THE USE OF LOGARITHMS IN THE INTERPRETATION OF CERTAIN ENTOMOLOGICAL PROBLEMS

By C. B. WILLIAMS

Chief Entomologist, Rothamsted Experimental Station, Harpenden

(With 4 Text-figures)

In many experiments involving a comparison of the numbers of insects present under different conditions—for example, the number of insects on alternative host plants, or in field plots under different treatments; the number of insects attracted to different baits or lights, or attracted to these under different conditions—it is frequently necessary to group together the catches in several cases under one set of conditions and to compare them with a number of cases under different conditions. One wishes, in fact, to compare the "average" catch under the two conditions.

In most cases one does this by adding together the numbers obtained under similar conditions and dividing it by the number of cases, thus obtaining an average which is an "arithmetic" mean. It is the purpose of this note, however, to show that in some experiments a more exact interpretation of the results can be obtained by the use of a "geometric" mean instead of an "arithmetic" mean, and indeed that conclusions drawn from the latter may at times be erroneous. The "geometric" mean is most simply obtained by adding together the logarithms of the numbers in question, finding the "arithmetic" mean of these logarithms and reconverting back from this (when necessary) to numbers again.

During the past 3 years we have had working at Rothamsted a light trap for catching insects at night. The number of insects caught per night has varied from zero to 72,000, and the total for the 3 years, March 1933 to March 1936, is over 600,000.

If we compare the "arithmetic" mean of the captures on nights with one set of weather conditions with the "arithmetic" mean of nights with another set of weather conditions a serious error becomes apparent, owing to the very great variability of the numbers and the swamping effect that a single large catch can have on the "arithmetic" mean.

For example, on 1 January 1935 more insects were captured in the trap than in the whole of January, February, March and April of the previous year. Any comparison which includes these two sets of figures is liable to be considerably biased by the one night. Again, the number of noctuid moths caught on the successive nights of the full-moon week in October 1933 was 0, 0, 1, 62, 0, 0, 0. In the corresponding no-moon week the numbers were 2, 4, 0, 0, 10, 3, 3. If we compare the total capture in the full-moon week (63) with that of the no-moon week (22) it is obvious that the former is unduly biased by the single large catch.

Reasoning *a priori* one might expect that similar differences in environment would produce similar percentage increases in a catch rather than similar numerical increases. Thus if the catches were on two nights under one set of conditions 100 and 1000 insects; and if in another set of conditions (for example, a second trap with a more powerful light) the catch on the first night was 200, one might expect the catch on the second night to be 2000 and not 1100, i.e. a similar percentage increase and not the addition of 900 insects to each catch. The addition would be in proportion to the basic catch.

If this reasoning is correct the proper mean to take for the comparison of two series of figures would be the geometric mean and not the arithmetic mean. This, however, would be a cumbersome piece of work if done arithmetically, and exactly the same result can be obtained by using the logarithm of the number caught in each case instead of the actual numbers; then for a series of nights one could use either the sum of the logarithms of the individual nights (not the logarithm of the sum of the numbers), or the average logarithm; or the latter could be reconverted back to the geometric mean by taking the anti-logarithm.

For example, if a series of catches under two different conditions are

Series A	5	15	47	1000	2	8
Series B	4	19	22	99	50	17

the comparison of the total numbers is 1077 : 211, or of the arithmetic mean catch 180 : 35. Series A therefore has the higher average.

The logarithms of the above series (to two decimal places, which has been found sufficient for all practical purposes) are

Series A	0·70	1·18	1·67	3·00	0·30	0·90
Series B	0·60	1·28	1·34	2·00	1·70	1·23

So the comparison of the logarithmic sum is 7·75 : 8·15, the comparison of the mean logarithm is 1·29 : 1·36, and the comparison of the geometric mean (anti-logarithm of above) is 19·6 : 22·8. It will be seen that the large catch of 1000 on one night of the first series no longer swamps the proportion, and series B has a higher mean than series A.

A complication ensues if any value in the series is zero, for the logarithm of zero is minus infinity and the geometric mean of any series containing a zero is itself zero. It has been found possible in these cases to add a unit to all values in the series before taking logarithms, i.e. to deal with log $(n+1)$ instead of log n. If this is done however it is necessary to subtract the unit from the final result when it is reconverted back into number from logarithms.

