LABORATORY REPORT, No. 1, Vol. VIII.

THE APPLICATION OF THE "LAW OF ERROR" TO THE WORK OF THE BREWERY.

Brewhouse Report, dated 3rd November, 1904.

WITH BOARD ENDORSEMENT, 9th March, 1905.

BOARD ENDORSEMENT, No. 62,

ON

REPORT, dated 3rd November, 1904, by
Mr. GOSSET, on

THE APPLICATION OF THE "LAW OF ERROR" TO THE WORK OF THE BREWERY.

Mr. Case will make arrangements for Mr. Gosset to have an interview with Prof. Karl Pearson.

C. D. La TOUCHE.

9th March, 1905.

THE APPLICATION OF THE "LAW OF ERROR" TO THE WORK OF THE BREWERY.

3rd November, 1904.

The following report has been made in response to an increasing necessity to set an exact value on the results of our experiments, many of which lead to conclusions which are probable but not certain. It is hoped that what follows may do something to help us in estimating the Degree of Probability of many of our results, and enable us to form a judgment of the number and nature of the fresh experiments necessary to establish or disprove various hypotheses which we are now entertaining.*

When a quantity is measured with all possible precision many times in succession, the figures expressing the results do not absolutely agree, and even when the average of results, which differ but little, is taken, we have no means of knowing that we have obtained an actually true result, and the limits of our powers are that we can place greater or lesser odds in our favour that the results obtained do not differ more than a certain amount from the truth.

Results are only valuable when the amount by which they probably differ from the truth is so small as to be insignificant for the purposes of the experiment. What the odds selected should be depends—

1. On the degree of accuracy which the nature of the experiment allows, and
2. On the importance of the issues at stake.

It may seem strange that reasoning of this nature has not been more widely made use of, but this is due—

1. To the popular dread of mathematical reasoning.
2. To the fact that most methods employed in a Laboratory are capable of such refinement that the results are well within the accuracy required.

Unfortunately, when working on the large scale, the interests are so great that more accuracy is required, and, in our particular case, the methods are not always capable of refinement. Hence the necessity for taking a number of inexact determinations and of calculating probabilities.

In any series of determinations of a simple quantity there are three kinds of errors which prevent all the results being the same. These three kinds of errors are—

1. Mistakes, often of a whole number of integers, as, for instance, of reading a saccharometer 5° too high, or of multiplying two numbers together and getting the wrong result. These can often be found out by checking, or can be conjecturally corrected, and need not concern us further, as they can be eliminated by taking sufficient care.

* A supplement is given at the end of this report bringing some matters in it up to a later date.

2. Constant errors, where, for some reason or other, some instrument is invariably read wrong in the same direction; thus, in determining the gravity of a sample of fermenting wort which has not been "tossed," the gas in the liquid will always cause the saccharometer to read too low. Or again, an incorrect saccharometer will always give a result too high or too low, as the case may be. This form of error can be corrected by taking a sufficient number of instances, and finding out the average amount wrong, by checking against another method, in this case by weighing "tossed" samples. This, therefore, need not concern us either, even though the process of making the correction may be difficult in some cases.

3. When all corrections have been applied the results will still not be quite uniform, though the more carefully the determinations are carried out the nearer will the results be to each other, and to the truth.

Thus, any number of weighings of an unfermented wort with a saccharometer will probably, if sufficient care be taken, come within one degree of one another. If weighed with a bottle they will very likely come within ·1 of a degree, *e.g.* a wort at about 1035 is weighed with a saccharometer, and an endeavour is made to read as accurately as possible: the results will mostly lie between 1034·5 and 1035·5, with a bottle, between 1034·95 and 1035·05. It will be impossible with the first instrument to say this weighs 1034·8, or with the second, 1035·02, though, perchance, these may actually be correct.

Hence we see that according to the precision of the instruments employed, and the care with which they can be used, the accuracy obtained is greater or less. The use of the theory of error is, primarily, to find a measure of the accuracy of a given method; when this is found, the probability that the true value lies within a given amount of an observed value can easily be calculated.

