Ind. J. mycol. Res. **1**, 31–38 (1955).

HETEROGENEITY IN THE LABORATORY DATA OF SYMBIOTIC NITROGEN FIXATION

By SATINATH BHADURI

Department of Botany, Presidency College, Calcutta

(With 2 Text-figures)

Abstract

Data on symbiotic nitrogen fixation are generally non-normal owing to different factors both inherent and environmental. Uniformity trials clearly indicate such non-normality of distribution of data on the number of nodules and the dry weight of plants.

In replicated well designed experiments in the laboratory and in the field, environmental factors are controlled; but the pooled estimate of error and consequential tests of significance of treatment effects are invalidated by such inherent non-normality which is often indicated in the correlation between means and variances.

This non-normal distribution can be effectively controlled by the logarithmic transformation of the data and this statistical tool can be successfully used in the laboratory data on symbiotic nitrogen fixation.

Introduction

In recent years replicated experiments and factorial designs have been successfully used in various types of field and laboratory experiments and Fisher's '*t*' and '*z*' tests are especially widely employed. These tests, however, assume the data to be normal or only slightly non-normal: but in strongly non-normal data such tests become invalid. Even when the data are non-normal the comparisons of the treatment differences can be made, provided that the standard deviation is not generally more than 30% of the mean, but if the deviations are of a higher order, non-normality vitiates the results to a great extent.

Experiments on symbiosis between clover plants and nodule bacteria always appear to show great heterogeneity and call for an examination of the distribution of the principal parameters of interest and the application of ordinary statistical methods. In a great variety of different experiments with clover on strain variation and related problems, a large body of data has been accumulated on the number and size of the nodules produced, dry weights of nodulated plants etc. The nodule number, on plants grown in tubes, has been shown to depend upon the concentration of bacteria in the inoculum within limits (Bhaduri, 1951), the volume of the media, rates of sowing and genetical characters of the plants themselves (Nutman, 1946, 1949). In majority of experiments, the type of inoculum, the volume and the nutritive nature of the media and rates of sowing the plants are largely controlled but great heterogeneity remains in the data owing to the practical difficulty in limiting the genetical variability of the plant. Further, heterogeneity results from the interactions of these plant factors with treatment effects. The raising of pure

line of plant material having particular genetical characteristics in unpractical although variability can somewhat be reduced by simple selection. Even when this is done, plant interaction with the bacteria is only partially controlled and the heterogeneity of the data is generally such that the ordinary tests of significance and methods of analysis of the data cannot give reliable and satisfactory information. The following work was undertaken in order to bring out the nature of such difficulties and to suggest means of dealing with them.

Uniformity trials and heterogeneity in experiments on symbiosis

A study of distribution was made utilising a variety of data obtained in experiments with clover plants. Each experiment by itself may be regarded as a uniformity trial, difference between the experiments lying in the age of the plants at which they were harvested and the time of the year during which they were grown. Typical examples of distribution of nodule numbers are presented in Table 1.

Table 1. **k** *and* **g** *statistics of experiments* (Dr. P. S. Nutman's data)

