logical observations available. Now that the long-sought foundation for a reliable system of pressure reduction is laid out, further developments in aerological work and the establishment of truer free-air normals, hold out hopes of future refinements in the system to the point of unreserved acceptance of the results.

2. As brought out in the text of this paper, there are a number of difficulties in the way of representing free-air trajectories by means of the free balloon, or of accepting such representation as valid. The chief of these are the difficulties, in practice, of maintaining constant altitude and determining the location of the balloon, and considerations of vertical component of air movement. The question as to whether further practice of free ballooning might lead to closer approximation to the facts on this particular point, vies with the question as to whether the additional knowledge gained would justify the effort and expense involved.

3. Referring to the general problem of free-ballooning with relation to meteorology, there can be no question as to the reciprocal benefits of one to the other; in fact, the very dependence of ballooning on meteorology compels the belief that meteorology can not help but add to its fund of knowledge from a pursuit to which it is indispensable. There are a number of problems in meteorology peculiarly adaptable to free balloon investigation, some of sufficient importance to make this means of attempt at solution well worth the effort. Suggestions of such problems, that incidentally were accessorial reasons for Meisinger's flights, are contained in the text of this paper, as, for example, in the fifth and sixth flights.

Any program of free-ballooning for meteorological purposes must necessarily contemplate flights in unsettled weather, rains, and snows, as well as in fair weather. To pronounce against flights in any but fair weather would go a long way toward admitting the futility of aeronautics. The great danger lies, of course, in thunderstorms. The tragic denouement of this project, following so soon after the toll of lives exacted by the conditions during the National and International Balloon races of 1923, emphasizes the menace of thunderstorms to inflammably charged balloons. It is a matter of record that in all instances in recent years of disasters to manned balloons attributable to thunderstorms the voyagers were aware beforehand of the risks they were facing. It is sufficient testimonial to the aid rendered by meteorology in this field that it can warn of danger even though it can not prophesy disaster. Safety in free-ballooning will be realized when precaution is no longer made subordinate to loyalty to purpose and when it is conceded that "safety first" is as applicable and justifiable in this line of scientific work as in any other peacetime endeavor.

ON THE MEAN VARIABILITY IN RANDOM SERIES [1]

By EDGAR W. WOOLARD

[U. S. Weather Bureau, Washington, D. C., March 27, 1925]

The human mind is so constituted that, when confronted by an extensive array of numerical data, it is incapable of adequately grasping the significance of the figures or of detecting and comprehending the relations and laws exhibited by the data. To overcome this defect, special methods, known as statistical, have been devised for the scientific treatment of collections of data pertaining to mass phenomena. Statistics accomplishes its object by displaying data in forms such that their significance can be more readily grasped, and by inventing special analytical processes designed to reveal the laws and relations concealed in the figures.

One common statistical procedure for rendering data amenable to our mental faculties is that of replacing the original large body of raw numerical material by a very small and compact set of summary coefficients which concisely, yet adequately and comprehensively, resume in themselves all the essential features of the complete data. Of course such a replacement is necessarily made at the expense of detail—no summary can, by virtue of its very nature, contain *all* the facts—but the effort is made to retain in the summary all the facts and features *essential or relevant to the purposes in hand.*

An extensive body of numerical data may thus be succinctly described or characterized by, and, at least for many purposes of statistical analysis, may be replaced by, a brief set of statistical coefficients or indices, one coefficient for each of the important and relevant features of the data. The comparative analysis of two different sets of data, and of the respective phenomena to which they pertain, in respect to each of their essential characteristics, then becomes largely a matter of the comparison of corresponding statistical indices; two different phenomena may be identical in that aspect characterized by the arithmetic mean of the data, and yet differ widely in respect to the feature characterized by, say, the standard deviation. Obviously, in any given case it is a matter of very great importance to be sure we have included in the set of coefficients an index for each and every important aspect of the phenomena under consideration.

