actions are not neglected, but are well brought out in the teaching.

The value of such centres as these is hardly calculable. They serve to show to the general public the importance that the Education Authorities attach to the training of girls in domestic subjects; they act as models in the districts surrounding them, and they point the moral that good housekeeping, like most other things, needs to be learned to be done well, and that it does not come by instinct, but that when thoroughly understood and carried out in a proper manner it is a work full of interest and enjoyment.

THE OPEN WINDOW.—Thousands of visitors flock annually to this borough, eager to exchange the murky atmosphere of the industrial towns for the pure invigorating air of Scarborough. But it may be seriously questioned if the residents of Scarborough adequately appreciate the advantages they enjoy in this respect. Judged by the comparative infrequency of open windows, especially during the colder months, it would appear that the primary function of a window (as the name implies), the admission of air, had been lost sight of, and that the sole use of windows was the admission of light. It is only fair to state that some of the girls' boarding schools give the inhabitants at large a valuable daily object lesson regarding the double function which a window subserves. No doubt this common neglect of the cult of the open window is due to an abnormal fear of draughts, but it may be confidently stated that the atmosphere of the ordinary dwelling house cannot be maintained at a proper standard of purity unless one or more windows be kept constantly open. A draught is simply a current of air moving so rapidly as to make its presence disagreeably evident. Ventilation aims at the adequate renewal of air without perceptible draught, and this can be readily accomplished in the ordinary dwelling-house by judiciously opening windows. Even in the smallest houses all the rooms are not occupied day and night, and the simplest plan of securing proper ventilation is to open the windows of the unoccupied rooms. Thus during the day all the bedroom windows should remain open, or on stormy days those on the lee side of the house. The force of the entering air is reduced by friction at the angles formed by walls and doors, while the temperature is gradually raised by the indoor warmth. By the time the incoming air reaches the occupied living rooms the draught is no longer perceptible, and a sufficient supply of pure warmed air is secured. It cannot be too strongly stated that there is no surer way of contracting ill-health, increasing susceptibility to colds, and greatly enhancing the risk of consumption, than by habitually living in a vitiated atmosphere. The remedy is simple, and that is probably the reason why it is so widely neglected.—*Annual Report of Dr. John Knight, Medical Officer of Health, Scarborough.*

VARIATIONS IN THE CHEMICAL AND BACTERIOLOGICAL COMPOSITION OF WATERS CONSIDERED FROM A STATISTICAL POINT OF VIEW.

By R. J. EWART, M.D., M.Sc., F.R.G.S., D.P.H.

IN a series of analyses, both into chemical composition and bacterial fauna of any water, variations are seen which at times may border on, or even exceed, such limits as are considered hygienically safe. Change within certain limits must of necessity occur quite irrespective of the addition of harmful matter, even in those constituents which are used as indicators of pollution. It is quite obvious that a great advance in the interpretation of a water report would be achieved if we could express the possible normal change of composition by some factor. In the present paper I venture to suggest certain methods by means of which the stability of any individual water with respect to any particular constituent may be expressed by a figure, which when considered along with the mean for any series of estimations will enable the formation of an opinion which would scarcely be possible with the present methods.

If a series of analyses of any inorganic constituent of a water are placed in order of magnitude, irrespective of the sequence of examination, and then plotted graphically, it will be seen that a curve is evolved, the smoothness of which depends on the frequency of the observations—the differences at the extremes being large and in the middle much smaller. Such an arrangement suggests that any series of analyses might be treated in the same manner, and obey the same mathematical laws, as any other series of figures selected at random.

If such estimations when arranged in this way produced some form of understandable curve then a factor is available which is eminently suitable to express their variability—namely, the standard deviation. Such is much more valuable than a mere empirical figure, as from it certain information can be obtained if the nature of the curve can be expressed by some known formula. The curves formed in only a few cases could be described as normal; in most they were markedly askew—that is to say, the distribution about the mean was not equal. The “skewness” and ambiguity of the curves rob this standard of much of its mathematical significance, and in the sense in which I have used it it can only be regarded in its

empirical meaning. To express the evenness of distribution about the mean perhaps the two quartiles would have been more useful, and their excess or defect from the mean expressed as some form of fraction.