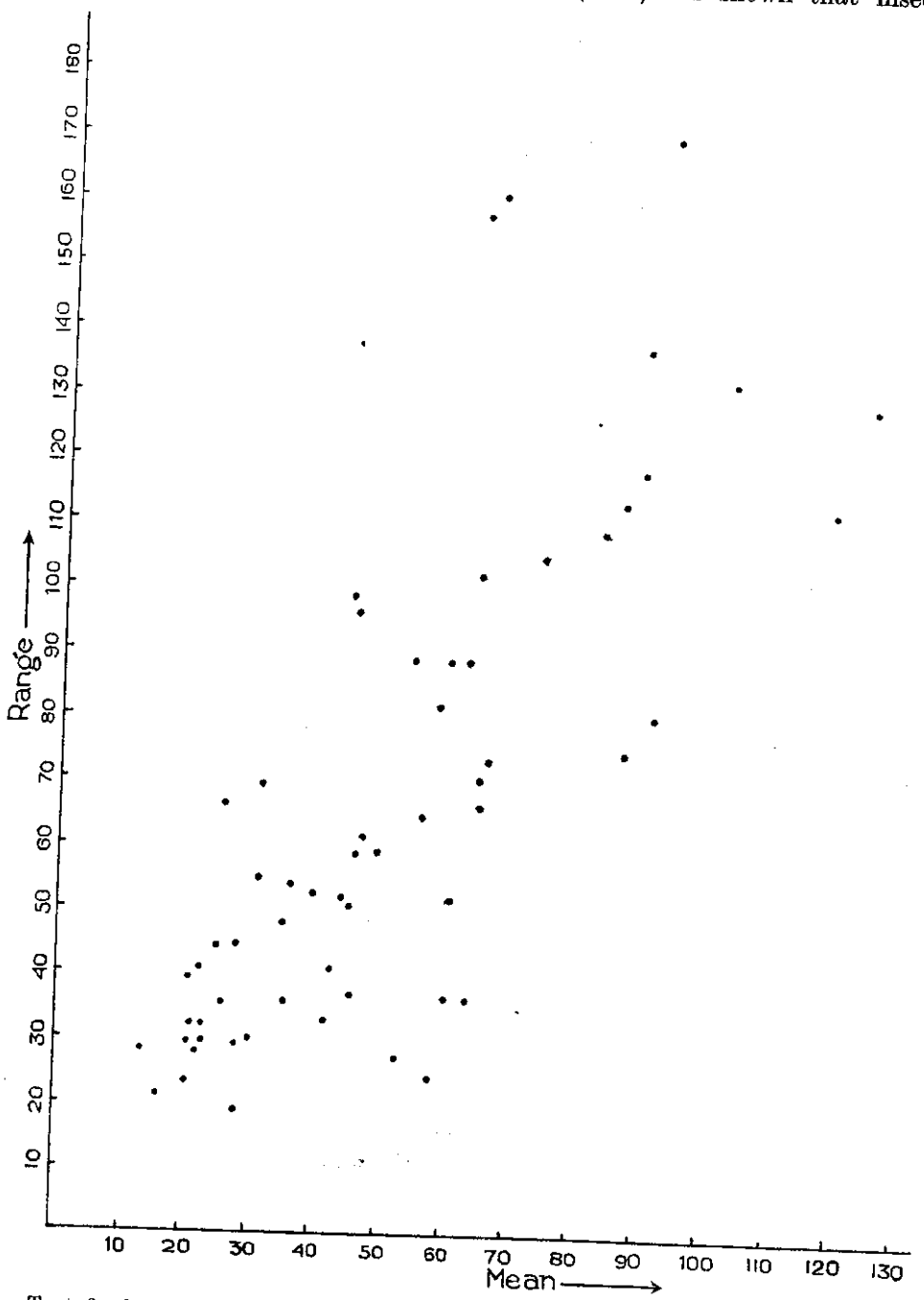
Experiment	D.F.	k-statistics	g-statistics	S.E. of g
I	58	$k_1 =$ 27.4		
		$k_2 =$ 87.0	$g_1 =$ −46.96	±0.095
		$k_3 =$ −38103.5		
		$k_4 =$ +21371363.0	$g_2 =$ −7569.00	±0.613
II	61	$k_1 =$ 24.5		
		$k_2 =$ 104.7	$g_1 =$ −15.42	±0.308
		$k_3 =$ −16530.5		
		$k_4 =$ +1206040.0	$g_2 =$ −109.94	±0.190
III	138	$k_1 =$ 23.6		
		$k_2 =$ 128.4	$g_1 =$ −17.33	±0.205
		$k_3 =$ +25214.1		
		$k_4 =$ −1102690.1	$g_2 =$ +168.01	±0.381

Table 1 shows the k-statistics and g-statistics calculated from the data of each experiment according to Fisher (1944). The highly significant g_1 and g_2 values of the experiments I and II shows that the distribution is markedly skew in nature and those of the experiment III kurtotic.

The correlation between standard errors and means of nodule numbers in 60 dissociation experiments are shown by plotting the ranges against the latter (text-fig. 1). In such cases, the analysis of ordinary data by standard methods is meaningless and transformation becomes necessary (Cochran, 1938). It can be shown algebraically that log x's are normally distributed when ranges or standard errors are proportional to the means, and the ordinary methods are available for the analysis of such transformed data. In text-fig. 2, the means and ranges of transformed data of the same dissociation experiments are shown to be quite independent.