If we consider a number of determinations of a quantity, all corrections having been made, it is obvious that it is more likely that a result will lie close to the real value than far off, so that most of the results will be grouped around the true value, and the numbers further out will tail off on either side. *E.g.* if the real specific gravity of the sample of worts be 1035·02, more of the determinations by the specific gravity bottle will give results between 1035·02 and 1035·03, than between 1035·22 and 1035·23, and few, if any, will lie between 1036·02 and 1036·03, though it is not impossible that such might occur.

Mathematicians have discovered an equation * which defines a curve such

* The equation is $$y = \frac{1}{\sqrt{\pi}} \cdot \frac{dx}{c} \cdot e^{-\frac{x^2}{c^2}}$$

where y is the frequency with which a given error of size occurs π and e are well-known constants, the former being the ratio of the circumference of a circle to the diameter (3·14159), the latter the base of the Napierian system of logarithms (2·718 . . .), dx * is the unit in which x is expressed. c * is the "Modulus of Error," a number which varies with the accuracy of the methods under consideration.

As most of the terms in books are expressed in terms of the modulus, it is not possible to avoid mention of the modulus, nor, beyond saying that it is a measure of the precision of the observation, is it possible to define it in popular language. See Airy, *Theory of Errors of Observations*, p. 17 *et seq.*; Lupton, *Notes on Observations*, p. 75 *et seq.*; Merriman, *Method of Least Squares*, p. 16 *et seq.*

* In measuring a mile dx might be 1 yard and c 20 yards. By a more accurate method, c might be brought down to 5 yards.

that if any given error be the abscissa, the corresponding ordinate represents, very approximately, the frequency with which the error in question will occur, and have found that, though the shape is similar for all series of determinations, the size of this curve depends on what is known as the Modulus of Error. This number is constant for determinations made by the same method with the same care, but larger for coarse, and smaller for more delicate measurements.

In the case above given, the modulus of error for a series of saccharometer weighings might be ten times the modulus of error of a system of bottle weighings.

In diagram No. 1 are three curves representing the frequency of error of systems whose modulus of error are 2 : **1** : $\frac{1}{3}$. These moduli of error are represented by the lines M_1O, $\mathbf{M_2O}$, M_3O.

(They may be imagined as belonging to three methods of estimating a quantity. No. 1 belongs to a rough method, like pacing a distance. **No. 2** belongs to a better method, such as chaining the same distance. No. 3 to a still more refined method, such as measuring with a tape.)

Since the curves are drawn to scale, the total area enclosed by each curve is the same, and if we draw an ordinate, XY_1 Y_2 Y_3 to cut all three curves, the probability of obtaining the error OX in the three systems are as $XY_1 : XY_2 : XY_3$.

Again, the probability of obtaining an error between OX and Ox in the

first curve is $\dfrac{\text{the area } Xxy_1Y_1}{\text{the area enclosed by the whole No. 1 curve}}$,

in the 2nd curve is $\dfrac{\text{the area } Xxy_2Y_2}{\text{the area enclosed by the whole No. 2 curve}}$,

in the 3rd curve is $\dfrac{\text{the area } Xxy_3Y_3}{\text{the area enclosed by the whole No. 3 curve}}$.

Or, since the areas enclosed by the three curves are the same, the probability of obtaining an error between OX and Ox in the three systems is as the area Xxy_1Y_1 : the area $Xxy_2Y_2 : Xxy_3Y_3$. By inspection of the diagram we see that, roughly speaking, it represents the fact that a small error, such as OX, is much more likely to occur in the series with the small modulus of error; whereas the larger the error the greater, in proportion, is the chance of its occurring in the series with the large modulus of error, though in each series the smaller the area the greater the chance of its occurring.

I have indicated in the three curves the points P_1P_1, P_2P_2, and P_3P_3 such that the odds are 20 : 1 that an error will lie between P_1P_1 in the first curve, between P_2P_2 in the second curve, and P_3P_3 in the third curve.

Again, the nature of some experiments obliges us to work with rough measurements, and so a large error, and yet to know the result with some certainty. This is effected by taking a mean of a number of observations. The means thus obtained are of course subject to similar, though smaller, errors, and the greater the number of observations of which the means are taken, the smaller the error, and the curve which represents their frequency of error becomes taller and narrower. In these curves, if No. 1 represents the frequency of error of a single observation, No. 2 will represent the frequency of error of the average of a group of four, and No. 3 will represent the frequency of error of the average of a group of 36.