Now, the statistical coefficients pertaining to a single variable to which nearly all the attention of statisticians has thus far been directed relate entirely to the various characteristics of the frequency distribution.[1] In most cases, perhaps, this is sufficient, but in some problems, including many important meteorological applications, at least one other feature of the data must be taken into consideration, viz, the *order of succession.* If the statistical data in hand relate to, say, biometric measurements, it is immaterial in what order the data are presented; but if the data relate to the successive values taken on by a time-variable, the order in which the values occur may be quite relevant.

The order of succession is, in fact, one of the many peculiar problems encountered [2] when one seeks to apply the Theory of Probability and the ordinary Theory of "Errors" to meteorological data; the meteorological variables frequently do not conform to the conditions under which the mathematical theories are valid. Statistical

[1] In view of the usefulness of the so-called Goutereau ratio in meteorological investigations, Mr. Woolard was asked to examine the question as to whether there could not be developed a generalization of Goutereau's theorem, which as we understand it applies strictly to numbers in a Gaussian distribution. From a very superficial examination on my own part, I am impressed with the fact that this ratio, while not constantly equal to the square root of 2, (1.41), nevertheless has a value differing but little from that value for very widely differing frequency distributions. For example, the U-shaped distribution of a table of sines seems to lead to a ratio of about 1.25.

The problem might be stated as follows:

Given a limited series of numbers, a, b, c, . . . k, of which f_a, f_b . . . f_k represent the relative frequencies of these numbers. Regardless of the order of succession, the mean deviation of these numbers may be expressed as *md*. If the average value of the mean variation of the numbers in a sequence of unrelated numbers is v, what is the ratio of v÷*md*?—*C. F. Martin.*

[1] See G. U. Yule, *Introduction to the Theory of Statistics*, chap. vii. 7 ed., London, 1924.

[2] See, e. g., V. H. Ryd, On Computation of Meteorological Observations. Danske Meteorologiske Institut, *Meddelelser Nr. 5.* Copenhagen, 1917.

meteorology thus furnishes examples wherein it is highly desirable to have some index depending upon, and completely characterizing, the order of succession of the values in a statistical series.

Suppose, for example, that we have an observed sequence of values of a time-variable, say a series of daily mean temperatures for a number of consecutive days,

$$t_1\, t_2\, t_3 \cdots\cdots\cdots t_k \cdots\cdots \tag{1}$$

If the numbers t_k were written on balls, and the balls drawn one at a time at random from an urn, the resulting sequence of values might, in some important respects, differ widely from the observed sequence (1)—e. g., in the case of daily temperatures, long series of successively increasing or decreasing values would be less frequent in the series obtained by the chance drawings than in the sequence produced by nature—yet the frequency distributions of the two, and the values of all the statistical constants pertaining thereto, would be identical. In one case we have a sequence brought about by the operation of pure chance only, whereas in the other case consecutive values may not be completely independent of one another; yet none of the tests ordinarily applied[3] to determine, when it is essential to know, whether or not the values of a variable are due to fortuitous causes, would distinguish between the two cases, because these tests relate only to the frequency distributions.

Very little work has been done on this matter. Here, as in the case of so many other statistical questions, the problem was first encountered as a special case in the theory of errors of precision measurements, viz, in connection with testing observations for the presence of systematic errors.[4]

The first investigation of the general statistical problem seems to have been that of Grossmann;[5] recently, however, some errors in Grossmann's reasoning have been pointed out by Besson, who has corrected and extended[6] the work; but nothing in the nature of a statistical coefficient has been provided by any of these investigations.

A statistical index characterizing the order of succession was first devised by Goutereau, in 1906.[7] He defined a *variability* as the absolute value of the difference between any number in a sequence and the next consecutive number; and with the aid of Maillet he showed that, provided the frequency distribution were Gaussian, the ratio of the mean of the variabilities to the mean deviation must be equal to $\sqrt{2}$ if the deviations from the mean were legitimately to be likened to fortuitous errors. The ratio is actually but about half this value in general in the case of daily temperatures.