If this system is used it is found that the sum of log $(n+1)$ for the values quoted above for the full-moon week captures of Noctuidae become 2·10 while that for the no-moon week is 3·89. Thus the geometric mean for the no-moon week is higher than the full-moon week.

Another way in which the use of the logarithm is found to be more suitable is in the distribution of departures from a mean. If an arithmetic mean value be taken for a series of captures on the number basis and if each day's value is expressed as a departure from the mean, then in the case of actual numbers (see examples 3 and 4 below) the values are made up of a large number of small negative departures and a small number of large positive departures which give a skew curve which does not lend itself to treatment by the normal formulae of standard deviation, etc. If on the other hand the departures of the logarithms of $n+1$ from a mean logarithm are used, the number of values on either side of the mean is almost equal and their distribution near to normal.

In one case (see example 3 below) the square root of each number was taken as a test and this was found to give a skew distribution, less asymmetrical than number curve but definitely not so good as the logarithm distribution.

It is thought that the best way of explaining the methods and results more fully would be to give a few examples from actual calculations that have occurred in the analysis of the trap records.

Example 1. *A comparison of the catches in a single trap on alternate nights*

There is no reason to suppose that with the same trap in the same spot there should be any consistent difference between captures on two series of alternate nights. Such differences as occur are due to accidental alterations of temperature, wind and other weather conditions, superimposed upon which is the experimental error.

To test the error that might occur from these causes in a long series of nights, the captures in the light trap were added together in two series: (*a*) those on the odd nights of the year (1, 3, 5, etc.), and (*b*) those caught on the even nights (2, 4, 6).

Table I shows the results when the numbers themselves are summed; when the square roots of the numbers are summed; and when the log $(n+1)$ are summed for each of the three years 1933–5. The number of nights on which the trap was working was 306 in 1933 (March to December), 364 in 1934 and 358 in 1935.

Table I

	1933 (9 months)		1934		1935		All 3 years	
	Odd	Even	Odd	Even	Odd	Even	Odd	Even
Σn	61,368	47,385	56,689	40,260	243,556	160,468	361,613	248,113
Ratio n	129	100	141	100	152	100	145	100
$\Sigma \sqrt{n}$	2,131	1,975	2,198	1,897	—	—	—	—
Ratio n	117	100	135	100	—	—	—	—
$\Sigma \log (n+1)$	266·62	265·70	301·73	288·39	322·25	327·77	890·60	881·86
Ratio n	101·4	100	118·5	100	100	107·4	104·2	100

It will be seen that by the accidental distribution of weather conditions and experimental error the odd nights in each year, when summed on a number basis, have differed very considerably from the even nights. The total catches on the odd nights were greater by 29 per cent in 1933, by 41 per cent in 1934, and by 52 per cent in 1935: with a total difference on all 3 years (including over 1000 nights) of 45 per cent. Thus if a different trap or a different intensity of light had been used for one of the two series a difference of 45 per cent after 3 years would have demonstrated no real difference in efficiency.

When the $\sqrt{n}$ was used as a basis of summation (this was not done for the third year as no advantage seemed to be gained) the ratios between the mean catches were lower but still large, 17 per cent in 1933 and 35 per cent in 1934.

When log $(n+1)$ was used the difference was reduced to 1·4 per cent in 1933, to 18·5 per cent for 1934, and to 7·4 per cent in favour of the *even* nights in 1935. The difference in the total captures on the odd and even series in 1935 was about 83,000 insects but of these 72,000 occurred on one night, so that the great effect produced by the use of logarithms is not so unexpected.

For the 3 years together the use of logarithms have reduced the difference to 4·2 per cent from the 45 per cent difference in actual numbers.

Example 2. *A comparison of the catches of Noctuidae in weeks of full-moon and of no-moon*

Table II shows the captures of Noctuidae in the full-moon and no-moon weeks of six summer months in each of the 3 years 1933–5. The totals are based first on a sum of the actual numbers caught (n) and secondly on a sum of $\log(n+1)$. The larger number in each pair is in heavy type (Williams, 1936).

