INSTYTUT MATEMATYCZNY POLSKIEJ AKADEMII NAUK

# ROZPRAWY MATEMATYCZNE



V

STEFAN DROBOT i MIECZYSŁAW WARMUS

## Dimensional Analysis in sampling inspection of merchandise

WARSZAWA 1954

PRINTED IN POLAND

Polskie Towarzystwo Matematyczne, Warszawa, Śniadeckich 8.
Nakład 600 Papier druk. satyn., kl. III, 70×100, 80 g. Arkuszy druk. 3³/₈
Oddano do składu w maju 1953 r. Cena 5 zł Druk ukończono w lutym 1954 r.
Zam. 439/53 Wrocławska Drukarnia Naukowa, Wrocław, Świerczewskiego 19. F-4-19193

## I. Methodological remarks

Sampling inspection of merchandise consists in examining a sample drawn from a lot and thus establishing a characteristic of the lot.

The theory of sampling inspection is usually based on the theory of probability and mathematical statistics. In order to solve the most important problems met with in the theory of sampling inspection we accept certain hypotheses of mathematical or economical nature and then awail ourselves of the circumstance that the problem concerns a large number of objects. The mass character of the phenomenon is the necessary premise for the application of the theory of probability with its well-developed apparatus.

This method, though sanctioned by tradition, has some disadvantages. First, if we want to apply it, we must accept at the very outset certain definite statistical, economical, and sometimes other hypotheses, and then, using the methods of the theory of probability we solve the problem and obtain definite, strictly determined, results. If the results or the hypotheses on which they are based give rise to any doubts, the whole solution is useless, and in order to correct the results the whole theory must be rebuild and problems corresponding to a different set of hypotheses must be solved.

Besides — and this is the second disadvantage of the methods used hitherto — the method of solving those problems is often fairly complicated and sometimes requires strónger mathematical means. Therefore, for technical considerations, it is not always possible to introduce into the calculations all hypotheses which sufficiently characterize the phenomenon in question, and it becomes necessary to adopt certain schemes which are often unduly simplified. The results obtained in this way are usually rigid, and if not wholly confirmed by practice they become useless as a whole, so that no part of them can be accepted.

The third disadvantage consists in the circumstance that in the theories of sampling inspection used hitherto the part played by experiment and the experience of expert practicians has not been sufficiently taken into account, because it has not been clearly seen. The methods used hitherto do not make it plain whether and how the results obtained from theory can be compared with reality. The differences of opinion between individual authors, concerning various details of theory are usually not settled by means of experiment but left unquestioned as the authors' personal convictions.

The aim of the present paper is to present a theory of sampling inspection free from the disadvantages mentioned above. This theory is not based on the theory of probability. Mathematical statistics plays a part in it but it is a different one to that it has played in the theories mentioned above. The theory presented here is *phenomenological.*

It's a matter of fact, any scientist who investigates reality, and any one who applies the theory of probability to problems of natural science, technology, economy, or other sciences, base their work on the conviction that there exists an objective relation between the elements of the phenomenon investigated. This relation can be discovered by the methods of the theory of probability and mathematical statistics, but it can also be done without them.

In order to express this fact and its consequences in a clearer and more definite way let us dwell for a while upon a special example. Let us consider, by way of comparison, a theory describing certain real phenomena, *e.g.* thermodynamics. Roughly speaking the historical development of that science has been the following. First certain facts were discovered and the connections between them were described phenomenologically. Thus, *e. g.* Boyle and Mariotte's law of gases was discovered and then the first and the second law of thermodynamics. The phenomenological theory of those phenomena did not make use at all of the atomic properties of matter and of the probability methods. All the same, it was a consistent theory, and — what is more important — its results agreed with the results of experience to a sufficient degree. It is a fact that the application of the theory of probability and mathematical statistics to thermodynamics has enabled us to approach well-known facts from a different angle and has given us an insight into certain more subtle phenomena, but the facts described in phenomenological thermodynamics have remained true, although in a statistical sense. For practical purposes, in physics and technology, phenomenological thermodynamics has lost almost nothing of its value and its theoretical foundations have even been strengthened.

In the theory of sampling inspection statistical methods have been used from the very beginning. In this respect the history of its development differs from the history of thermodynamics. However, the methods of the theory of probability and mathematical statistics often prove too subtle for such gross problems as those encountered in sampling inspection. Besides, they have the disadvantages mentioned above. Therefore it seems important — at least for practical purposes — to construct a *phenomenological theory* of sampling inspection.



The principal idea on which we base this theory is the conviction that *sampling inspection is a phenomenon determined to a sufficient degree.* We assume that there exists a definite objective connection between the investigated characteristic of the sample and the corresponding characteristic of the lot. The assumptions concerning this connection are formulated in Chapter III in a general way, which — we think — well reflects reality. General principles of this theory, formulated in Chapter III, will resemble in character the principles of classical thermodynamics or mechanics.

On the basis of the principles thus formulated we shall be able, without applying the methods of probabilistic solutions, to solve in a sufficiently exact way the most important problems encountered in the theory of sampling inspection, *e. g.* we shall be able to deduce sufficiently determined formulae for the sample size. The calculations necessary to solve those problems are quite simple and seem to require only the knowledge of elementary algebra. The method which we shall use is called the *dimensional analysis.* It will enable us to solve many fundamental questions in a way that is general, uniform, and at the same time very simple.

The dimensional analysis has been known in physics and technology for a long time, though it has not been sufficiently determined and fully utilized. This method makes it possible to solve in a very efficient way many concrete problems by elementary means. *E. g.* in mechanics it is possible to deduce the formula of the period of pendulum swing merely on the ground of general considerations concerning the units of the quantities in question. Owing to the method of the dimensional analysis it is even sometimes possible to give a quantitative description of a phenomenon for which no sufficiently developed other theory exists.

Ending this chapter we should like to point out a circumstance important for non-mathematicians, viz. that both the dimensional analysis and the theory of sampling inspection, which we shall develop, are abstract theories constructed axiomatically. Abstraction in mathematics does not at all mean a breach with reality; it consists in singling out common characteristics of different phenomena and studying the connections between those characteristics. Owing to this the abstract method has extraordinary power and generality. But there are dangers to be avoided. Fundamental notions and postulates of abstract mathematical theories are usually formulated in a very simple way and so as to evoke intuitive understanding. This intuitiveness and the misleading effect of the everyday meaning of the terms used may sometimes obscur to a considerable degree the actual generality of the theory and lead to false conclusions.

## II. Dimensional analysis

Let us briefly state those principles of the dimensional analysis which will be necessary in the sequel[1]).

In the dimensional analysis we use certain *quantities* $A, B, C, \ldots$, among which are also *positive* numbers $\alpha, \beta, \gamma, \ldots$ The quantities which are not numbers are called *dimensional quantities*. Such quantities are *e. g.* 1 gram, 1 liter, 3 cubic meters, 4 amperes, 3 eggs, 4 dozen buttons, 17 wagon-loads of coal, 7 pieces, etc.

Ordinary numbers will sometimes be called *dimensionless quantities*.

We assume that the following operations can be performed on dimensional quantities. We can

1⁰ multiply the quantity $A$ by the quantity $B$; the result of the multiplication is written: $AB$,

2⁰ raise the quantity $A$ to a power of an arbitrary real exponent $a$; the result of this operation is written: $A^a$.

The results of these two operations are, again always, dimensional or dimensionless quantities. The rules of multiplication and involution are the same as for ordinary numbers. For every dimensional quantity $A$

$$1A = A \tag{1}$$

and

$$A^0 = 1. \tag{2}$$

We say nothing of addition for the present. It will be discussed at the end of this Chapter.

The quantities $A_1, A_2, \ldots, A_m$ will be called *dimensionally independent*, if the equality

$$A_1^{a_1} A_2^{a_2} \ldots A_m^{a_m} = \alpha \tag{3}$$

[1]) Cf S. Drobot, *On the foundations of Dimensional Analysis*, Studia Mathematica 14, to appear.



holds if and only if

$$a_1 = a_2 = \ldots = a_m = 0, \qquad \alpha = 1. \tag{4}$$

Otherwise those quantities will be called *dimensionally dependent*.

We assume that in the phenomena to which dimensional analysis can be applied there always exist a certain number $m$ of dimensionally independent quantities, any $m+1$ quantities being dimensionally dependent. Every set $X_1, X_2, \ldots, X_m$ of such dimensionally independent quantities will be called a *system of units*. *E. g.* in classical mechanics the system of units cm, g, sec is adopted. It is easy to prove that among the quantities $X_1, X_2, \ldots, X_m$, which constitute a system of units there can be no ordinary number, *i. e.* a dimensionless quantity.

We shall give four theorems without proof.

THEOREM 1. *Every dimensional quantity $A$ can be presented in a unique manner in a chosen system of units $X_1, X_2, \ldots, X_m$ as follows:*

$$A = \alpha X_1^{a_1} X_2^{a_2} \ldots X_m^{a_m}, \tag{5}$$

*where $\alpha$ is a positive number, and $a_1, a_2, \ldots, a_m$ are real numbers.*

We say that two quantities $A$ and $B$ have the *same dimension* if there exists a number $\alpha$ such that $B = \alpha A$. The class of quantities of the same dimension is called the *dimension* of those quantities. The dimension of the quantity $A$, given in the system of units $X_1, X_2, \ldots, X_m$, is

$$[A] = [X_1^{a_1} X_2^{a_2} \ldots X_m^{a_m}]. \tag{6}$$

THEOREM 2. *In order that $m$ dimensional quantities $A_1, A_2, \ldots, A_m$, which in the system of units $X_1, X_2, \ldots, X_m$ have the following dimensions:*

$$\begin{aligned} [A_1] &= [X_1^{a_{11}} X_2^{a_{12}} \ldots X_m^{a_{1m}}], \\ [A_2] &= [X_1^{a_{21}} X_2^{a_{22}} \ldots X_m^{a_{2m}}], \\ &\ldots\ldots\ldots\ldots\ldots \\ [A_m] &= [X_1^{a_{m1}} X_2^{a_{m2}} \ldots X_m^{a_{mm}}], \end{aligned} \tag{7}$$

*should be dimensionally independent, it is necessary and sufficient that*

$$\begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1m} \\ a_{21} & a_{22} & \ldots & a_{2m} \\ \ldots & \ldots & \ldots & \ldots \\ a_{m1} & a_{m2} & \ldots & a_{mm} \end{vmatrix} \neq 0. \tag{8}$$

*E. g.* let $X_1=\mathrm{cm}$, $X_2=\mathrm{g}$, $X_3=\mathrm{sec}$. In classical mechanics the following dimensional quantities are used: velocity $V$, whose dimension is $[V]=[\mathrm{cm\,sec^{-1}}]$; acceleration $A$, whose dimension is $[A]=[\mathrm{cm\,sec^{-2}}]$; density $D$, whose dimension is $[D]=[\mathrm{g\,cm^{-3}}]$. These dimensional quantities: velocity, acceleration, and density, are dimensionally independent, because condition (8) is fulfilled, viz.

$$\begin{vmatrix} 1 & 0 & -1 \\ 1 & 0 & -2 \\ -3 & 1 & 0 \end{vmatrix} = 1 \neq 0.$$

In the sequel we shall use the notion of *dimensional function* $\Phi(A_1, A_2, \ldots, A_p)$, whose arguments are the dimensional quantities $A_1, A_2, \ldots, A_p$, and whose values are also dimensional quantities. Functions of *any* kind cannot be used here, they must satisfy special conditions. We are not going to state these conditions here in its precise form. We only explain intuitively that it is required that a dimensional function shall have the same form in every system of units.

All functions used in physics, technology and natural science in general, satisfy this condition, which is also satisfied without any restrictions by all functions whose arguments and values are ordinary numbers (dimensionless quantities). The formulae which we shall introduce in the theory of sampling inspection also satisfy this condition, because only such functions will be used in deducing those formulae. Thus the form of the formulae for the sample size which we shall deduce will not depend, for example, upon the circumstance whether the size of the lot is measured in dozens or in packages, and the sample size in pieces or in boxes; even the number of pieces which make up a dozen is inessential.

THEOREM 3. *If the quantities $A_1, A_2, \ldots, A_m$ are dimensionally independent, then every dimensional function $\Phi(A_1, A_2, \ldots, A_m)$ which satisfies the condition mentioned above must have the form*

$$\Phi(A_1, A_2, \ldots, A_m) = \varphi A_1^{f_1} A_2^{f_2} \ldots A_m^{f_m}, \tag{9}$$

*where $\varphi$ is a certain constant numerical factor, and $f_1, f_2, \ldots, f_m$ are real numerical constants.*

THEOREM 4[2]). *If $B_1, B_2, \ldots, B_r$ are dimensional quantities dimensionally dependent on the quantities $A_1, A_2, \ldots, A_m$, which are dimensionally independent, i. e. if*

[2]) Theorem 4 is known in the dimensional analysis under an odd name of "Π-theorem".



$$\begin{aligned} B_1 &= \beta_1 A_1^{b_{11}} A_2^{b_{12}} \ldots A_m^{b_{1m}}, \\ B_2 &= \beta_2 A_1^{b_{21}} A_2^{b_{22}} \ldots A_m^{b_{2m}}, \\ &\ldots\ldots\ldots\ldots \\ B_r &= \beta_r A_1^{b_{r1}} A_2^{b_{r2}} \ldots A_m^{b_{rm}}, \end{aligned} \tag{10}$$

*where $\beta_1, \beta_2, \ldots, \beta_r$ are positive numbers and $b_{kl}$ ($k=1,2,\ldots,r$; $l=1,2,\ldots,m$) are real numbers, then every dimensional function*

$$\Phi(A_1, A_2, \ldots, A_m; B_1, B_2, \ldots, B_r)$$

*which satisfies the conditions mentioned above must have the form*

$$\Phi(A_1, A_2, \ldots, A_m; B_1, B_2, \ldots, B_r) = \varphi(\beta_1, \beta_2, \ldots, \beta_r) A_1^{f_1} A_2^{f_2} \ldots A_m^{f_m} \tag{11}$$

*where $\varphi(\beta_1, \beta_2, \ldots, \beta_r)$ is an ordinary numerical function (i. e. a function that assumes dimensionless numerical values) of numerical (dimensionless) arguments $\beta_1, \beta_2, \ldots, \beta_r$, while $f_1, f_2, \ldots, f_m$ are certain real numbers independent of $A_1, A_2, \ldots, A_m$ and of $\beta_1, \beta_2, \ldots, \beta_r$.*

Theorems 3 and 4 form the basis of determining the shope of dimensional functions. We shall use them repeatedly.

The following two problems will complete this chapter.