In practice, the first step is to find what is known as the mean error,* and, as the modulus of error has been found to be 1·77 times this, a measure is obtained of the size of the curve of frequency of error.

To obtain the mean error we must first find the mean (average) of all the determinations under consideration, then the difference between this and each of the determinations, and add together all the positive differences and divide by their number. This will give the mean positive error. We likewise add together all the negative differences and divide by their number, which will give the mean negative error. If sufficient determinations have been made, these will probably be the same, but if they are not, we must take half their sum as the mean error.

The mean error may therefore be described as the average amount of difference between the various results and the grand average.

Extracts in Ascending Order.	No. of Cases.	Difference from Mean.	Total.	Squares.
122·9	1	−1·7	1·7	2·89
123·1	1	−1·5	1·5	2·25
123·2	1	−1·4	1·4	1·96
123·6	1	−1·0	1·0	1·00
123·7	1	− ·9	·9	·81
123·9	1	− ·7	·7	·49
124·0	8	− ·6	4·8	2·88
124·1	2	− ·5	1·0	·50
124·2	6	− ·4	2·4	·96
124·3	5	− ·3	1·5	·45
124·4	3	− ·2	·6	·12
124·5	5	− ·1	·5	·05
No. of negative cases 35 + 3 = 38.			38)18·0(**·47** Mean negative **error.**	14·36
Mean 124·6	5	0	0	
124·7	6	+ ·1	4·6	·06
124·8	9	+ ·2	1·8	·36
124·9	3	+ ·3	·9	·27
125·0	7	+ ·4	2·8	1·12
125·1	3	+ ·5	1·5	·75
125·2	6	+ ·6	3·6	2·16
125·3	1	+ ·7	·7	·49
125·4	4	+ ·8	3·2	2·56
125·6	1	+1·0	1·0	1·00
125·7	1	+1·1	1·1	1·21
126·3	1	+1·7	1·7	2·89
No. of positive cases 42 + 2 = 44.			44)18·9(**·43** Mean positive **error.**	12·87

44)18·9(·43
·47
2)·90
·45 mean error.

14·36
12·87

27·23 ÷ 81 = 3·36 mean square of error.

$\sqrt{3{\cdot}36} = {\cdot}58 \times {\cdot}8$

= ·46 mean error.

* *Vide*, however, supplement.

As a check on this, it is usual to find the "error of mean squares," a number which should be about ·71 times the modulus, and from which another value of the mean error may be found by multiplying by ·8. This number is found by squaring each of the differences from the mean, dividing the sum of the squares obtained by the number of determinations less 1, and taking the square root.

(The value thus obtained is the better value of the two in the proportion 114 : 100.)

E.g. 82 analyses of a malt were done in the Brewery Office Laboratory with the intention of defining the accuracy with which the produce could be found. The results are displayed in the tables on previous page, and the mean error obtained by both methods.

We see then that the mean error is defined by the 82 cases with sufficient accuracy, thus, we have—

Mean of negative error . . .	·47	
,, positive ,, . . .	·43	Average ·46.
,, calculated from errors of mean squares	·46	

Diagram No. 2 compares the theoretical and actual distribution of these results.*

From the mean error we can pass, as mentioned above, to the modulus by multiplying by 1·77 and can construct the curve of the frequency of error, and thus (or more simply by the use of tables) calculate the chances of our observation having an error between any given limits. For this purpose we must first decide—

1. Within what limits of accuracy we desire to know the result.
2. What certainty we require that it will fall within those limits.

E.g. it might be maintained that a Laboratory produce should be within ·5 of the true result with a probability of 10 to 1. The mean error being ·46, the modulus is (·46 × 1·77) = ·8 roughly. Our ·5 error $= \frac{\cdot 5}{\cdot 8}$ of the modulus, say ·6. By consulting the table † we see that 60·4 per cent. of the observations (say 3 out of 5) have an error less than this, *i.e.* the odds are not 10 : 1 but 3 : 2. Anyone demanding 10 to 1 in favour of an error less than ·5 could not therefore rely on a single observation.

