On the Error of Counting with a Haemacytometer
Author(s): Student
Source: *Biometrika*, Vol. 5, No. 3 (Feb., 1907), pp. 351-360
Published by: Biometrika Trust
Stable URL: http://www.jstor.org/stable/2331633
Accessed: 03/04/2013 02:52

ON THE ERROR OF COUNTING WITH A HAEMACYTOMETER.

By Student.

When counting yeast cells or blood corpuscles with a hæmacytometer there are two main sources of error: (1) the drop taken may not be representative of the bulk of the liquid; (2) the distribution of the cells or corpuscles over the area which is examined is never absolutely uniform, so that there is an "error of random sampling."

With the first source of error we are concerned only to this extent; that when the probable error of random sampling is known we can tell whether the various drops taken show significant differences. What follows is concerned with the distribution of particles throughout a liquid, as shewn by spreading it in a thin layer over a measured surface and counting the particles per unit area.

Theoretical Consideration.

Suppose the *whole* liquid to have been well mixed and spread out in a thin layer over N units of area (in the hæmacytometer the usual thickness is ·01 mm. and the unit of area $\frac{1}{400}$ sq. mm.).

Let the particles subside and let there be on an average m particles per unit area, that is Nm altogether. Then assuming the liquid has been properly mixed a given particle will have an equal chance of falling on any unit area.

i.e. the chance of its falling in a given unit area is $1/N$ and of its not doing so $1-1/N$.

Consequently considering all the mN particles the chances of 0, 1, 2, 3 ... particles falling on a given area are given by the terms of the binomial $\left\{\left(1-\frac{1}{N}\right)+\frac{1}{N}\right\}^{mN}$, and if M unit areas be considered the distribution of unit areas containing 0, 1, 2, 3 ... particles is given by $M\left\{\left(1-\frac{1}{N}\right)+\frac{1}{N}\right\}^{mN}$.

Now in practice N is to be measured in millions and may be taken as infinite.

Let us find the limit when N is infinite of the general term of this expansion. The $(r+1)$th term is:

$$\left(1-\frac{1}{N}\right)^{mN-r}\cdot\left(\frac{1}{N}\right)^r\frac{mN(mN-1)(mN-2)\dots(mN-r+1)}{r!}$$

$$=\left(1-\frac{1}{N}\right)^{mN-r}\frac{m\left(m-\frac{1}{N}\right)\left(m-\frac{2}{N}\right)\dots\left(m-\frac{r-1}{N}\right)}{r!}$$

$$=\left(1-\frac{mN-r}{N}+\frac{(mN-r)(mN-r-1)}{N^2\,.\,2!}-\dots\right.$$

$$\left.+(-1)^s\frac{(mN-r)\dots(mN-r-s+1)}{N^s\,.\,s!}+\dots\right)$$

$$\times\, m\frac{\left(m-\frac{1}{N}\right)\left(m-\frac{2}{N}\right)\dots\left(m-\frac{r-1}{N}\right)}{r!}.$$

But when we proceed to the limit $\frac{1}{N}, \frac{2}{N}\dots\frac{r-1}{N}$ and $\frac{r}{N}, \frac{r+1}{N}\dots\frac{r+s-1}{N}$ are all negligeably small compared to m so that the expression reduces to

$$\left(1-m+\frac{m^2}{2!}-\dots+(-1)^s\frac{m^s}{s!}\dots\right)\times\frac{m^r}{r!}=e^{-m}\times\frac{m^r}{r!}.$$

That is to say that the expansion is equal to

$$e^{-m}\left\{1+m+\frac{m^2}{2!}+\dots+\frac{m^r}{r!}+\dots\right\}.$$

Hence it is this distribution with which we are concerned.