The absolute quantity or any individual constituent is, of course, of paramount importance, but if compared with a second supply from a different source, beyond certain wide limits little can be learnt. The variability, however, in each case may enable us to classify the supply more accurately. For example, a variability of a certain magnitude which in a well water might be significant in the case of a surface water might be neglected.

In the following calculations the formulæ used are taken from Davenport's "Statistical Methods."

The average or mean "A" may be described as the abscissa of the centre of gravity of the frequency curve, and is obtained by:—

$$A = \frac{\sum (V f)}{n}$$

V = magnitude of any class; *f* its frequency.

Σ = sum of the products for all classes into which frequency is to be got, and *n* is the number of analyses made.

The index of the variability of the variates when they group themselves in some form of frequency curve is found by adding the squares of the differences between the mean and the class, multiplied by its frequency, and dividing the total by the number of observations made and extracting the square root of the quotient, thus:—

$$\sigma = \sqrt{\frac{\sum (x^2 f)}{n}} \times \lambda$$

x = the deviation of the class from the mean.

λ = number of units in the class range.

To apply the index of variability in a mathematical sense to any series of variates it is essential, as has been stated, that the values when grouped about the mean should fall into some form of frequency polygon. It was found that the data, when grouped into classes in such a way that no obvious blank was left and the class range was large enough to remove extreme roughness, and were sufficient in number, that curves were obtained whose development could be understood; *e.g.*, there were 71 observations on the amount of chlorine in a water made at stated intervals during the year, and these were grouped, and the frequency

with which certain values were found was as follows:—

PARTS PER 100,000.		
1.40 to 1.49	1.50 to 1.59	1.60 to 1.69
6	21	26
1.70 to 1.79	1.80 to 1.89	
17	1	

It is seen that these figures practically coincide with the normal frequency curve, which is symmetrical: the mode, median and mean being equal. The actual mean is not 1.65 (the middle value of the class 1.60 to 1.69), but to the left, and agrees with the average. This type of frequency curve, which can be understood without any elaborate mathematical knowledge, covers some of the curves found for the saline constituents of a water, though not all, as in some cases a second mode makes its appearance (and in others there is a marked skewness), and greatly complicates the development. The solution of such a curve belongs to the sphere of higher mathematics.

The following is the equation to a normal curve:—

$$Y = \frac{n}{\sigma \sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

This gives the value of any ordinate *Y* at any distance *x* along the abscissa from the mode. *e* is a constant 2.71828, the base of the Naperian system of Logarithm. *n* is the total area of the curve or number of variates, and σ is the standard deviation, which is a constant for any curve, and measures its variability.

The organic constituents, though sometimes giving a normal curve, from the graphical appearance, often resemble Type III (Prof. Karl Pearson).

It may be assumed that simple observations, if made often enough throughout the year, will arrange themselves with certain frequencies which can be plotted in some form of curve, and that the steepness of such a curve is indicated by the standard deviation, a figure expressed in the same units as those in which its measurements have been made.

Prof. Karl Pearson has suggested the use of what he terms "the co-efficient of variability" found by multiplying the standard deviation by 100 and dividing by the mean; this gives a figure which is useful for comparative purposes, but is fictitious in cases where only small quantities of the variates are found, with absolutely small but relatively great differences. Thus in the case of saline ammonia a mean of .0002 with a

standard deviation of '0002 would give a coefficient of 100, whilst the same deviation with a mean of '001 would give a coefficient of 20, though in the latter case the significance of the deviation might be as great as in the former.

As we are considering variability only, the standard deviation is used as a measure of the characteristic.

