

Biology. — *Mathematics of pollen diagrams.* II. By J. WESTENBERG.
(Communicated by Prof. M. W. WOERDEMAN.)

(Communicated at the meeting of March 29, 1947.)

The lower extremities of the tail error lines are determined by the conditions: $n < N - n$; $n - \Delta = 0$ and $P_{n;n,N} =$ tail error value. By graphical interpolation between several values of $P_{n;n,N}$ we find

$N = 150$	Tail error %	n
	$\frac{1}{10}$ %	4,9
	$\frac{1}{2}$ %	3,7
	1 %	3,2
	2 %	2,7
	3 %	2,4
	4 %	2,2
	5 %	2,1

Similarly we find for the upper extremities ($N - n < n$; $N - n - \Delta = 0$)

Tail error %	n
$\frac{1}{10}$ %	145,1
$\frac{1}{2}$ %	146,3
1 %	146,8
2 %	147,3
3 %	147,6
4 %	147,8
5 %	147,9

In testing the significance of the difference of two countings of the same species in two different strata, we have to plot these countings on the same horizontal axis in the pollen diagram (*a* and *b*), and insert their mean value in the middle between them (*c*) (fig. 5). Then we place the transparent graph over the pollendiagram, its sides covering the sides of the pollen diagram. Next we shift the graph upward or downward until the mean of the countings (*c*) is covered by the oblique line. If the position of the two counting marks (*a* and *b*) is outside the two tail error lines, we are led to conclude that it is likely, that the difference $d (= 2 \Delta)$ of the two countings considered, should have been significant, because the probability that a difference so great or greater should have arisen through random sampling is even less than the chosen tail error value.

In order to examine how the reliability of the method depends on the amount of work, we compute similar data for other values of N , viz. 10, 20, 50, 100, 300 and 450.

$N = 10$	Tail error %	$n = \Delta$	$n = 3$	$n = 4$
		$n = 7$	$n = 6$	$n = 5$
	$\frac{1}{10}$ %	3,6	—	3,7
	$\frac{1}{2}$ %	3,0	—	3,3
	1 %	2,7	2,8	3,0
	2 %	2,3	2,6	2,7
	3 %	2,1	2,4	2,5
	4 %	2,0	2,3	2,4
	5 %	1,9	2,2	2,3

$N = 20$	Tail error %	$n = \Delta$	$n = 5$	$n = 8$
		$n = 15$	$n = 12$	$n = 10$
	$\frac{1}{10}$ %	4,3	4,5	5,2
	$\frac{1}{2}$ %	3,4	3,9	4,5
	1 %	3,0	3,6	4,1
	2 %	2,6	3,2	3,6
	3 %	2,4	3,0	3,3
	4 %	2,2	2,8	3,1
	5 %	2,0	2,7	3,0