So far the only operations on dimensional quantities have been multiplication and involution. However, for practical purposes it is convenient to introduce also addition, subtraction and limit; these operations can be performed on quantities of the same dimension only. The following equalities:

$$\alpha A + \beta A = (\alpha + \beta) A,$$

$$\alpha A - \beta A = (\alpha - \beta) A,$$

$$\lim_{n\to\infty} (\alpha_n A) = (\lim_{n\to\infty} \alpha_n) A,$$

are, formally, to be considered as the definition of these operations. In order to make subtraction always possible it is natural to introduce also the elements $\alpha A$ with non-positive coefficients. It must be emphasized that the quantities $\alpha A$ with non-positive coefficients $\alpha$ do not belong to the original set of dimensional quantities.

This convention enables us to "add" dimensional quantities of the same dimension and even to introduce the notion of integral and that of infinite series. With quantities of the same dimension we can make all operations that are known in mathematics, without restriction.

### III. General principles of the theory

We proceed to formulate the fundamental notions and postulates of the theory of sampling inspection. They will form a basis for further reasonings, necessary to solve individual, concrete problems. The notions and postulates of the theory of sampling inspection can be introduced in various ways, according to the degree of generality required. Here we shall confine ourselves to one of the possible ways, which is usually sufficient for practical purposes; for we primarily intend to show how a consistent phenomenological theory of sampling inspection can be constructed, and not to develop this theory in its greatest generality. Therefore the notions and postulates that we are going to accept serve as examples, and by no means exhaust all possibilities. Given opportunity we shall point out only some possibilities of generalization, not very significant ones; deeper and more essential generalizations should form the subject of a separate paper.

A *lot of merchandise* is a set $\Omega$ of objects which has the following properties:

If $\Omega_1, \Omega_2, \ldots$ are subsets of the set $\Omega$, *i. e.* parts of the lot, then we assume that we are able to add these parts of the lot and that the result of the addition is always again a part of the lot (or the whole lot). The addition will be denoted by the symbol $\cup$. We assume that for each part of any lot there exist two definite measures, denoted by $N$ and $W$ and satisfying the following conditions:

$1^0$ The measures $N$ and $W$ are dimensional quantities.

$2^0$ The measures $N$ of all parts of a lot have the same dimension, and the measures $W$ of all parts of a lot have the same dimension.

$3^0$ The quantities $N$ and $W$ are dimensionally independent.

$4^0$ If $\Omega_1$ and $\Omega_2$ are disjoint parts of the same lot $\Omega$, then

$$N(\Omega_1 \cup \Omega_2) = N(\Omega_1) + N(\Omega_2), \tag{4'}$$

$$W(\Omega_1 \cup \Omega_2) = W(\Omega_1) + W(\Omega_2). \tag{4''}$$

The quantity $N(\Omega)$ is termed the *size of the lot* $\Omega$; the quantity $W(\Omega)$ is termed the *value of the lot* $\Omega$. Analogously the quantity $N_1 = N(\Omega_1)$ is the *size of the part* $\Omega_1$ of the lot, and the quantity $W_1 = W(\Omega_1)$ is the *value of the part* $\Omega_1$ of the lot.

We shall clarify these notions and the postulates accepted by means of examples. According to our assumptions, 20 wagon-loads of coal, for example, constitute a lot. Let us assume that the addition $\cup$ of two parts of this lot, *e. g.* 3 wagon-loads and 5 wagon-loads consists in heaping the coal together. Let us regard the weight of the coal composing the lot (or its part) as the size $N$ of this lot (or its part). The dimension of the size $N$ is, for instance, kg or ton. The money-equivalent, *i. e.* the value in the everyday sense of the word, will be regarded as the value $W$ of the lot or its part. *E. g.* 20 wagon-loads of coal are worth 500 guineas. Thus the dimension of the value $W$ is guinea. We verify the fact that the size $N$ of the lot and its value $W$, thus accepted, satisfy all the conditions $1^0$-$4^0$.



Let us take another example. Let us regard as a lot 1000 barrels of wine, not all of which contain wine *of the same sort.* Let us assume that the addition $\cup$ of two parts of this lot, *e. g.* 20 liters and 30 liters, not necessarily from the same barrel, consists in pouring the wine together. The number of litres will be the size $N$ of the lot (or its part); litre is therefore the dimension of the size $N$. The money-equivalent, *i. e.* the value in the everyday sense of the word, is the value $W$ of the lot (or its part). *E. g.* let 1000 barrels of wine be worth 25 000 guineas. Then guinea is the dimension of the value $W$. Now, the quantities $N$ and $W$, determined in this way, do *not* fulfil all the conditions $1^0$-$4^0$, viz. condition (4″) is not satisfied. For in wine-trade the amount paid for a mixture of two sorts of wine is not always equal to the sum of the amounts paid for the component parts of that mixture. Often by mixing two brands of wine of high quality we obtain a poor sort of wine, of little worth. Hence, the theory which we are going to present does not apply to such goods as the above mentioned 1000 barrels of wine of *different brands.*

However, the postulates of the theory could be formulated in such a way as to comprise also such cases. This is not done here, because, as has been mentioned above, we are concerned only with presenting a method and not with increasing its generality. Besides, it seems that in many practical instances of sampling inspection only such goods are concerned (*e. g.* coal) as satisfy all postulates of our theory.

Before formulating further notions and postulates let us make a methodological remark. In every axiomatic theory the *terms* of the notions are purely conventional and one should not be misled by their everyday meaning. Thus the term "value" may be used in this theory both in its everyday and metaphorical sense. One thing is essential, namely that the notion designated by the term should satisfy the conditions that we have accepted. Thus the value of a lot may be for instance its money-equivalent (*e. g.* a lot of iron-ore is said to be worth 3000 guineas), or another quantity, for instance the amount of pure product (*e. g.* the total amount of iron in the ore is 1000 t).

In most practical problems it is the money-equivalent that is regarded as the value of the lot. This circumstance may account for the choice of the term, but detracts nothing from the generality of the notion.

The term "size", too, is conventional. *E. g.* we may regard the weight of a lot of coal as its size, volume as the size of a lot of timber, number of pieces as the size of a lot of bricks. The term "size" has been adopted for the sake of generality, regardless of particular interpretations in concrete cases.

The dimension of the size $N$ of a lot (or its part) will be denoted by the symbol "PIECE" (capital letters). Thus

(12) $$[N] = \text{PIECE}.$$

The dimension of the value $W$ of a lot (or its part) will be denoted by the symbol "GUINEA" (capital letters). Thus

(13) $$[W] = \text{GUINEA}.$$

The terms "PIECE" and "GUINEA" are of course entirely conventional.

*E.g.* if we take the weight of a lot of coal as its size, it would be more convenient perhaps to use the term "ton" for the dimension of size. Or if we take the volume of a lot of timber as its size, it would be more convenient perhaps to use the term "$m^3$" for the dimension of size. But if we take the number of bricks in a lot as its size, it is convenient to use the term "PIECE" for the dimension of size. For the sake of generality of reasoning we accept the term "PIECE" regardless of particular interpretations in concrete cases.

The same applies to the term "GUINEA". *E. g.* if we take the money-equivalent of a lot of eggs as its value, it is convenient to use the term "GUINEA" for the dimension of value. But if we take the number of good eggs in a lot as its value, it would be more convenient perhaps to use the term "good egg" for the dimension of value. However, for the sake of generality of reasoning we accept the term "GUINEA" regardless of particular interpretations in concrete cases.

These names are only necessary to fix the terminology in solving various problems. Having solved the problems, we may use in practical computations any convenient names of units, in accordance with the character of the problem considered; or we may even use no units at all.

Let us now introduce further notions. A *sample* is a singled out subset $\omega$ of a lot. A sample must have the following properties:



If $\omega_1, \omega_2, \ldots$ are subsets of the set $\omega$, *i. e.* are parts of a sample, we assume that we are able to add them, and that the result of the addition is again a part of the sample (or the whole sample). This addition will be denoted by the symbol $\vee$.

$\vee$-addition of parts of a sample need not be of the same kind as $\cup$-addition of parts of a lot.

We assume that for each part of each sample of a lot two measures have been defined, denoted by $n$ and $w$, which satisfy the following conditions:

$5^0$ The measures $n$ and $w$ are dimensional quantities.

$6^0$ The measures $n$ of all parts of a sample have the same dimension and the measures $w$ of all parts of a sample have the same dimension.

$7^0$ The quantities $n$, $w$, $N$, $W$, are dimensionally independent.

$8^0$ If $\omega_1$ and $\omega_2$ are disjoint parts of the sample $\omega$, then

(8′) $$n(\omega_1 \vee \omega_2) = n(\omega_1) + n(\omega_2),$$

(8″) $$w(\omega_1 \vee \omega_2) = w(\omega_1) + w(\omega_2).$$

The quantity $n(\omega)$ is called the *size of the sample* $\omega$, and the quantity $w(\omega)$ is called the *value of the sample* $\omega$.

We shall clarify these notions and postulates by means of examples. According to our conditions a quantity of coal, 1 $cm^3$ in volume, drawn in a suitable manner from a lot of 20 wagon-loads of coal, ground and compressed into briquettes, is a sample. If two briquettes are made into one, this may be considered as $\vee$-addition of two parts of this sample. Let us regard the volume of this briquette as the size $n$ of the sample. The dimension of the size $n$ is for instance $cm^3$. Let us regard the quantity of heat obtained when the sample is burnt, as the value of the sample. Then the dimension of the value $w$ is for instance calorie. We verify the fact that if we take the weight of a lot of coal as the the size $N$ of this lot, so that for instance ton is taken as the dimension of the size $N$, and if we take the money-equivalent of the lot of coal as the value $W$ of this lot, so that for instance GUINEA is taken as the dimension of the value $W$, then the size $n$ and the value $w$ of the sample satisfy all the postulates $5^0$-$8^0$.

It seem doubtful whether postulate $7^0$ must always be satisfied in this example; for we could regard as the size $n$ of the sample not the *volume* of the briquette, measured in $cm^3$, but its *weight* measured, for instance, in grams; then the quantities $N$ (measured in tons) and $n$ (measured in grams) would be dimensionally dependent, because a ton $= 10^6$ g, and postulate 7⁰ would not be fulfilled. Now, even in that case we shall regard the size $n$ of the sample as a quantity ***dimensionally independent*** from the size $N$ of the lot. In order to avoid misunderstandings we could denote the dimension $N$ of the lot size by the symbol Ton (capital T), and the dimension $n$ of the sample size by the symbol g (small g), so as to emphasize the dimensional *independence* of those quantities. The essential circumstance in the reasoning is that in such cases we shall never make use of the fact that 1 Ton$=10^6$ g. The form of the formulae that we shall deduce will not depend upon the circumstance whether a sample of coal is measured in cm³ or in g, and whether the lot is measured in tons or in kg; nor will it depend upon the number of grams in a ton of merchandise. Practically speaking we do not know exactly how many grams of coal there are in 20 wagon-loads and such calculations are never made. We shall not make such calculations as a matter of *principle* because we regard the size $N$ of the lot and size $n$ of the sample as dimensionally independent quantities, which is required by postulate $7^0$.

The dimension of the size $n$ of a sample will be denoted by the symbol "piece" (small letters). Thus

(14) $$[n] = \text{piece}.$$

The dimension of the value $w$ of a sample will be denoted by the symbol "guinea" (small letters). Thus

(15) $$[w] = \text{guinea}.$$

The terms "piece" and "guinea" are entirely conventional and one should not be misled by their everyday meaning.

*E. g.* if we take the length of the wire contained in a sample drawn from a lot of 1500 kg of wire, as the size of the sample, then it would be more convenient perhaps to use the term "cm" for the dimension of size. If we take the number of electric-light bulbs in a sample drawn from a lot of 10 000 bulbs, *e. g.* 70 bulbs, as the size of the sample, it is convenient to use the term "piece" (small letters) for the dimension of size. For the sake of generality of reasoning we choose the term "piece" regardless of particular interpretation in concrete cases.

The same holds for the term "guinea". *E. g.* in the case of bulbs the time of burning of a bulb (until it burns out) is expressed in hours. If we take that time as the value of a sample drawn from a lot of 10 000 bulbs (of equal strength), it would be more convenient perhaps to use the term "hour" for the dimension of the value of that sample. If we take value in the ordinary sense of the word, *i. e.* the money-equivalent, measured in guineas, as the value of a sample drawn, *e. g.* from 7000 m of cloth, then it is convenient to use the term "guinea" (small letters) for the dimension of value of that sample.



In most cases the assumption of the dimensional independence of sample value and lot value finds expression in the circumstance that generally in testing sample a different characteristic is directly determined from that which interests us in the whole lot. *E. g.* in the sampling inspection of coal the heat value of the sample is established in the laboratory and not its money worth, which is established by the commercial staff. In the sampling inspection of electric bulbs the laboratory measures how long the bulbs burn and not the price at which they will be sold.

But even in the cases where the investigated characteristic of the sample is the same as that which interests us in the whole lot (*e. g.* when we examine the acetylen content in a sample of calcium carbide in order to ascertain its content in a lot of calcium carbide) we shall consistently and scrupulously distinguish the dimension of lot value from the dimension of sample value. We do it as a matter of principle since it is required by postulate $7^0$. Thus in the case of calcium carbide the dimension of acetylene content in a sample must be given a different notation, *e. g.* gram (with a small g) from the dimension of acetylene content in the lot, *e. g.* Kilogram (with capital K), in order to point out that we shall deduce the formulae without making use of the fact that $1\,\text{Kg} = 10^3$ g.

Practically the most important assumption made hitherto is that the lot size $N$, the lot value $W$, the sample size $n$, and the sample value $w$, are dimensionally independent. We assume, besides, that the four quantities: $N$, $W$, $n$, $w$, form a system of units of the theory of sampling inspection in the sense defined in Chapter II.

Before introducing further notions and postulates let us make a remark of mathematical nature. Postulates $1^0$-$8^0$, by means of which we have introduced size and value of sample and of lot, are known in mathematics in the abstract theory of measure. Lot and sample may even be understood as something more general than set, they may be elements of a Boolean algebra. For didactical reasons, however, it is not expedient to introduce those generalizations, which are obvious for a mathematician, although — and this is noteworthy — they are not without importance for practice. *E. g.* it may be a question of some importance, whether the impurities found in the coal should be considered as belonging to the lot or not, if a lot of goods is regarded as a set. But it makes no difference if we regard a lot as an element of a Boolean algebra.

Using the notions introduced above and the postulates accepted let us define further notions necessary to the theory.