It is obvious that the measure of relationship, namely, the correlation factor between sequence of observation, time and variability, or between one constituent and any other, is the natural step to the work that has been already done.

It is found that the correlation between saline and albuminoid ammonia for a raw river water is 0'33, with a probable error of approximately 0'1, hence a significant relationship exists. On filtration and storage the correlation increases by about one-tenth the value already given. Considerations on these lines would take us much too far for our present purpose; sufficient be it to show that a field is open for much inquiry.

Series of analyses were collected from different sources and standard deviations calculated, the nature of the sources being as follows:—

- I. *Natural raw river water* with (a) much pollution; (b) with some pollution, as judged from the collecting area.
- II. *Well water* from an unimpeachable source, and
- III. From a large, artificial lake, supplying a large town.
- IV. *Stored waters* from (a) polluted, and (b) slightly polluted sources.
- V. *Filtered water* from (a) polluted, and (b) slightly polluted sources.
- VI. *Two filtered waters*, (a) surface and (b) from a slightly polluted river with a possibility of tidal influence:—

So far as the data allowed, the deviation was estimated for the following four constituents: Saline and albuminoid ammonia, chlorides and nitrates in the chemical groups, and the number of organisms growing on agar and gelatine, presumptive *Bacillus Coli*, and confirmed in the bacterial groups.

Great difficulty was experienced in finding a standard deviation for the bacteriological counts, both, as has already been said, on account of the extreme variability of the enumerations and also on account of the different criteria used by various observers, especially with regard to the *Bacillus Coli*.

The actual sources of data upon which the following are based cannot be named for many reasons. I may say, however, that the standards are taken from Dr. A. C. Houston's detailed Report on the Water Supply of the Metropolitan Board.

Assuming, then, that the frequency curve is normal, limiting values can easily be found. The maximum and minimum observed for saline ammonia in No. I were '0141 and '0030; theoretically the range would approximately be from '005 to '0130. Thus it is seen that the actual observations probably missed the minimum value and slightly exceeded the maximum. It is obvious, then, that by a number of random estimations any particular one can be found. A method by means of which we can approximate to the largest estimate of any constituent if experiment has missed that object, and whether the greater value found is really the maximum or not, is of extreme value.

From the following table it is seen that the standard deviation is a numerical number, which, though having a relationship to the mean, and accentuating its significance, cannot be considered separate from it:—

STANDARD DEVIATION OF CHEMICAL CONSTITUENTS.

SOURCE.	Saline Ammonia.		Albuminoid Ammonia.		Nitrates.		Chlorides.	
	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
1. A polluted river water used for drinking supply	'0065	'0036	'0132	'0029	'260	'034	1'72	'120
2. Another river about the same as No. 1...	'0107	'0041	'0140	'0037	'290	'064	2'06	'146
3. A river free from sewage contamination	'0027	'0012	'0072	'0023	'280	'042	1'59	'075
4. A large lake (artificial) used for town supply (uncontaminated)	'0001	'0001	'0006	'0002	'000	'000	0'87	'044
5. Well water of unimpeachable source ...	'0002	'0005	'0013	'0001	'550	'021	2'38	'059

From the analyses, and also from general observations, Water No. 2 is the least desirable, with No. 1 next in order. Neither of these would be safe for drinking purposes. No. 3 is doubtful, whilst Nos. 4 and 5 would be considered quite wholesome.

The standard deviation may be regarded from two points of view: (1) its absolute value, and (2) its relationship to the mean. In the four examples given the absolute value of the factor varies synchronously with the mean. The relative value of saline and albuminoid ammonia varies a little; it is synchronous except when the estimates approach zero and the experimental error becomes large in comparison with the values observed.

teristic is influenced. No. 1 is from the River Thames, and No. 2 is a river water of fair purity.