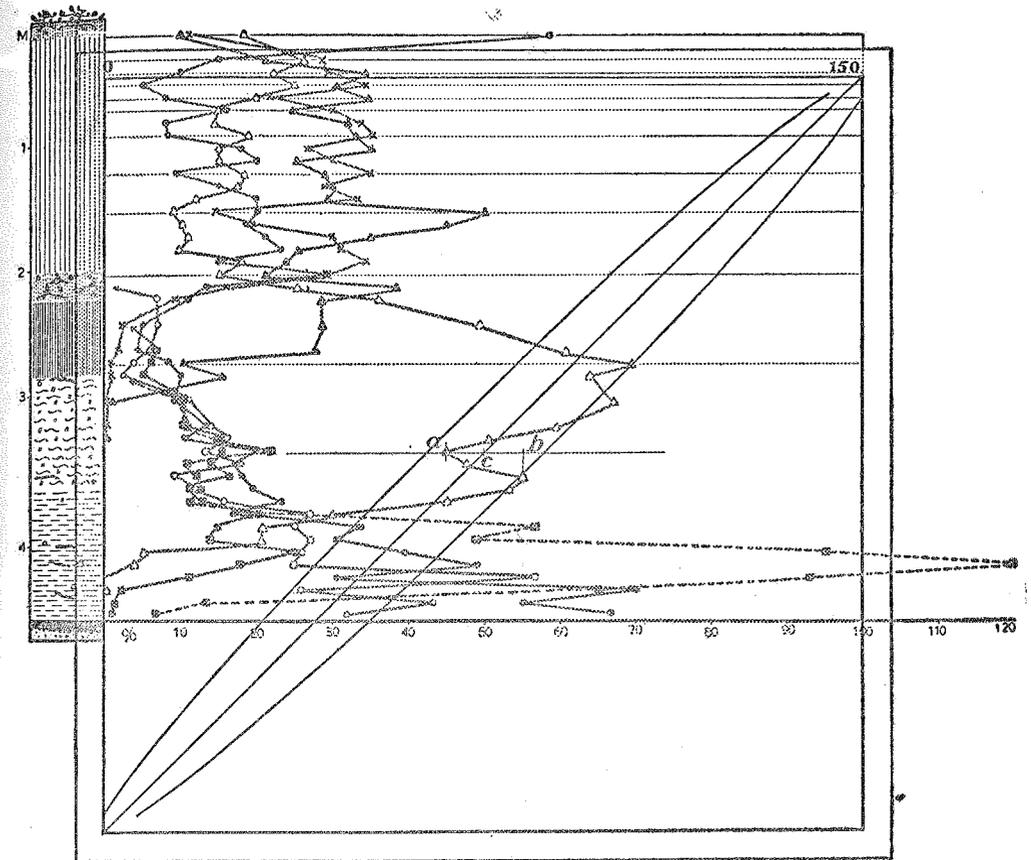


Fig. 5. Tail error line superposed on pollen diagram.

$N = 50$	Tail error %	$n = \Delta$	$n = 5$	$n = 10$	$n = 15$	$n = 20$
		$n = 45$	$n = 40$	$n = 35$	$n = 30$	$n = 25$
	$\frac{1}{10}$ %	4,6	4,8	6,5	7,5	8,0
	$\frac{1}{2}$ %	3,6	4,2	5,6	6,3	6,7
	1 %	3,2	3,9	5,1	5,8	6,1
	2 %	2,8	3,5	4,5	5,1	5,4
	3 %	2,5	3,3	4,2	4,8	5,1
	4 %	2,3	3,1	4,0	4,5	4,7
	5 %	2,1	3,0	3,8	4,2	4,5

N = 100	Tail error	N - n = Δ	n = 5	n = 10	n = 20	n = 30	n = 40	n = 50	
			n = 95	n = 90	n = 80	n = 70	n = 60		
	1/8 %		4,8	4,9	6,8	9,1	10,4	11,1	11,4
	1/4 %		3,8	4,3	5,8	7,7	8,8	9,3	9,5
	1 %		3,3	4,0	5,4	7,0	8,0	8,4	8,7
	2 %		2,8	3,5	4,8	6,2	7,1	7,6	7,7
	3 %		2,5	3,3	4,5	5,8	6,6	7,0	7,1
	4 %		2,3	3,1	4,2	5,4	6,1	6,6	6,7
	5 %		2,1	3,0	4,0	5,1	5,8	6,2	6,3

N = 150	Tail error	N - n = Δ	n = 5	n = 10	n = 20	n = 30	n = 40	n = 50	n = 60	n = 75
			n = 145	n = 140	n = 130	n = 120	n = 110	n = 100	n = 90	n = 75
	1/8 %		4,9	4,9	6,9	9,4	11,1	12,2	13,0	13,5
	1/4 %		3,7	4,3	5,9	8,1	9,3	10,3	11,0	11,4
	1 %		3,2	4,0	5,4	7,3	8,5	9,3	10,0	10,4
	2 %		2,7	3,6	4,9	6,5	7,6	8,3	8,9	9,2
	3 %		2,4	3,3	4,5	6,0	7,0	7,7	8,2	8,4
	4 %		2,2	3,1	4,3	5,6	6,6	7,2	7,6	7,9
	5 %		2,1	3,0	4,0	5,3	6,2	6,8	7,2	7,5

N = 300	Tail error	N - n = Δ	n = 10	n = 20	n = 40	n = 60	n = 80	n = 100	n = 125	n = 150
			n = 290	n = 280	n = 260	n = 240	n = 220	n = 200	n = 175	n = 150
	1/8 %		4,9	7,0	9,8	13,3	15,5	17,2	18,2	19,1
	1/4 %		3,7	6,0	8,3	11,2	13,1	14,4	15,8	16,0
	1 %		3,2	5,5	7,6	10,1	11,9	13,1	13,9	14,5
	2 %		2,8	5,0	6,7	9,0	10,5	11,6	12,4	12,9
	3 %		2,5	4,6	6,2	8,3	9,7	10,7	11,4	11,8
	4 %		2,3	4,3	5,8	7,8	9,1	10,0	10,6	11,1
	5 %		2,1	4,1	5,5	7,3	8,6	9,4	10,0	10,4