The 1st moment about the origin, O, taken at zero number of particles is

$$e^{-m}\left\{m+\frac{2m^2}{2!}+\frac{3m^3}{3!}+\dots+\frac{rm^r}{r!}+\dots\right\}$$

$$=me^{-m}\left\{1+\frac{m}{1!}+\frac{m^2}{2!}+\dots+\frac{m^{r-1}}{(r-1)!}+\dots\right\}$$

$$=m\times\text{total frequency.}$$

Hence the mean is at m.

The 2nd moment about the point O is

$$e^{-m}\left\{m+\frac{2^2m^2}{2!}+\frac{3^2m^3}{3!}+\dots+\frac{r^2m^r}{r!}+\dots\right\}$$

$$=e^{-m}\left\{m+\frac{2m^2}{1!}+\frac{3m^3}{2!}+\dots+\frac{rm^r}{(r-1)!}+\dots\right\}$$

$$=e^{-m}\left\{m+\frac{m^2}{1!}+\dots+\frac{m^r}{(r-1)!}+\dots+m^2+\frac{2m^3}{2!}+\dots+\frac{(r-1)\,m^r}{(r-1)!}+\dots\right\}$$

$$=(m+m^2)\times\text{total frequency.}$$

Hence the second moment-coefficient about the mean

$$\mu_2 = m + m^2 - m^2 = m.$$

By similar* methods the moment-coefficients up to μ_6 were obtained, as follows:

$$\mu_1' = m.$$
$$\mu_2 = m.$$
$$\mu_3 = m.$$
$$\mu_4 = 3m^2 + m.$$
$$\mu_5 = 10m^2 + m.$$
$$\mu_6 = 15m^3 + 25m^2 + m.$$

Hence $$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{1}{m},$$

and $$\beta_2 = \frac{\mu_4}{\mu_2^2} = 3 + \frac{1}{m}.$$

It will be observed that the limit to which this distribution approaches as m becomes infinite is the normal curve with its β_1, β_3, β_5, etc., all equal to 0, and $\beta_2 = 3$, $\beta_4 = 15$, etc.

Further, any binomial $(p + q)^n$ can be put into the form $(p + q)^{nq/q}$, and if q be small and nq not large it approaches the distribution just given.

Thus if 1000 $(\frac{99}{100} + \frac{1}{100})^{500}$ be expanded the greatest difference between any of its terms and the corresponding term of 1000 $e^{-5}\left(1 + 5 + \frac{5^2}{2!} + \dots + \frac{5^r}{r!} + \dots\right)$

* The evaluation of the moments about the point O will be found to depend on the expansion of r^n in the form

$$r^n = r\left\{\frac{(r-1)!}{(r-n-2)!} + a_1\frac{(r-1)!}{(r-n-1)!} + a_2\frac{(r-1)!}{(r-n)!} + \dots + a_{n+1}\frac{(r-1)!}{(r-1)!}\right\}$$
$$= r\left\{\frac{1}{(r-n-2)!} + \frac{a_1}{(r-n-1)!} + \frac{a_2}{(r-n)!} + \dots + \frac{a_{n+1}}{(r-1)!}\right\}(r-1)!$$

Then if we form the series for $n+1$ from this it will be found that the following relations hold between a_1, a_2, a_3 etc. and the corresponding coefficients for $n+1$, A_1, A_2, A_3 etc.

$$A_1 = a_1 + n,$$
$$A_2 = a_2 + (n-1)\,a_1,$$
$$A_p = a_p + (n-p+1)\,a_{p-1}.$$

From these equations we can write down any number of moments about the point O in turn, and from these may be found the moments about the mean by the ordinary formulae.