We observe at once how filtration steadies, and at the same time slightly modifies, the ratios of some of its constituents. The relationship of the standard deviation to the mean appears to be somewhat variable, and depends on too many factors for accurate analysis. It is worthy of note that the deviation of the nitrates is increased. It may be accounted for on the ground that varying quantities of ammonia appear in the filtrate as oxidised nitrogen. Sand filtration is probably more than a mere mechanical process, the bacterial film on the surface of the bed has a

					Saline Ammonia.		Albuminoid Ammonia.		Nitrates.		Chlorides.	
					Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
1. RAW RIVER WATER.												
Unfiltered	...	...	...	...	'0065	'0036	'0132	'0029	'26	'034	1'72	0'12
Filtered, 1st Series	...	...	...	...	'0001	'00007	'0054	'0011	'25	'045	1'65	0'13
Filtered, 2nd Series	...	...	...	...	'0002	'00009	'0051	'0006	'21	'042	1'60	0'08
2. RIVER WATER.												
Unfiltered	...	...	...	...	'0027	'0012	'0072	'0023	'28	'042	1'59	0'07
Filtered	...	...	...	...	'0003	'0002	'0056	'0007	'22	'078	2'08	0'115

In the case of the chlorides and nitrates this factor has some relative as well as an absolute significance, and it is here, perhaps, that Prof. Karl Pearson's co-efficient of variability could be usefully employed.

THE INFLUENCE OF FILTRATION AND STORAGE.

The influence of these two methods of purification of water on the stability is naturally of considerable interest. Various series of analyses were, with this object, collected from different sources (mainly from the Metropolitan Water Board's report) to ascertain how this charac-

potent influence on the filtrate, and the relationship of the nitrates and ammonias might be used as a chemical indicator of this vital action. The chlorides show a variation which is not modified much by the process the water has undergone, but differences in the standard are seen which will be referred to in another connection.

Simple storage for a short period (approximately one month) might be expected to trend to steady the composition of any water, in so far as the supply over a period of time becomes more or less mixed.

					Saline Ammonia.		Albuminoid Ammonia.		Nitrates.		Chlorides.	
					Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
RAW RIVER WATER.												
Before storage	...	...	...	...	'0066	'0038	'0153	'0063	'26	'062	1'67	'102
After storage	...	...	...	...	'0046	'0025	'0172	'0066	'23	'055	1'62	'086

The foregoing table, however, shows that very little change occurs in constituents under consideration either as regards the absolute amount present or as regards their variation.

The two following series are of interest:—

	Saline Ammonia.		Albuminoid Ammonia.		Nitrates.		Chlorides.	
	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
1. Filtered river water	·0007	·0002	·0073	·0014	...	...	2·3	·5634
2. Mixed surface water (filtered)	·0022	·0007	·0167	·0034	·04	·028	0·84	·2789

Both these analyses show rather high deviations arising from different causes. This is particularly well marked in the case of chlorides in No. 1. On this river a number of small steamers run from a portion of the river which is occasionally tidal to the highest reaches where the intake is situated, and no doubt the seaweed on the boats carries a certain amount of chlorides into the upper parts of the river. No. 2 is a water arising partly from surface drainage and from a fairly polluted river after filtration; the mixing is not very thorough, hence a large deviation is observed.

We may conclude that so far as the chemical constituents of a water are concerned we have in the "standard deviation from the mean" a very accurate and reliable measure of the stability of a water. The criteria should, of course, be to make this factor as small as possible. This object is achieved by (1) long storage and (2) filtration. Storage over short periods of time does not appear to influence this co-efficient very materially.

BACTERIOLOGICAL.

As already remarked, there is some doubt as to whether we are justified in applying the same formulæ to the bacteriological fauna as that used for the chemical constituents.