N = 450	Tail error	N - n = Δ	n = 15	n = 30	n = 60	n = 90	n = 120	n = 150	n = 180	n = 225
			n = 435	n = 420	n = 390	n = 360	n = 330	n = 300	n = 270	n = 225
	1/8 %		4,9	8,6	11,9	16,2	19,0	20,9	22,3	23,7
	1/4 %		3,8	7,3	10,1	13,5	15,9	17,6	18,7	19,4
	1 %		3,3	6,7	9,1	12,3	14,4	15,9	16,9	17,5
	2 %		2,8	6,0	8,1	10,9	12,8	14,1	15,0	15,5
	3 %		2,5	5,5	7,5	10,1	11,8	12,9	13,8	14,3
	4 %		2,3	5,2	7,0	9,4	10,9	12,1	12,8	13,4
	5 %		2,1	4,9	6,6	8,9	10,3	11,4	12,1	12,6

The above tables allow the construction of tail error diagrams, as described in preceding lines; in our illustrations the values of Δ pertain to a (unilateral) tail error value of $\frac{1}{8}\%$. If we draw these diagrams just as broad as that one for $N = 150$, the significant difference $d (= 2\Delta)$ will appear in these diagrams as $\delta = d \frac{150}{N}$.

As a measure for the precision of the method we choose δ for $n = N - n$. Next we unite these diagrams into one three dimensional graph, plotting N along the third axis, as shown in fig. 6. The tail error lines are connected by a pair of smooth surfaces, the so-called tail error surfaces. The

construction is then completed by connecting the horizontal axes for $n = N - n$ by a plane ¹⁾.

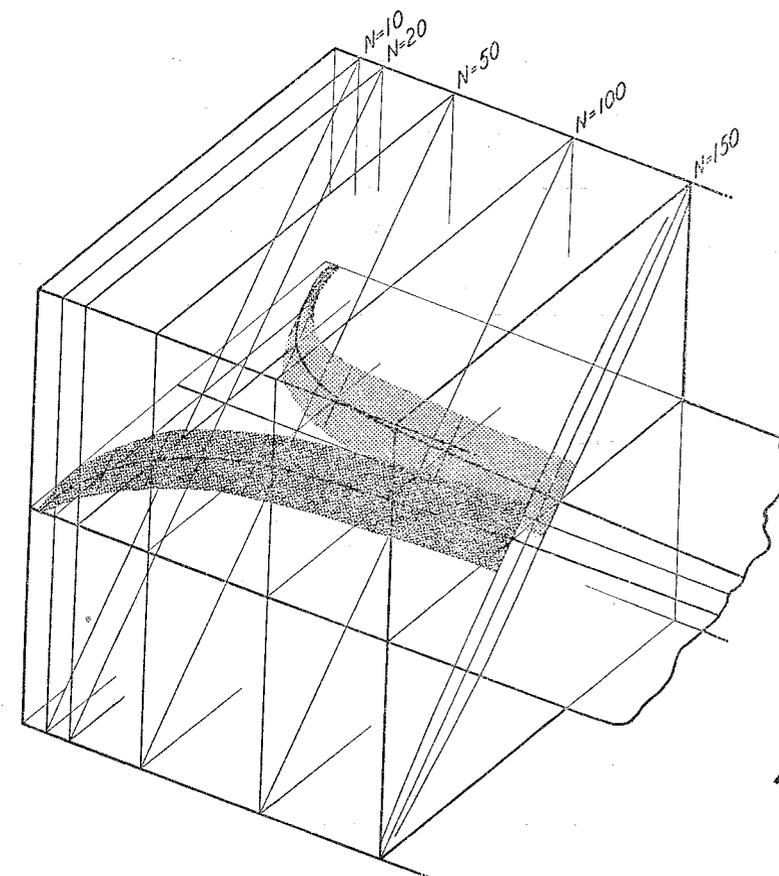


Fig. 6. Three dimensional arrangement of tail error diagrams pertaining to increasing values of N .