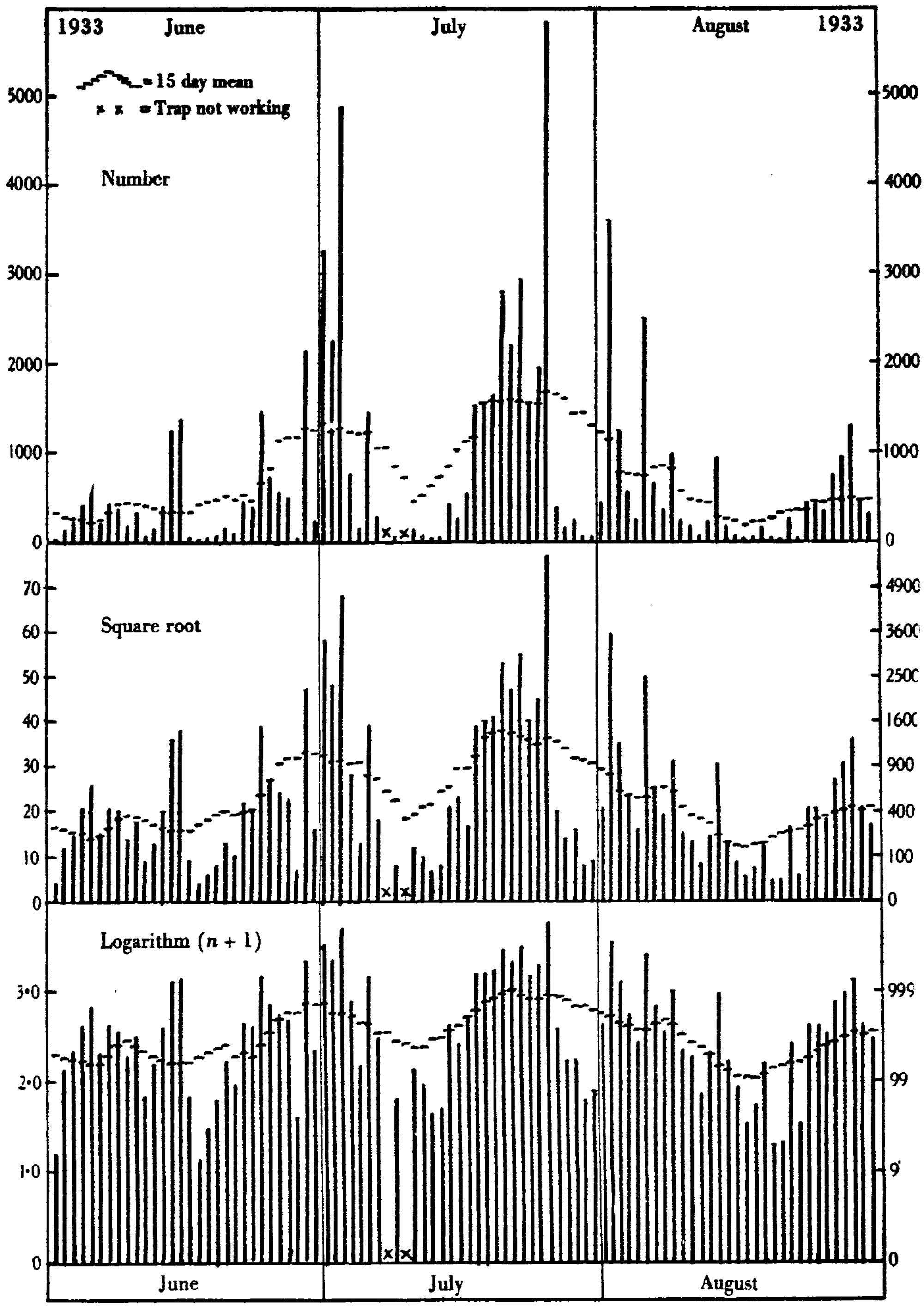
Table II

		May	June	July	Aug.	Sept.	Oct.	Total	All 3 years
				Numbers $\Sigma(n)$					
1933	Full	15	55	58	64	29	**63**	284	
	No	**19**	**95**	**73**	**72**	**76**	22	**363**	
1934	Full	2	52	37	69	61	1	222	1242
	No	**25**	**56**	**140**	**76**	**204**	**71**	**575**	**2859**
1935	Full	2	15	477	168	64	10	736	
	No	**23**	**179**	**917**	**385**	**267**	**133**	**1904**	
				Logarithms $\Sigma\log(n+1)$					
1933	Full	2·76	**7·51**	6·49	6·03	4·25	2·10	29·14	
	No	**3·59**	7·36	**7·14**	**7·14**	**7·34**	**3·89**	**36·46**	
1934	Full	0·60	4·20	5·31	6·63	5·77	0·30	22·81	83·41
	No	**3·64**	**5·86**	**9·01**	**7·11**	**10·20**	**6·95**	**42·77**	**133·20**
1935	Full	0·60	2·75	10·87	8·99	6·32	1·93	31·46	
	No	**3·49**	**7·93**	**13·86**	**11·70**	**9·92**	**7·07**	**53·97**	

In each case 17 out of the 18 weeks give values in favour of no-moon, indicating that there is undoubtedly a consistent difference between the full- and no-moon weeks.

With the numbers the mean difference per week between the two is 87·6 with a standard deviation of $\pm 27{\cdot}3$. This gives a "t" test (that is, the mean difference divided by the standard deviation) of 3·2.