The moments may also be deduced from the point binomial $(p+q)^{nq/q}$ when q is small and n large and $nq=m$, i.e. $p=1$, $q=0$, $nq=m$. We have

$$\mu_1' = nq = m,$$
$$\mu_2 = npq = m,$$
$$\mu_3 = npq\,(p-q) = m,$$
$$\mu_4 = npq\,\{1 + 3\,(n-2)\,pq\} = m\,(1+3m) = 3m^2 + m.$$

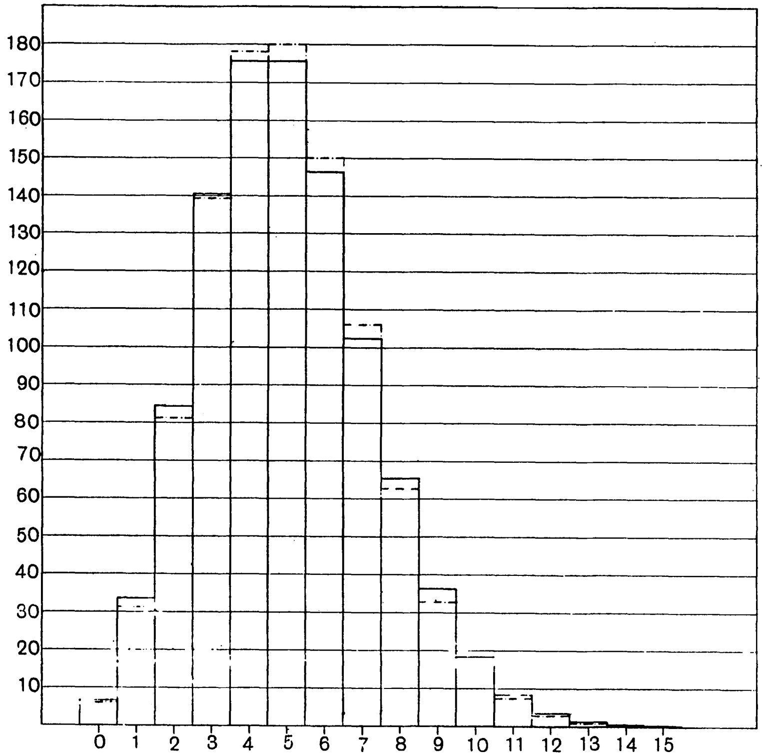
is never as much as 1, being about ·8 for the term 1000 $e^{-5}\frac{5^5}{5!}$ which is 175·5 against 176·3 from the binomial.

Diagram I compares 1000 $e^{-5}\left(1+5+\frac{5^2}{2!}+\ldots+\frac{5^r}{r!}+\ldots\right)$ with the binomial 1000 $(\frac{19}{20}+\frac{1}{20})^{100}$ which of course differ, but not by very much.

DIAGRAM I. Comparison of the exponential and binomial expansions.

Firm line represents $1000e^{-5}\left\{1+5+\ldots+\frac{5^r}{r!}+\text{etc.}\right\}$.

Broken line represents $1000\left\{\frac{19}{20}+\frac{1}{20}\right\}^{100}$.



In applying this to actual cases it must be noted that we have not taken into account any "interference" between the particles; there has been supposed the same chance of a particle falling on an area which already has several particles as on one altogether unoccupied. Clearly if m be large this will not be the case, but with the dilutions usually employed this is not of any importance.

It will be shewn that the actual distributions which were tested do not diverge widely from this law, so we will consider the probable error of random sampling on the supposition that they follow it.

We have seen that $\mu_2 = m$.

Hence the standard deviation $= \sqrt{m}$.

So that if we have counted M unit areas the probable error of our mean (m) is $\cdot 67449\sqrt{\frac{m}{M}}$.

If we are working with a hæmacytometer in which the volume over each square is $\frac{1}{4000000}$ mm. there will be 40,000,000 m particles per c.c. and the probable error will be $40{,}000{,}000 \times \cdot 67449 \times \sqrt{\frac{m}{M}}$.

Suppose now that we dilute the liquid to q times its bulk, we shall then have $\frac{m}{q}$ particles per square, and if we count M squares as before, our probable error for the number of particles per c.c. in the original solution will be 40,000,000 $\times \cdot 67449 \times q\sqrt{\frac{m}{q} \times \frac{1}{M}}$. That is $40{,}000{,}000 \times \cdot 67449\sqrt{\frac{mq}{M}}$.

That is we shall have to count qM squares in order to be as accurate as before.