The section lines of this plane and the tail error surfaces demonstrate the decrease of δ with increasing N . This is depicted separately in fig. 7, the plane being extended to $N = 450$. From this figure we see, that δ decreases rapidly up to $N = 50$, still appreciably up to $N = 150$. On from $N = 150$, the decrease becomes very slight, so that we may conclude, that pollen analysts did well to choose $N = 150$.

Sometimes, however, the pollen analyst might need more precision in order to reach a conclusion. This might be the case when a series of successive strata yields samples, that suggest a slight increase or decrease of a certain species, but do not allow any decision of the kind by comparing the countings of single strata. Then the difficulty might be surmounted,

¹⁾ In order to avoid a confusing complexity, fig. 6 has been simplified: the tail error surfaces have been drawn only near $n = N - n$.

by pooling the countings of two or three successive samples and comparing these with the pooled countings of another set of two or three successive

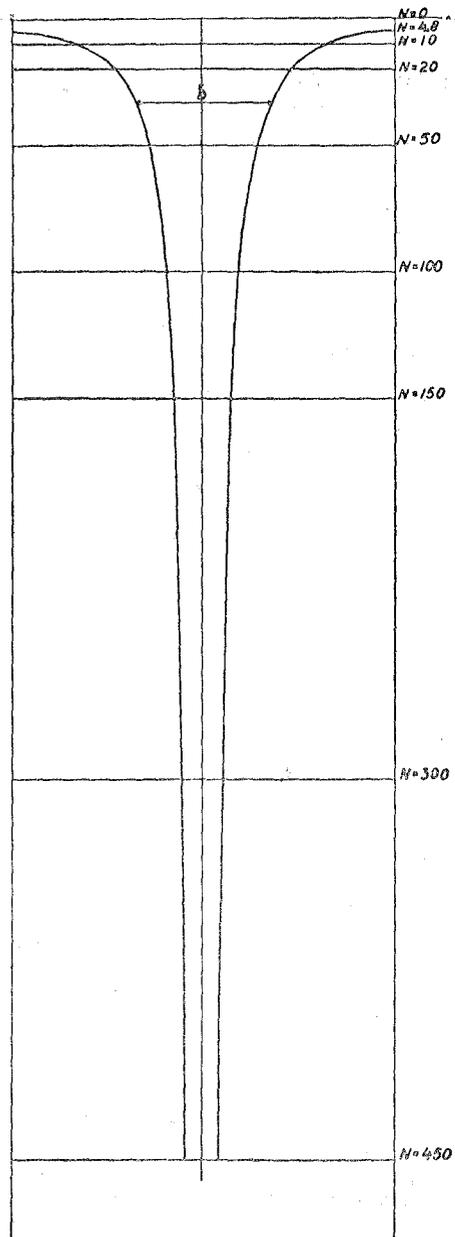


Fig. 7. Decrease of δ with increase of N .

samples. For this purpose we have to make use of the tables for $N = 300$ or $N = 450$.

As has been mentioned before, the extra pollen requires a different treatment. In the comparison of two extra pollen countings $n + \Delta$ and $n - \Delta$, the total numbers of pollen grains will amount to $N_1 + n + \Delta$ and $N_2 + n - \Delta$, when we keep to N_1 and N_2 for the marker species.

Hence the probability of $n + \Delta$ is:

$$P_{n+\Delta}^* = \frac{(N_1 + n + \Delta)!}{(n + \Delta)! N_1!} \cdot p^{*n+\Delta} \cdot (1-p^*)^{N_1}$$

Similarly

$$P_{n-\Delta}^* = \frac{(N_2 + n - \Delta)!}{(n - \Delta)! N_2!} \cdot p^{*n-\Delta} \cdot (1-p^*)^{N_2}$$

The probability of the simultaneous occurrence of these two events will therefore be

$$P_{\Delta; n, N_1, N_2, p^*}^* = \frac{(N_1 + n + \Delta)! (N_2 + n - \Delta)!}{(n + \Delta)! (n - \Delta)! N_1! N_2!} \cdot p^{*2n} \cdot (1-p^*)^{N_1+N_2}$$

Since p^* is not known, this probability can not be computed. In this formula, the powers of p^* are independent of Δ , and therefore the unknown factor

$$p^{*2n} \cdot (1-p^*)^{N_1+N_2}$$

is the same for all possibilities with fixed values of n, N_1, N_2 and p^* . If these values are kept constant, the probability of any value of Δ occurring, is proportional to

$$F^* = \frac{(N_1 + n + \Delta)! (N_2 + n - \Delta)!}{(n + \Delta)! (n - \Delta)! N_1! N_2!}$$