On the logarithmic basis the mean difference per week is 2·77 with a standard deviation of $\pm 0{\cdot}44$ which gives $t = 6{\cdot}3$. Thus, as t is a measure of the significance of the results, the use of the logarithms has given a result of very much higher significance than the use of numbers. The mean catch per night calculated from the numbers (that is arithmetic mean) is 9·9 insects for the full-moon nights and 22·7 for the no-moon nights or a ratio of 100 to 229. The geometric means (obtained by reconverting the mean logarithm per night back to an anti-logarithm and subtracting one) is 3·59 insects for the full-moon nights and 10·40 for the no-moon nights or a ratio of 100 to 289. It is therefore seen that as the differences were consistent they have not been reduced (but actually increased) by the use of logarithms, in direct comparison with the results obtained in the first example where non-consistent differences are very much reduced.



1933
June
July
August
1933
= 15 day mean
x x = Trap not working
Number
Square root
Logarithm (n + 1)
5000
4000
3000
2000
1000
0
70
60
50
40
30
20
10
0
3·0
2·0
1·0
0
4900
3600
2500
1600
900
400
100
0
999
99
9
0
June
July
August

Fig. 1.

Example 3. *Captures of all insects in June to August* 1933 *and the departures from a mean*

Fig. 1 shows the captures day by day for the months of June, July and August 1933 of all insects, first on a number basis, then on the square root, and finally on the basis of log $n+1$; with a running 15-day mean of the values in each case.

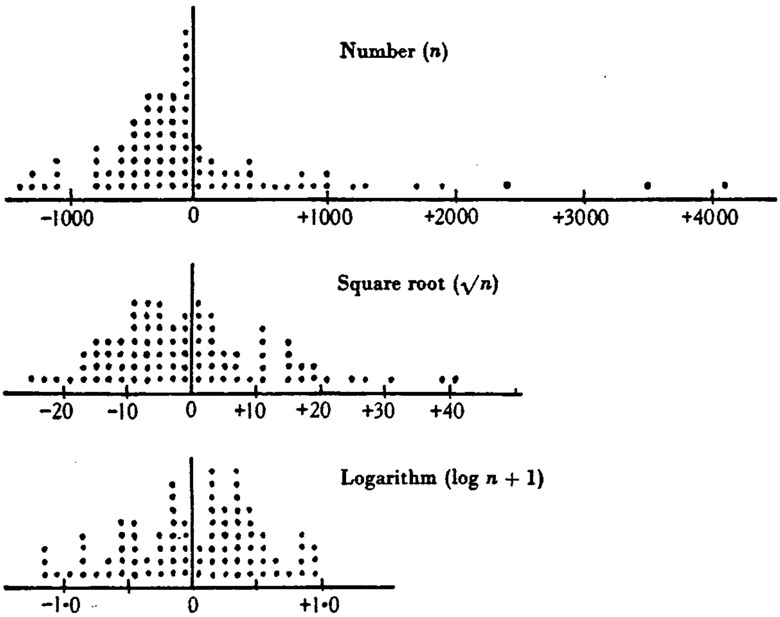


Fig. 2.