The ratio of the value $W$ of a part $\Omega_0$ of the lot (or of the whole lot) to the size $N$ of the same part $\Omega_0$ is called the *price $C$ of the lot* or *commercial price*,

$$\frac{W}{N} = C. \tag{16}$$

The ratio of the value $w$ of the part $\omega_0$ of a sample (or of the whole sample) to the size $n$ of the same part $\omega_0$ is called the *price $c$ of the sample* or *laboratory price*

$$\frac{w}{n} = c. \tag{17}$$

The definitions of the prices $C$ and $c$ could be generalized. Instead of the total price we could introduce "density" of value and assume, for instance, that

$$W = \int_{\Omega_0} C\,dN,$$

$$w = \int_{\omega_0} c\,dn,$$

where $C(N)$ and $c(n)$ are certain assumed dimensional functions dependent on $N$ or $n$, respectively. But we shall not dwell upon these generalizations, because they can easily be made if need be. The essential thing here is the *dimension* of both those prices.

According to definitions (16) and (17) we establish the dimensions of the prices $C$ and $c$, viz.,

$$[C] = \text{GUINEA} \cdot \text{PIECE}^{-1}, \qquad [c] = \text{guinea} \cdot \text{piece}^{-1}. \tag{18}$$

The notions of the price of a lot and that of a sample, thus defined, are much more general than is suggested by the terms used. We exemplify this as follows.

*E. g.* if we regard the weight of a lot as its value and its volume as its size, then its specific gravity will be its price. If we regard the money worth of a lot as its value, and its weight as its size, then its price in the everyday sense of the word will be its price. If we take the amount of heat produced when a sample of coal is burnt as the value of this sample, and the volume of a briquette as the size of the sample, then the amount of heat per unit of volume will be the laboratory price. If we regard the number of good eggs in a sample as its value, and the total number of eggs as its size, then the fraction of good eggs in the sample will be the price of the sample. In the last example the dimension of sample value might be called, for instance, "good egg" instead of "guinea", and the dimension of size might be called, for instance, "egg" instead of "piece", and then the dimension of sample price would be good egg$\cdot$egg$^{-1}$. But in the general theory we retain the accepted terms: PIECE, GUINEA, piece, guinea. In practical applications those dimensions may be termed in a different way if need be.



Let us formulate further assumptions. Namely we assume that there is a fixed method of calculating the commercial price $C$ when the laboratory price $c$ is known. For simplicity of reasoning we assume, for instance, that the commercial price $C$ is a linear function of the laboratory price $c$, *i. e.*

$$C = qc + C_0 \tag{19}$$

where $q$ and $C_0$ are constant quantities for a given lot. From definition (19) it will be seen at once that the quantity $C_0$ has the dimension of the commercial price. The quantity $C_0$ may be interpreted, for instance, as a constant investment cost. The dimension of the quantity $q$ is

$$[q] = \text{GUINEA} \cdot \text{PIECE}^{-1} \cdot \text{guinea}^{-1} \cdot \text{piece}. \tag{20}$$

The quantity $q$ defined by formula (19) we shall call *conversion coefficient.*

*E. g.* let the heat value of 1 cm$^3$ of coal, established in the laboratory, be 30 calories, so that the laboratory price $c$ of the sample of coal is 30 cal$\cdot$cm$^{-3}$; let the constant investment cost $C_0$ be

$$0{,}1\,\frac{\text{GUINEA}}{\text{ton}}$$

(*i. e.* 0,1 GUINEAS are added to the price of each ton of coal, *e. g.* for the amortisation of laboratory equipment), let the conversion coefficient $q$ be

$$0{,}05\ \text{GUINEA} \cdot \text{ton}^{-1} \cdot \text{cal}^{-1} \cdot \text{cm}^3.$$

Then the commercial price of the lot of coal is

$$C = 0{,}05\ \text{GUINEA} \cdot \text{ton}^{-1} \cdot \text{cal}^{-1} \cdot \text{cm}^3 \cdot 30\ \text{cal} \cdot \text{cm}^{-3} + 0{,}1\ \text{GUINEA} \cdot \text{ton}^{-1} = 1{,}6\ \text{GUINEA} \cdot \text{ton}^{-1}.$$

The assumption expressed by formula (19) is, of course, not the only one that is possible and corresponds to all cases encountered in practice. Other assumptions could also be made in this matter. *E. g.* if we consider the concentration of sulphuric acid as its laboratory price $c$, then it is assumed in practice that the commercial price $C$ of the acid is *not* a linear function of the concentration $c$. Instead of the relation (19) another and more general relation could be introduced, of the form

$$C = q_0 c^{m_0} + q_1 c^{m_1} + \ldots + C_0 \tag{19'}$$

where $q_0, q_1, \ldots$ and $C_0$ are certain constant dimensional quantities, which could be termed conversion coefficients, and $m_0, m_1, \ldots$ are certain real numbers. A formula of the type (19′) can always be adapted to any practical case.

However, we shall not develop the theory with assumption (19′); for the sake of concreteness we shall confine ourselves in the sequel to the relation (19), which corresponds to many cases encountered in practice.

Let us now introduce further notions. We are going to define the *mean value in the lot* and the *mean value in the sample.*

Let $\Omega_1$ be a part of the lot $\Omega$. If $N_1$ is the size of the part $\Omega_1$, $N$ — the size of the whole lot $\Omega$, whose value is $W$, and $C=W/N$ — the price of the lot, then the quantity

$$\overline{W} = W\frac{N_1}{N} = CN_1 \tag{21}$$

is termed the *mean value of the part* $\Omega_1$ in the lot. In virtue of this definition the dimension of the mean value $\overline{W}$ in the lot is

$$[\overline{W}] = [W] = \text{GUINEA}.$$

It will be seen from formula (21) that the parts whose size is equal have the same mean value in the lot.

We define analogously the mean value $\overline{w}$ of the part $\omega_1$ of size $n_1$ in the sample $\omega$ of size $n$ and value $w$, when the price of the sample is $c=w/n$:

$$\overline{w} = w\frac{n_1}{n} = cn_1. \tag{22}$$

Hence the dimension of the mean value $\overline{w}$ in the sample is

$$[\overline{w}] = [w] = \text{guinea}.$$



It will be seen from formula (22) that the parts whose size is equal have the same mean value in the sample.

Let us discuss one more notion, of fundamental importance in the theory of sampling inspection, viz., the quantity which is the measure of the dispersion of value in the lot or in the sample.

In mathematical statistics various measures of dispersion are used. If the random variable $\xi$ may assume only the numerical values $\xi_1, \xi_2, \ldots, \xi_\mu$ and their probabilities are equal, then the number

$$\bar{\xi} = \frac{1}{\mu}(\xi_1 + \xi_2 + \ldots + \xi_\mu)$$

is termed the mean value of the random variable $\xi$, and the following measures of dispersion, for instance, are assumed:

$$\text{standard deviation} = \sqrt{\frac{1}{\mu}[(\xi_1-\bar{\xi})^2 + (\xi_2-\bar{\xi})^2 + \ldots + (\xi_\mu-\bar{\xi})^2]},$$

$$\text{average deviation} = \frac{1}{\mu}[|\xi_1-\bar{\xi}| + |\xi_2-\bar{\xi}| + \ldots + |\xi_\mu-\bar{\xi}|], \tag{23}$$

$$\text{range} = \operatorname*{Max}_\nu \xi_\nu - \operatorname*{Min}_\nu \xi_\nu.$$

The quantities used in this theory are not numbers but dimensional quantities, so that the measure of dispersion of value in the lot or in the sample will also be a dimensional quantity. One naturally asks what the dimension of this measure of dispersion should be. It might seem that it is sufficient to regard, in definition (23), $\xi_\varkappa$ as the value $W_\varkappa$ of a PIECE in the lot, whose dimension we know to be GUINEA, and $\mu$ as the size $N$ of the lot, whose dimension is PIECE; we should then obtain the dimension of the dispersion of value for a PIECE in the lot. But it is easy to verify the fact that the dimension of standard deviation would then be GUINEA·PIECE$^{-1/2}$, that of average deviation would be GUINEA· PIECE$^{-1}$, and that of range would be GUINEA. On the other hand, if a different measure of dispersion were used from those quoted, as examples, in formulae (23), then that measure could also have another dimension. In that case the theory which we are constructing would depend upon convention and would have little scientific value.

Therefore we shall attempt to impose upon the measure of the dispersion of value in the lot or in the sample certain conditions, which will enable us to determine the dimension of dispersion in a unique manner. The measure of dispersion will be defined axiomatically by means of postulates. The axioms will be formulated so as to be abstractions of certain essential properties of the notions that are being defined. In this way we shall only explain why we have chosen such and such postulates instead of others, but, of course, this will not constitute a "proof" of those postulates, because in an axiomatic theory the postulates are not objects of proof.

Let the part $\Omega$ of a lot of goods be a sum of $\mathfrak{N}$ disjoint portions $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$, *i. e.*

$$\Omega = \Omega_1 \cup \Omega_2 \cup \ldots \cup \Omega_{\mathfrak{N}}.$$

To simplify our reasoning, but without any essential limitation of generality, we assume that all portions $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$ have an equal size $N_0$. In virtue of definition (21) all those portions have an equal mean value, viz.,

$$\overline{W} = W\frac{N_0}{N} = CN_0 \tag{24}$$

where $N$ is the size of the lot, $W$ the value of the lot, and $C$ the price of the lot. The size of the part $\Omega$ is, of course, $N_\Omega = \mathfrak{N} N_0$.

Let the value of the portion $\Omega_\gamma$ be $W_\gamma$ ($\gamma = 1, 2, \ldots, \mathfrak{N}$). Let

$$S = S_\Omega(W_1, W_2, \ldots, W_{\mathfrak{N}}, \overline{W}, N_\Omega, N_0)$$

be a dimensional function of the dimensional arguments: $W_1, W_2, \ldots, W_{\mathfrak{N}}$; $\overline{W}, N_\Omega$ and $N_0$, which depends on the set $\Omega$ and which satisfies certain postulates to be formulated below. We call this function the *measure of the dispersion* $S$ of value of the portions $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$ in the part $\Omega$ of a lot. In virtue of Theorem 4 (Chapter II) we find that

$$S = \Psi_\Omega(\mathfrak{W}_1, \mathfrak{W}_2, \ldots, \mathfrak{W}_{\mathfrak{N}}, \mathfrak{N})\overline{W}^a N_0^b, \tag{25}$$

where $\Psi_\Omega$ is a numerical function of the numerical arguments $\mathfrak{W}_\gamma = W_\gamma / \overline{W}$ ($\gamma = 1, 2, \ldots, \mathfrak{N}$) and $\mathfrak{N}$, $a$, $b$ are real numerical constants. The form of the function $\Psi_\Omega$ is dependent on the set $\Omega$.

Now we proceed to develop the postulates mentioned above. Let us observe that all measures of dispersion mentioned, for instance, in formulae (23) are homogeneous functions of the first degree with respect to the variables $\xi_\varkappa$, *i. e.* if any of these measures of dispersion is denoted by the symbol $F(\xi_1, \xi_2, \ldots, \xi_\mu)$, and $\lambda$ denotes an arbitrary number, then

$$F(\lambda\xi_1, \lambda\xi_2, \ldots, \lambda\xi_\mu) = \lambda F(\xi_1, \xi_2, \ldots, \xi_\mu).$$

We make a postulate of this property and *we assume that in formula* (25)

$$a = 1. \tag{26}$$



In order to determine the number $b$ in formula (25) we assume further postulates.

Let $\Omega^{\mathrm{I}}, \Omega^{\mathrm{II}}, \Omega^{\mathrm{III}}$ be parts of a lot of merchandise. Let the part $\Omega^i$ ($i$=I, II, III) be the sum of the same number, *e. g.* $\mathfrak{N}$, of disjoint portions $\Omega_1^i, \Omega_2^i, \ldots, \Omega_{\mathfrak{N}}^i$, each of which has the size $N_0^i$. Hence the size of the part $\Omega^i$ is

$$N^i = \mathfrak{N} N_0^i \qquad (i = \mathrm{I, II, III}).$$

In virtue of (24) the mean value of each of the portions is

$$\overline{W}_i = CN_0^i, \tag{27}$$

where $C$ is the price of the lot. Let the value of the portion $\Omega_\gamma^i$ be $W_\gamma^i$ ($i$=I, II, III; $\gamma = 1, 2, \ldots, \mathfrak{N}$). Let

$$S_i = \Psi_i(\mathfrak{W}_1^i, \mathfrak{W}_2^i, \ldots, \mathfrak{W}_{\mathfrak{N}}^i, \mathfrak{N})\overline{W}_i(N_0^i)^b \quad \text{and} \quad \mathfrak{W}_\gamma^i = \frac{W_\gamma^i}{\overline{W}_i^\gamma} \tag{28}$$

be the measure of dispersion of value of the portions $\Omega_1^i, \Omega_2^i, \ldots, \Omega_{\mathfrak{N}}^i$ in the part $\Omega^i$ of the lot.

Finally, let the parts $\Omega^{\mathrm{I}}, \Omega^{\mathrm{II}}$ be disjoint, and the part $\Omega^{\mathrm{III}}$ be their sum consisting of $\mathfrak{N}$ portions $\Omega_\gamma^{\mathrm{III}}$ formed by joining the respective portions $\Omega_\gamma^{\mathrm{I}}$ and $\Omega_\gamma^{\mathrm{II}}$ in pairs, *i. e.*

$$\Omega_\gamma^{\mathrm{III}} = \Omega_\gamma^{\mathrm{I}} \cup \Omega_\gamma^{\mathrm{II}} \qquad (\gamma = 1, 2, \ldots, \mathfrak{N}).$$

In virtue of postulate 4⁰, formulated at the beginning of this chapter, we have

$$N_0^{\mathrm{III}} = N_0^{\mathrm{I}} + N_0^{\mathrm{II}}. \tag{29}$$

We assume now that the disjoint parts $\Omega^{\mathrm{I}}$ and $\Omega^{\mathrm{II}}$ satisfy the condition

$$S_{\mathrm{I}}^2 + S_{\mathrm{II}}^2 = S_{\mathrm{III}}^2. \tag{30}$$

We shall call the disjoint parts $\Omega^{\mathrm{I}}$ and $\Omega^{\mathrm{II}}$, which satisfy the condition (30) *uncorrelated random parts* (*i.e.* uncorrelated in regard to the measure $W$).

Postulate (30) replaces in our theory the conditions of random independence of parts of the lot, required in a theory based on mathematical statistics. It is noteworthy that our postulate (30) is weaker than the condition of random independence.