The Goutereau Ratio, as it may be called, was applied by its author only to normal frequency distributions. Moreover, it seems to the writer that the derivations of the formulae, as given by Goutereau, are not as clear and satisfactory as they might be made, and furthermore the equations as printed contain several serious errors. Therefore it seems worth while to remedy these defects in Goutereau's presentation, and, if possible, to extend the work to include distributions that are not normal.

Let a time-variable t, a sequence (1) of n of whose values we have observed, have the following frequency distribution:

$$\begin{matrix} \textit{Values:} & x_1\, x_2\, x_3 \cdots\cdots\cdots x_i \cdots\cdots\cdots x_s \\ \textit{Frequencies:} & a_1\, a_2\, a_3 \cdots\cdots\cdots a_i \cdots\cdots\cdots a_s \end{matrix} \tag{2}$$

$$\sum_i a_i = n \tag{3}$$

If the variable t be a continuous one, (2) gives the ordinary histogram, the x_i being the mid-points of the classes.

The Arithmetic Mean of t is

$$\mathrm{M} = \frac{\sum_k t_k}{n}, \tag{4}$$

while the standard deviation and the mean deviation are, respectively,

$$\sigma = \sqrt{\left\{ \frac{\sum_k (t_k - M)^2}{n} \right\}}; \tag{5}$$

$$\theta = \frac{\sum_k |t_k - M|}{n}. \tag{6}$$

The mean, the standard deviation, and the mean deviation are indices which characterize certain features of the frequency distribution (2), and they would have the same values in whatever order the t_k were observed to occur. In the variabilities and their mean, however, we have something depending upon the order in which the t_k present themselves in the sequence. The variabilities in a sequence such as (1) are given by

$$v_k = |t_{k+1} - t_k|, \tag{7}$$

and the mean variability in a series of N values is

$$V_m = \frac{\sum_{k=1}^{N-1} v_k}{N-1}. \tag{8}$$

Now, if we assume that the observed sequence (1) is a representative sample of the results that will follow the operation of the causes producing the phenomenon under observation, then the frequency ratios a_i/n may be taken to be the *a posteriori* empirical probabilities of the individual values x_i, and these may in turn be identified with the postulational *a priori* mathematical probabilities of the x_i. The production of the observed series through the operation of the complex of causes determining the phenomenon may then be simulated by drawing balls from an urn in which either (A) there is an infinite number of balls marked with the x_i in such proportions that for any n balls there are on the average a_1 marked x_1, a_2 marked x_2, and so on, the proportions of the different kinds remaining the same no matter how many may be drawn out; or (B) in which there are n balls marked with the various x_i in the proper proportions, the balls being returned after each drawing before the next drawing is made. (This assumes, of course, that we are dealing with a Bernoullian Series).

If we make a number, N, of successive drawings from the urn, we obtain a so-called random or chance sequence

---

[3] Goutereau, *Annuaire de la Soc. Mét. de France*, 54, 122–123, 1906; Woolard, MO. WEATHER REV., 49, 132–133, 1921.

[4] E. Abbe, Ueber die Gesetzmässigkeit in der Verteilung der Fehler bei Beobachtungsreihen, Jena, 1863 (*Habilitationsschrift*); *Ges. Abh.*, Bd. II, Jena, 1906, pp. 55–81. "Abbe's Criterion" has been modified slightly by Helmert; see F. R. Helmert, *Die Ausgleichungsrechnung nach der Methode der Kleinsten Quadrate*, 2te Aufl., Leipzig, 1907, pp. 341–345. Another method of dealing with the same question was used by F. R. Helmert and W. Seibt, Das Mittelwasser der Ostsee bei Swinemünde. 2 Mitt. *Veröff. d. Kön. Preuss. Geod. Inst.*, 1890; *cf. Jahresbericht d. Direktors*, 1889–90, p. 26–27.

[5] L. Grossmann. Die Aenderung der Temperature von Tag zu Tag an der deutschen Kuste. *Aus dem Archiv der Deutschen Seewarte*, XXIII Jahrgang, 1900, pp. 34–37.

[6] L. Besson. On the Comparison of Meteorological Data with results of Chance. Translated by Edgar W. Woolard. MO. WEATHER REV, 48, 89–94, 1920.