Fig. 2 shows the distribution of the departures from the mean in each of the three cases. It will be seen that on the number basis the departures consist of a larger number (60) of smaller negative departures and a much smaller number (29) of larger positive departures, giving a skew distribution.

The square roots give a less skew curve with 51 negative and 39 positive departures, the latter having nearly twice the range of the former. The logarithm gives 41 negative and 47 positive departures with almost the same range.

Example 4. Time of flight during the night of all Diptera in the years 1933, 1934 and 1935, and month by month in 1935–6

The light trap which is in use at Rothamsted is fitted with a mechanism so that eight killing bottles pass under the light in succession each night. The timing is so arranged that each bottle contains the insects caught in one-eighth of the night, starting from half an hour after sunset and ending half an hour before sunrise. Four of the bottles are before, and four after midnight.

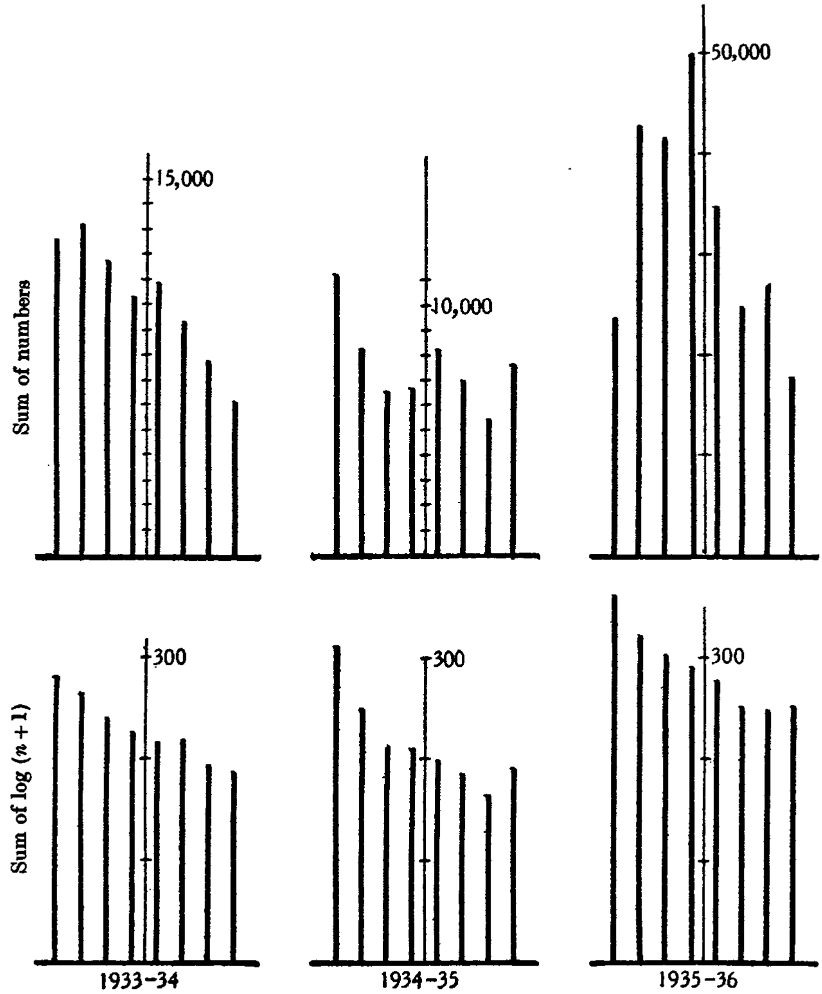
If over any series of nights the captures in bottle 1 or bottle 2 etc. are added together, a general measure of the abundance of insects caught at that time of the night can be obtained, and hence, from all the bottles, the distribution of the captures during the night.