The total probability of all possible values of Δ , with fixed values of n, N_1, N_2 and p^* is represented by

$$\sum_{\Delta} P_{\Delta; n, N_1, N_2, p^*}^* = \frac{(N_1 + n + \Delta)! (N_2 + n - \Delta)!}{(n + \Delta)! (n - \Delta)! N_1! N_2!} \cdot p^{*2n} \cdot (1-p^*)^{N_1+N_2}$$

which may also be written as

$$\sum_{\Delta} P_{\Delta; n, N_1, N_2, p^*}^* = p^{*2n} \cdot (1-p^*)^{N_1+N_2} \cdot \sum_{\Delta} F^*$$

In this summation $|\Delta| \leq n$ and $\leq N_1$ and $\leq N_2$.

The probability of a certain Δ , the values of n, N_1, N_2 being fixed, is given by

$$P_{\Delta; n, N_1, N_2}^* = \frac{P_{\Delta; n, N_1, N_2, p^*}^*}{\sum_{\Delta} P_{\Delta; n, N_1, N_2, p^*}^*} = \frac{F^*}{\sum_{\Delta} F^*}$$

whatever the value of p^* might be.

In order to carry out the summation of $\sum F^*$, we use the following method, which was again suggested by VAN DANTZIG (private communication).

By the binomial series of NEWTON we have:

$$\sum_0^{\infty} \frac{(N_1 + k)!}{N_1! k!} \cdot x^k \equiv (1 - x)^{-(N_1+1)}$$

$$\sum_0^{\infty} \frac{(N_2 + l)!}{N_2! l!} \cdot x^l \equiv (1 - x)^{-(N_2+1)}$$

Multiplication yields

$$\begin{aligned} \sum_0^{\infty} \sum_0^{\infty} \frac{(N_1 + k)! (N_2 + l)!}{N_1! N_2! k! l!} \cdot x^{k+l} &\equiv (1 - x)^{-(N_1+1) - (N_2+1)} \equiv \\ &\equiv \sum_0^{\infty} \frac{(N_1 + N_2 + 1 + h)!}{(N_1 + N_2 + 1)! h!} \cdot x^h. \end{aligned}$$

Equating the coefficients of x^h , ($h = k + l$) gives:

$$\sum_0^h \frac{(N_1 + k)! (N_2 + h - k)!}{N_1! N_2! k! (h - k)!} = \frac{(N_1 + N_2 + 1 + h)!}{h! (N_1 + N_2 + 1)!}$$

Putting $h = m_1 + m_2$, $k = m_1 + \Delta$ and $h - k = m_2 - \Delta$ we find

$$\sum_{-m_1}^{+m_2} \frac{(N_1 + m_1 + \Delta)! (N_2 + m_2 - \Delta)!}{N_1! N_2! (m_1 + \Delta)! (m_2 - \Delta)!} = \frac{(N_1 + N_2 + m_1 + m_2 + 1)!}{(m_1 + m_2)! (N_1 + N_2 + 1)!}$$

For $m_1 = m_2 = n$, we arrive at:

$$\sum_{-n}^{+n} \frac{(N_1 + n + \Delta)! (N_2 + n - \Delta)!}{(N_1! N_2! (n + \Delta)! (n - \Delta)!} = \frac{(N_1 + N_2 + 2n + 1)!}{(2n)! (N_1 + N_2 + 1)!}$$

Finally we may write

$$P_{\Delta; n, N_1, N_2}^* = \frac{F^*}{\sum_{\Delta} F^*} = \frac{(N_1 + n + \Delta)! (N_2 + n - \Delta)!}{(n + \Delta)! (n - \Delta)! N_1! N_2!} \cdot \frac{(N_1 + N_2 + 2n + 1)!}{(2n)! (N_1 + N_2 + 1)!}$$

For equal values of N , we arrive at the simplified form

$$P_{\Delta; n, N}^* = \frac{(N + n + \Delta)! (N + n - \Delta)!}{(n + \Delta)! (n - \Delta)! (N!)^2} \cdot \frac{(2N + 2n + 1)!}{(2n)! (2N + 1)!}$$

In order to make the method fit for use, we have to compute the values of $P_{\Delta; n, N}^*$ for a certain value of n and $N = 150$, varying Δ only. Next we calculate the cumulative chances of Δ , and by means of graphical interpolation we find the values of Δ , belonging to (unilateral) tail error values of 0,001, 0,005, 0,01 &c. The same is done for other values of n .