So that the same accuracy is obtained by counting the same number of particles whatever the dilution, or, to look at it from a slightly different point of view, whatever be the size of the unit of area adopted.

Hence the most accurate way is to dilute the solution to the point at which the particles may be counted most rapidly, and to count as many as time permits: then the probable error of the mean is $\cdot 67449\sqrt{\frac{m}{M}}$ where m is the mean and M is the number of unit areas counted over, squares, columns of squares, microscope fields, or whatever unit be selected.

But owing to the difficulty of obtaining a drop representative of the bulk of the liquid the larger errors will probably be due to this cause, and it is usual to take several drops: if two of these differ in their means by a significant amount compared with the probable error (which is $\cdot 67449\sqrt{\frac{m_1 + m_2}{M}}$ where m_1, m_2 are the means and M the number of unit areas counted), it is probable that one at least of the drops does not represent the bulk of the solution.

Experimental Work.

This theoretical work was tested on four distributions * which had been counted over the whole 400 squares of the hæmacytometer. The particles counted were yeast cells which were killed by adding a little mercuric chloride to the water in which they had been shaken up. A small quantity of this was mixed with a 10 °/ₒ solution of gelatine, and after being well stirred up drops were put on the hæmacytometer. This was then put on a plate of glass kept at a temperature just above the setting point of gelatine and allowed to cool slowly till the gelatine had set. Four different concentrations were used.

* One of these is given in Table I.

In this way it was possible to count at leisure without fear of the cells straying from one square to another owing to accidental vibrations. A few cells stuck here and there to the cover glass, but as they appeared to be fairly uniformly distributed and were very few compared with those that sank to the bottom they were neglected: had the object of the experiment been to find the number of cells present they would have been counted by microscope fields, and correction made for them; but in our case they were considered to belong to a different "population" to those which sank.

Those cells which touched the bottom and right-hand lines of a square were considered to belong to the square; a convention of this kind is necessary as the cells have a tendency to settle on the lines.

There was some difficulty owing to the buds of some cells remaining undetached in spite of much shaking. In such cases an obvious bud was not counted, but sometimes, no doubt, a bud was counted as a separate cell, which slightly increases the number of squares with large numbers in them.

In order to test whether there was any local lack of homogeneity the correlation was determined between the number of cells on a square and the number of cells on each of the four squares nearest it; if from any cause there had been a tendency to lie closer together in some parts than in others this correlation would have been significantly positive.

Distributions 3 and 4 were tested in this way (Table II), with the result that the correlation coefficients were $+\cdot016 \pm \cdot037$ and $\cdot015 \pm \cdot037$. This is satisfactory as shewing that there is no very great difficulty in putting the drop on to the slide so as to be able to count at any point and in any order; as good a result may be expected from counting a column as from counting the same number of squares at random.

The actual distributions of cells are given below, and compared with those calculated on the supposition that they are random samples from a population following the law which we have investigated: the probability P of a worse fit occurring by chance is then found.

I. Mean $=\cdot6825$: $\mu_2=\cdot8117$: $\mu_3=1\cdot0876$.

Containing	0	1	2	3	4	5 cells
Actual	213	128	37	18	3	1
Calculated	202	138	47	11	1·84	·24
					2 (4 and 5 together)	

Whence $\chi^2=9\cdot92$ and $P=\cdot04$.

Best fitting binomial $(1\cdot1893-\cdot1893)^{-3\cdot6054}\times400$ for which $P=\cdot52$.

II. Mean $=1\cdot3225$: $\mu_2=1\cdot2835$ μ_3 : $=1\cdot3574$.

	0	1	2	3	4	5	6
Actual	103	143	98	42	8	4	2
Calculated	106	141	93	41	14	4	1

Whence $\chi^2=3\cdot98$ and $P=\cdot68$.

Best fitting binomial $(\cdot97051+\cdot02949)^{46\cdot2084}\times400$ for which $P=\cdot72$.