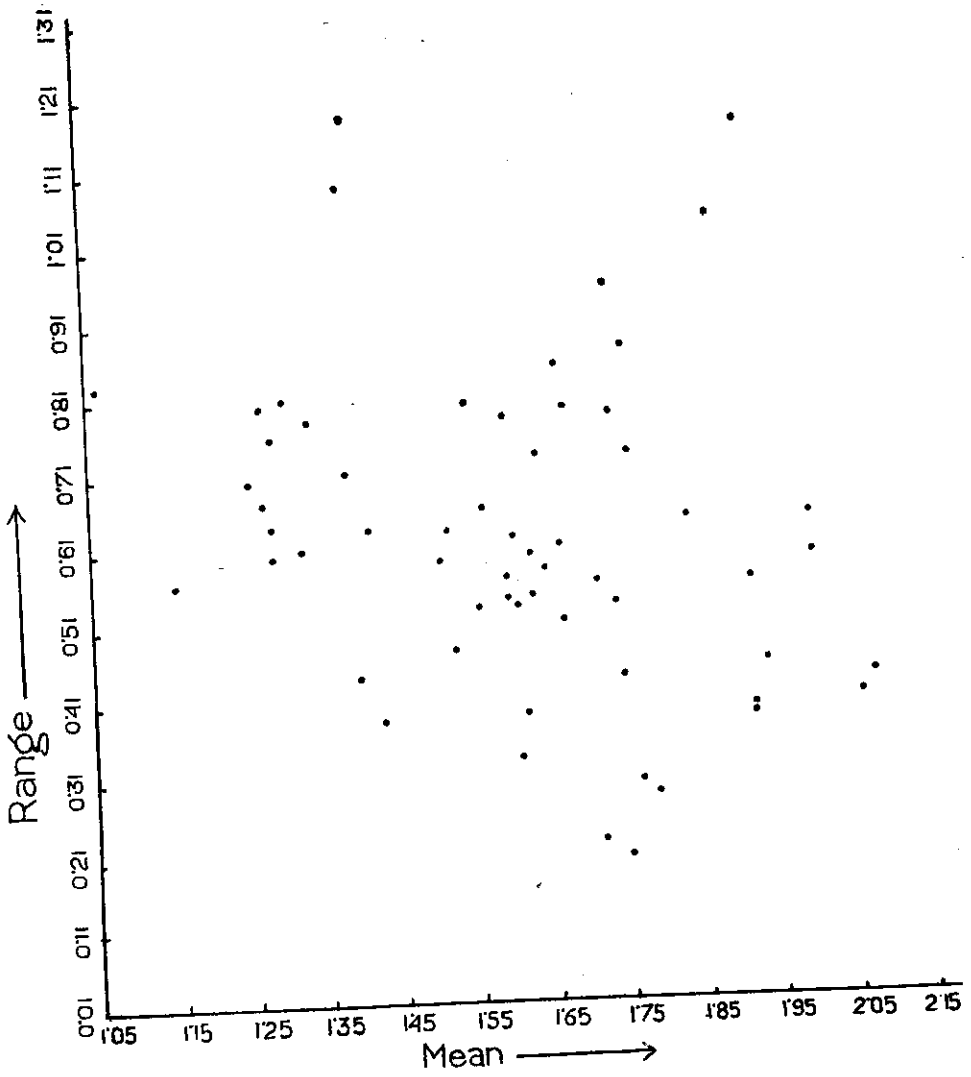
Logarithmic transformation of data has been successfully used in horticultural experiments where different treatments had very large effects

on the diameter of trees (Pearce, 1943). Cochran (1938) has given a summary of methods to be applied when peculiar data accumulates in the laboratory or in the field. Williams (1937) has shown that insect



Text–fig. 1. Diagram showing correlation between means and ranges in the raw data

catches had to be transformed logarithmically as variations in catches were in the order of several thousands from night to night.



Text-fig. 2. Diagram showing the absence of correlation between means and ranges in the log. transformed data

As far as can be ascertained, these methods have not been applied to experiments on symbiosis (Nutman, 1952), though Wilson (1940). when referring to objections that data furnished in experiments on the influence of pN_2 on nitrogen fixation were non-normal, states that similar conclusions can be drawn from logarithmically transformed data.

In comprehensive replicated experiments, the loss of orthogonality arises when one or more samples of the experiment may be lost by unforeseen circumstances, as premature death of the plant, presence of a resistant variety, albinoes, etc. It is very laborious to make necessary analysis after fitting constants to the data owing to the arithmatic involved, and in such cases, orthogonality may be restored by calculating the missing values (Yates, 1933) with necessary changes in the degrees of freedom and standard errors. It may be noted that such missing values should be calculated from the transformed data.

Application of logarithmic transformation to a factorial experiment

The experiment was designed to study the effect of the type of inoculum, influence of the variety of plants and the rate of sowing on the numbers of nodules produced and the effectivity of symbiosis as reflected on the dry weight of plants.