By graphical interpolation we also find, for which values of n , $P_{n; n, N}^*$

will correspond to the chosen tail error values. Thus we arrive at the following table:

Tail error	$n = \Delta$	$n = 5$	$n = 10$	$n = 20$	$n = 30$	$n = 40$	$n = 50$	$n = 60$
$\frac{1}{10}$ %	5,1	—	7,3	10,6	13,4	15,8	18,1	20,2
$\frac{1}{2}$ %	3,9	4,4	6,2	9,0	11,3	13,3	15,3	17,1
1 %	3,4	4,1	5,7	8,2	10,2	12,1	13,8	15,4
2 %	2,9	3,7	5,1	7,3	9,1	10,8	12,3	13,7
3 %	2,6	3,4	4,8	6,8	8,4	9,9	11,3	12,6
4 %	2,3	3,2	4,4	6,3	7,9	9,3	10,5	11,8
5 %	2,2	3,1	4,2	6,0	7,4	8,7	10,0	11,1
	$n = 75$	$n = 90$	$n = 110$	$n = 150$	$n = 200$	$n = 300$		
	23,4	26,3	30,4	38,0	47,2	65,4		
	19,7	22,2	25,5	31,8	39,5	54,8		
	17,8	20,1	23,0	28,8	35,8	49,5		
	15,8	17,9	20,5	25,5	31,7	43,8		
	14,6	16,3	18,8	23,5	29,1	40,2		
	13,6	15,3	17,5	21,9	27,1	37,4		
	12,8	14,4	16,5	20,6	25,5	35,3		

From this table we can proceed to the construction of a tail error diagram, to be used as previously described.

Since the extra pollen species are considered as such, because of their great variability in abundance, the pollen analyst is not likely to need more precision for the interpretation of extra pollen countings. For this reason we refrain from computing tables for the treatment of pooled extra pollen data.

Acknowledgements.

Many years ago, I took up the plan to deal with the present subject, but it was not possible for me, to go through the matter all by myself. Therefore I wish to make grateful acknowledgement to those, who enabled me to carry out my plan.

To Prof. Dr. M. J. VAN UVEN, I feel indebted for his kind advice in questions of computation.

To Prof. Dr. D. VAN DANTZIG I wish to express my thanks for criticising the manuscript and for giving the mathematical solution of the last part of the problem.

To Dr. E. F. DRION particularly, I feel deeply indebted for repeatedly discussing the matter. Without his kind help, this paper would not have been written.

To my friend A. J. D. VEEN, I owe the careful drawing of the illustrations.

Summary.

In stating the significance of the difference between two frequency distributions, we have to test our samples by means of statistical methods imposing certain properties on the frequency curves of the populations.

In a vast majority of cases the normal law is assumed, including two parameters, i.e. the mean of the population, and the standard deviation. In the classical treatment of data, we have to substitute the parameters by so-called statistics, these being their estimates as computed from the samples. This is tolerable as long as large samples are considered. In the case of small samples, this procedure will appreciably affect the reliability of our conclusions. In later years, this difficulty has been surmounted by designing the so-called *t*-test. Since no unknown parameters appear in the formula, this test enables us to get exact information on the significance of the difference between two small samples, taken from normal populations. The application of this method, however, is still limited by the assumption of the normal law. In many cases of counting work, as in pollen analysis, the data may be supposed to follow the binomial law. Since the standard deviation, or the initial probabilities do not appear in the formula for the distribution of the mean of two samples, we have no need for computing their estimates. A suitable treatment is then carried out by applying R. A. FISHER's binomial test, which is exact, even for small samples. The derivation of its mathematical basis is given in full, in order to attract the attention of biological workers. The test is tabulated as far as needed for pollen analytical research. The practice of the method resolves itself into a graphical application, as is exemplified with a pollen diagram. The custom of counting 150 pollen grains is justified on theoretical grounds. The current theory does not hold for extra pollen countings. In order to meet the needs, a special method is developed for a similar treatment of extra pollen data.