Table III shows the number of Diptera caught in each period of the night in each month of 1935–6 on the basis of the sum of numbers and the sum of log $(n+1)$. The maxima in each month are in heavy type and the minima in italics.

Table III

Period of night ...	1	2	3	4	5	6	7	8
				Number				
1935 Mar.	**361**	214	140	139	117	166	107	*72*
Apr.	**187**	109	84	78	38	34	*9*	14
May	**2,204**	**2,204**	807	1,081	511	443	*206*	293
June	*4,190*	**16,859**	15,034	14,700	14,319	7,169	8,751	4,430
July	5,893	12,568	14,489	**23,356**	9,254	8,652	7,112	*5,771*
Aug.	5,746	6,592	**7,958**	6,950	5,015	3,717	6,599	*3,166*
Sept.	**2,580**	1,922	1,277	1,641	2,356	2,334	1,091	*1,018*
Oct.	**3,396**	1,462	954	1,176	1,552	787	*574*	1,354
Nov.	933	497	*238*	316	1,135	1,382	**2,601**	1,456
Dec.	113	81	69	*63*	79	67	**189**	128
1936 Jan.	198	233	**413**	111	201	131	112	*72*
Feb.	**49**	10	62	85	100	*6*	10	*5*
1935–36 Total	25,850	42,751	41,525	**49,696**	34,677	24,888	27,361	*17,779*
1933–34 „	12,564	**13,184**	11,756	10,304	10,810	9,341	7,798	*6,034*
1934–35 „	**11,193**	8,221	6,571	6,657	8,224	6,956	*5,433*	7,647
				Logarithm $(n+1)$				
1935 Mar.	**22·38**	15·90	13·30	10·79	9·10	9·71	10·60	*8·04*
Apr.	**14·50**	11·44	9·65	8·43	6·06	5·67	*2·28*	2·95
May	**29·76**	26·15	22·84	23·44	19·00	18·20	*15·07*	18·44
June	42·99	**49·42**	43·41	40·42	41·12	*34·68*	37·61	40·01
July	58·20	**62·55**	62·51	59·33	55·45	53·22	49·39	*45·37*
Aug.	**59·49**	57·31	55·58	52·40	49·35	*45·44*	46·63	45·57
Sept.	**46·22**	40·29	35·15	33·46	35·96	30·70	*26·57*	30·31
Oct.	**27·58**	22·68	*22·44*	25·55	24·78	24·48	23·75	*22·46*
Nov.	**31·56**	19·53	*15·65*	18·17	17·39	16·47	21·39	20·40
Dec.	**12·48**	7·04	7·30	8·35	7·99	*5·69*	8·47	11·01
1936 Jan.	**12·06**	6·72	10·33	8·41	6·14	*5·32*	6·17	6·94
Feb.	**5·04**	2·28	3·02	2·23	3·18	*0·85*	1·82	1·38
1935–36 Total	**362·26**	321·3	301·2	291·0	275·5	250·4	*249·8*	252·9
1933–34 „	**280·9**	264·4	242·1	226·3	216·8	219·9	194·3	*187·2*
1934–35 „	**311·8**	247·7	219·2	209·6	199·0	184·0	*164·5*	190·2

It will be seen that the logarithmic results are much more regular than the number results. In the latter case the maximum is in six months in the first period, twice in the second, twice in the third, once in the



All diptera

Fig. 3.

fourth and twice in the seventh. On the logarithm the maximum is in the first period in ten months and in the second in two.

When the annual totals are compared (Table III and Fig. 3) the numbers show a maximum once in the first period, once in the second and once in the third. The logarithm shows all 3 years with the maximum in the first period. The reason that the sum of numbers generally gives a maximum flight later in the night than the logarithmic sum is that nights with exceptionally high catches tend to be nights in which the activity is late. This can be tested by separating the nights in each month into classes based on whether the catches are above or below normal. When this is done both numbers and logarithms show similar

