting of distinct units, each one of which is the bearer of a characteristic

TABLE 6

Calculation of Sampling Constants for Materials of Known Composition and Distribution
(Binomial Variates Only)

Class of Material	I			II	III
	Material consisting of separate items characterized by (X) and (non-X) in gaseous, liquid or solid form, or in mixtures of same (suspensions, emulsions, pulps or pastes). Items (X) can be separated from items (non-X) by physical or chemical methods.			Material consisting of separate aggregates of (X) and (non-X). The aggregates are characterized by "high-X" and "low-X" and are separable.	Other materials. 1. Variate (X) is dispersed without being accumulated in separate physical units. 2. (X) occurs in units that cannot be identified or separated.
	Items are countable.	The number of items in the sample is too large to be counted.			
Material-Group No.	1	2	3	4	5
Method of Evaluating average grade of consignment	The average grade is determined by counting the number of items (X) and (non-X) in the sample, either directly or after separating items (X) from (non-X).	The average grade is determined by separating the sample by suitable physical or/and chemical methods into two fractions, (X) and (non-X). Fractions are measured by a parameter, expressed in a suitable unit of measurement. Items (X) have same specific gravity as items (non-X).	Items (X) differ significantly in specific gravity from items (non-X).	The average grade is determined directly, by suitable chemical or/and physical analytical methods. Units may have different size and/or specific gravity.	Standard specimen of the material may be required for specific tests.
Parameter used for measuring average grade	Variate (X)	A dimension of the items-length (width, height, depth, diameter, thickness, etc.); surface area; volume.	Weight of fractions (X) and (non-X).	Weight of fractions "high-X" and "low-X".	A length (diameter, depth, expansion, etc.), time; load (force) or other parameters used in the test.
Unit of Measurement	Number	Unit of weight, volume, length, area; surface area per unit of weight; etc.	A unit of weight.	A unit of weight.	A unit of weight, force, time, length, surface area, suitable for measuring the parameter.
Examples	1. Sampling for public opinions. 2. Proportion of defectives (X) in the manufacturing of mass-produced goods.	1. Size analyses. 2. The fineness of hydraulic cement, by surface area (turbidimeter). 3. Sampling of textiles for wool content.	1. Light-weight pieces in aggregate. 2. Float-sink analysis of coal.	1. Ash content (X) of a consignment of broken coal. 2. Sampling of sands for heavy minerals.	1. Sampling of ores in place. 2. The abrasion of crushed gravel, by weight loss. 3. Ductility of bitumen, by elongation.
Sampling Constants	$A = p(1-p)$ p = average fractional number of items (X) known by approximation. $B = Az^2$ z = degree of segregation (known).	$A = p(1-p)/m$ p = average proportional amount of (X) fraction. m = average number of items per unit of measurement. $B = Amz^2$ z = as in (1).	$A = p(1-p)d/Dm$ p = as in (2). d = specific gravity of items (X) or (non-X). D = average specific gravity of material. m = as in (2). $B = Amz^2$ z = as in (1).	$A = p(1-p)(a_1-a_2)^2\ d_1d_2/D\ m$ p = as in (2). $a_{1,2}$ = X-values of fractions (1,2). $d_{1,2}$ = specific gravity of fractions (1,2). D = specific gravity of material. m = as in (2). $B = Amz^2$ z = as in (1).	1. (X) is chemically separable. $B = p(1-p)dz^2/D$ p = average proportional amount of chemical constituent. d,D = as in (3). z = as in (1). 2. (X) is not separable chemically. $B = s^2$ s = standard deviation of (X) from available data.

quality (X) or (non-X). Variability in the values of samples drawn from a consignment of such a material is caused by the fact that these elementary units can move relative to each other; they can be either randomly mixed or can cause a certain degree of segregation in the consignment. It is generally easy to separate the units (X) from units (non-X) in these substances by physical or chemical methods. Most gases, fluids, and mixtures of these with solids (amalgams, suspensions, pastes) belong to this class. Applications of the method can be found in the fields of microchemistry and assaying. Likewise, the sampling of mass-produced items and similar "discrete populations" also belongs in this first class.

The second class of substances comprises materials in which variability is caused as above by the free movement of elemental units, but the variate (X) is not localized to certain units; it is spread over all the elemental units in varying degrees. Granular solids such as broken coal and ore, wheat, and many other materials fall into this class. The units can be separated into two fractions characterized by "high-X" and "low-X"; the variability caused by the relative movement of the units of these two fractions is reflected in the variations of the sample drawn from such material.