Plants were grown on slopes of agar in test tubes under bacteriologically aseptic condition as described by Thornton (1930). The volume of the medium has been shown to influence considerably on the number of nodules produced (Nutman, 1946), and hence was strictly controlled. The two varieties of clover plants used were '*g*' and '*x*' and the three bacterial strains were A121111, 211 and f12 (Nutman, 1946). The plant variety '*g*' was known to give an effective symbiosis with strain A121111 and 211, and the variety '*x*' which differed essentially from '*g*' only on being homozygous for a recessive gene gave an effective symbiosis with A121111 and ineffective symbiosis with 211. Both varieties, however, gave ineffective symbiosis with f12. The application of missing plot technique became necessary owing to the unavoidable seggregation in this material of plants totally resistant to nodule formation.

The experiment was set up with two varieties in three combinations of spacing viz.—(*i*) a plant growing alone, (*ii*) a plant growing in the presence of one of its own vareity and (*iii*) a plant growing in presence of the other variety, and when inoculated with each of the three strains. The whole scheme of treatments in this $2\times3\times3$ factorial design is summarised in Table 2.

Table 2. *Symbols used for different treatments*

Strains	Plant alone		Plant with own variety		Plant with the other variety	
	g	x	gg	xx	Gx	gX
A121111	Ag	Ax	Agg	Axx	AGx	AgX
211	Bg	Bx	Bgg	Bxx	BGx	BgX
f12	Cg	Cx	Cgg	Cxx	CGx	CgX

N.B.—A, B and C stand for the strains A121111, 211 and f12 respectively.

Instead of making two sets of tubes, for the last two columns of Table 2, with the same plants growing under identical conditions, it was worth while making only one set of tubes and counting the nodules of both '*g*' and '*x*' plants from the same tube, thus affecting a considerable reduction in the bulk of the experiment. In such a case a serious loss of efficiency is liable to occur, if there were a correlation between the symbiotic effects of the two plants in the same tube. No such correlation was, however, evident in the data.

Externally sterilised seeds were grown in test tubes containing 10 cc. of nitrogen defficient nutrient agar (Bhaduri, 1951) and were inoculated with bacterial strains when the plants were 21-days-old. Each test tube was supplied with 2 cc. of appropriate inoculum standardised by haemocytometer count to contain the same number of bacteria per cc. The water level was marked on the test tube and the volume of the liquid in the test tube was kept constant by adding requisite amount of sterile water periodically. Eight replicate tubes of each treatment were made and randomised in blocks. Plants were grown for three months after which nodule counts were made and dry weights of plants were determined.

The particular interest in this experiment lay in the study of the influence of spacing, plant vareity and the types of strains on the interaction which occurs between plants occupying the same tube.

On harvesting, it was found that as many as 6 tubes with '*x*' plants were without nodules in the whole experiment owing to plants being resistant. When such resistant plants were present in tubes containing both '*g*' and '*x*' plants, numbers of nodules and dry weights of both plants were calculated by the missing plot technique (Yates, 1933).

Table 3. *Comparison of different transformations of* nodule number *and* dry weight *data*

	Nodule numbers			Dry weights	
	Raw data	Square root transformation	Log. transformation	Raw data	Log. transformation
Replicate variance (7 d.f.)	1476.03*	4.797*	0.06817*	205.19**	0.0427*
Treatment variance (17 d.f.)	2764.79***	11.384***	0.14754***	1246.11***	0.3562***
Error variance (113 d.f.)	531.98	2.109	0.03177	72.15	0.01567
General mean	64.05	7.801	1.7591	24.35	1.3181
S.E. as percentage of the mean	±36.01%	±18.62%	±10.13%	±34.88%	±9.50%
Z, for Replicate variance (n_1= 7 d.f.) (n_2=113 d.f.)	0.50942	0.41099	0.39163	0.52190	0.50215
Z, for Treatment variance (n_1= 17 d.f.) (n_2=113 d.f.)	0.82433	0.84302	0.76779	1.4245	1.5394

* asterisk indicates 5% point of significance.
** and *** asterisks indicate 1% and 0.1% points of significance.