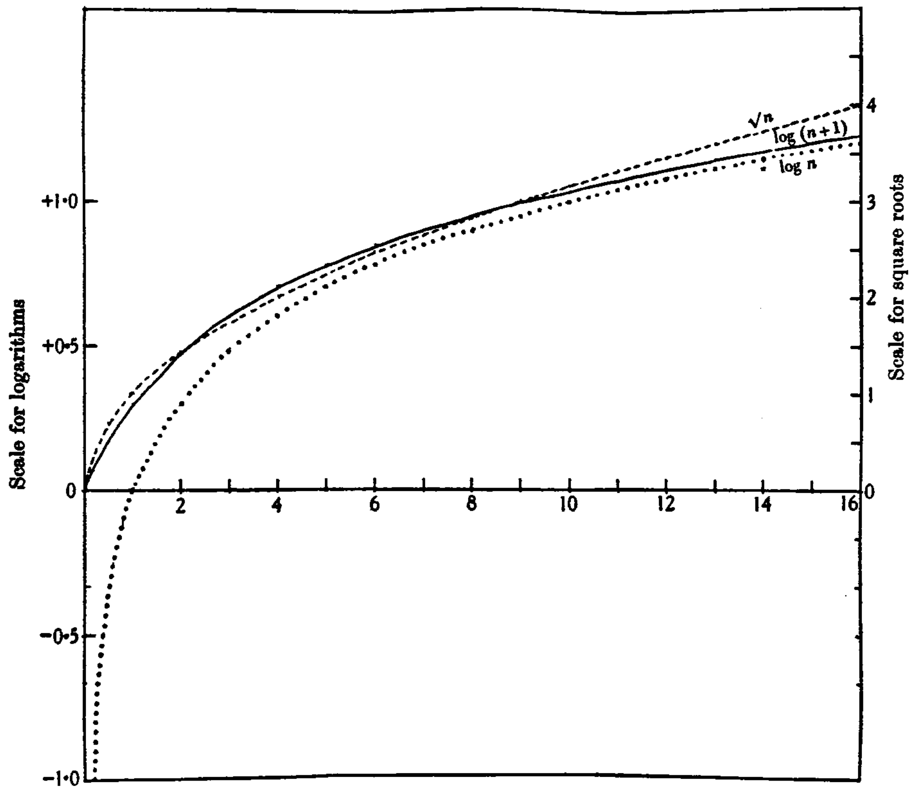


Fig. 4. ——— log $(n+1)$; ---- square root n; ..... log n.

differences, indicating that the effect is real and consistent (Williams, 1935, p. 533).

Note. Mr W. Yates of the Statistical Department at Rothamsted Experimental Station has drawn my attention to the fact that the curve of the relation $y = \log(n+1)$ closely approximates to the curve $y = \frac{1}{3}\sqrt{n}$ for low numbers, being identical when $n = 0$, 1·8 and 9. At values of n above 10 it gradually departs from the square-root curve and approaches more and more closely the curve $y = \log n$ from which it is practically indistinguishable (in the second decimal place) at values above 100.

Fig. 4 shows the relations graphically. The result is that by using the value log $(n+1)$ in analysis as above, we are substantially using a value proportional to $\sqrt{n}$ for low numbers and one proportional to log n for high numbers.

Summary

Evidence is brought forward that in comparing the numbers of insects caught under varying conditions, with particular reference to captures in a light trap, more consistent results are obtained if the geometric means are compared than by the use of arithmetic means. This is most conveniently done by summing the logarithms of the numbers instead of the numbers themselves. If any of the numbers in the series is zero it has been found practical to add one unit to all the captures in the series and so deal with log $(n+1)$ instead of log n.

The use of the logarithms prevents the swamping of the results in a series of observations by very high numbers on a single night. It also gives a more normal distribution of departures from a mean. As a result of the latter it is possible to apply the statistical formulae for standard deviation etc. which are not applicable to the skew curve obtained by the use of the departures of the numbers themselves from an arithmetic mean.

REFERENCES

Williams, C. B. (1935). The times of activity of certain nocturnal insects, chiefly Lepidoptera, as indicated by a light trap. *Trans. ent. Soc. Lond.* LXXXIII, 523–55.

—— (1936). The influence of moonlight on the activity of certain nocturnal insects. *Philos. Trans.* B, CCXXVI, 357–89.

(*Received* 12 *August* 1936)