A third class of materials is distinguished in which variability is caused by an uneven dispersion of the variate "X" throughout the consignment. Essentially, these materials differ from the above ones only in that the elemental units "X" and "non-X", which may be real or imaginary, cannot move relative to one another; this reduces random variation. Many physical properties such as the tensile strength of wax or the abradability of gravel fall under this category. Distribution of such a variate over the consignment can be attributed to segregation of elementary units, characterized by either "X" or "non-X", that cannot be separated and often not even identified.

All three classes are seen as binomial populations; samples collected from material consignments belonging to the third class have a variance that is substantially determined by segregation.

Five categories of materials are recognized under this main classification; these will now be described in some more detail.

Group No. 1 (see Table 6) deals with substances that occur in the form of separate units, each characterized by either (X) or (non-X). Another feature of this group of materials is that the samples are analyzed by counting the individual units (X) and (non-X).

Groups Nos. 2 and 3 include materials consisting of separate units too numerous to be counted individually and are consequently measured by some dimension of the items (length, surface area, volume or weight) expressed in a suitable unit of measurement (inch, square foot, gallon, pound, etc.).

Group No. 2 includes materials for which the items characterized by variate (X) have the same specific gravity as items (non-X); for instance, granular materials sampled for size analysis.

Group No. 3 deals with materials consisting of items (X) that differ significantly in specific gravity from items (non-X). These are the materials that are sampled for specific gravity analysis (e.g., by float-sink analysis).

Groups Nos. 4 and 5 deal with materials in which the variate (X) is dispersed without being necessarily accumulated in separate physical units of the material.

Group No. 4 includes all materials consisting of separate aggregates that are characterized by either a high percentage of variate (X) or a low percentage of variate (X), the two components being separable.

Group No. 5 includes other materials. Variate (X) is dispersed without being accumulated in separate physical units or it occurs in units that cannot be identified or separated.

The following examples 7 to 12 may serve to illustrate the use of Table 6:

<u>Group 1 (Table 6)</u>

<u>Example 7</u>

A mass-produced item is known to contain about 4% defectives. Therefore, $p = 0.04$ and sampling constant $(A) = 0.0384$ or approximately 0.04. It follows from Equation 4 that the effect of any segregation can be eliminated by collecting sample items one by one $(w' = 1)$. The number (N) of items required for determining the percentage of defectives to the nearest 1% nineteen times out of twenty now follows from

$$N = A/s^2, \text{ where } s^2 = 26 \times 10^{-6}.$$

Consequently, N = 1,500.

<u>Example 8</u>

The results of a general election are used to determine the number of investigators to be employed in a poll to survey the changes in political popularity, and the number of persons to be interviewed by each investigator. The party whose election returns were closest to 50% was party (X), its vote amounting to 61% of the total returns; this figure is subject to the greatest variations and is used as a yardstick for evaluating sampling precision of the poll. Consequently $p = 0.61$ and the sampling constant $(A) = 0.24$. The degree of segregation for (X) is known to be $z = 0.13$; it follows that the sampling constant (B) is 0.0041. From the many possible combinations of (w) and (N), a value w = 20 is chosen as a reasonable figure for the number of persons that can be interviewed by one investigator in one day.

It follows from Equation 7 that, by employing 155 investigators, the results of the poll will indicate political popularities with a precision of 2%, nineteen out of twenty times. The total number of persons interviewed would thus be: wN = 3,100.

Group 2 (Table 6)

Example 9

It is required, for the operational control in an ore beneficiation plant, that a daily sample of minus 14 mesh sand be collected for sieve analysis. The precision of the sieve curve is important, especially with regard to the silt fraction, which should be determined with a precision of 1% nineteen out of twenty times. The sand is segregated (z = 0.20); the average amount of silt (minus 200 mesh material) is 3%.

The accidental interchange of silt particles with sand particles during sampling is determined by the size of the particle. Errors thus caused depend primarily on the size and relative abundance of the coarse particles; that is, on the sand fraction. The weighted average particle weight of the sand fraction (14 x 200 mesh) of this ore is known to be 0.010 gram. Therefore, m = 100, when expressing the sample weight in grams. It follows that:

$$A = p(1-p)/m = 0.0003$$

$$B = Amz^2 = 0.0012.$$

Samples in this plant are collected automatically by increments weighing 30 grams each. The minimum number of increments required now follows from Equation 7:

$$N = 47.$$

Group 3 (Table 6)

Example 10

A non-uniform lightweight aggregate is tested by a float-sink analysis for determining the percentage of lightweight pieces. The material is known to contain approximately 10% by weight of lightweight pieces floating on bromotrichloromethane (sp. gr. 2.00); the average specific gravity is d = 1.6. The average specific gravity of the entire aggregate is D = 2.3. The degree of segregation is known to be z = 0.3. The size of the lightweight aggregate is minus $1\frac{1}{2}$ inch; the rated average particle weight is 15 grams; hence m = 1/15 = 0.067. The sampling constants A = 0.934 and B = 0.0056 are found from the equations given in Table 6 under Group No. 3.