As will be shown later, the modulus varies inversely as the square root of the number of observations, the mean of which is considered. Thus with—

		New modulus.	Limit of error desired (·5) ÷ modulus.	Odds in favour of smaller error than ·5.
2 observations	$= \frac{\cdot 80}{\sqrt{2}}$	·57	·9	4 : 1
3 ,,	$= \frac{\cdot 80}{\sqrt{3}}$	·46	1·09	7 : 1
4 ,,	$= \frac{\cdot 80}{\sqrt{4}}$	·40	1·25	12 : 1
5 ,,	$= \frac{\cdot 80}{\sqrt{5}}$	·36	1·40	19 : 1
82 ,,	$= \frac{\cdot 80}{\sqrt{82}}$	·09	5·50	practically infinite.

* Besides the 82 analyses of the same malt, there were three other series of analyses each with its own malt, giving 77 more analyses, or 159 in all; these were all corrected to the same mean and plotted in red ink on Diagram II to show how increasing the number of observations tends to smooth off the curve.

† Appendix I.

In order to get the accuracy we require, we must, therefore, take the mean of four determinations. Practically, we have held the position that the Laboratory extract gives a result within one degree of the truth, and the odds in our favour that this is the case are about 12 : 1. This theoretical conclusion is almost exactly justified by the series quoted above, in which 6 out of 82, or $\frac{6}{76}$, $12\frac{1}{2}$ to 1, exceeded this limit.

In the daily calculation of Laboratory produce, for comparison with the Brewery, several determinations are brought into play, on account of several malts being used, and the error is much less. The experimental brewings, resting on the malt drawn from several bins, have also more than one determination made for the calculation of their Laboratory produce.

In all experiments it is a question to be decided by the comparative labour how far we should repeat experiments, or refine the method, as it is by taking the mean of the first few experiments, that the accuracy is most increased. Thus, to halve the error, four, but to reduce it to $\frac{1}{4}$, 16 experiments are necessary. The table of odds that the error will not exceed ·1 in the Laboratory, is as follows :—

Number of observations.	Modulus.	Odds in favour of error smaller than ·1.
1	·8	1 : 6
4	·4	1 : 3
9	·27	2 : 3
16	·2	13 : 12
64	·1	5 : 1
82	·09	7 : 1
100	·08	12 : 1

A further proposition dealt with by mathematicians * is that connecting the errors affecting a quantity with the known errors of two quantities of which it is composed.

Suppose the mean error of determining a quantity X to be e, and that of determining another quantity Y (quite distinct from X), f, and suppose Z to be the sum or the difference of X and Y, then if E be the mean error of Z, e, f and E are connected by the equation $E^2 = e^2 + f^2$. *E.g.* it has lately been shown that the mean error of an O.G., as determined by the Laboratory, is, roughly, ·3. Again, the Brewery aims at a constant O.G. subject to an error altogether unconnected with the error of the Laboratory determination, which we will suppose to be ·2. Our Laboratories now return us results giving us the so-called O.G.'s of brewings subject, of course, to both these errors. The mean error of such a series will be found, according to the above equation, to be E when $E^2 = (\cdot 2)^2 + (\cdot 3)^2$; $\therefore E = \sqrt{\cdot 13} = \cdot 36$, not much greater than the larger of the constituent errors.

Another example was furnished while considering the figures in the report "The Grind Experiments in 1904." The question arose as to the increase in the error introduced by a correction for half the difference in the Laboratory produce so as to allow for inequalities in malt. The mean error of an average of two Laboratory determinations = °·32 : only half this came into the calculation = say ·2. Now the error of a single brewing = say ·6, and the calculation therefore becomes

$$E^2 = \cdot 6^2 + \cdot 2^2 = \cdot 36 + \cdot 04 \,;$$
$$\therefore E = \sqrt{\cdot 40} = \cdot 63,$$

the Laboratory correction has, therefore, only introduced an additional error of ·03.

* Airy, *Theory of Errors of Observations.* Part II.

Again, in the same report it became important to determine the mean error of the Brewery extract. In this work every result was determined in duplicate, and we may consider that we have two series, an A series and a B series; the first of two determinations belonging to the A series, the second to the B series. We can thus obtain a third series by taking the differences between the A's and the B. Now the mean errors of the A and B determinations are the same, and as they are quite independent, the proposition $E^2 = e^2 + f^2$ applies. In this case e (mean error of the A determinations) $= f$ (mean error of the B determinations), E^2 (the mean error of the difference series) $= 2e^2$;

$$\therefore E = e\sqrt{2}; \therefore e = \frac{E}{\sqrt{2}}.$$

Now E has been determined directly,* and so we deduce e.

