DISTRIBUTION OF SUGARCANE CLUMPS WITH REGARD TO TILLER NUMBER AND ITS TRANSFORMATION FOR ANALYSIS OF VARIANCE

By R. S. Srivastava

Indian Institute of Sugarcane Research, Lucknow

1. Introduction

Among the various characters of sugarcane plant taken into account in varietal selection the number of tillers in a clump is an important one. A variety which within limits, other factors being equal characteristically gives a large number of tillers per clump forming good canes at the time of harvest is preferred. The object of the present paper is to study the distribution of clumps according to the number of tillers. This has a twofold purpose; one is to evolve a probability model consistent with the physiology of tiller formation in the clump, and the other is to find a transformation for making the method of analysis of variance appropriate for comparing the varieties.

A clump starts its formation from a single bud. The bud, when placed in the soil, sends its mother shoot above the soil surface after germination. In the portion of this shoot below the soil surface many a small node are formed, upon each of which a bud similar to that of the parent type develops. Many of these buds germinate and send their respective shoots, commonly known as B shoots or first tillers above the ground level. Upon the underground portion of each of these tillers, again buds similar to original develop and it is probable that many of them may germinate and send their tillers above the ground. It may be pointed here that in this vegetative scheme every bud at any step is similar to the original type and one bud gives rise to only one tiller. Barber[2] has observed that under ideal conditions this process of tillering would continue indefinitely. In nature, however, this process seldom goes up to three or four steps, and all the tillers coming out of the ground developing from one parent bud belong to one clump.

2. Distribution of Clumps according to the number of Tillers

An attempt can be made to fit a theoretical model to this process. Let us assume that throughout this process the probability of germination of any bud at any step be p and of failing to germinate be $q = (1 - p)$; and the germination of buds on any one shoot be mutually exclusive of one another. Further, let the probability that each germinated bud reproduces a number of buds on the underground portion of its shoot capable of germination at the next step be c. Then if the first n steps results in r tillers only in a clump, the conditional probability of getting another tiller at $(n + 1)$-th step is $(p + rc)/(1 + nc)$ and of not getting is $q + (n - r)c/1 + nc$. If this addition of tiller in the clump is considered irrespective of the step at which this occurs, then the absolute probability of getting exactly r tillers in the first n steps will be given by

$$P(r; n) = (n_r) \frac{p(p+c)(p+2c)\ldots(p+\overline{r-1}c)\, q(q+c)(q+2c)\ldots(q+\overline{n-r-1}c)}{1(1+c)(1+2c)\ldots(1+\overline{n-1}c)}. \tag{2.1}$$

Since the process is to continue indefinitely, we have the condition that $n \to \infty$. Further let $\lim_{n\to\infty} np = m$ and $\lim_{n\to\infty} nc = 1/k$. Under these conditions let $P(r; n) \to P(r)$ then it has been shown by Polya[5] that

$$P(r) = \binom{k+r-1}{r} \frac{\left(\frac{k}{m}\right)^k}{\left(1+\frac{k}{m}\right)^{r+k}}$$

$$= \binom{k+r-1}{r} \left(1+\frac{m}{k}\right)^{-k} \left(\frac{m}{m+k}\right)^r. \tag{2.2}$$

Thus (2.2) gives the probability of finding exactly r tillers in a clump. This distribution depends on two parameters m and k. Anscombe[1] has observed that there is theoretical evidence to show that k depends upon the intrinsic character of the buds whereas m depends on external factors.

The distribution given by (2.2) is Negative Binomial and it is found to occur frequently in biological problems. Since k may vary from 0 to ∞ depending on the intrinsic nature of the buds, the above distribution would become identical with Fisher's Logarithmic distribution when $k \to 0$ and with Poisson when $k \to \infty$. Thus if the sugarcane tiller formation corresponds in reality to the above model then the observed data should fit to the distribution of (2.2) in general or in extreme case to the Poisson or Fisher's Logarithmic distribution.

Anscombe[1] has considered the estimation of m and k from a large sample of one set. The best estimate of m is $\bar{r}$ and it is always fully efficient. He has recommended the following two methods for the estimation of k and has pointed out the circumstances in which they provide efficient estimates:

Method A.—If s^2 is the sample variance then

$$k = \frac{\bar{r}^2}{s^2 - \bar{r}}. \qquad (2.3)$$

Method B.—From (2.2) the probability of $r = 0$ or the chance of the clumps having only mother shoot is $(1 + m/k)^{-k}$. If the size of sample be N and n_0 the observed frequency of clumps having only mother shoot, then

$$n_0 = N\left(1 + \frac{\bar{r}}{k}\right)^{-k} \qquad (2.4)$$

gives the estimate of k.