Table 4. *Computation of χ^2 of data on nodule number per plant*

No.	Treat-ment	Sum of Squares	D.F. n	$\frac{1}{n}$	Variance v	log. v
1	BgX	932.875	7	0.1429	133.3	2.1248
2	AgX	2094.875	7	,,	299.3	2.4761
3	Axx	2804.720	7	,,	400.7	2.6028
4	Bxx	1579.220	7	,,	225.6	2.3533
5	Cxx	4124.500	7	,,	574.9	2.7596
6	Bx	3988.000	6	0.1667	664.7	2.8226
7	Cx	1239.200	4	0.2000	309.8	2.4911
8	Ax	4850.520	6	0.1667	808.4	2.9076
9	CgX	7432.860	6	,,	1061.8	3.0258
10	Agg	879.470	7	0.1429	125.6	2.0990
11	Bgg	1010.875	7	,,	144.4	2.1596
12	BGx	2867.875	7	,,	409.7	2.6125
13	AGx	6156.000	7	,,	879.4	2.9442
14	Cgg	4523.375	7	,,	648.2	2.8117
15	Ag	4694.500	7	,,	670.6	2.8265
16	Bg	6662.875	7	,,	951.8	2.9785
17	CGx	4628.875	7	,,	661.3	2.8204
18	Cg	10281.500	7	,,	1464.8	3.1658
	Total	70752.120	120	—	—	—

(*N.B.*—For expalnation of treatment symbols see Table II).

$S(n.\ \log v) = 319.6440;$ $\quad S^2 = \frac{70752.120}{120} = 589.6;$ $\quad n.\log s^2 = 332.4720;$

$\therefore\ \chi^2 = 2.3020\ (n.\log s^2 - S(n.\log v)) = 29.54$, with 17 degrees of freedom

Corrected $\chi^2 = 28.05$; this is significant with a p between 0.05 and 0.02.

The analysis of variance of nodule numbers and dry weights of plants in Table 3 show the standard errors as high as 36.01% and 33.88% of the general mean respectively. Such high percentages of standard errors occur in non-normal data. Eighteen different treatment variances have also been tested for homogeneity by the method evolved by Bartlett (1936). In the computations shown in Table 4 the corrected $\chi^2 = 28.05$ for 17 degrees of freedom is clearly significant so that there are real differences between the variances of the treatments and as such the comparison of means with pooled estimate of error is incorrect. Although the Standard Error of nodule numbers is reduced to 18.62% by square root transformation, yet the correlation between means and variances is evident. The correlation is, however, completely eliminated by the logarithmic transformation and the pooled estimate of error, then, can be used successfully for testing treatment effects.

Acknowledgments

The author is indebted to Dr. H. G. Thornton, F.R.S. and Dr. P. S. Nutman for valuable criticism and encouragement during the work done at Rothamsted Experimental Station, England.

This paper was read before the Statistical Section of Indian Science Congress Association, Baroda Session, 1955.

References

Bartlett, M. S. (1936). Square root transformation in the analysis of variance. *J. R. Statist. Soc. Suppl.*, **3**, 68.

Bhaduri, S. N. (1951). Influence of the numbers of *Rhizobium* supplied on the subsequent nodulation of the host plant. *Ann. Bot., Lond.*, N.S., **15**, 219.

Cochran, W. G. (1938). Some difficulties in Statistical analysis of replicated experiments. *Emp. J. exp. Agric.*, **6**, 157.

Fisher, R. A. (1944). *Statistical Methods for Research Workers.* Oliver & Boyd, Edinburgh.

Nutman, P. S. (1946). Variation within strains of clover nodule bacteria in the size of the nudules produced and in the effectivity of the symbiosis. *J. Bact.*, **51**, 411.

Nutman, P. S. (1949). Nuclear and cytoplastic inheritance of resistance to infection by nodule bacteria in red clover. *Heridity*, **3**, 263.

Nutman, P. S. (1952). Studies on the physiology of nodule formation. III. Experiments on the excision of root tips and nodules. *Ann. Bot., Lond.*, N.S., **16**, 79.

Pearce, S. C. (1943). Statistical interpretation of vigour measurements of fruit trees. *J. Pomol.*, **20**, 111.

Thornton, H. G. (1930). The early development of the root nodule of lucerne (*Medicago sativa*, L.). *Ann. Bot., Lond.*, **44**, 385.

Williams, C. B. (1937). Use of logarithm in interpretation of nightly catches of insects. *Ann. appl. Biol.*, **24**, 404.

Wilson, P. W. (1940). *The Biochemistry of Nitrogen Fixation.* Univ. Wis. Press, Madison, Wisconsin.

Yates, F. (1933). The analysis of replicated experiments when field results are incomplete. *Emp. J. exp. Agric.*, **1**, 129.

(*Received for publication August 4, 1955*)