[7] Ch. Goutereau. Sur la variabilité de la température. *Annuaire de la Soc. Mét. de France*, 54, 122–127, 1906; Edgar W. Woolard, The Mean Variability as a Statistical Coefficient, MO. WEATHER REV., 49, 132–133, 1921.

of numbers; and it is possible under such circumstances, as we shall show, to compute the mathematical expectation[8] of the value of a variability. A comparison of this theoretical or "expected" value with the actual mean of the observed variabilities may show whether or not the observed sequence constitutes a chance sequence, i. e., whether the order of succession in Nature is a random one, controlled by pure chance alone, or is controlled by some *systematic* influence.

An individual variability in any sequence made up from the frequency distribution (2) may happen to have any one of the various possible values of

$$|x_i - x_j|,\ i,\ j, = 1,\ 2,\ 3, \ldots\ldots, s. \tag{9}$$

Obviously, the total number of values which the expression (9) may take on is given by the number of "combinations with repetitions" or "complete combinations" of s things two at a time, which is[9]

$$\frac{(s+1)!}{(s-1)!\,2!} = \frac{s(s+1)}{2}; \tag{10}$$

but by no means all these values are numerically distinct. The same value zero, e. g., which occurs whenever $i=j$, is produced by s of these combinations, and in general the remaining

$$\frac{s(s+1)}{2} - s = \binom{s}{2} \tag{11}$$

combinations will not all produce different numerical values.

Now, the total number of possible ways in which variabilities may be produced is given by the number of "permutations with repetitions" or "complete arrangements"[9] of n things two at a time, which is

$$n^2, \tag{12}$$

each of which is an "equally probable event." By equation (3)

$$\begin{aligned} n^2 = a^2_1 + a_2^2 + \ldots\ldots\ldots\ldots + a^2_s \\ + 2a_1a_2 + 2a_1a_3 + 2a_1a_4 + \ldots\ldots\ldots + 2a_1a_s \\ + 2a_2a_3 + 2a_2a_4 + \ldots\ldots\ldots + 2a_2a_s \\ + \ldots\ldots\ldots\ldots\ldots + 2a_{s-1}a_s \\ = \sum_{i=1}^{s} a^2_i + 2\sum_{m=1}^{s-1}\sum_{i=1}^{s-m} a_i a_{i+m}. \end{aligned} \tag{13}$$

It is easily seen that the $\sum a^2$ ways comprise those of the n^2 permutations which give a zero value to (9), while the term $2a_qa_r$ corresponds to the permutations which make (9) equal to $|x_q - x_r|$.

Hence the probability of a zero variability in a sequence drawn at random from the distribution (2) is

$$\frac{\sum_i a^2_i}{n^2}, \tag{14}$$

while the probability of a variability $|x_q - x_r|$ is

$$\frac{2a_qa_r}{n^2}. \tag{15}$$

[8] A good exposition of the nature and significance of mathematical expectation will be found in G. Castelnuovo, *Calcolo delle Probabilità*, Milan, 1919, capit. iii; see also Arne Fisher, *Mathematical Theory of Probabilities*, Vol. 1, 2 ed., pp. 102-103, New York, 1922.

[9] For the combinatorial formulae needed in this paper, see E. Netto and H. Vogt, Analyse Combinatoire et Théorie des Déterminants, *Encyc. des Sci. Math.*, Tome I, vol. 1, Fasc. 1, Paris, 1904; or E. Netto, *Lehrbuch der Combinatorik*, Leipzig, 1901.