Among these two methods, method A should be used if k is greater than one and the Method B if k is less than one. He has also found that the errors of estimation of m and k are independent when N is large.

3. Material

(*a*) For investigating into the nature of the population, one replication of the Standard Varietal Trial at the farm of the Indian Institute of Sugarcane Research, Lucknow, comprising of twelve varieties, each in a plot of 55 ft. length and 15 ft. width resulting in 275 ft. row length to each variety was taken. At the time of planting, three budded sets were placed end to end and when it was assured that the tillering phase had been completed actual tiller counts of each clump were made. While counting it was made sure that no two clumps overlapped and the whole counting was completed within three days—11th, 12th and 13th June 1958.

(*b*) For considering the relationship between the mean and the variance, the data of an earlier experiment were employed. In it the total germinated clumps of variety Co. 453 were ranked in a serial order and forty clumps among them were selected on random basis and tagged. For twenty consecutive weeks tiller counts were made on these clumps commencing from 25th April 1957 and ending on 11th November 1957.

(*c*) In the replication considered above for population study, from the plots of six different varieties, samples of eight clumps from each row, forty in all for each variety were selected on random basis. The data of the tiller counts of these clumps were utilised to examine if the transformation made the analysis of variance appropriate for comparing the variation with regard to the number of tillers.

4. Fitting of the Distribution on Actual Data

The frequency chart of clumps having tillers 0, 1, 2, $\cdots$ for all the twelve varieties as found from Section 3 (*a*) has been given in Table I. It also gives the total number of clumps, the mean and the variance of tillers. The number of clumps is sufficiently large so that the methods valid for large samples may be used. The fitting of the Poisson distribution to this data does not arise because the variance in each case happens to be larger than the mean.

The last column of Table I gives the over-dispersion which is (1-mean/variance) as defined by Quenouille.[6] When the over-dispersion is significant number of distributions have been suggested. Few of these are Negative Binomial, Polya, Aeppli, Neyman's Contagious and Fisher's Logarithmic distributions. Among these perhaps the easiest to fit and widely applicable to the biological data is Negative Binomial. Incidentally, our theoretical approach has also led us to this distribution in general, and only as extreme cases to Poisson or Logarithmic distribution. We have, therefore, tried to fit Negative Binomial as indicated in Section 2. Table II gives for each variety the expected frequencies of clumps having different number of tillers. The values of the parameters m and k have also been given along with the Chi-square value for each series.

The tests other than Chi-square described by Anscombe[1] and based upon the expected second and third moments were not applied because the tail end of each series consisted only small values. The Chi-square value in each case is significantly low and the expected frequencies very closely resemble the observed ones. The k's are neither very small nor very big so that it can safely be concluded that the Negative Binomial model is quite appropriate and the Poisson or Logarithmic distribution is not expected to occur with the clump distribution.

5. Transformation of the Data

The knowledge that the distribution of clumps is non-normal requires that the analysis of variance should not be used without examining the conditions for its validity. Bartlett[5] has suggested $k^{-\frac{1}{2}} \sinh^{-1} \sqrt{kx}$ transformation where x varies according to the Negative Binomial distribution.