This proposition carries with it the important corollary that a series of small errors can be added to a large one without materially increasing the original error. *E.g.* the slide rule introduces a small error of ·03 into the O.G., but the error of the O.G. is ·2, which gives a combined error of only ·202—a quite immaterial increase.

As a further extension of this proposition, we have the cases—

(1) When there are more than two sources of errors, for example e, f, h, etc., when the equation becomes $E^2 = e^2 + f^2 + h^2$, etc.

(2) Where several, say n, determinations are made subject to the same error when $e = f = h$, etc., then $E^2 = ne^2$; $\therefore E = e\sqrt{n}$, *i.e.* the error of a sum of n experiments is n times the error of a single experiment. But, since to find the average we divide by n, we get the formula—

$$\text{Error of average} = \frac{E}{n} = \frac{e\sqrt{n}}{n} = \frac{e}{\sqrt{n}},$$

i.e. the error of an average is equal to the error of a single determination divided by the square root of the number of experiments made, as exemplified above.

The proposition † cannot be extended to observations which have any mutual dependence; but we have, by investigation of the errors, a method of estimating the amount of this dependence, if any, and of discovering if two phenomena, which are supposed to be mutually related, are actually so. *E.g.* the determination of the O.G. on Settling-back in the Brewery Office Laboratory was inaugurated in the hope of showing that high O.G. was a result of under-calculation of the Brewery Produce, and, consequently, that a low Laboratory difference should go with a high O.G., and *vice versâ*. The methods, however, carry a high experimental error, while the error common to the two is comparatively small. Hence ordinary attempts at correlation gave negative results.

When attacked with the aid of the above proposition, the connection can be shown by means of the following argument:—

Any increase in the O.G. due to miscalculation of the "quantity," will cause a decrease of about twice as much in produce, and, consequently, in the Laboratory difference, and *vice versâ*. Therefore, if the O.G. be added to approximately half the Laboratory difference, *this* error will disappear. There are, of course, other errors in the O.G. and Laboratory difference.

* Appendix II.

† The full formula is $E^2 = e^2 + f^2 + 2ref$ where r is the correlation between the quantities X and Y which have mean errors e and f. Hence the above is only true if $r = o$. E, e, and f would generally be taken not as the mean error, but as the Standard Direction or error of mean squares, and the proposition is then true for any type of distribution, and not only for the "normal" type.

Now if we find the mean error of—

1. Half the Laboratory difference.
2. The Original gravity.
3. Of a series obtained by adding them together, day by day, we should find, if there had been no connection between them, the square of this last error equal to the squares of the other two. As a matter of fact, it is found to be considerably less, thus showing that the error common to the two is appreciable. There is, therefore, a connection between them. By taking *sufficient* observations, the extent of this connection can be ascertained.

There are two further points which may be here considered. The first is the rejection of doubtful observations, and the second the degree of accuracy of the number found as the mean error.

It is a matter of common knowledge that often, in the midst of many concordant observations, one occurs very far from the mean. Sometimes it is possible to assign a known cause for this extreme variation, and then, if the cause is found to be adequate, and one peculiar to that observation, we can either reject the observation, or correct it for the extraordinary error. Very frequently, however, no particular reason can be found to account for the discordant observation, and then the question arises, whether we are to accept the result or no.

In the first place, it is clear that if we have only a few observations, one with a very large error will be apt to introduce altogether too large an error into the mean; whereas the mean of a great number of observations will be but little influenced by a single error, even if much greater than the others.

On the other hand, it is generally agreed that to leave the rejection of experiments entirely to the discretion of the experimenter is dangerous, as he is likely to be biassed. Hence it has been proposed to adopt a criterion depending on the probability of such a wide error occurring in the given number of observations.

The criterion proposed by Chauvenet is the simplest of these, and is as follows:—

If, in the number of observations made, an error occurs of such a size that, *à priori*, less than half an observation, with as large an error, might be expected to occur, the observation is to be rejected. In practice, this is applied as follows:—

If n be the number of observations and P the probability that all but half an observation occurs, then

$$\mathrm{P}=\frac{2n-1}{2n} \text{ (since } n-n\mathrm{P}=\tfrac{1}{2}).$$

If, then, we look out in the table (Appendix I) the size of the error corresponding to P, this will give us the limit within which all but half an observation may be expected to occur, and all observations with larger errors may be rejected.