The probabilities (14) and (15) may also be found by noting that the respective probabilities of the x_i in a single drawing are

$$p_1 = \frac{a_1}{n},\ p_2 = \frac{a_2}{n},\ \ldots\ldots,\ p_i = \frac{a_i}{n},\ \ldots\ldots,\ p_s = \frac{a_s}{n}; \tag{16}$$

so that the probability of x_r coming adjacent to x_q in a random series is

$$\frac{a_q}{n}\cdot\frac{a_r}{n} + \frac{a_r}{n}\cdot\frac{a_q}{n} = \frac{2a_qa_r}{n^2}, \tag{17}$$

whereas the probability of two identical values x_q being adjacent is merely

$$\frac{a_q}{n}\cdot\frac{a_q}{n} = \left(\frac{a_q}{n}\right)^2 \tag{18}$$

If the x_i are the midpoints of the classes into which the t_k are grouped in forming the frequency distribution (2), and h the (constant) class interval, then we may always write

$$x_i = c + ih,\quad i = 1,\ 2, \ldots\ldots\ldots s, \tag{19}$$

where c is some constant, positive, negative, or zero. The same is true if the x_i are actual values of a discrete variable, h being the unit of measurement. Then the *numerically distinct* values of the expression (9), any one of which an individual variability may happen to have, are

$$0,\ h,\ 2h,\ 3h,\ \ldots\ldots\ldots\ldots,\ (s-1)h. \tag{20}$$

Of the $s(s+1)/2$ complete combinations to each of which corresponds a value of (9), $(s-m)$ produce the *same* numerical value, viz, mh.[10]

If we have

$$\begin{aligned} |x_q - x_r| &= |(q-r)h| \\ &= mh, \end{aligned} \tag{21}$$

then we must have

$$|q - r| = |m; \tag{22}$$

and the $(s-m)$ combinations all of which result in the same value mh for (9) are

$$|x_{i+m} - x_i|,\ i = 1,\ 2,\ 3, \ldots\ldots, (s-m), \tag{23}$$

the respective probabilities of which are, by (15),

$$\frac{2}{n^2}a_ia_{i+m}.\ \begin{cases} m \neq 0 \\ i = 1,\ 2,\ 3, \ldots, (s-m) \end{cases} \tag{24}$$

(It is not necessary to take into account the cases in which $m=0$, since they would not contribute anything to our final result).

Therefore, the mathematical expectation of the variability—the expected, probable, or mean, variability in an unlimited sequence of random drawings—is given by the equation

$$E(v_k) = \sum_{m=1}^{s-1} mh \sum_{i=1}^{s-m} \frac{2}{n^2} a_ia_{i+m} = \frac{2h}{n^2}\left\{\sum_{m=1}^{s-1} m\left[\sum_{i=1}^{s-m} a_ia_{i+m}\right]\right\} \tag{25}$$

[10] Since $\sum_{m=0}^{s-1}(s-m)$ is an arithmetic progression (the sum of the first s natural numbers in fact), its sum is, as it should be, $\frac{s(s+1)}{2}$, the total number (10) of combinations.

A check on the accuracy of formula (25) is afforded by the following somewhat different derivation: The value x_i, with probability a_i/n, having been drawn, the variability mh will result if the next value drawn be

$$x_i \pm mh = c + (i \pm m)\,h; \qquad (26)$$

the probability of such an event is of course

$$\frac{a_i}{n} \cdot \frac{a_{i\pm m}}{n}. \qquad (27)$$

Now, in order that (26), the second value drawn, may possibly differ from the first, x_i, by as much as $\pm mh$, it is necessary and sufficient that i have any one of the values

$$i = \begin{cases} 1,\ 2,\ 3, \ldots\ldots, (s-m), \\ m+1,\ m+2, \ldots\ldots, s. \end{cases} \qquad (28)$$

Hence, by the addition theorem in the Calculus of Probabilities, the probability of the variability mh, $m = 1, 2, 3, \ldots\ldots, (s-1)$, is [11]

$$\frac{1}{n^2}\left[\sum_{i=1}^{s-m} a_i a_{i+m} + \sum_{i=m+1}^{s} a_i a_{i-m}\right]; \qquad (29)$$

and the mathematical expectation becomes

$$\begin{aligned} \mathrm{E}\,(v_k) &= \frac{1}{n^2}\sum_{m=1}^{s-1} mh \left[\sum_{i=1}^{s-m} a_i a_{i+m} + \sum_{i=m+1}^{s} a_i a_{i-m}\right] \\ &= \frac{2h}{n^2}\left\{\sum_{m=1}^{s-1} m \left[\sum_{i=1}^{s-m} a_i a_{i+m}\right]\right\}, \end{aligned}$$

the same as (25). This formula lends itself very readily to numerical computation, as the examples to be given below will show.