In order to determine the exponent $b$ in formula (25) in a unique manner we formulate another postulate, viz. we postulate that if the disjoint parts $\Omega^{\mathrm{I}}$ and $\Omega^{\mathrm{II}}$ of a lot of merchandise are uncorrelated, then there is a finite limit, $\lim\limits_{\mathfrak{N}\to\infty} \Psi_\Omega$, equal for all parts $\Omega^{\mathrm{I}}, \Omega^{\mathrm{II}}, \Omega^{\mathrm{III}} = \Omega^{\mathrm{I}} \cup \Omega^{\mathrm{II}}$, *i. e.*

(31) $$\lim_{\mathfrak{N}\to\infty} \Psi_{\mathrm{I}} = \lim_{\mathfrak{N}\to\infty} \Psi_{\mathrm{II}} = \lim_{\mathfrak{N}\to\infty} \Psi_{\mathrm{III}} = \Psi.$$

It would be useful perhaps to state more exactly the meaning of $\lim\limits_{\mathfrak{N}\to\infty} \Psi_\Omega$; we shall not do this, because it is done in the same way as in the theory of probability.

Though postulate (31) replaces the limit theorem of the theory of probability, still it may also be accepted as an empirical law, according to the well-known saying of Lippman quoted by Poincaré[3]: "Everybody believes in the law of errors, the experimenters because they think it is a mathematical theorem, and the mathematicians, because they think it is an experimental fact". In order to construct consistently a phenomenological theory of sampling inspection we accept postulate (31) as an experimental fact, particularly since there are mathematicians who consider applied mathematics not without reason as an empirical science.

Postulates (30) and (31) enable us to determine the exponent $b$ in formula (25); for, in virtue of (27) and (28) we have

$$S_i = \Psi_i C (N_0^i)^{b+1} \qquad (i = \mathrm{I, II, III}),$$

and it follows from postulate (30) that

$$\Psi_{\mathrm{I}}^2 C^2 (N_0^{\mathrm{I}})^{2(b+1)} + \Psi_{\mathrm{II}}^2 C^2 (N_0^{\mathrm{II}})^{2(b+1)} = \Psi_{\mathrm{III}}^2 C^2 (N_0^{\mathrm{III}})^{2(b+1)}.$$

Passing on both sides of this equality to the limit for $\mathfrak{N}\to\infty$, we find from the equality (29) and postulate (31) that

(32) $$(N_0^{\mathrm{I}})^{2(b+1)} + (N_0^{\mathrm{II}})^{2(b+1)} = (N_0^{\mathrm{I}} + N_0^{\mathrm{II}})^{2(b+1)}$$

for all $N_0^{\mathrm{I}}$ and $N_0^{\mathrm{II}}$. Therefore we substitute in particular $N_0^{\mathrm{I}} = N_0^{\mathrm{II}}$. Then

$$2 (N_0^{\mathrm{I}})^{2(b+1)} = 2^{2(b+1)} (N_0^{\mathrm{I}})^{2(b+1)},$$

whence it follows that $2(b+1) = 1$, *i. e.*

(33) $$b = -\frac{1}{2}.$$

It is easily verified that for $b = -1/2$ equality (32) is true not only in the particular case under consideration, but in all cases.

[3]) H. Cramer, *Mathematical methods of statistics*, Princeton 1946, p. 232.



In virtue of equalities (26) and (33) formula (25) assumes the following form:

(34) $$S = \Psi_\Omega (\mathfrak{W}_1, \mathfrak{W}_2, \ldots, \mathfrak{W}_{\mathfrak{N}}, \mathfrak{N}) \overline{W} N_0^{-1/2}.$$

Therefore a *dimensional function whose dimension is*

GUINEA·PIECE$^{-1/2}$

*and which has the form* (34), where $\Psi_\Omega$ is an ordinary numerical function of the numerical arguments $\mathfrak{W}_1, \mathfrak{W}_2, \ldots, \mathfrak{W}_{\mathfrak{N}}, \mathfrak{N}$, with the finite limit $\lim\limits_{\mathfrak{N}\to\infty} \Psi_\Omega$, will be termed the *measure of the dispersion of value of the portions $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$ in the part $\Omega$ of a lot of merchandise.*

The most important for the sequel of our theory is the *dimension of the dispersion* $S$, which, for every function $\Psi_\Omega$, is always

(35) $$[S] = [\overline{W}] \cdot [N_0^{-1/2}] = \mathrm{GUINEA \cdot PIECE}^{-1/2}.$$

We define analogously the measure of the dispersion $s$ of value in the sample.

According to this definition the *dimension of the measure of the dispersion* $s$ is always

(36) $$[s] = [\overline{w}] \cdot [n_0^{-1/2}] = \mathrm{guinea \cdot piece}^{-1/2},$$

where $\overline{w}$ is the mean value of a portion in the sample, and $n_0$ the size of each of these portions.

If we give special forms to the function $\Psi_\Omega$, we shall obtain the measures of dispersion used in mathematical statistics. Viz., if

$$\Psi_\Omega = \Psi_\Omega (\mathfrak{W}_1, \mathfrak{W}_2, \ldots, \mathfrak{W}_{\mathfrak{N}}, \mathfrak{N})$$
$$= \sqrt{\frac{1}{\mathfrak{N}} [(\mathfrak{W}_1 - 1)^2 + (\mathfrak{W}_2 - 1)^2 + \ldots + (\mathfrak{W}_{\mathfrak{N}} - 1)^2]},$$

then the measure of dispersion of the form (34) will be called the *standard deviation* of value of the portions $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$ in the part $\Omega$ of the lot. In that case formula (34) may be written in the form

(37) $$S = \sqrt{\frac{1}{N_\Omega} [(W_1 - \overline{W})^2 + (W_2 - \overline{W})^2 + \ldots + (W_{\mathfrak{N}} - \overline{W})^2]},$$

where $N_\Omega$ is the size of the part $\Omega$, and $W_\gamma$ the value of the portion $\Omega_\gamma$ $(\gamma = 1, 2, \ldots, \mathfrak{N})$.

If

$$\Psi_\Omega=\Psi_\Omega(\mathfrak{W}_1,\mathfrak{W}_2,\ldots,\mathfrak{W}_{\mathfrak{N}},\mathfrak{N})=\frac{1}{\mathfrak{N}}[|\mathfrak{W}_1-1|+|\mathfrak{W}_2-1|+\ldots+|\mathfrak{W}_{\mathfrak{N}}-1|],$$

then the measure of dispersion will be called the *average deviation* of value of the portions $\Omega_1,\Omega_2,\ldots,\Omega_{\mathfrak{N}}$ in the part $\Omega$ of a lot of merchandise.

In that case formula (34) may be written in the form

(38) $$S=\frac{1}{\mathfrak{N}\sqrt{N_0}}[|W_1-\overline{W}|+|W_2-\overline{W}|+\ldots+|W_{\mathfrak{N}}-\overline{W}|],$$

where $N_0$ is the size of each of the portions $\Omega_1,\Omega_2,\ldots,\Omega_{\mathfrak{N}}$.

If

$$\Psi_\Omega=\Psi_\Omega(\mathfrak{W}_1,\mathfrak{W}_2,\ldots,\mathfrak{W}_{\mathfrak{N}},\mathfrak{N})=\operatorname*{Max}_{\nu}\mathfrak{W}_\nu-\operatorname*{Min}_{\nu}\mathfrak{W}_\nu$$

then the measure of dispersion will be termed the *range* of value of the portions $\Omega_1,\Omega_2,\ldots,\Omega_{\mathfrak{N}}$ in the part $\Omega$ of a lot of merchandise. In that case formula (34) may be written in the form

(39) $$S=(\operatorname*{Max}_{\nu}W_\nu-\operatorname*{Min}_{\nu}W_\nu)\frac{1}{\sqrt{N_0}}.$$

In practice only those three measures of dispersion are used in most cases.

If

$$N_0=\text{PIECE}$$

then the form of the formulae (37), (38), (39) is identical with that of formulae (23), used in mathematical statistics; for in formula (37) we then have

$$N_\Omega=\mathfrak{N}\ \text{PIECE}.$$

Standard deviation, average deviation, and range of value in a sample of merchandise are defined analogously.

Thus it makes no difference for the theory presented here which form of the function $\Psi_\Omega$, appearing in the definition of the measure of dispersion, is chosen. This is a conclusion drawn from the central limit theorem of the theory of probability, expressed in a different terminology.

In connection with the above considerations concerning dispersion let us make some methodological remarks. It might seem that since we quote the central limit theorem of the theory of probability, our theory is not a phenomenological theory, but a statistical one. Now, we must repeat that postulate (31) is accepted here as an experimental fact. It is not difficult to imagine a mental experiment, or even a practical experiment, that would confirm this fact. Every phenomenological theory must have not only definitions of certain notions but also laws obtained by abstraction from experience. Similarly, for instance, in order to construct such a phenomenological theory as classical mechanics it is not sufficient to introduce the notions of length, time, and mass, or even to define the notions of velocity, acceleration, and force. If we restricted mechanics to definitions it would be a barren nomenclature. It acquires scientific value by the fact that it postulates, as an experimental fact, Newton's law expressed by the formula: force = mass · acceleration. A similar role to that of Newton's law in mechanics is played in our theory by postulate (31), which we accept as an abstracted result of experiments. It will not be out of place to remark, should anyone ask about it, that we consider mathematical theorems, particularly those of the theory of probability, also as facts. Postulate (30) is only necessary to define the meaning of the statement that the parts of the lot are uncorrelated. We cannot expect a phenomenological theory of sampling inspection to work without the notion of such stochastic independence. Postulate (30) requires that no selection shall be made in drawing samples. Practicians know well how to achieve this, and our theory does not restrict this practice in any way; it is even less exigent for condition (30) is weaker than that of stochastic independence, which is fairly easy to secure in practice.



It is true that postulate (31) could be verified directly by experiment, but it would not be a reasonable way of confronting our theory with reality. It is much more advantageous to verify the consequences of the postulates that we have accepted, *e. g.* the formulae of sample size, which will be deduced in Chapter IV. Similarly, in mechanics we verify experimentally not the principles of Newton's mechanics but their consequences. In some phenomenological theories a direct verification of postulates is extremely difficult, *e. g.* a verification of Hooke's law in the theory of elasticity; in other theories it is downright impossible, *e. g.* in thermodynamics a verification of the first and the second law, formulated by means of a negation. In those theories it is only conclusions resulting from the accepted axioms that are verified experimentally. We shall deal with the experimental verification of the consequences of our theory in Chapter V.

The most essential remark is that postulates (30) and (31) have been necessary only to establish the *dimension of dispersion* of value in the lot or in the sample. Therefore instead of those postulates we could set up other ones, *e. g.* we could accept equalities (26) and (33) as postulates, and then it would not be necessary even to mention the limit theorems and statistics, because the postulates (30) and (31) would be inherent in formula (33) in a disguised form. That would be quite correct logically, but it would not be intuitive. Therefore, when formulating postulates (30) and (31) we preferred to say explicitly what intuitions underlie our assumptions. Those intuitions, borrowed from the theory of probability, are not a "proof" of any kind, for, as we have mentioned, in a deductive theory axioms are accepted without proof.

Finally, let us point out that in the sequel we shall no more mention the limit laws and mathematical statistics. We shall need no more intuitions from those theories. In statistical theories of sampling inspection, on the other hand, the limit laws are applied in the solving of almost every problem, and the theory of probability is an instrument of constant use.

In practical applications it is usually the standard deviation that is taken as the measure of dispersion. That is why we shall consider here a certain particular case of calculating the standard deviation of value in the lot or in the sample, which is important in applications.

Formula (37) may be written in a simpler form if we assume that the values $W_1, W_2, \ldots, W_{\mathfrak{N}}$ may be equal only to a definite value $W_0$ or to zero. Let $\varDelta$ denote in this case the number of all those parts among $\Omega_1, \Omega_2, \ldots, \Omega_{\mathfrak{N}}$ whose value is $W_0$. Formula (37) assumes the form

$$S=\sqrt{\frac{1}{N_\Omega}[\varDelta(W_0-\overline{W})^2+(\mathfrak{N}-\varDelta)\overline{W}^2]}\,. \tag{40}$$

In virtue of formulae (24) and (16) we have

$$\overline{W}=CN_0, \qquad C=\varDelta\frac{W_0}{N_\Omega}.$$

Besides

$$N_\Omega=\mathfrak{N}N_0.$$

Further, it follows that

$$\overline{W}=\frac{\varDelta}{\mathfrak{N}}W_0.$$

Hence, formula (40) may be written



$$S=\frac{W_0}{\sqrt{N_0}}\sqrt{\frac{\varDelta}{\mathfrak{N}}\left(1-\frac{\varDelta}{\mathfrak{N}}\right)^2+\left(1-\frac{\varDelta}{\mathfrak{N}}\right)\left(\frac{\varDelta}{\mathfrak{N}}\right)^2}\,.$$

Let $\Gamma=\varDelta/\mathfrak{N}$. Then

$$S=\sqrt{\Gamma(1-\Gamma)}\frac{W_0}{\sqrt{N_0}}.$$

*E. g.* let $\Omega$ be a lot of merchandise and $N_0=1$ PIECE. Let the number of "GOOD PIECES" be the measure of value of the lot and of its parts as well. Hence the values $W_1, W_2, \ldots, W_{\mathfrak{N}}$ may be equal only to the value $W_0=1$ GOOD PIECE or to zero. Let $\Gamma$ be the fraction of GOOD PIECES in the lot, so that the price of the lot, which will be called in this case the "goodness of the lot", is

$$C=\Gamma\frac{\text{GOOD PIECE}}{\text{PIECE}}.$$

Then the standard deviation of value of one PIECE in the lot $\Omega$ can be calculated from the formula

$$S=\sqrt{\Gamma(1-\Gamma)}\frac{\text{GOOD PIECE}}{\sqrt{\text{PIECE}}}. \tag{41}$$

This formula has a wide application in that kind of sampling inspection where the rule is either to accept or to reject. Sometimes instead of the value whose dimension is GOOD PIECE another value is used, whose dimension is BAD PIECE. The price of the lot is then called deficiency.

Analogously, if the measure of value both of the sample $\omega$ of merchandise and of its parts is the number of "good pieces" and $\gamma$ is the fraction of good pieces in that sample, *i. e.* if the price of the sample, which in that case will be called the "goodness of the sample" is

$$c=\gamma\frac{\text{good piece}}{\text{piece}},$$

then the standard deviation of value of one piece in the sample $\omega$ can be calculated from the formula

$$s=\sqrt{\gamma(1-\gamma)}\frac{\text{good piece}}{\sqrt{\text{piece}}}.$$

Finally let us formulate certain fundamental notions and postulates of economical nature.