Increments are collected by an automatic sample cutter, each cut weighing approximately 400 grams. The minimum number of increments required to attain a sample precision of 1% follows from Equation 7:

$$N = 303.$$

The weighted average particle weight can be determined from a sieve analysis, using the following equation (11, 17, 18):

$$V = \Sigma q l^3 / \Sigma q$$

where V = weighted average particle volume, in cu cm.,

q = weight of individual size fraction, and

l = central value of individual size fraction, in cm.

Group 4 (Table 6)

Example 11

A minus-$1\frac{1}{2}$-inch mine-run slack coal with an average ash content of about 30% is sampled for ash by an automatic sampler collecting increments of 5 lb. This coal is known to contain approximately 64% ($p = 0.64$) floats at 1.60 sp. gr. with 5% ash ($a_2 = 0.05$), and 36% sinks with approximately 80% ash ($a_1 = 0.80$). The specific gravity of these two fractions are known to be $d_2 = 1.30$; $d_1 = 2.35$; the overall specific gravity $D = 1.60$.

The weighted average particle weight (Example 10) of this coal is 5.26 grams. As the weight of sample is expressed in pounds (1 lb = 454 grams), the ratio $m = 454/5.26 = 86$. The degree of segregation of the mine-run slack is known to be $z = 0.13$. From this it follows that the sampling constants (see Table 6, Group 4) are:

$$A = 0.00288$$

$$B = 0.004186 .$$

The minimum number of increments required to determine the ash content with a precision of 1% ash, nineteen out of twenty times, is $N = 183$. The gross sample weight is therefore 915 pounds.

Group 5 (Table 6)

Materials in this group occur as a solid or fluid mass in which the variate (X) is dispersed without being accumulated in separate physical units; or, the variate occurs in units that cannot be identified or separated and is measured in some indirect manner.

Under these circumstances there can be no accidental interchange of units (X) and (non-X) during sample collection, except at the molecular level, as in the sampling of fluids. Therefore while sampling constant (A) may have a distinct value for molecular units or similar, very small aggregates, its value for any practical unit of measurement becomes negligibly small as the ratio (m) approaches infinity. While the binomial distribution is inoperative with regard to chance variations occurring during sample collection, it is still the prime cause of all segregation.

In materials under this group where variate (X) is a constituent that can be extracted by chemical means, (A) can generally be calculated for molecular units and sampling constant (B) can then be estimated as before, from the average composition of the material and its degree of segregation (z).

In other materials under this group, where (X) does not refer directly to units that can be determined or separated by chemical extraction (such as the compressive strength of briquets, the ductility of bitumen, etc.), sampling constant (B) can only be found from available variance data.

Example 12

The sampling of ore in place will be used as an example to illustrate the use of the equations mentioned in Table 6 under Group 5.

Channel samples are collected from a zinc vein containing 10% metallic zinc in the form of smithsonite ($ZnCO_3$); the degree of segregation of the metal is known to be z = 0.20. As the zinc occurs in the form of the carbonate, it follows that the proportional amount of this constituent is p = 0.20; the specific gravity of smithsonite is d = 4.4; the average specific gravity of the ore is D = 2.8. It follows, for sampling constant (B), that

$$B = p(1 - p)\ dz^2/D = 0.010.$$

The total sample variance:

$$s^2 = 0.010/N.$$

This variance is independent of sample weight. The number of increments required to attain a sampling precision of 1% zinc is found to be

$$N = 384.$$

Non-Binomial Variates

In actual sampling practice many instances are found where the variate has a non-binomial parent distribution. For instance, in the sampling for the number of defectives the variate has a parent distribution of the Poisson type. In many other cases the parent distribution is a normal curve, but frequently curves of irregular shape are encountered as well.

While the parent frequency curves of variates may differ, they have one common property: the difference between the true value of any sample and the true mean of the material lot from which such a sample originates can be expressed as the algebraic sum of two deviations, one caused by random variation, the other by segregation. The efficiency of this distinction lies in the fact that it applies to any variate and any material.