If the x_i are all equally probably, $a_i = const. = a$, $n = sa$, and (25) reduces to

$$\mathrm{E}\,(v_k) = \frac{2h}{s^2}\sum_{m=1}^{s-1} m\,(s-m) = \frac{h\,(s-1)\ (s+1)}{3s}. \qquad (30)$$

By equations (6) and (25), the expected value of the Goutereau Ratio in a random sequence

$$t_1\, t_2\, t_3 \ldots\ldots\ldots t_k \ldots\ldots\ldots t_N \qquad (31)$$

of N values drawn from the frequency distribution (2) is

$$\mathrm{G_c} = \frac{\mathrm{E}\ (v_k)}{\theta} = \frac{2h}{n}\left\{\frac{\sum_{m=1}^{s-1} m \left[\sum_{i=1}^{s-m} a_i a_{i+m}\right]}{\sum_{i=1}^{s} a_i\,|\,x_i - \mathrm{M}\,|}\right\}. \qquad (32)$$

The actual value will be

$$\mathrm{G_o} = \frac{\frac{1}{\mathrm{N}-1}\left[\sum_{k=1}^{N-1} |\,t_{k+1} - t_k\,|\right]}{\sum_{k=1}^{N}\left|\,t_k - \frac{\sum t_k}{\mathrm{N}}\,\right|}. \qquad (33)$$

Now, if random drawings are made from an urn which contains the distribution (2), then, just as each of the *individual* variabilities in the sequence obtained may have *any* one of the values (20), so if a *number* of sequences like (31) are drawn, the actual mean variabilities of these different sequences will range over a number of different values, none of which may happen to coincide with each other or with the expected value; so, too, the mean deviations of the individual sequences will depart from each other and from the value (6). In other words, the actual mean variabilities, mean deviations, and Goutereau ratios of different individual sequences drawn perfectly at random from the same unvarying universe will, like all statistical coefficients, be subject to fluctuations of sampling, of magnitudes dependent on the size of the sample; and in any specific case, before any conclusions as to the presence or absence of systematic control can be drawn from a comparison of the two values given by (32) and (33), it is necessary to know whether or not the difference, if any, can be ascribed to errors of sampling.

From the known formula for the standard error of the mean of an unbiased sample, and the theorem that the standard deviation of the difference of two uncorrelated quantities is equal to the square root of the sum of the squares of the standard deviations of the quantities,[12] we see that the *standard error* of the *observed* mean variability in a *random* series is

$$\epsilon = \frac{\sigma\sqrt{2}}{\sqrt{\mathrm{N}-1}}, \qquad (34)$$

where σ is the standard deviation of (2). (The standard error of a mean, it will be recalled, is independent of the form of the frequency distribution).

However, the usual difficulties are of course encountered when we come to apply our formulæ to actual cases: When drawing sample sequences (31) from an urn of known composition (2), we know *a priori* the true values of the probabilities (16) and of σ, θ, etc.; but any sequence actually presented to us by Nature is a sample drawn from a universe of unknown composition, and in general the best we can do is to adopt for these quantities the values given by this sample itself. All these adopted values are then subject to errors of sampling,[13] and in the case of small samples are quite unreliable; nor are we sure in general that the true values remain constant. Therefore, in actual practice, the value which we compute from (32) is itself in error because of our ignorance of the true composition of the universe from which the observed sequence was drawn; and, furthermore, if we do not know whether or not the observed sequence is a random one, we can not tell whether or not (34) would be applicable even if the real value of σ were known.