LITERATURE.

- BAAS BECKING, L. G. M. and E. F. DRION. 1935. On the origin of frequency distributions in biology. *Acta Biotheoretica*, Vol. 1, Pars III, p. 133—150.
- BARKLEY, A. 1934. The statistical theory of pollen analysis. *Ecology*, Vol. 15, p. 283—289.
- CAIN, S. A. 1939. Pollen analysis as a palaeo-ecological research method. *The Botanical Review*, Vol. 5, no. 12, p. 627—651.
- DRION, E. F. 1939. Eenige opmerkingen over den vorm van de frequentiekromme in de biologie. *Natuurk. Tijdschr. v. Nederlandsch Indië*, Deel 99, Aft. 4, p. 166—174.
- DUARTE, F. J. 1927. Nouvelles tables de $\text{Log}_n!$ à 33 décimales depuis $n = 1$ jusqu'à $n = 3000$. Edit. Gauthier Villars, Genève et Paris, 8°, XXIV + 136 pp.
- GODWIN, H. 1934. Pollen-analysis. *The New Phytologist*, Vol. 33, Part. I, p. 278—305; Part II, p. 325—358.
- FISHER, R. A. 1930. Inverse probability. *Proc. of the Cambridge Phil. Soc.* Vol. 26, Part 4, p. 528—535. (See especially p. 529.)
- , 1935. The logic of inductive inference. *Journ. of the Royal Statistical Soc. Nw. Series*, Vol. 98, p. 39—54. (See especially: Ancillary statistics, Example I, p. 48—51.)
- , 1946. *Statistical methods for research workers*. Oliver & Boyd, London, 10th edition, 354 pp. (See especially p. 96—97.)
- WHITWORTH, W. A. 1901. *Choice and change with 1000 exercises*. Deighton Bell & Co., Cambridge; George Bell & Sons, London, 5th edition, 8°, 342 pp.
- WOODHEAD, N. and L. M. HODGSON. 1935. A preliminary study of some Snowdonian peats. *The New Phytologist*, Vol. 34, p. 263—282. (Estimation of sampling errors, p. 267—268.)

Geology. — *Some data on the Muriah volcano (Java), and its leucite-bearing rocks.* By L. BOOMGAART. (Communicated by Prof. J. H. F. UMBGROVE.)

(Communicated at the meeting of May 31, 1947.)

The Muriah volcano complex, East of Semarang on the North-coast of Java, protrudes into the Java-Sea. It comprises the Muriah volcano (inactive), 1602 m, and North of it the small ashvolcano Genuk, 717 m (formerly called Tjilering). East of this Genuk volcano we find some small isolated hills; from North to South the Bako, 157 m, the Ragas, 122 m and the Tempur, 49 m.

During an investigation of the northern sector by the present author 83 rocksamples have been collected (60 of solid rocks, 23 of boulders) and a microscopical study made, especially of the leucite-bearing rocks. Of 67 samples the content of K_2O and Na_2O were chemically determined by the Laboratory of the Geological Survey at Bandung.

I. The Muriah volcano.

The whole of the northern aspect of the mountain is controlled by the imposing triangular peak of the Sutorenggo, 1604 m, which stands out on the Southside of the Tempur-cauldron. This northern crater is surrounded by steep walls from 1100—1600 m, through which the Gelis-river forces an outlet to the North. The village of Tempur, situated in the crater, lies at an altitude of 600 m.

The northern part of the Muriah is mainly built up by breccias. Flows of basalto-andesitic rocks and of leucite-bearing rocktypes lie upon the breccias, or are found as intercalations in breccias. Some tufflayers have been observed also. No definite sequence of deposition could be reconstructed.

Leucite-bearing rocks.

This type of rock has been found as flows, as boulders in superficial block-fields, in one case as a dike (Tempur-cauldron), as constituent of breccias and as boulders in tuff. The following leucite-bearing rocktypes can be distinguished.

1. Leucite-tephrites without orthoclase (sanidine), except a little groundmass-orthoclase.
2. Leucite-tephrites with orthoclase (sanidine) as phenocrysts. The leucite is mostly altered. All these samples are from the crater-area.
3. Leucite-tephrites with nephelinite. Only one sample.
4. Leucite-basanites (characterized by olivine). Plagioclase as phenocryst is present in 46% of the examined rocks. In the other specimen plagioclase is generally sporadically present.