The law of propagation of errors applies (see derivation in the Appendix), provided these two individual deviations are independent of each other for any sample or increment. It is impossible to prove, by mathematical analysis, the correctness of this assumption for all materials and all variates. From tests on the sampling board and results of field trials (5) it can, however, be understood intuitively that here the law of propagation of errors has a general application, which means that Equations 8 and 9 apply, independent of the type of frequency distribution of the variate (X). It may be noted here that in cases where the mean value and the standard deviation of a variate are related it is often possible to transform the variate by substitution with a variate whose mean (M) and standard deviation (s) are approximately independent of each other.

Generally, if (s) is a function f(M) of the mean (M), the appropriate transformation to stabilize the variance of (X) is:

$$Y = \int \frac{1}{f(X)} dX$$

Examples:

Relationship	Transformation
(s) proportional to M^2	Take reciprocals of observations
(s) proportional to M	Take logarithms of observations
(s) proportional to $\sqrt{M}$	Take square roots of observations

Such transformation variates can be used in extreme cases where the above conclusions would not apply.

SAMPLING TO A PRE-ASSIGNED ACCURACY

The main motive for this report has been to formulate a common basis for evaluating the precision of the average grade of a material in simple terms, regardless of the type of material or variate and of the state of segregation of the consignment.

The guiding principle has been to determine the causes of variability in any material and to find general equations rather than to adapt a statistical technique to a given class of materials and/or a certain type of variate.

The conclusion from this study is that in any sampling experiment the difference between the true sample value and the true mean of the lot can be expressed as the sum of the random deviation and a remaining deviation which is caused by the fact that the material is not randomly mixed. Consequently, two variance components can be distinguished that are common to incremental sampling experiments with all types of materials and variates, and these can be expressed in an equation that relates sampling variance to the number of increments and the size of the gross sample. Tests with a model population confirm that the variance estimates found from this equation hold for the systematic sampling of segregated populations.

In the method presented here, use has been made of early work done by Mika (12), followed by Kassel and Guy (10), Landry (11), Deming (8) and, more recently, de Wijs (21). The sampling variance can be forecast for materials of known composition and distribution when the variate has a binomial parent distribution. In cases other than this the variance components are found from a duplicate test with small and large samples.

Great value has been attached to clarity, because the statistics of sampling stands in need of simplification lest it remain a specialist's domain. The work of Moroney (14) has been very stimulating in presenting statistics in ordinary language. It is recognized that existing methods such as representative sampling have their place in certain fields as far as they are useful in calculating "intra-class" and "inter-class" variances. In other respects the practical limitations of these methods are obstructing a broader application of sampling statistics that ought to cover the forecasting of the precision of sampling, including systematic sampling; the latter is an accepted practice that has thus far remained a controversial subject amongst statisticians.

The proper collection of samples is a matter of training and strict adherence to good specifications, rather than of theory. The sampler should know how to avoid bias (systematic errors) during sample collection and how to avoid having his increments get "in step" with the periodicity of the variate. Equations 6, 7, 8 and 9 can be used effectively only if these conditions are met.

The sampling board is recommended to experimenters as an effective and inexpensive device for testing the quantitative aspects of sampling theory. It has already proved to be useful in testing the quantitative importance of some theoretical objections voiced by statisticians. The main objective is, and should be, to estimate the precision of an average value by first-order approximation, rather than to argue the precision of that precision.

The literature cited indicates that, by carefully sorting out what is significant from what is trivial, the obstacles to a unified method of evaluating sampling precision can be removed. The present report is intended as a contribution to that end.

ACKNOWLEDGMENT

The writer is indebted to Dr. R.P. Charbonnier, Senior Scientific Officer of the Fuels and Mining Practice Division, Mines Branch, Department of Mines and Technical Surveys, Ottawa, and to Dr. J.R. McGregor, Assistant Professor, Department of Mathematics, University of Alberta, Edmonton, for reading the manuscript and supplying valuable comments.

REFERENCES

(1) Becker, R.M. and Hazen, Scott W., Jr., Probability in Estimating the Grade of Ore; Ninth Annual Drilling Symposium (1959), Penn State University, University Park, Pa. (Preprint)

(2) Bertholf, W.M., The Analysis of Variance in A Sampling Experiment; A.S.T.M., S.T.P. 114 (1951).

(3) Bertholf, W.M., The Design of Coal Sampling Procedures; A.S.T.M., S.T.P. 114 (1951).