Now, suppose that we have for the frequency distribution not a histogram (2) but an analytical expression, i. e., the equation to the frequency curve, so that

$$\left.\begin{aligned} a &= n\varphi\ (x) \\ p &= \ \ \varphi\ (x) \end{aligned}\right\},\ \mathrm{L_1} \leqq x \leqq \mathrm{L_2}. \qquad (35)$$

Then, following the second method by which we deduced equation (25), we have for the probability of a variability of magnitude h

$$\int_{\mathrm{L_1}}^{\mathrm{L_2}-h} \varphi(x)\,\varphi(x+h)\,dx + \int_{\mathrm{L_1}+h}^{\mathrm{L_2}} \varphi(x)\,\varphi(x-h)\,dx,\ 0 \leqq h \leqq |\mathrm{L_2} - \mathrm{L_1}|; \quad (36)$$

and for the expected variability

$$\begin{aligned} \mathrm{E}\,(v) = \int_0^{|\mathrm{L_2}-\mathrm{L_1}|} h \Bigg[&\int_{\mathrm{L_1}}^{\mathrm{L_2}-h} \varphi(x)\,\varphi(x+h)\,dx \\ &+ \int_{\mathrm{L_1}+h}^{\mathrm{L_2}} \varphi(x)\,\varphi(x-h)\,dx \Bigg] dh. \end{aligned} \qquad (37)$$

Thus, if the distribution be normal, and $x = 0$ is the mean, then

$$a = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right), -\infty \leqq x \leqq +\infty; \qquad (38)$$

[11] The case $m=0$ has again been excluded; including it, the sum of the probabilities (29) is seen, by equation (13), to be unity, as of course it must.

[12] G. U. Yule, *Introduction to the Theory of Statistics*, 5 ed., pp. 344 *ff*., and 210–211; London, 1919. British M. O., *Computer's Handbook*, Sec. V, subsec. 2 chap. v; London, 1915.

[13] See *Handbook of Mathematical Statistics*, ed. by Rietz, chap. v, especially pp. 77, 78. Boston, 1924.

and

$$\mathrm{E}(v)=\int_0^\infty h\left\{\frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty}\exp\left(-\frac{x^2}{2\sigma^2}\right)\left[\exp\frac{-(x+h)^2}{2\sigma^2}+\exp\frac{-(x-h)^2}{2\sigma^2}\right]dx\right\}dh$$

$$=\frac{1}{2\pi\sigma^2}\int_0^\infty hdh\int_{-\infty}^{+\infty}\left\{\exp\frac{-[x^2+(x+h)^2]}{2\sigma^2}+\exp\frac{-[x^2+(x-h)^2]}{2\sigma^2}\right\}dx. \tag{39}$$

The integrations may be carried out by means of the substitutions

$$\begin{aligned} x+(x\pm h)&=2x\pm h=z,\\ x&=\tfrac{1}{2}(z\mp h),\\ x^2+(x\pm h)^2&=\tfrac{1}{2}(z^2+h^2)\\ dx&=dz/2; \end{aligned} \tag{40}$$

whence

$$\mathrm{E}(v)=\frac{1}{2\pi\sigma^2}\int_0^\infty hdh\int_{-\infty}^{+\infty}\exp\frac{-(z^2+h^2)}{4\sigma^2}dz$$

$$=\frac{1}{\sigma\sqrt{\pi}}\int_0^\infty h\exp\left(-\frac{h^2}{4\sigma^2}\right)dh=\frac{2\sigma}{\sqrt{\pi}}. \tag{41}$$

The mean deviation in a normal distribution is given by [14]

$$\vartheta=\frac{\sqrt{2}}{\sigma\sqrt{\pi}}\int_0^\infty x\exp\left(-\frac{x^2}{2\sigma^2}\right)dx=\sigma\sqrt{\frac{2}{\pi}}. \tag{42}$$

Hence, in a random sequence drawn from a normal distribution, the value of the Goutereau Ratio becomes

$$\mathrm{G}_N=\frac{2\sigma}{\sqrt{\pi}}\cdot\frac{\sqrt{\pi}}{\sigma\sqrt{2}}=\sqrt{2}, \tag{43}$$

the result of Goutereau and Maillet.[15]