We postulate first that sampling inspection involves expense. If the inspection costs nothing, it would be more reasonable to inspect the whole lot than to make sampling inspection. Therefore we introduce a fundamental notion of *inspection cost B*. The meaning of this term in our theory is broader than its everyday sense. Any economical loss occasioned by the process of inspection will be regarded as inspection cost $B$. Further, we assume that inspection cost is covered in the same currency as the lot of merchandise in question. The explanation of this assumption is taken from practice. Inspection often damages the merchandise, *i. e.* diminishes its value. Inspection sometimes involves loss of time, and in economical relations the principle "time is money" is usually accepted; this principle may give rise to doubts from the point of view of the dimensional analysis (time and money are dimensionally independent), but in practical life loss of time often causes loss of money. Inspection cost may also comprise other expenses, *e. g.* the payment of persons engaged in the inspection, amortisation of the inspection equipment, etc. All these expenses and the value of merchandise will be measured with the same unit. In other words, we assume that the *dimension of the inspection cost B is* GUINEA (capital letters); thus

$$[B] = \text{GUINEA}.$$

We shall say, for instance, that the value of the whole lot of merchandise is 5000 GUINEAS, and the inspection costs 20 GUINEAS, the unit GUINEA being the same as that used in paying for the merchandise.

The price $k$ of the inspection of a sample whose size is $n$ is

$$k = \frac{B}{n}. \tag{42}$$

Therefore, the dimension of the price $k$ of sample inspection is

$$[k] = \frac{[B]}{[n]} = \text{GUINEA} \cdot \text{piece}^{-1}. \tag{43}$$

In this chapter we have dealt with fundamental notions, postulates, and definitions of notions, necessary to formulate general principles of sampling inspection of merchandise. It is easy to see that here merchandise is only a concrete interpretation of certain experimental material and that the theory which we are setting forth may be applied without any essential modifications to every *method of representation*.



## IV. Sample size

One of the most important problems in the theory of sampling inspection is to establish what should be the sample size. The question formulated in this way is vague of course, so long as we have not stated the conditions which the size should satisfy. Those conditions will be dealt with in Chapter V. However, in this chapter we are going to show how our theory enables us, in various concrete cases, to obtain the form of formulae expressing sample size, provided we know on what quantities the size $n$ is to depend, regardless of other conditions which the size must satisfy.

We give again the symbols of the notions which we shall use and their dimensions.

TABLE I

| Number | Name of quantity | Symbol | Number of formula in text | Dimension $\text{PIECE}^{a_1} \cdot \text{piece}^{a_2} \cdot \text{GUINEA}^{a_3} \cdot \text{guinea}^{a_4}$ | | | |
|---|---|---|---|---|---|---|---|
| | | | | $a_1$ | $a_2$ | $a_3$ | $a_4$ |
| 1 | Size of the lot | $N$ | (12) | 1 | 0 | 0 | 0 |
| 2 | Size of the sample | $n$ | (14) | 0 | 1 | 0 | 0 |
| 3 | Value of the lot | $W$ | (13) | 0 | 0 | 1 | 0 |
| 4 | Value of the sample | $w$ | (15) | 0 | 0 | 0 | 1 |
| 5 | Price of the lot | $C$ | (18) | −1 | 0 | 1 | 0 |
| 6 | Price of the sample | $c$ | (18) | 0 | −1 | 0 | 1 |
| 7 | Conversion coefficient | $q$ | (20) | −1 | 1 | 1 | −1 |
| 8 | Dispersion of value of one PIECE in the lot | $S$ | (35) | −1/2 | 0 | 1 | 0 |
| 9 | Dispersion of value of one piece in the sample | $s$ | (36) | 0 | −1/2 | 0 | 1 |
| 10 | Price of the inspection of the sample | $k$ | (42) | 0 | −1 | 1 | 0 |

We are going to solve, by way of example, certain problems concerning the sample size $n$. In those problems we shall assume only that it is known on what quantities rising the lot or the sample the sample size $n$ depends. We assume, for example, that the quantities on which the sample size $n$ can depend are some of the quantities shown in Table I. Of course, we could also define other quantities, besides those mentioned in Table I, and make the sample size $n$ dependent on them, but we shall not do this here, because — as mentioned before — we only want to illustrate the method and not to make a detailed discussion of its assumptions and results. That should be the subject of separate studies.

The problems which we are going to formulate and solve will be divided into two groups, according to the choice of different quantities which are contained in Table I and on which the sample size will be made to depend.

To the first group will belong those problems in which the sample size $n$ depends on the dispersions $s$ or $S$, and — of course — on some other quantities from Table I perhaps.

To the second group belong those problems in which the sample size $n$ depends neither on the dispersion $s$ nor on $S$, but may depend on other quantities from Table I.

There are deeper reasons for this division. It follows from the considerations in Chapter III that in the problems of the first group we use postulates concerning the measure of dispersion, and in the problems of the second group we do not make use of those postulates. The problems of the first group are called *problems of statistical type*, and the problems of the second group are of *non-statistical type*.

In the practice of sampling inspection different formulae for the sample size $n$ are used. Some of those formulae are based on methods of mathematical statistics and the theory of probability; others are said to be taken from experience.

Our aim in this chapter is to deduce by the methods of our theory various formulae for the sample size $n$, among which there will also be all well-known formulae used in practice, both those which are said to be taken from experience and those which are based on methods of mathematical statistics; we shall also deduce other formulae.

We are going now to formulate and solve, by way of example, certain problems of the first group, *i.e.* those in which the dispersion $s$ or $S$ is among the quantities on which the sample size $n$ depends.

1.1. Let us assume that the sample size $n$ depends only on the size $N$ of the lot, the price of the inspection $k$, the dispersion $s$ of value of a piece in the sample, and the conversion coefficient $q$. Thus

$$n=\Phi(N,k,s,q) \tag{44}$$

where $\Phi$ is a dimensional function. We ask what is the form of the function $\Phi$.



Before solving this problem, we shall discuss the meaning of the assumptions made.

When we assume that the sample size $n$ depends on the size $N$ of the lot, we apparently do so in order to inspect a large lot in a different way from a small one. Perhaps it would be better to make the sample size $n$ dependent not only on the size $N$ of the lot, but also on the price $C$ of the lot (and thus — on the value of the lot $W=CN$). But we shall do so in the next example.

When we assume that the sample size $n$ depends on the price of the inspection $k$, we apparently do so because we anticipate the inspection cost $B=kn$, and do not want this cost to be too high and to destroy the economical effect of the inspection. We may have other reasons besides.

When we make the sample size $n$ depend on the measure of the dispersion $s$ of value of a piece in the sample, this is apparently done so, because from the results of the inspection of a *sample* we want to obtain information concerning the whole *lot*. This may be done as a precautionary measure or because we do not trust the producer or for other reasons.

Finally, the assumption that the conversion coefficient $q$ is an argument of the function $\Phi$ is important for the very technique of inspection. But we should not forget that assumption (19) in Chapter III was made as an example only.

Let us now determine the form of the function $\Phi$. We verify that the arguments $N,k,s,q$, of this function are dimensionally independent. Using Table I, we construct the determinant mentioned in Theorem 2, Chapter II. We obtain

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & -1/2 & 0 & 1 \\ -1 & 1 & 1 & -1 \end{vmatrix} = \frac{3}{2} \neq 0.$$

Hence, in virtue of Theorem 3, Chapter II, we have

$$n=\alpha\, N^{a_1}k^{a_2}s^{a_3}q^{a_4}, \tag{45}$$

where $\alpha$ is a constant numerical factor, $a_1, a_2, a_3, a_4$ are real numbers, which must be calculated. It follows at once that

$$[n]=[N]^{a_1}[K]^{a_2}[s]^{a_3}[q]^{a_4}.$$

Substituting for $[n],[N],[k],[s],[q]$ the corresponding dimensions taken from Table I and comparing the exponents of the units of the system PIECE, piece, GUINEA, guinea, we obtain in virtue of Theorem 1, Chapter II, the following equations:

$$a_1-a_4=0, \qquad -a_2-\frac{1}{2}a_3+a_4=1,$$

$$a_2+a_4=0, \qquad a_3-a_1=0,$$

whose solutions are

$$a_1=a_3=a_4=\frac{2}{3}, \qquad a_2=-\frac{2}{3}.$$

Substituting them in (45) we finally obtain

$$n=\alpha\left(\frac{Nqs}{k}\right)^{2/3} \tag{46}$$

where $\alpha$ is a constant numerical factor, independent of $N,q,s,k$; our theory does not enable us to state anything concerning this factor. Formula (46) in the case of $q=1\,\mathrm{GUINEA}\cdot\mathrm{PIECE}^{-1}\cdot\mathrm{guinea}^{-1}\cdot\mathrm{piece}$ has been deduced by the methods of mathematical statistics by H. Steinhaus[4]), who obtained

$$n=\left(\frac{Ns}{5k}\right)^{2/3}.$$

Thus the question how to determine the constant coefficient $\alpha$ remains unsettled. This problem, which concerns all formulae deduced by the methods of this theory, will be dealt with in Chapter V.

1.2. Let us discuss another problem, more general than 1.1. Let us assume that the sample size $n$ depends not only on the size $N$ of the lot, the price of the inspection $k$, the dispersion $s$ of value of a piece in the sample, and the conversion coefficient $q$, but also on the price $C$ of the lot. Thus

$$n=\Phi(N,k,s,q,C), \tag{47}$$

where $\Phi$ is a dimensional function, whose form must be determined.

[4]) Cf. H. Steinhaus, *Statistical appraisal,* Colloquium Mathematicum II. 3-4 (1951), p. 313, and *Wycena statystyczna jako metoda odbioru towarów produkcji masowej,* Studia i Prace Statystyczne 2 (1950), p. 3.



The assumption that the sample size $n$ depends also on the price $C$ of the lot is made, apparently, because we want to inspect more expensive lots in a different way from less expensive ones; this has been mentioned in problem 1.1. As far as we know such an assumption is not introduced in the theory of sampling inspection based on mathematical statistics, probably because the calculation difficulties involved in solving such a problem by the methods of mathematical statistics would be too great.

We proceed to determine the form of the function $\Phi$. We have verified by calculation that the quantities $N,k,s,q$, are dimensionally independent; therefore in virtue of Theorem 1, Chapter II, the price $C$ of the lot must be expressed by means of $N,k,s,q$, and using Table I we obtain by easy calculations

$$C=\beta_1\left(\frac{kq^2s^2}{N}\right)^{1/3},$$

where $\beta_1$ is a dimensionless quantity, namely

$$\beta_1=CN^{1/3}(kq^2s^2)^{-1/3}. \tag{48}$$

Hence, in virtue of Theorem 4, Chapter II,

$$n=\varphi(\beta_1)\left(\frac{Nqs}{k}\right)^{2/3}, \tag{49}$$

where $\varphi(\beta_1)$ is a numerical function of the numerical variable $\beta_1$; our theory does not enable us to say anything about this function; we can only say that $\varphi(\beta_1)>0$. However, even without knowing the form of the function $\varphi(\beta_1)$, we have found that it depends on a special product of variables written out in (48). The number

$$\beta_1=CN^{1/3}k^{-1/3}q^{-2/3}s^{-2/3}$$

is the *characteristic parameter* for this method of inspection. In other words, if in the inspection of various lots of various merchandise the quantities $C,N,k,q,s$, are different, but their parameters $\beta_1$ are equal, then the coefficient $\varphi(\beta_1)$ will be equal for all the lots concerned. Number $\beta_1$ is called the *parameter of similitude* of lots of merchandise for a given method of inspection, *i. e.* when it is settled on which quantities (*e. g.* $C,N,k,q,s$) the method of inspection depends.

However, we can make fairly general additional assumptions concerning the function $\varphi(\beta_1)$ and then we shall find its more precise form. *E. g.* we can assume that the function can be developed into a power series in the neighbourhood of the point $\beta_1=0$, and we can confine our considerations to a few terms of this series. *E. g.* if we confine our considerations to the first two terms, *i. e.* if we assume that

$$\varphi(\beta_1)=\alpha_0+\alpha_1\beta_1,$$

where $\alpha_0$ and $\alpha_1$ are certain dimensionless constants, then in virtue of (48) and (49) we shall obtain

$$n=\alpha_0\left(\frac{Nqs}{k}\right)^{2/3}+\alpha_1\frac{CN}{k}. \tag{50}$$

Formula (50) is a generalization of formula (46) and it is transformed into (46) for $\alpha_1=0$, $\alpha_0=\alpha$. The second right-hand term of formula (50) is noteworthy. Using definition (16) we have

$$W=CN,$$

so that the additional term $\alpha_1 W/k$ is proportional to the value $W$ of the lot, and inversely proportional to the price $k$ of the inspection, what agrees with intuition.

The assumption that the function $\varphi(\beta_1)$ can be developed into a power series is, of course, not the only one possible. The form of the function $\varphi(\beta_1)$ may be found from experience. This will be dealt with in Chapter V.

In the same way as in problems 1.1 and 1.2, we could also assume other quantities besides $N,k,q,s,C$, as arguments on which the sample size $n$ is to depend, *e. g.* the nominal price $C_1$ of the lot, which is generally different from the price $C$, established by means of sample inspection, or other quantities still. By the same method we should obtain formulae which would be further generalizations of formula (49). But instead of doing this we shall proceed to another kind of problems of the first group.

1.3. Let us assume that the sample size $n$ depends on the size $N$ of the lot, the price $k$ of the inspection, and the dispersion $S$ of value of a PIECE in the lot.

This assumption differs from the assumptions of problem 1.1 first of all in the following point: we take into account not the dispersion $s$ of value in the sample but the dispersion $S$ of value in the whole lot. This assumption is economically less cautious than the assumption of problem 1.1, because by means of an inspection of a sample we can measure the dispersion $s$ in the sample but not the dispersion $S$ in the lot. Some authors, however, assume that the dispersion $S$ of value of a PIECE in the lot is known from other data, *e. g.* from the inspection of previous supplies of merchandise received from the same producer. This assumption is made for instance in the acceptance method used in the United States.



Another difference between the assumptions of problem 1.3 and those of problem 1.1 lies in the fact that in 1.3 we do not assume any dependence of $n$ on the conversion coefficient $q$.

We could explain this intuitively reasoning that, as a matter of fact, in those assumptions we do not consider an inspection of individual pieces in the sample, since we do not investigate the dispersion $s$ of value of these pieces.

Such an explanation, however, might seem unsatisfactory; in order to avoid doubtful points let us introduce the coefficient $q$ as an argument on which, *a priori*, the sample size $n$ may also depend, and let the calculation mechanism itself settle the doubts.