There can be no doubt that the bacteriological balance depends on certain factors, but they appear to be so numerous and conflicting in influence that their effects can only be guessed in a general way. It appears to be established that the content is greatest during the winter, rather more in the early months of the year than in the late autumn. The influences at work during the winter months are conflicting. A low temperature hinders bacteriological growth, while at the same time it favours surface drainage. A high temperature, even when wet, owing to the enormous evaporation both from the ground and vegetation, lessens the surface drainage so that what water gets

into the streams is to a great extent filtered. On the other hand, bacteriological growth is increased. The natural conditions affecting this aspect of a water are most complicated and too intricate for complete analysis, and it is not surprising then that the findings should be so bizarre. Each count has its meaning and should be judged on its own merits, depending on the conditions extant at the time. Bacteriological examination is essentially suitable for individual samples, whilst the chemical examination lends itself particularly well to grouping and application of frequency curves.

One criteria of statistical data is not fulfilled by a series of bacteriological findings, namely, that laws of chance and probability scarcely apply, as there is not the necessary element of selection at random; still, the conditions are so conflicting that although a strict application is not justifiable, results may be obtained which may be hints as to the lines on which future inquiries should be made. A difficulty arises in the different standards various observers take as to what is *Bacillus Coli* and what is not, and, further, the different methods that have been adopted to state results. In the present paper the criteria is the "Flaginac" one, laid down by Dr. A. C. Houston, and as far as possible the results are stated as a percentage of samples per month, showing such organism in 1·0 c.c. or less. The same sources of information were used as in the case of the chemical data—the standard counts being from the Metropolitan Water Board's reports.

Although the application of a standard deviation may not be justifiable on theoretical grounds, practically it does appear to give some information. Further, the differences are so large that a relative as well as an absolute value can be assigned to it.

In Nos. 1 and 2 the deviation is greater than the mean; in Nos. 3 and 4 it is less. No. 2 would be considered the most dangerous of the four series.

	No. of Organisms in 1 c.c. growing on Gelatine at 20% C.		No. of Organisms in 1 c.c. growing on Agar at 37% C.		No. of Samples in which B. Coli was found in 1 c.c. or less.	
	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
1. Raw river water (polluted) Series A	2740	2705	319	140	87.5	10.1
2. Raw river water (polluted) Series B	8444	8714.8	521	310	93.9	4.1
3. Raw river water (slightly polluted)	959	590.1	93	25.8	47.8	16.55
4. Well water (good source) ...	5.9	2.26	...	...	1.7	0.001?

Turning now to the *Bacillus Coli* we see the reverse state of affairs regarding relative variability. In No. 2 we have the smallest absolute and smallest relative value of the first three of the table. This series represents a rather curious condition of balance. The gross counts are large and extremely variable; the *Bacillus Coli* content is large and constant. It appears as if the pollution was so heavy that attempts at self-purification are in abeyance. This characteristic increases the weight of evidence against its use for household purposes, and this point exemplifies the fact that consideration apart from the mean would be of little consequence.

If we substitute the co-efficient of variability for the standard deviation the table would read as follows (the mean must be considered along with this co-efficient) :—

	No. of Organisms in 1 c.c. growing on Gelatine at 20% C.		No. of Organisms in 1 c.c. growing on Agar at 37% C.		Percentage of Samples in which B. Coli was found in 1 c.c. or less.	
	Mean.	Co-efficient of Variability.	Mean.	Co-efficient of Variability.	Mean.	Co-efficient of Variability.
1. Raw river water (polluted) Series A	2740	91	319	22	87.5	8
2. Raw river water (polluted) Series B	8444	103	521	17	93.9	4
3. Raw river water (slightly polluted)	959	61	93	36	47.8	4
4. Well water (good source) ...	5.9	39	...	...	1.7	Less than 1.

THE INFLUENCE OF FILTRATION.

Filtration has somewhat the same influence as that already shown with reference to the chemical analyses. The following table gives an idea of this influence. The co-efficient of variability is given as well as the standard deviation :—

	No. of Organisms in 1 c.c. growing on Gelatine at 20% C.			Percentage of Samples in which B. Coli was found in 1 c.c. or less.		
	Mean.	Standard Deviation.	Co-efficient of Variability.	Mean.	Standard Deviation.	Co-efficient of Variability.
Raw river water	2740	2705	91	87.5	10.1	8
Series A (filtered)	202	194.5	96	0.3	.001	...
Series B (filtered)	90	151.25	161	0.2	.002	...
Another river water	8444	8714.8	103	93.9	4.1	4
Filtered	162	130.3	86	0.7	.002	...