In the case of the 82 observations of Laboratory produce, $n=82$;

$$\therefore \mathrm{P}=\frac{163}{164}=\cdot 994.$$

This corresponds to 1·94 times the modulus (which is ·80). Hence any observation with a greater error than 1·55 must be rejected in determining the mean of the series. Two would be rejected, one positive and one negative, so that the mean would not be affected.

In the case of the whole 159 observations, $n=159$,

$$\mathrm{P}=\frac{317}{318}=\cdot 997.$$

This corresponds to 2·09 times the modulus, or 1·67. Hence, any observation with a greater error than 1·67 should be rejected. As the greatest is 1·7 these would probably be retained, but might, strictly, be rejected.

Now, suppose we had only 5 observations—

$$P = \frac{9}{10} = \cdot 9.$$

This corresponds to 1·17 times the modulus, or, in the Laboratory produce series, ·94 of a degree. Therefore, in a series of 5 observations of a Laboratory produce, one which varied a whole degree from the mean should be rejected.

If this test be applied to the grind experiments, only one pair is rejected, since the limit found for the difference between a pair is 2·6 degrees, and only one difference is higher than this (3·0).

This pair included one experiment which was, in other ways, doubtful; but as it was not one of the important results, and it was not possible to repeat the experiment, the earlier criterion was used, viz., that any experiment not called in question before the result was obtained, was to be included.

(This rule is practically being applied by Mr. Jackson to the case of two, three, and four determinations of extracts.)

The other point is the uncertainty of the number found as the mean error. This is found to have an uncertainty of such a kind that a modulus of error can be found for it, and odds given that it lies within given limits.

The formula by which this is determined is as follows:—If C be the modulus of error found from n experiments, then $\frac{C}{\sqrt{2n}}$ will be the modulus of error for this figure. Or, assuming that 1000 : 1 is certainty, the modulus of error will lie between $C\left(1 + \frac{2\cdot 32}{\sqrt{2n}}\right)$* and $C\left(1 - \frac{2\cdot 32}{\sqrt{2n}}\right)$, and the mean error between $\left(1 + \frac{2\cdot 32}{\sqrt{2n}}\right)$ and $\left(1 - \frac{2\cdot 32}{\sqrt{2n}}\right)$ times its value. This is the uncertainty attached to the error found by the method of mean squares: the other method gives a larger error $\left(1 \pm \frac{2\cdot 48}{\sqrt{2n}}\right)$.

As our results are taken as the mean of these two, it is clear that our accuracy will be greater than either. It will lie somewhere between $\left(1 \pm \frac{2\cdot 32}{\sqrt{2n}}\right)$ and $\left(1 \pm \frac{1\cdot 70}{\sqrt{2n}}\right)$, say $\left(1 \pm \frac{2}{\sqrt{2n}}\right)$.*

In the 82 Laboratory produce experiments, the mean error, ·45, is accurate (with 1000 : 1 certainty), to $\frac{2}{\sqrt{164}} \times \cdot 45$ or ·07, *i.e.*, it is 1000 : 1 that it lies between ·38 and ·52. In the 159 experiments, between ·40 and ·50.

If the probable error (even chances) be taken, the limits are divided, roughly, by 4.

The odds are even that the mean error of a Laboratory produce lies between ·43 and ·47 from the first series, from all 159 experiments between ·44 and ·46.

Similarly, the figure found for the mean error of a produce in the grind experiments, ·67, has a probable error of ·04, *i.e.*, it is an even chance that it lies between ·63 and ·71.

* *Vide* Appendix I.

In conclusion, then, we have gone through the principal methods by which the Law of Error can be applied to our work in the Brewery. We believe that it may become valuable in many directions.

We may point out that, although the proof of the law rests on higher mathematics, the application of it only demands quite simple algebra. We have been met with the difficulty that none of our books mentions the odds, which are conveniently accepted as being sufficient to establish any conclusion, and it might be of assistance to us to consult some mathematical physicist on the matter.

W. S. GOSSET.

CONCLUSIONS.

Discusses the utility of applying the Law of Error to our hypotheses as a measure of their probability.

Classifies Errors.