Thus, we assume that

$$n=\Phi(N,S,k,q). \tag{51}$$

In order to determine the form of the function $\Phi$ we state that the quantities $N,S,k,q$ are dimensionally independent; for, using Table I, we find the determinant

$$\begin{vmatrix} 1 & 0 & 0 & 0 \\ -1/2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 1 & 1 & -1 \end{vmatrix} = -1\neq 0.$$

Thus, analogously to the solution of problem 1.1, we have

$$[n]=[N]^{b_1}[S]^{b_2}[k]^{b_3}[q]^{b_4},$$

where the real numbers $b_1,b_2,b_3,b_4$ will be found by substituting for $[n],[N],[S],[k],[q]$ the corresponding dimensions taken from Table I, and by comparing the exponents of the units of the system PIECE, piece, GUINEA, guinea. We obtain the following system of equations:

$$b_1-\frac{1}{2}b_2-b_4=0, \qquad -b_3+b_4=1,$$
$$b_2+b_3+b_4=0, \qquad b_4=0,$$

whose solution is the following:

$$b_1=\frac{1}{2}, \qquad b_2=1, \qquad b_3=-1, \qquad b_4=0.$$

Hence

(52) $$n=\delta\frac{SN^{1/2}}{k},$$

where $\delta$ is a constant numerical factor. We have found that the sample size $n$ really does not depend on the conversion coefficient $q$.

Formula (52) is used in practice, and — as far as we know — the authors do not explain it by the methods of mathematical statistics.

Let us observe that the sample size $n$, calculated from formula (46) is *caeteris paribus* of a higher order with respect to $N$ than the sample size $n$, calculated by means of formula (52), and the ratio of these two sizes is of the order $N^{1/6}$. Therefore, if we do not know the dispersion $S$ of value of a PIECE in the lot and know only the dispersion $s$ of value of a piece in the sample, then — generally speaking — we must draw samples of greater size. This agrees with intuition.

1.4. Formula (52) can be generalized if we assume — similarly to example 1.2 — that $n$ is dependent not only on $N,s,k$, but also on the price $C$ of the lot, *i. e.*

(53) $$n=\Phi(N,S,k,C).$$

The quantities $N,S,k,C$, are dimensionally dependent, because the respective determinant

$$\begin{vmatrix} -1 & 0 & 0 & 0 \\ -1/2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ -1 & 0 & 1 & 0 \end{vmatrix}=0.$$

Therefore we express $C$ by $N,S,k$, and find

(54) $$CN^{1/2}S^{-1}=\beta,$$

where $\beta$ is a dimensionless quantity. Thus the calculation made in problem 1.3 gives a generalization of formula (52):

(55) $$n=\psi(\beta)\frac{SN^{1/2}}{k},$$

where $\psi(\beta)$ is a numerical function of the characteristic parameter $\beta$. If we assume that the function $\psi(\beta)$ can be developed into a power series and if we then confine our considerations to the first two terms of that series, we shall obtain



(56) $$n=\delta_0\frac{SN^{1/2}}{k}+\delta_1\frac{CN}{k}$$

where $\delta_0$ and $\delta_1$ are numerical constants.

In practice it is sometimes required that the formula of sample size $n$ shall be such that 1° if the size $N$ of the lot increases infinitely, then $n$ is finite, and 2° if the value $W=CN$ of the lot is equal to zero, then $n$ is equal to zero, which is natural, because there is no sense in inspecting worthless merchandise. From formulae (54) and (55) it follows that these conditions can be written in the following form:

$$\lim_{N\to\infty}\beta\psi(\beta)=\text{const}\neq 0,$$

$$\psi(0)=0.$$

One of the simplest functions satisfying the first condition is the function

$$\beta\psi(\beta)=\frac{\delta_4\beta^2+\delta_5\beta+\delta_6}{\delta_1\beta^2+\delta_2\beta+\delta_3}$$

where $\delta_1,\delta_2,\ldots,\delta_6$ are certain numerical constants. From the second condition we find that there must be

$$\delta_5=\delta_6=0$$

and we can always assume $\delta_4=1$. Hence we choose

$$\psi(\beta)=\frac{\beta}{\delta_1\beta^2+\delta_2\beta+\delta_3}.$$

Substituting for $\beta$ the value from formula (54) we obtain, in virtue of (55),

$$n=\frac{CNk^{-1}}{\delta_1C^2NS^{-2}+\delta_2CN^{1/2}S^{-1}+\delta_3},$$

which may also be written in the form

$$\frac{1}{n}=\delta_1\frac{kC}{S^2}+\delta_2\frac{k}{SN^{1/2}}+\delta_3\frac{k}{CN},$$

where $\delta_1,\delta_2,\delta_3$ are numerical constants. If we assume $\delta_1=\delta_3=0$, we obtain formula (52); and if we assume $\delta_2=0$, we obtain formula

$$\frac{1}{n}=\delta_3\frac{k}{CN}+\delta_1\frac{kC}{S^2}.$$

J. Oderfeld has printed out that this formula is used in practice in a slightly less precise form.

Finally, if we substitute $\delta_2=\delta_3=0$, we obtain formula

$$n=\delta_0\frac{S^2}{kC},$$

where $\delta_0=1/\delta_1$ is a numerical constant. This formula is noteworthy on account of the circumstance that it does not contain the size $N$ of the lot at all.

In assumptions (51) or (53) we could also introduce other quantities as arguments, according to the point of view adopted, and obtain various formulae by the method shown above. We shall not do this, as it is only a matter of simple calculations.

1.5. As an example, we shall now deduce a formula, using greatly simplified assumptions of the statistical type. Let us assume that

(57) $$n=\Phi(s,c),$$

*i. e.* that the sample size $n$ depends on the dispersion $s$ of value of a piece in the sample and on the price of the sample; and that this size $n$ does not depend on the size $N$ of the lot, its price, or the price $k$ of the inspection.

Then

(58) $$n=\varepsilon\left(\frac{s}{c}\right)^2,$$

where $\varepsilon$ is a certain constant numerical factor.

Sometimes formula (58) is also proved by the methods of mathematical statistics.

1.6. We are going to consider certain particular cases of the formulae deduced hitherto. If the "goodness" or the "deficiency" of the sample, defined in Chapter III, is taken as its price, then in all the examples so far considered we should substitute

$$s=\sqrt{\gamma(1-\gamma)}\,\frac{\text{good (bad) piece}}{\sqrt{\text{piece}}},$$

where $\gamma$ is the fraction of good pieces or bad pieces, respectively, in the sample. Analogously, if the "goodness" or the "deficiency" of the lot is taken as the price in the lot, we should substitute

$$S=\sqrt{\Gamma(1-\Gamma)}\,\frac{\text{GOOD (BAD) PIECE}}{\sqrt{\text{PIECE}}},$$



where $\Gamma$ is the fraction of GOOD PIECES or BAD PIECES, respectively, in the lot, according to the reasoning in Chapter III.

Those cases occur in practice when the principle adopted in the sampling inspection is *either to accept* the lot *or to reject* it.

We proceed now to consider one problem of the second group, *i. e.* a problem in which the measure of dispersion $s$ or $S$ is not among the quantities on which the sample size $n$ depends.

2.1. Let us assume that the sample size $n$ depends on the size $N$ of the lot, the price $C$ of the lot, and the price $k$ of the inspection, *i. e.*

(59) $$n=\Phi(N,C,k).$$

These assumptions can be interpreted in the same way as the assumptions in the preceding problems chosen as examples. Besides, the conversion coefficient $q$ could be introduced as an argument of the function $\Phi$, but calculation, as in example 1.3, would show that with these assumptions $n$ does not depend on $q$. The formula:

(60) $$n=\alpha_1\frac{CN}{k},$$

where $\alpha_1$ is a constant numerical factor, follows from the assumption (59) in a way illustrated in the preceding examples. Formula (60) may also be regarded as a particular case of formulae (50) and (56). Formula (60) is used in practice.

Finally, let us make some methodological remarks. It was not been our aim to make the assumptions formulated in the problems which we have solved cover all possible cases. Neither have we given a detailed interpretation of those assumptions. It is even possible that some of the interpretations outlined above do not agree with reality. That is because we have not aimed at developing the theory of sampling inspection and discussing its various problems and their solutions; we have only intended to illustrate the method of solving those problems by means of concrete examples.

It can be seen from those examples that this method is very simple, as it requires very easy calculations. We think that in this respect it is superior to probabilistic methods, which sometimes require considerable mathematical apparatus. Owing to this simplicity it is possible to put forth and tackle problems which would be very difficult to dealt with by means of the theory of probability and mathematical statistics. Besides, in our method the set of assumptions accepted when a problem is formulated stands out very clearly. Practice or economical requirements may very often suggest assumptions or interpret them; in a method based on the theory of probability the complicated mathematical apparatus sometimes obscures the view of the whole.

Our method has also disadvantages, the most important of which is the appearance of undetermined numerical coefficients or numerical functions of numerical arguments in the formulae obtained. Those numbers cannot be given by our theory; they can be and should be taken from experience. This problem is dealt with in Chapter V. However, in the cases where formulae deduced in our theory contain undetermined numerical functions, our method enables us to determine characteristic parameters of the problem. An example of such a parameter and its role has been discussed, for instance in problem 1.2.

### V. Aim of inspection and accordance with experience

All formulae that we have hitherto deduced in the theory of sampling inspection of merchandise contain numerical coefficients or numerical functions dependent on numerical arguments. Those coefficients and functions cannot be calculated by means of the dimensional analysis. But in order to make the formulae obtained practically useful it is necessary to determine those coefficients or functions. It is the aim of this chapter to present an idea which can and should be realized in order to determine the numerical value of those coefficients or functions.

In order to clarify this idea let us observe a circumstance which seems rather surprising at first, viz. that we have deduced all formulae for sample size assuming only on which quantities the size should depend; we have made no more assumptions stating what we require from the size in question. Therefore we can and should impose on all those formulae additional conditions, enabling us to determine numerical coefficients.

We shall give a general outline of those conditions. First of all they should state more precisely the *aim* of sampling inspection; then they must lead to formulae agreeing with reality. This very vague statement requires, of course, detailed explanations.

We shall first consider the notion of the aim of sampling inspection. Let us take an example. When we receive a lot of coal we make a sampling inspection. Why is it done? We may do so, for instance, in order to know, more or less accurately, how much the consumer is to pay for the lot to the producer, so that neither the former nor the latter should incur too heavy a loss. The acceptance of merchandise inspected by sampling



is therefore a game for a certain stake. This stake is always a quantity measured with the same units as the value of the lot.

From these observations, made by way of example, we draw by abstraction the following common characteristic, which we shall regard as a postulate:

If $N,n,W,w,C,c,q,k,S,s$ are quantities from Table I, then in every sampling inspection there is a defined function $R$, dependent on at least some of the enumerated arguments

$$R=\Phi(N,n,W,w,C,c,q,k,S,s), \tag{61}$$

whose dimension is the same as that of the value $W$ of the lot, *i. e.*

$$[R]=\text{GUINEA}. \tag{62}$$

The function $R$ will be called the *risk of inspection.*

Dimensional analysis, however, enables us to deduce certain characteristics of the risk $R$. Let us first observe that in (61) we can omit $W$ and $w$ as arguments without loss of generality, because by definitions (16) and (17)

$$W=CN \qquad \text{and} \qquad w=cn.$$

Let us now choose among the remaining arguments four dimensionally independent quantities, *e. g.* $N,n,S,s$, and let us regard them as a system of units. In virtue of the theorems of Chapter II, which we have repeatedly applied, we can express the remaining quantities by means of $N,n,S,s$. In this way we obtain the following characteristic numerical parameters of the risk $R$:

$$\begin{aligned}\xi_1&=CN^{1/2}S^{-1},\\ \xi_2&=cn^{1/2}s^{-1},\\ \xi_3&=qN^{1/2}sn^{-1/2}S^{-1},\\ \xi_4&=knN^{-1/2}S^{-1}.\end{aligned} \tag{63}$$

Moreover

$$[R]=[N^{1/2}S].$$

Hence, in virtue of Theorem 4, Chapter II,

$$R=\varrho(\xi_1,\xi_2,\xi_3,\xi_4)N^{1/2}S, \tag{64}$$

where $\varrho$ is a numerical function of numerical arguments $\xi_1,\xi_2,\xi_3,\xi_4$.

The numerical function $\varrho$ should be given by the economists. However, let us make, by way of example, the fairly general assumption that the function $\varrho$ can be developed into a power series of the variables $\xi_1, \xi_2, \xi_3$, and let us confine our considerations to the terms of first degree, *i. e.*

$$\varrho = \varrho_0 + \varrho_1\xi_1 + \varrho_2\xi_2 + \varrho_3\xi_3 + \varrho_4\xi_4, \tag{65}$$

where $\varrho_0, \ldots, \varrho_4$ are certain numerical constants, whose numerical values should be established in each case by the economists.

Let us substitute the values (63) in (65) and the result in (64). We obtain

$$R = \varrho_0 SN^{1/2} + \varrho_1 CN + \varrho_2 cn^{1/2}s^{-1}N^{1/2}S + \varrho_3 Nqsn^{-1/2} + \varrho_4 kn. \tag{66}$$

Let us illustrate this assumption by examples.

1. Let us consider a method of sampling inspection in which only the following quantities are considered: the sample size $n$, the size $N$ of the lot, the dispersion $S$ of value of a PIECE in the lot, and the price $k$ of the inspection. Let us assume that the risk $R$ also depends on the quantities $s, N, S, k$, only. Therefore in formula (66) we substitute

$$\varrho_1 = \varrho_2 = \varrho_3 = 0$$

and obtain

$$R = \varrho_0 SN^{1/2} + \varrho_4 kn.$$

We have seen in problem 1.4, Chapter IV, that if the sample size $n$ depends on $N, S$, and $k$, only, then

$$SN^{1/2} = \delta kn,$$

where $\delta$ is a constant numerical factor, *i.e.* in the method of sampling inspection described in problem 1.4, Chapter IV, the risk $R$ is proportional to the inspection cost $kn$.

2. Let us now consider a method of sampling inspection described in problem 1.1, Chapter IV, *i. e.* a method in which the quantities $n, N, S, k, q$ alone are considered. From formula (66) we obtain

$$R = \varrho_3 Nqsn^{-1/2} + \varrho_4 kn. \tag{67}$$

Substituting formula (46), Chapter IV,

$$n = \alpha\left(\frac{Nqs}{k}\right)^{2/3}$$

we observe that also in this case the risk $R$ is proportional to the inspection cost $kn$, *i. e.*

$$R = \gamma kn,$$

where $\gamma$ is a constant numerical factor.



In the same way we could also examine other methods of inspection considered in Chapter IV. But we shall not do this, because these examples make it obvious that the notion of risk $R$ is noteworthy.