III. Mean $=1\cdot80$: $\mu_2=1\cdot96$: $\mu_3=2\cdot529$.

	0	1	2	3	4	5	6	7	8	9
Actual	75	103	121	54	30	13	2	1	0	1
Calculated	66	119	107	64	29	10	3		1	

Whence $\chi^2=9\cdot03$ and $P=\cdot25$.

Best fitting binomial $(1\cdot0889-\cdot0889)^{-20\cdot2473}\times 400$ for which $P=\cdot37$.

IV. Mean $=4\cdot68$: $\mu_2=4\cdot46$: $\mu_3=4\cdot98$.

	0	1	2	3	4	5	6	7	8	9	10	11	12
Actual	0	20	43	53	86	70	54	37	18	10	5	2	2
Calculated	4	17	41	63	74	70	54	36	21	11	5	2	1

Whence $\chi^2=9\cdot72$ and $P=\cdot64$.

Best fitting binomial $(\cdot9525+\cdot0475)^{98\cdot53}\times 400$ for which $P=\cdot68$.

These results are given graphically in Diagram II. on the next page.

It is possible to fit a point binomial from the mean and the 2nd moment according to the two equations $\mu_1=nq$, $\mu_2=npq$ and these point binomials fit the observations better than the exponential series, but the constants have no physical meaning except that $nq=m$. And since the exponential series is a particular form of the point binomial and is fitted from one constant, while two are used for the "ad hoc" binomial, this better fit was only to be expected.

It will be noticed that in both I and III the 2nd moment is greater than the mean, due to an excess over the calculated among the high numbers in the tail of the distribution. As was pointed out before, the budding of the yeast cell increases these high numbers, and there is also probably a tendency to stick together in groups which was not altogether abolished even by vigorous shaking.

In any case, the probabilities ·04, ·68, ·25 and ·64, though not particularly high, are not at all unlikely in four trials, supposing our theoretical law to hold, and we are not likely to be very far wrong in assuming it to do so.

Let us now apply it to a practical problem: for some purposes it is customary to estimate the concentration of cells and then dilute so that each two drops of the liquid contain on an average one cell. Different flasks are then seeded with one drop of the liquid in each, and then "most of those flasks which show growths are pure cultures."

The exact distribution is given by

$$e^{-\frac{1}{2}}\left(1+\frac{1}{2}+\frac{(\frac{1}{2})^2}{2\,!}+\frac{(\frac{1}{2})^3}{3\,!}+\dots\right),$$