States, A, The Law of Error, and defines "modulus," "mean error," etc.

Gives instances—

1. Of the calculation of the mean error of the determination of Laboratory Extract.

2. Of the number of determinations of Laboratory Extract necessary to determine this figure within certain limits, with certain degrees of probability.

B, the equation $E^2 = e^2 + f^2$ where E is the error of a quantity compounded of two quantities having errors e and f, the two quantities being determined quite independently, showing that this equation is useful in various ways, *e.g.*—

1. As proving that errors can be added with comparatively slight increase of total error.

2. When e and f are the errors incidental to similar determinations of the same constant $e = f$, leading to the law that where n determinations are made the mean is more accurate than a single determination inversely as the square root of the number of determinations made.

3. To prove the error of the Brewery produce.

4. As a means of investigating phenomena which are supposed to be connected, but in which the connection is obscured by large experimental errors. Such cases follow the equation if not connected, but depart from it if connected, but many instances must be averaged.

Gives a rule for rejecting a doubtful observation.

Explains that we have no information of the degree of probability to be accepted as proving various propositions, and suggests referring this question to a mathematician.

APPENDIX I*

Limit of Error in terms of Modulus.	Per cent. of cases within the given Error.	Limit of Error in terms of Modulus.	Per cent. of cases within the given Error.
0·1	11·2464	1·9	99·2790
0·2	22·2702	2·0	99·5322
0·3	32·8626	2·1	99·7020
0·4	42·8392	2·2	99·8136
0·5	52·0500	2·3	99·8856
0·6	60·3856	2·4	99·9310
0·7	67·7802	2·5	99·9592
0·8	74·2102	2·6	99·9764
0·9	79·6908	2·7	99·9866
1·0	84·2700	2·8	99·9924
1·1	88·0206	2·9	99·9958
1·2	91·0314	3·0	99·9977
1·3	93·4008	3·1	99·9988
1·4	95·2286	3·2	99·9994
1·5	96·6106	3·3	99·9997
1·6	97·6348	3·4	99·9998
1·7	98·3790	3·5	99·9999
1·8	98·9090		

If we consider only the + or the − cases the percentage is halved.

* Adapted from Lupton, *Notes on Observation,* Appendix Table I, p. 122.

APPENDIX II.

TABLE SHOWING THE DIFFERENCES BETWEEN THE PAIRS OF OBSERVATIONS IN THE WHOLE SERIES OF BREWINGS, WITH THEIR SQUARES. TO CALCULATE THE MEAN ERROR OF THE DIFFERENCE, AND SO OF THE BREWING.

No. 1. Brewery.		No. 2. Brewery.		Experimental Brewery.	
Difference.	Square.	Difference.	Square.	Difference.	Square.
·7	49	·5	25	·9	81
1·3	169	·4	16	1·3	169
·3	9	2·1	441	1·0	100
·2	4	1·2	144	·4	16
1·4	196	·8	64	1·3	169
1·8	324	·6	36	1·4	196
·7	49	·7	49	0	...
1·1	121	·0	...	7	49
·6	36	2·2	484	8)70	7)780
·0	...	·8	64	·87	111
·9	81	1·5	225	(Mean error of difference.)	=10·5
2·0	400	·0	...		× ·8
·4	16	·7	49		= ·84
·6	36	·7	49		
·9	81	14)122	13)1646		
·0	...	·87	127		
·5	25	(Mean error of difference.)	=11·3		
1·2	144		× ·8		
·2	4		= ·89		
·9	81				
·7	49				
2·0	400				
1·5	225				
·0	...				
1·4	196				
3·1	961				
·3	9				
1·4	196				
1·4	196				
1·6	256				
30)29·1	29)4313				
·97	149				
	12·2				
	× ·8				
	= ·97				

∴ Mean error of single brewing, No. 1 Brewery, = ·69.
" " No. 2 " = ·62.
" " Experimental " = ·62.

SUPPLEMENT.

The two statistical reports included in this volume were both of an interim character, and in consequence it may be as well to point out a few exaggerations such as are inherent in essays of that kind. This supplement is written, then, to correct the wrong impressions which the reports might be expected to produce.

To consider first "The Application of the Law of Error to the work of the brewery."