When the risk $R$ is established it is necessary to define the condition which that risk is to satisfy in the given method of sampling inspection. It may be, for instance, the condition that the risk $R$ should be constant or that it should not exceed a certain value fixed in advance, or that it should be reduced to minimum, etc. The condition imposed upon the risk $R$ is called the *economical condition* $\mathfrak{E}$; it can and should be defined by the producer and the consumer or by economists in general.

In the light of the preceding considerations we can determine the aim of inspection more precisely. The *aim of sampling inspection* is to determine the value of the lot by inspecting a sample so that the risk of inspection should satisfy the economical condition $\mathfrak{E}$.

We shall now explain what we mean by saying that the formulae obtained in the theory of sampling inspection should agree with reality. Since one of the most important problems of sampling inspection is to determine the sample size $n$, we shall confine ourselves to taking as examples the formulae deduced in Chapter IV.

It has been mentioned in Chapter I and III that in our opinion the theory of sampling inspection, as a science of objective and determined phenomena, can and should be based on experience. This opinion is certainly shared by many practicians who are able to determine correctly at first sight the sample size in accordance with the aim of inspection. They say that they can do that owing to long years of experience. But we should clearly understand what is meant by experience and formulate rules of procedure, which would enable us to ascertain objectively the agreement with reality of a given formula of sample size, without referring to the authority of practicians and their vaguely defined knack of "first sight" decisions.

In order to make our point clearer we shall confine ourselves to an example. Let us suppose that we are making a sampling inspection of a lot of coal, using the method described in problem 1.1, Chapter IV, and in example 2 of the present Chapter. Let us consider the following problem.

We are given a lot of coal, whose size $N$ is known. We draw from this lot a sample of size $n$ and we measure the dispersion $s$ of value of one piece in the sample. The conversion coefficient $q$ and the price $k$ of inspection are known. We also know the inspection risk $R$, *i. e.* we not only know that the risk $R$ is expressed by formula (67), but we also know the numerical values of the coefficients $\varrho_3$ and $\varrho_4$, established by a statement of the producer and the consumer or of economists in general. The economical condition $\mathfrak{E}$ is also established, *e. g.* the condition that the risk $R$ shall be reduced to minimum.

We ask what the sample size $n$ should be in order to satisfy all the requirements given above.

In virtue of formula (46), Chapter IV, we know that regardless of the definition of the risk $R$ and regardless of the economical condition $\mathfrak{E}$, the sample size $n$ must be expressed by the formula

$$n = a\left(\frac{Nqs}{k}\right)^{2/3},$$

where $a$ is a constant numerical factor. The problem therefore consists in finding the numerical value of the coefficient $a$.

This problem could be solved in the following way. Since the risk $R$ is expressed by formula (67), in which we know the coefficients $\varrho_3$ and $\varrho_4$, and since the economical condition requires that $R$ shall be reduced to minimum, we calculate the minimum of the function $R(n)$. Hence, we must have

$$\frac{dR}{dn} = 0. \tag{68}$$

An essential difficulty of this solution is that the dispersion $s$ is also a function of $n$. If we assumed that $n$ would be sufficiently large and that then we could assume $\frac{ds}{dn} = 0$, then the condition (68) would give

$$-\frac{1}{2}\varrho_3 Nqn^{-3/2} + \varrho_4 k = 0,$$

whence

$$n = \frac{\varrho_3}{2\varrho_4}\left(\frac{Nqs}{k}\right)^{2/3}$$

and the coefficient $a$ in the formula would be determined, because the numbers $\varrho_3$ and $\varrho_4$ are known. By a similar reasoning H. Steinhaus has established[5]) formula analogous to (46).

However, certain doubtful points still remain unsettled. Firstly, are we entitled to assume that $s$ does not depend upon $n$? Secondly, can we be sure that the obtained formula agrees with reality? Thirdly, what

[5]) l. c.



does it mean that this formula agrees with reality? K. Wiśniewski[6]) has made experiments towards the solution of these problems. It appears, however, that the sample size $n$ in those experiments did not exceed a few pieces. In view of this can we regard as justified the assumption that $s$ does not depend on $n$?

We believe that the coefficient $a$ can and should be determined by experiment. The reasoning underlying that experiment is the following. From a lot of coal of a given size $N$ we draw as many samples of various sizes $n$ as possible. For each sample we measure the dispersion $s$ of value of a piece in the sample. The values obtained from the measurements are entered on a diagram in the system of coordinates, whose axis of abscissae is $n$, and the axis of ordinates is

$$R = \varrho_3 Nqsn^{-1/2} + \varrho_4 kn. \tag{69}$$

We thus get a "graph" determined by the points representing the values entered (Fig. 1).

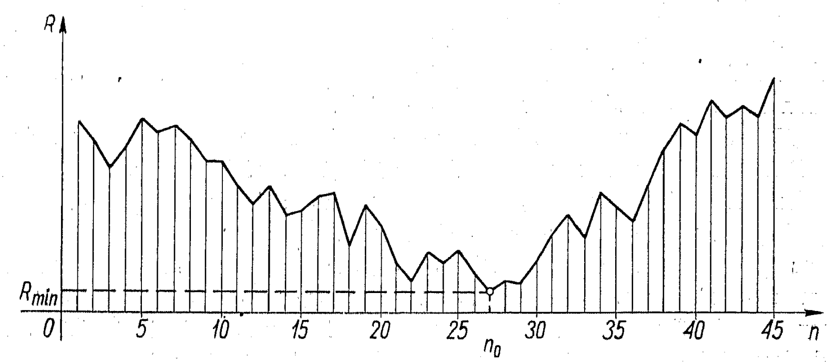


Fig. 1.

If this diagram proves to have a distinct minimum for $n = n_0$, then in formula (46) we substitute for $n$ the quantity $n_0$ and we compute the coefficient $a$.

We cannot expect to determine the number $a$ by means of a single measurement. Such measurements must be repeated for a lot of coal of a different size $N$ and the coefficient $a$ for that lot must be established in the same way. We should make as many such measurements with lots of coal of different sizes $N$ as possible. If we get the same number $a$ each time, regardless of the size $N$, the conversion coefficient $q$, the

[6]) K. Wiśniewski, *Metody statystycznej kontroli jakości w świetle doświadczeń,* Wiadomości PKN 10/51, Warszawa 1951.

measure of the dispersion $s$, and the price $k$ of the inspection, we shall be able to say that formula (46) with the constant $\alpha$, thus established and *universal* for all lots of coal, agrees with reality.

Simultaneously we shall obtain an experimental verification of all assumptions that had led to formula (46). Among them are the assumptions which we have made in Chapter III, when establishing the dimension of the dispersion $s$, the assumptions concerning the coefficients $\varrho_3$ and $\varrho_4$, necessary to determine risk, and finally the assumption that the sample size depends on $N, k, s, q$ alone.

If we find that the number $\alpha$ is not the same for different $N, q, s, k$, then the only conclusion that can and should be drawn is that formula (46) does not agree with reality; therefore some of the assumptions on which it is based are false. Still, the method of dimensional analysis enables us to obtain other formulae by means of easy calculation.

Apparently it is not difficult to make such experiments in practice. And even if it required more time, this would be amply repaid by the result: if formula (46) were confirmed by experiments, the numerical constant $\alpha$, established in this way, would best agree with reality. As a matter of fact it is on this that the experience of the practicians is based, although they are not always aware of it.

We cannot expect that the results of the measurements made to ascertain whether the number $\alpha$ is independent of $N, q, s, k$, will present an "orderly" line and give one exact value $\alpha$ for all $N, q, s, k$. Every experimenting scientist, even dealing with phenomena as well determined as physical phenomena, knows that the results of a number of measurements of one and the same quantity are not really invariable, but show certain fluctuations. The methods of mathematical statistics can and should be used in treatment the results of experiments. Thus mathematical statistics in the phenomeonlogical theory of sampling inspection which we have presented has the same role to play as in any other empirical science, viz. it is used in treatment the results of experiments. Thus the role of mathematical statistics in this theory has been shifted but not eliminated altogether. It would be preposterous to attempt this.

Finally, let us point out again that the main intention of this paper is not the discussion of various problems of sampling inspection but the presentation of a method. That is why certain problems have not been treated, perhaps, with sufficient exhaustiveness or generality.

---

## *АНАЛИЗ РАЗМЕРНОСТИ В ВЫБОРОЧНОМ ИСПЫТАНИИ ТОВАРОВ*

С. Дробот и М. Вармус (Вроцлав).

I. Методологические замечания. Выборочное испытание товаров состоит в том, чтобы на основании испытания выборки из партии товара определить некоторую характеристику всей партии. Теория выборочного испытания обычно основана на теории вероятностей и математической статистике. Хотя методы теории вероятностей дают ответы на многие практические вопросы выборочного испытания, но страдают некоторыми недостатками: часто они громоздкие, трудоёмкие, заставляют схематизировать явление и не всегда дают результаты вполне согласные с действительностью. Впрочем роль опыта в этих методах недостаточно ясна.

Так как вероятностные суждения выражают собой некоторые объективные свойства изучаемых явлений, то можно попытаться сформулировать феноменологическую теорию выборочного испытания. Аналогично, например, феноменологическая и статистическая термодинамика описывают один и тот-же класс явлений. Статистическая теория описывает явления более тонко, но в столь грубых вопросах, какие встречаются в выборочном испытании такая тонкость кажется излишней, тем более, что она требует сложных математических средств.

Предметом настоящей работы и есть *опыт феноменологической теории выборочного испытания.*

II. Об анализе размерности. Математическим инструментом этой теории является *анализ размерности*. Сформулируем понятия и теоремы анализа размерности, которые используются в дальнейшем.

Размерные величины $A, B, C, \ldots$ считаются элементами некоторого линейного пространства, в котором определено коммутативное и ассоциативное произведение $AB$ величин $A$ и $B$, а также потенцирование $A^a$ с действительным показателем $a$, имеюще следующие свойства:

$$A^{a+b} = A^a A^b,$$

$$(AB)^a = A^a B^a,$$

$$(A^a)^b = A^{ab},$$

$$A^1 = A.$$

Положительные числа (безразмерные величины) $\alpha, \beta, \gamma, \ldots$ тоже считаются элементами этого пространства. Величины $A_1, \ldots, A_m$ называются *размерно независимыми*, если из равенства

$$A_1^{a_1} \ldots A_m^{a_m} = \alpha$$

следует $a_1 = \ldots = a_m = 0$ (и $a=1$). При этом предполагается, что существует точно $n$ размерно независимых величин. Всякая система $X_1, \ldots, X_n$ размерно независимых величин называется *системой единиц*. Каждая величина $A$ выражается в системе единиц $X_1, \ldots, X_n$ единственным образом в виде

$$A = aX_1^{a_1} \ldots X_n^{a_n},$$

где $a$ — безразмерная величина и $a_1, \ldots, a_n$ — вещественные числа. Величина

$$[A] = [X_1^{a_1} \ldots X_n^{a_n}],$$

называется *размерностью* $A$ в системе $X_1, \ldots, X_n$.

Рассматриваются функции $\Phi(Z_1, \ldots, Z_r)$, определённые для размерных аргументов $Z_1, \ldots, Z_r$ и принимающие размерные значения. На эти функции налагаются требования, чтобы вид этих функций не зависел от выбора системы единиц и чтобы их размерность сохранялась вместе с размерностью аргументов. Тогда справедлива следующая

Теорема П. *Если $P_1, \ldots, P_q$ размерно зависимы от размерно независимых величин $A_1, \ldots, A_m$, т. е. если существуют такие безразмерные $\pi_1, \ldots, \pi_q$, что*

$$P_1 = \pi_1 A_1^{p_{11}} \ldots A_m^{p_{1m}},$$
$$\cdots\cdots\cdots$$
$$P_q = \pi_q A_m^{p_{q1}} \ldots A_m^{p_{qm}},$$

*где $p_{kl}$ $(k = 1, \ldots, q; l = 1, \ldots, m)$ — действительные числа, то*

$$\Phi(A_1, \ldots, A_m; P_1, \ldots, P_q) = \varphi(\pi_1, \ldots, \pi_q) A_1^{f_1} \ldots A_m^{f_m},$$

*где $\varphi(\pi_1, \ldots, \pi_q)$ — безразмерная (числовая) функция безразмерных (числовых) переменных $\pi_1, \ldots, \pi_q$, а $f_1, \ldots, f_m$ — действителные числа, не зависящие ни от $A_1, \ldots, A_m$ ни от $\pi_1, \ldots, \pi_q$.*

Ради формального удобства вводим ещё следующие определения:

$$\alpha A + \beta A = (\alpha + \beta) A,$$
$$\alpha A - \beta A = (\alpha - \beta) A,$$
$$\lim_{n \to \infty} (\alpha_n A) = (\lim_{n \to \infty} \alpha_n) A.$$

Благодаря этому можно на величинах одной размерности производить формальные вычисления как на обыкновенных действительных числах.

III. Общие принципы теории. В формулировке основных аксиом и понятий теории выборочного испытания мы ограничиваемся такой степенью общности, которая вполне достаточна для большинства практических случаев.

*Партией* товара пусть называется множество $\Omega$ предметов, которое имеет следующие свойства.

Если $\Omega_1, \Omega_2, \ldots$ подмножества (части) партии $\Omega$, то определено *сложение* $\cup$ этих частей, а результат $\Omega_1 \cup \Omega_2$ — тоже часть партии. Для всех частей партии существуют две меры $N$ и $W$, удовлетворяющие аксиомам:

1° Меры $N$ и $W$ — размерные величины и размерно независимы.



2° Меры $N$ всех частей одной и той-же партии имеют общую размерность — мы её называем „ШТУКОЙ" — и меры $W$ всех частей одной и той-же партии тоже имеют общую размерность — мы её называем "РУБЛЁМ".

3° Если $\Omega_1$ и $\Omega_2$ не имеют общей части, то

$$N(\Omega_1 \cup \Omega_2) = N(\Omega_1) + N(\Omega_2),$$
$$W(\Omega_1 \cup \Omega_2) = W(\Omega_1) + W(\Omega_2).$$

Меру $N$ мы называем *объёмом партии* $\Omega$, а меру $W$ — *стоимостью партии* $\Omega$.

*Выборкой* из партии товара мы называем некоторое подмножество $\omega$ множества $\Omega$, имеющие следующие свойства.

Если $\omega_1, \omega_2, \ldots$ части выборки $\omega$, то определено *сложение* $\vee$ этих частей (необязательно совпадающее со сложением $\cup$), а результат $\omega_1 \vee \omega_2$ — тоже часть выборки. Для всех частей выборки существуют две меры $n$ и $w$, удовлетворяющие следующим аксиомам:

5° Меры $n$ и $w$ размерные величины и вместе с $N$ и $W$ образуют систему единиц.