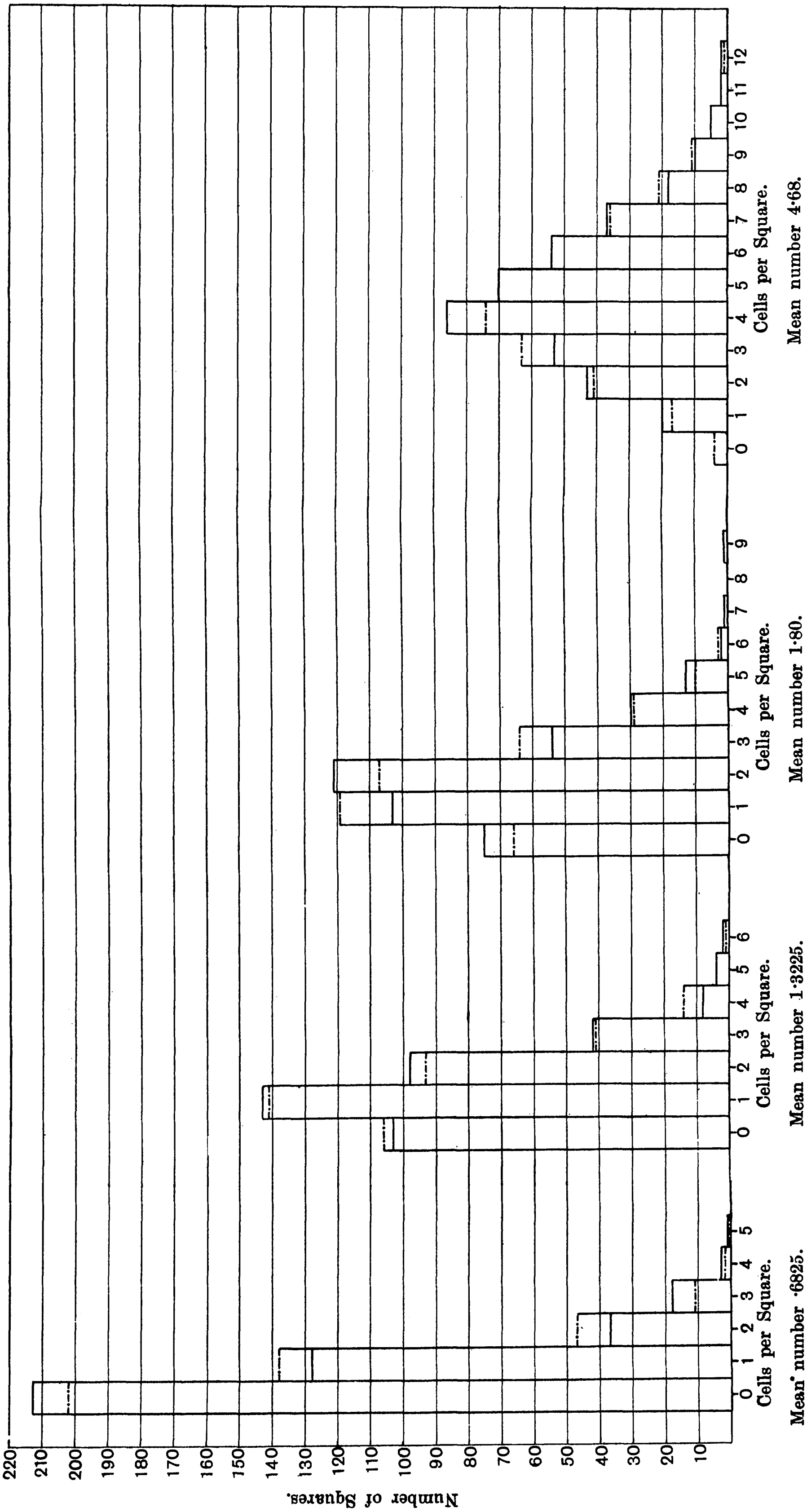
which is

No. of Yeast cells	*0*	*1*	*2*	*3*	*4*
Percentage Frequency	60·65	30·33	7·58	1·26	·16

or approximately three-quarters of those which show growth are pure cultures.

DIAGRAM II. Distribution of 400 Squares.

Firm lines: Actual Observations. Broken lines: Calculated from the Exponential Series. Where they coincide the firm line alone is given.



Conclusions.

We have seen that the distribution of small particles in a liquid follows the law

$$e^{-m}\left\{1+m+\frac{m^2}{2!}+\ldots+\frac{m^r}{r!}+\ldots\right\}$$

where m is the mean number of particles per unit volume * and the various terms in the series give the chances that a given unit volume contains 0, 1, 2, ... r, ... particles. We have also seen that this series represents the limit to which any point binomial $(p+q)^n$ approaches when q is small, insomuch that even $(\frac{19}{20}+\frac{1}{20})^{100}\times 1000$ is represented by $e^{-5}\left(1+5+\frac{5^2}{2!}+\ldots+\frac{5^r}{r!}+\ldots\right)\times 1000$ with a maximum error of about 4·5 in 180.

For the rough calculation of odds with n small compared to $\frac{1}{q}$ the exponential series may be used instead of the binomial as being less laborious.

Finally, we have found that the standard deviation of the mean number of particles per unit volume is $\sqrt{\frac{m}{M}}$ where m is the mean number and M the number of unit volumes counted, so that the criterion of whether two solutions contain different numbers of cells is whether m_1-m_2 is significant compared with $\cdot 67449\sqrt{\frac{m_1}{M_1}+\frac{m_2}{M_2}}$.