Nowhere in this report is there any mention of any other distribution of errors than that given by the "Normal" curve of errors discovered by La Place and Gauss, and a few words seem necessary to explain why it is suitable for representing errors of observation.

The assumptions made in the first place to reach this curve were that an indefinite number of independent sources of error are present, all of which produce an equal deviation from the correct result and all of which are equally likely to be positive or negative. It was subsequently shown that the number of sources of error need not be very great if only they are equal, independent, and equally likely to be positive or negative.

For example, the distribution of heads if only ten coins be tossed up repeatedly will be very closely given by the normal curve even though there are but ten causes present.

But it is quite clear that the necessary conditions of equality, independence, and equal likelihood of being positive or negative are very rarely absolutely fulfilled.

Let us consider, by way of illustration, the curve representing the frequency of the average time of sparging. Let the time of sparging be plotted against the occurrence during the year. There would be a point of time, say $5\frac{1}{2}$ hours, below which the kieves never sparged over, then a few cases would occur, then rapidly more and more, till at, say, $6\frac{3}{4}$ hours there would be more cases than before or after, then the number would decrease again, probably more slowly, till at 11 or 12 hours there would again be no cases. The point where most cases occur is called the mode (the fashionable point), and must be distinguished from the mean or average, which may not coincide with it.

The causes which contribute to a day's average time of sparging occurring in any one five minutes are many, but we may instance—the modification of the various malts, the fineness of the grind, the temperature of the mashing liquor, the occurrence of breakdowns at first mash, the attention bestowed on the kieves, the demand for worts in the copper, etc. etc. It will at once be noticed that the effect on the average time of sparging of these various factors is likely to be quite unequal.

Next, they are not independent, the malts are mixed and ground with an eye to their modification, more attention is bestowed on the kieves on a bad day, the kieves are run more slowly than they could be on a good day owing to their being ahead of the coppers, and so forth.

Finally, they are not equally likely to act either way; in point of fact the larger and rarer causes act in the direction of increasing the time of sparging over.

And so instead of a symmetrical distribution we get one in which the mode is less than the mean, and there is a long "tail" stretching out towards relatively large average times of sparging, which, however, occur but rarely.

This is an instance of what may be called an unsymmetrical "cocked hat" type of Frequency curve, but it is clear that there are other possibilities, such as—(1) The mode being right at one end, *e.g.* cricket scores, or the number of publicans returning 0 1 2 3 etc. casky casks in a month. (2) There may be two modes, which may be due either to a mixture of observations or to an inherent relative absence of the mediocre, *e.g.* the amount of cloudiness present, which has a mode at "no clouds" and another at "all clouds."

But there is another point: it is only when very large numbers are taken that any frequency curve becomes anything like smooth: there are always irregularities due to "random sampling," which only decrease slowly relatively to the number of the observations, absolutely of course they continually increase, as we take larger and larger samples.

Now, owing to these irregularities of random sampling it is often impossible to say that a given distribution was not taken from a population distributed according to the normal law, although if we could analyse a larger sample we might find that this was not the case.

And so although all large samples which have been investigated have been found to deviate in some way or other from "normality," yet for small samples it is practically convenient to use a curve to describe them which has been thoroughly investigated, of which the values have been tabulated, and which in the majority of cases describes them "within the error of random sampling."

In the case of a large population distributed according to the normal law it is quite immaterial whether we fit the curve from the "mean error" or "standard deviation" (error of mean squares) or any other moment co-efficient, but since our samples are small it is better for us to use the "Standard Deviation" (tables have been published in terms of this in place of the "modulus"), as we obtain greater accuracy from the same number of observations. Further, the relation (mean error $= \cdot 8$ standard deviation) only holds in the case of the normal curve, and in that the values obtained from the mean error and the standard deviation are so closely correlated that we gain very little by determining the mean error at all, and it is best to work from the standard deviation alone.

Passing on to the second report on the Pearson Co-efficient of Correlation, what has been said above as to the error of random sampling and the consequent general use of the normal curve, which will approximately fit most small samples, applies also to the use of straight regression lines. We require, as a rule, a large number of cases in order to show that any given regression line is not straight, and so for practical purposes we can generally assume it to be straight.

But at the same time other possible kinds of correlation must be kept in mind in which the regression lines are not straight, and can be shown to be curved by the examination of large numbers of cases.

W. S. GOSSET.