The co-efficient of variation as a mere figure does not appear to have much relationship to filtration, as in the first series it is increased and in the second decreased. The standard deviation for *Bacillus Coli* is also enormously reduced, the reduction being in a

greater ratio to the mean than was observed for the total number of organisms present. It appears as if a well-managed filter bed had a selective action for *Bacillus Coli* and allied organisms.

This apparent selective action on the part

of the filter bed is probably more apparent than real, and arises in this way: Before the water passes on to the bed a certain bacterial balance between one organism and another is established. After passing through the filter bacteria and bits of debris on which colonies grow are removed, so that in the filtrate the same ratio is still present, or apparently so.

This can be shown by the analyses of samples

hence, although the gross aggregate falls, still there are times when they are able to hold their own against the more normal water organisms. This deduction, if correct, is interesting, for it indicates the desirability of a margin of safety even in stored waters. There can, however, be no doubt of the beneficial effect of storage in establishing a balance more akin to the normal.

	No. of Organisms in 1 c.c. growing on Gelatine.		No. of Organisms in 1 c.c. growing on Agar.		No. of Organisms in 1 c.c. growing in Bile-salt Agar.	
	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
Raw river water (before storage) ...	4465	10719	279	250	41	43
Raw river water (after storage) ...	362	800	52	100	7	34

of water immediately on leaving the filter-beds. It follows, then, that the increase in organismal content which is known to occur after filtration will be greatest amongst those organisms that are in their more natural environment, and less where the circumstances are untoward. Practically the whole of the organismal increase occurring after filtration is in the normal water organisms, so that the apparent effect of filtration, if the filtrate is examined some few days after leaving the beds, may give the impression that the beds had picked out the *Bacillus Coli* group of organisms only.

The comparison between the chemical composition before and after storage showed little change. In the bacteriological analyses we

The following are two interesting series of analyses which correspond to those already described as a surface water and a river water with possible tidal influence. Both are filtered.

Both these waters from their bacteriological findings are such as require careful supervision. In the latter, although the total number of organisms is low and not subject to much variation, still a large proportion of the samples contains the *Coli* group in fair numbers. The former is a peaty water—most moorland waters are of this type—and although it contains a large number of organisms of the *Bacillus Coli* group, less than 15 per cent. of these obey the confirmatory tests. They usually gave all reactions except the indol. It is an appreciated

	No. of Organisms in 1 c.c. growing on Gelatine.		No. of Organisms in 1 c.c. growing on Agar.		P.c. of Samples showing No. of <i>B. Coli</i> Organisms in 1 c.c. or less.	
	Mean.	Standard Deviation.	Mean.	Standard Deviation.	Mean.	Standard Deviation.
Surface water (filtered) ...	251	234	23	10	46	...
River water (possibly tidal) ...	16	12	...	...	16	...

have a very pronounced alteration in the balance between the various types of organismal life present. In all cases both the number and standard deviation is reduced to about one-tenth of its previous magnitude.

It is observed that although the mean for coliform organisms is decreased, the variability about the mean remains much the same. It would almost appear from this observation that the detrimental influence on the life of this group in stored water is not always the same, at some periods being greater than at others;

fact that *Bacillus Typhosis* has a short life (a few days only) in a peaty water, and the true *Bacillus Coli* itself lives a relatively short period of time. The aberrant forms may possibly live longer. This fact may account for the excursive percentage of typical forms of *Bacillus Coli* found in a surface water when fair pollution is possible.

DR. A. GIBB GLASS, Assistant Medical Officer of Health of Burnley, has been appointed Schools Medical Officer of the Counties of Selkirkshire and Roxburghshire.