TABLE I.

Distribution of Yeast Cells over 1 *sq. mm. divided into* 400 *squares.*

2	2	4	4	4	5	2	4	7	7	4	7	5	2	8	6	7	4	3	4
3	3	2	4	2	5	4	2	8	6	3	6	6	10	8	3	5	6	4	4
7	9	5	2	7	4	4	2	4	4	4	3	5	6	5	4	1	4	2	6
4	1	4	7	3	2	3	5	8	2	9	5	3	9	5	5	2	4	3	4
4	1	5	9	3	4	4	6	6	5	4	6	5	5	4	3	5	9	6	4
4	4	5	10	4	4	3	8	3	2	1	4	1	5	6	4	2	3	3	3
3	7	4	5	1	8	5	7	9	5	8	9	5	6	6	4	3	7	4	4
7	5	6	3	6	7	4	5	8	6	3	3	4	3	7	4	4	4	5	3
8	10	6	3	3	6	5	2	5	3	11	3	7	4	7	3	5	5	3	4
1	3	7	2	5	5	5	3	3	4	6	5	6	1	6	4	4	4	6	4
4	2	5	4	8	6	3	4	6	5	2	6	6	1	2	2	2	5	2	2
5	9	3	5	6	4	6	5	7	1	3	6	5	4	2	8	9	5	4	3
2	2	11	4	6	6	4	6	2	5	3	5	7	2	6	5	5	1	2	7
5	12	5	8	2	4	2	1	6	4	5	1	2	9	1	3	4	7	3	6
5	6	5	4	4	5	2	7	6	2	7	3	5	4	4	5	4	7	5	4
8	4	6	6	5	3	3	5	7	4	5	5	5	6	10	2	3	8	3	5
6	6	4	2	6	6	7	5	4	5	8	6	7	6	4	2	6	1	1	4
7	2	5	7	4	6	4	5	1	5	10	8	7	5	4	6	4	4	7	5
4	3	1	6	2	5	3	3	3	7	4	3	7	8	4	7	3	1	4	4
7	6	7	2	4	5	1	3	12	4	2	2	8	7	6	7	6	3	5	4

* The prism standing on unit area.

It must be noted, however, that the probable error will always be greater than that calculated on this formula when for any reason the organisms occur as aggregates of varying size.

In conclusion, I should like to thank Prof. Adrian J. Brown, of Birmingham University, for his valuable advice and assistance in carrying out the experimental part of the enquiry.

TABLE II.

"Centre" Squares.

"Adjacent" Squares	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	Totals
1	6	6	9	15	15	9	4	3	2	—	—	—	69
2	6	14	17	31	24	17	10	5	6	2	1	1	134
3	8	15	25	32	37	20	15	7	7	1	4	—	171
4	18	34	33	45	48	41	22	7	5	4	1	—	258
5	15	24	37	47	39	37	18	12	11	4	1	2	247
6	9	17	25	39	34	32	14	8	2	4	1	1	186
7	5	12	14	21	19	16	9	7	3	—	—	—	106
8	3	5	7	8	12	8	6	1	3	4	—	—	57
9	2	6	7	5	10	2	2	3	—	1	—	—	38
10	—	1	1	4	4	4	—	3	—	1	—	—	18
11	—	1	4	1	1	1	—	—	—	—	—	—	8
12	—	1	1	—	1	1	—	—	—	—	—	—	4
Totals	72	136	180	248	244	188	100	56	40	20	8	4	1296

Mean of "Centre" Squares, 4·6821 ; S. D., 2·139.

Mean of "Adjacent" Squares, 4·7014 ; S. D., 2·116.

$r = +\cdot016 \pm \cdot037$.

Correlation table between the number of cells in a square and the numbers of cells in the four adjacent squares taken all over Table I.