(4) Bertholf, W.M., The Development of the Theoretical Basis of Coal Sampling; A.S.T.M., S.T.P. 162 (1954).

(5) Bertholf, W.M., The Effect of Increment Weight on Sampling Accuracy; Symposium on Bulk Sampling (1958); A.S.T.M., S.T.P. 242, p. 30, Figure 2.

(6) Brown, R.L., British Coal Utilization Research Association (BCURA) Information Circular 39 (1950).

(7) Bruté de Rémur, M., Echantillonnage des gisements; Revue de L'Ind. Minérale, June 1959, pp.457-70.

(8) Deming, W.E., On the Sampling of Physical Materials; Review Intern. Statistical Institute (1950:1/2), p. 11 (reprint).

(9) Hansen, M.H., et al., Sample Survey Methods and Theory; Wiley, New York, 1953, 2 vols.

(10) Kassel, L.S. and Guy, T.W., Determining the Correct Weight of Sample in Coal Sampling; Ind. Eng. Chem., Anal. Ed. 7, 112-15 (1935). (C.A. 3134[5] [1935])

(11) Landry, B.A., The Fundamentals of Coal Sampling; U.S. Bureau of Mines, Bulletin 454 (1944).

(12) Mika, J., Theoretical Notes on Sample Taking; Z. Anal. Chemie 73, pp. 257-64 (1928). (C.A. 22, 1921[9] [1928])

(13) Moroney, M.J., Facts from Figures; Penguin Books (1954), p. 99 (Sampling of Blood for Counting of Blood Corpuscles).

(14) Tomlinson, R.C., The Routine Sampling of Coal; J. Inst. Fuel (August 1953).

(15) Tomlinson, R.C., A Note on Coal Sampling Theories; Fuel XXXVI (1957), pp.442-46.

(16) Yule, Udney G. and Kendall, M.G., An Introduction to the Theory of Statistics; 14th ed. (Hafner Publishing Co., London, 1950), pp. 533-39.

(17) Visman, J., Sampling of Coal and Washery Products; Trans. World Power Conference, The Hague, 1947.

(18) Visman, J., De Monsterneming van heterogene binomiale Korrelmengsels, in het byzonder Steenkool; Thesis, Delft University, Holland, 1947.

(19) Visman, J., Tests on the Binomial Sampling Theory for Heterogeneous Coals; A.S.T.M., S.T.P. 162 (1954), pp. 141-52.

(20) Visman, J., Sampling to a Pre-assigned Accuracy; Canada Department of Mines and Technical Surveys, Mines Branch Report F.R.L. 212 (1955).

(21) de Wijs, H.J., Statistics of Ore Distribution; Geologie en Mijnbouw 13 (1951), pp. 365-75; 15 (1953), pp. 12-24.

JV:(PES)

APPENDIX

LAW OF PROPAGATION OF ERRORS

Application to Random and Segregation Variations

The true value (x) of a sample (i) collected from a segregated population with true average value (y) can be written as follows:

$$x_i = y \pm t_{i1} \pm t_{i2}$$

where t_{i1} = random deviation, and

t_{i2} = deviation caused by segregation.

The total deviation for any sample (i) is, therefore:

$$x_i - y = t_i = \pm t_{i1} \pm t_{i2}$$

From this it follows, for a large number of samples, that:

$$t_1^2 = t_{11}^2 + t_{12}^2 \pm 2t_{11}t_{12}$$

$$t_2^2 = t_{21}^2 + t_{22}^2 \pm 2t_{21}t_{22}$$

$$\cdot\cdot \quad \cdot\cdot \quad \cdot\cdot \quad \cdot\cdot\cdot$$

$$\cdot\cdot \quad \cdot\cdot \quad \cdot\cdot \quad \cdot\cdot\cdot$$

$$t_n^2 = t_{n1}^2 + t_{n2}^2 \pm 2t_{n1}t_{n2}$$

$$\text{Average:} \quad \frac{\Sigma t_i^2}{n} = \frac{\Sigma t_{i1}^2}{n} + \frac{\Sigma t_{i2}^2}{n} + \frac{2\Sigma(\pm t_{i1}t_{i2})}{n}$$

From this it follows, by first-order approximation, that:

$$s^2 = s_1^2 + s_2^2$$

where s_1^2 = random variance, and

s_2^2 = segregation variance.

The mean value of the double products is of a lower order of magnitude owing to opposite signs, provided there is no correlation between t_{i1} and t_{i2}.

The derivation applies for any type of parent distribution and supports the general validity of Equation 5.

JV:(PES)