Table 1

x	Urn composition		Composition, sample of 1,001		$p'-p$
	a	p	p'	Stand. error	
1	5	.082	.089	.009	+.007
2	4	.065	.057	.007	−.008
3	8	.131	.123	.010	−.008
4	13	.213	.226	.013	+.013
5	22	.361	.349	.015	−.012
6	3	.049	.045	.006	−.004
7	2	.033	.047	.007	+.014
8	4	.066	.064	.008	−.002
	61	1.000	1.000		
	M=4.311 σ=1.714 ϑ=1.34		M=4.332 σ=1.739		

Equation (25) was subjected to experimental test in the following way: Sixty-one beans were marked with the numbers 1, 2, ..., 8 in the proportions given in Table I,[16] and put into a dish; 1001 random drawings were then made, the one drawn being returned each time before the next drawing, and the sequence of numbers thus obtained recorded. Table I also shows the observed frequency distribution for this sample sequence

The expected variability may be computed from equation (25) according to the scheme shown in Table II The frequency distribution is tabulated in the first two columns, following which are $s-1$ columns numbered 1, 2, ..., $s-1$; now, beginning with the last frequency, viz, 4, as multiplier, and taking each of the other frequencies in order—2, 3, ... 5—as a multiplicand, fill out the last row of the table—8, 12, 88 ...., 20; then, using the next to the last frequency—2—as multiplier, and each of the frequencies preceding it—3, 22, ..., 5—as multiplicand, fill out the next to the last row—6, 44, ..., 10; and so on. This can be done quite rapidly with a multiplication table or a calculating machine. Then add up each of the numbered columns, multiply the sum by the number of the column, add these products, and multiply this last sum by $2h/n^2$; the result is the required expectation.

Table 2

x	a	1	2	3	4	5	6	7	
1	5								
2	4	20							
3	8	32	40						
4	13	104	52	65					
5	22	286	176	88	110				
6	3	66	39	24	12	15			
7	2	6	44	26	16	8	10		
8	4	8	12	88	52	32	16	20	
	61	522	363	291	190	55	26	20	
		522	726	873	760	275	156	140	3,452

$$n=61;\ s=8;\ h=1.$$

$$\mathrm{E}(v_k)=\frac{2\times3452}{61^2}=1.855$$

The expected value of the variability for a random sequence drawn from the frequency distribution of Table I is found to be 1.855; if, as is the case in practise, we had not known the true composition of the universe from which the sequence was drawn, but had been forced to use in (25) the observed composition of the sample itself, we should have found 1.885 for the expected variability. The actually observed mean of the 1,000 variabilities was 1.84, with a *standard error*, according to (34) of .0766, and hence a *probable error* of .052.

The observed sequence of 1,000 variabilities was also cut up into 100 samples of 10 each, and the mean of each of these samples computed. According to (34) the standard error of a mean variability computed from 10 values would be .76; the 100 values were not enough to give a smooth frequency distribution, but after grouping them until a smooth distribution was obtained, they gave a mean of 1.862 and a standard deviation of .61.

Table 3 gives the results of 1,001 drawings from another frequency distribution. The results of these experiments fully confirm the theoretical formulae developed in this paper.

Table 3

x	Urn composition		Sample of 1,001	
	a	p	p'	$p'-p$
1	23	.354	.342	−.012
2	13	.200	.212	+.012
3	9	.138	.141	+.003
4	7	.108	.101	−.007
5	4	.062	.064	+.002
6	3	.046	.040	−.006
7	5	.077	.079	+.002
8	0	.000	.000	±.000
9	1	.015	.021	+.006
	65	1.000	1.000	
	M=2.86 σ=2.04		n=65; s=9; h=1 E(v_k)=2.164 V_m=2.150: Standard error, .091; probable error, .061	

[14] *Cf.* Arne Fisher, *Mathematical Theory of Probabilities*, vol. I. 2 ed., pp. 122–24, New York, 1922; G. U. Yule, *Introduction to the Theory of Statistics*, 5 ed., p. 304, London, 1919.
[15] Goutereau, *l.c.*
[16] *Cf.* C. F. Marvin, Mo. WEATHER REV., 52, 440–441, 1924.