6° Меры $n$ всех частей одной и той-же выборки имеют общую размерность — мы её называем „штукой" — и меры $w$ всех частей одной и той-же выборки имеют общую размерность — мы её называем „рублём".

7° Если $\omega_1$ и $\omega_2$ не имеют общей части, то

$$n(\omega_1 \vee \omega_2) = n(\omega_1) + n(\omega_2),$$
$$w(\omega_1 \vee \omega_2) = w(\omega_1) + w(\omega_2).$$

Меру $n$ называем *объёмом выборки* $\omega$, а меру $w$ — *стоимостью выборки* $\omega$.

Символы ШТУКА, РУБЛЬ, штука, рубль — конечно — условные наименования общих понятий, образующих систему единиц.

Если $N$ — объём, $W$ — стоимость партии товара, то *ценой партии* мы называем величину

$$C = W/N.$$

Аналогично, *ценой выборки* (или *лабораторной ценой*) мы называем величину

$$c = w/n.$$

Размерности цен следующие:

$$[C] = \text{РУБЛЬ} \cdot \text{ШТУКА}^{-1}, \qquad [c] = \text{рубль} \cdot \text{штука}^{-1}.$$

Предполагаем, что

$$C = qc + C_0,$$

где $q$ называется *бухгалтерским коэффициентом* и имеет размерность

$$[q] = \text{РУБЛЬ} \cdot \text{ШТУКА}^{-1} \cdot \text{рубль}^{-1} \cdot \text{штука},$$

а $C_0$ — постоянная величина размерности РУБЛЬ (накладные расходы). Пусть $N_1$ объём части $\Omega_1$ партии $\Omega$. Пусть $N$ — объём, а $W$ — стоимость партии $\Omega$. Тогда *средней стоимостью* $\overline{W}$ *в части партии* $\Omega_1$ мы называем

$$\overline{W} = W \frac{N_1}{N} = CN.$$

Аналогично определяем *среднюю стоимость $\overline{w}$ в части $\omega_1$ выборки.* Размерности этих средних следующие:

$$[\overline{W}]=[W]=\text{РУБЛЬ},\qquad [\overline{w}]=[w]=\text{рубль}.$$

*Рассеяние стоимости в партии* (и соответственно в выборке) определяем на основании некоторых аксиом, которые формулируем ниже.

Пусть какая-то часть $\Omega_0$ партии товара состоит из $\mathfrak{N}$ долей $\Omega_1,\Omega_2,\ldots,\Omega_{\mathfrak{N}}$ без общих частей, т. е.

$$\Omega_0=\Omega_1\cup\Omega_2\cup\ldots\cup\Omega_{\mathfrak{N}}.$$

Пусть (для определённости, но без нарушения общности) объём всех) долей $\Omega_1,\Omega_2,\ldots,\Omega_{\mathfrak{N}}$ одинаков и равен $N_0$, а стоимость доли $\Omega_\gamma$ равна $\overline{W}$ $(\gamma=1,2,\ldots,\mathfrak{N})$. Тогда средняя стоимость $W_\gamma$ в каждой из этих долей одинакова и равна

$$\overline{W}=W\frac{N_0}{N}=CN_0,$$

где $N$ — объём, а $W$ — стоимость всей партии $\Omega$. Объём части $\Omega_0$ равен $N_{\Omega_0}=\mathfrak{N}N_0$.

Мы считаем, во-первых, что *мера $S$ рассеяния стоимости в части $\Omega_0$ партии* является *размерной функцией* $S_{\Omega_0}$, которая зависит от множества $\Omega_0$ и от размерных величин $W_1, W_2,\ldots,W_{\mathfrak{N}}, N_{\Omega_0}, W, N_0$, т. е.

$$S=S_{\Omega_0}(W_1,\ldots,W_N,N_{\Omega_0},\overline{W},N_0).$$

Введём безразмерные аргументы

$$W_\gamma=\frac{W_\gamma}{W}\qquad(\gamma=1,2,\ldots,\mathfrak{N}).$$

Тогда, на основании теоремы II, получаем

$$S=\Psi_{\Omega_0}(W_1, W_2,\ldots, W_{\mathfrak{N}},\mathfrak{N})\,\overline{W}^a N_0^b,$$

где $\Psi_{\Omega_0}$ — безразмерная функция, вид которой зависит от множества $\Omega_0$, а показатели $a, b$ — действительные числа.

Мы полагаем, во-вторых, что $a=1$.

Чтобы определить показатель $b$, мы принимаем ещё одну аксиому. Пусть $\Omega^{\mathrm{I}}$, $\Omega^{\mathrm{II}}, \Omega^{\mathrm{III}}$, какие-то три части партии товара, удовлетворяющие всем аксиомам сформулированным для части $\Omega_0$. Пусть число $\mathfrak{N}$ будет общим для $\Omega^{\mathrm{I}}$, $\Omega^{\mathrm{II}}$, $\Omega^{\mathrm{III}}$; значит $\Omega^i$ $(i=\mathrm{I},\mathrm{II},\mathrm{III})$ состоит из $\mathfrak{N}$ долей $\Omega_1^i,\Omega_2^i,\ldots,\Omega_{\mathfrak{N}}^i$ (без общих частей), объём которых равен $N_0^i$, так что объём $N^i$ части $\Omega^i$ равен $N_i=\mathfrak{N}N_0^{(i)}$. Если стоимость доли $\Omega_\gamma^i$ $(i=\mathrm{I},\mathrm{II},\mathrm{III};\ \gamma=1,2,\ldots,\mathfrak{N})$ равна $W_\gamma^i$, то её средняя стоимость равна

$$\overline{W}_i=CN_0^{(i)}.$$

Пусть, далее, рассеяние $S_i$ стоимости в части $\Omega^i$ будет

$$S_i=\Psi_i(W_1^i,\ldots, W_{\mathfrak{N}}^i, \mathfrak{N})\,\overline{W}_i(N_0^{(i)})^b\qquad(i=\mathrm{I},\mathrm{II},\mathrm{III}),$$

где безразмерные величины $W_\gamma^i$ $(i=\mathrm{I},\mathrm{II},\mathrm{III};\ \gamma=1,\ldots,\mathfrak{N})$ равны

$$W_\gamma^i=\frac{W_\gamma^i}{W_i}.$$



Пусть $\Omega^{\mathrm{I}}$ и $\Omega^{\mathrm{II}}$ не имеют общих частей и $\Omega^{\mathrm{III}}$ состоит из $\mathfrak{N}$ долей $\Omega_\gamma^{\mathrm{III}}$ $(\gamma=1,2,\ldots,\mathfrak{N})$, возникших путём соединения в пары соответствующих долей $\Omega_\gamma^{\mathrm{I}}$ и $\Omega_\gamma^{\mathrm{II}}$, т. е.

$$\Omega_\gamma^{\mathrm{III}}=\Omega_\gamma^{\mathrm{I}}\cup\Omega_\gamma^{\mathrm{II}}\qquad(\gamma=1,2,\ldots,\mathfrak{N}).$$

Тогда, в силу аксиомы 3°, получаем

$$N_0^{\mathrm{III}}=N_0^{\mathrm{I}}+N_0^{\mathrm{II}}.$$

Части $\Omega^{\mathrm{I}}$ и $\Omega^{\mathrm{II}}$ партии называем *некорелированными*, если

$$S_{\mathrm{I}}^2+S_{\mathrm{II}}^2=S_{\mathrm{III}}^2.$$

Третья аксиома состоит в том, что если части $\Omega_{\mathrm{I}}$ и $\Omega_{\mathrm{II}}$ некорелированные, то существует общий для всех частей $\Omega_{\mathrm{I}},\Omega_{\mathrm{II}},\Omega_{\mathrm{III}}$ предел

$$\lim_{\mathfrak{N}\to\infty}\Psi_{\mathrm{I}}=\lim_{\mathfrak{N}\to\infty}\Psi_{\mathrm{II}}=\lim_{\mathfrak{N}\to\infty}\Psi_{\mathrm{III}}.$$

Из этой аксиомы уже следует, что $b=-1/2$.

Совершенно аналогично определяем рассеяние $s$ стоимости в выборке. Размерности рассеяний $S$ и $s$ получаются таким образом следующие:

$$[S]=\text{РУБЛЬ}\cdot\text{ШТУКА}^{-1/2},\qquad [s]=\text{рубль}\cdot\text{штука}^{-1/2}:$$

Если $N_0=\text{ШТУКА}$, то формулы на рассеяние, употребляемые в статистике, имеют следующий вид:

$$\text{Стандартное отклонение}=\sqrt{\frac{(W_1-\overline{W})^2+\ldots+(W_{\mathfrak{N}}-\overline{W})^2}{N}},$$

$$\text{Среднее отклонение}=\frac{|W_1-\overline{W}|+\ldots+|W_{\mathfrak{N}}-\overline{W}|}{\mathfrak{N}}\,\frac{1}{\sqrt{\text{ШТУКА}}},$$

$$\text{Широта}=(\operatorname*{Max}_\gamma W_\gamma-\operatorname*{Min}_\gamma W_\gamma)\frac{1}{\sqrt{\text{ШТУКА}}}.$$

Таким образом, размерность рассеяния стоимости одной ШТУКИ в партии (или соответственно одной штуки в выборке) одна и та-же для всех мер рассеяния, употребляемых в статистике.

В заключение этой главы даются определения и аксиомы экономического характера. Мы считаем, что расход $B$, затраченный на испытание, имеет размерность РУБЛЬ, а расценка $K$ испытания, т. е. расход на испытание одной штуки в выборке, равна

$$k=\frac{B}{n}$$

и, следовательно, имеет размерность

$$[k]=\text{РУБЛЬ}\cdot\text{штука}^{-1}.$$

IV. Объём выборки. Исходя из единственного предположения о том, от каких аргументов зависит объём выборки, можно определить более точный вид этой зависимости. Вот несколько примеров.

1.1. Предположим, что объём $n$ выборки зависит только от объёма $N$ партии, расценки $k$ испытания, рассеяния $s$ стоимости штуки в выборке и бухгалтерского коэффициента $q$, т. е.

$$n=\Phi(N,k,s,q).$$

Так как аргументы $N, k, s, q$ размерно независимы, то на основании теоремы II получаем

$$n=a\left(\frac{Nqs}{k}\right)^{2/3},$$

где $a$ — постоянный безразмерный коэффициент. Аналогичную зависимость получил другим путём H. Steinhaus.

1.2. Предположим, что

$$n=\Phi(N,k,s,q,C),$$

где $C$ — цена партии. На основании теоремы П получаем тогда

$$n=\varphi(\beta_1)\left(\frac{Nqs}{k}\right)^{2/3},$$

где $\varphi$ — безразмерная функция безразмерной переменной

$$\beta_1=CN^{1/3}(kq^2s^2)^{-1/3}.$$

Если кроме того предположить, что $\varphi(\beta_1)$ разлагается в степенной ряд и ограничиться рассмотрением только двух первых его членов, то

$$n=a_0\left(\frac{Nqs}{k}\right)^{2/3}+a_1\frac{W}{k},$$

где $a_0, a_1$ — постоянные безразмерные коэффициенты, а $W$ стоимость партии товара.

1.3. Предположим, что объём $n$ выборки зависит только от $N$, от $k$ и от рассеяния $S$ стоимости ШТУКИ в партии. Тогда

$$n=\delta\frac{S\sqrt{N}}{k},$$

где $\delta$ — постоянный безразмерный коэффициент. Эта формула считается практиками эмпирической.

Аналогичным образом можно получить другие формулы на величину $n$ выборки, в зависимости от принятых предположений.

V. Цель испытания и согласие с опытом. Анализом размерностей нельзя получить численные значения безразмерных коэффициентов (или функций), которые выступают в формулах на объём выборки. Так как эти формулы получены на основании единственного предположения о том, от каких аргументов зависит объём выборки, то на эти формулы можно и нужно наложить ещё другие дополнительные требования. Общая идея состоит тут в следующем.



Предполагаем, что во всяком выборочном испытании определена некоторая размерная функция $R$, называемая *риском испытания*, значения которой имеют размерность $[R]=$РУБЛЬ.

В качестве аргументов риска $R$ достаточно рассмотреть, в общем случае, только $N, n, W, w, q, k, S, s$. Итак

$$R=\Phi(N,n,W,w,q,k,S,s).$$

По теореме П получаем, что

$$R=\varrho(\xi_1,\xi_2,\xi_3,\xi_4)N^{1/2}S,$$

где $\varrho(\xi_1,\xi_2,\xi_3,\xi_4)$ безразмерная функция безразмерных переменных

$$\xi_1=WN^{-1/2}S^{-1},$$

$$\xi_2=wn^{-1/2}s^{-1},$$

$$\xi_3=qN^{1/2}n^{-1/2}sS^{-1},$$

$$\xi_4=knN^{-1/2}S^{-1}.$$

Вид функции $\varrho$ нужно определить по экономическим соображениям. Если считать, что числовая функция $\varrho$ разлагается в степенной ряд и ограничиться рассмотрением членов до первой степени, то

$$R=\varrho_0SN^{1/2}+\varrho_1W+\varrho_2wn^{-1/2}N^{1/2}s^{-1}S+\varrho_3qn^{-1/2}Ns+\varrho_4kn,$$

и тогда экономисты должны определить только пять безразмерных постоянных $\varrho_0,\varrho_1,\varrho_2,\varrho_3,\varrho_4$.

Кроме вида функции $\varrho$ нужно определить ещё условие, которому должен удовлетворять риск $R$, чтобы *цель* выборочного испытания была достигнута. Это условие — мы называем его *экономическим интересом* — может, например, состоять в том, чтобы риск $R$ был наименьший, или чтобы он не превосходил определённого значения. Могут быть ещё другие условия.

Если вид риска $R$ известен и экономический интерес указан, то безразмерные коэффициенты в формулах на величину выборки лучше всего определять *опытом*. В таком опыте нужно испытать всю партию определённого вида товара.

Таким опытом не только проверяется правильность принятых гипотез, но получаются тоже формулы, которые наиболее соответствуют действительности. Нужно конечно произвести несколько таких опытов, а их результаты можно обработать по методам математической статистики, которая играет роль только на этом этапе. На первый взгляд может казаться, что выполнение таких опытов слишком громоздко. На самом деле это однако не так, и можно с успехом использовать имеющиеся в каждом предприятии информации. „Опытный глаз" специалистов по выборочным испытаниям состоит именно в том, что они — хотя, может быть, несознательно — накопили много результатов таких опытов. Но даже в том случае, когда для некоторых сортов товара такие опыты моглибы оказаться трудоёмкими, то пользы от них получалось бы гораздо больше, чем расходов.

---

CONTENTS