TABLE I

Frequency chart of sugarcane clumps according to the number of tillers

Variety	Tillers																		Total (n)	Mean ($\bar{r}$)	Variance (s^2)	Over-Dispersion
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17				
Co. 453 ..	37	68	61	49	25	23	13	5	2	3	2	1	1	..	..	..	..	..	290	2·6	4·59	0·434
Co. 527 ..	22	44	41	30	26	19	8	9	6	1	0	1	1	0	1	..	..	..	209	3·0	5·47	0·452
Co. 963 ..	18	23	50	40	47	39	22	9	6	1	0	0	1	0	1	0	1	..	258	3·6	4·88	0·262
Co. 967 ..	21	32	56	36	35	32	18	11	3	5	3	1	1	..	..	..	..	..	254	3·4	5·38	0·362
Co. 975 ..	15	35	46	40	44	36	21	13	5	1	3	1	..	..	..	..	..	..	260	3·5	4·67	0·251
Co. 976 ..	15	28	58	53	47	38	28	7	6	3	..	..	..	..	..	..	..	..	283	3·4	4·05	0·160
Co. 1058 ..	8	21	31	44	45	44	36	28	22	11	7	6	4	2	3	3	1	2	318	5·2	8·96	0·420
Co. 1062 ..	13	27	42	45	48	36	26	18	10	8	1	0	3	1	..	..	..	..	278	4·0	5·85	0·316
Co. 1096 ..	33	34	52	37	31	23	15	7	2	5	1	..	..	..	..	..	..	..	240	2·9	4·76	0·391
Co. 1105 ..	35	48	39	33	37	18	14	5	9	2	0	0	1	..	..	..	..	..	241	2·8	5·33	0·475
Co. 1107 ..	14	28	42	39	33	25	23	21	9	9	8	1	0	1	..	..	..	..	253	4·1	7·02	0·416
Co. 1111 ..	18	33	42	43	42	34	23	15	7	4	2	3	1	..	..	..	..	..	267	3·7	5·49	0·326

TABLE II

Expected frequency chart of sugarcane clumps according to the number of tillers

Variety	Tillers													Parameters		Chi-square	Degrees of freedom
	0	1	2	3	4	5	6	7	8	9	10	11	Over 11	*k*	*m*		
Co. 453 ..	42	62	59	46	32	20	12	7	10	..	..	..	..	3·39	2·6	4·175	6
Co. 527 ..	23	39	40	34	26	18	12	7	10	..	..	..	..	3·64	3·0	3·140	6
Co. 963 ..	12	32	46	49	42	31	21	13	12	..	..	..	..	10·12	3·6	11·804	6
Co. 967 ..	17	37	47	46	37	27	18	11	14	..	..	..	..	6·00	3·4	6·619	6
Co. 975 ..	13	33	48	50	42	31	20	12	6	5	..	..	..	10·51	3·5	3·713	7
Co. 976 ..	13	36	54	57	47	33	20	11	12	..	..	..	..	17·78	3·4	8·826	6
Co. 1058 ..	6	19	33	42	45	43	36	29	21	15	10	7	12	7·19	5·2	4·059	10
Co. 1062 ..	10	28	43	49	45	36	26	17	11	6	7	..	..	8·65	4·0	2·906	8
Co. 1096 ..	26	45	49	41	30	20	12	7	10	..	..	..	..	4·51	2·9	7·180	6
Co. 1105 ..	33	48	47	38	28	19	12	7	9	..	..	..	..	3·10	2·8	6·991	6
Co. 1107 ..	11	27	38	42	39	31	23	16	11	7	8	..	..	5·57	4·1	6·600	8
Co. 1111 ..	18	33	42	43	42	34	23	15	7	10	..	..	..	7·65	3·7	3·189	7

The $\sinh^{-1}$ transformation has been used by Beall[3] but this involves k which either should be known from previous experience or estimated from the data before it is applied for analysis of variance. The data of the Section 3 (*b*) was employed to estimate the value of k from a small sample. The mean and the variance for twenty consecutive weeks are given in Table III. The time variation had to be kept because the experimenter may like to count tillers at any time when the crop is standing.

TABLE III

Mean and variance of tillers for twenty weeks

Week number	0	1	2	3	4	5	6	7	8	9
Mean ..	0·22	1·13	2·18	2·60	2·73	3·35	3·93	4·75	5·28	6·73
Variance ..	0·44	2·25	3·00	3·24	2·68	3·43	3·49	5·44	8·65	10·79
Week number	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
Mean ..	7·95	6·75	6·00	5·05	4·79	4·47	4·50	4·71	4·79	5·06
Variance ..	17·85	12·94	10·30	7·70	5·51	6·76	7·09	9·16	9·06	7·01

Both the mean and variance appear to vary with time and value of the partial correlation coefficient between mean and variance comes to 0·9074 which shows that 81% of the change in variance may be attributed to mean. To obtain the estimate of k, curve u (variance) $= ax$ (mean) $+ bx^2$ was fitted to the data of Table III by the Method of Least Squares. This relation was taken with the consideration that it gives the transformation

$$\theta(x) = \int \frac{c}{\sqrt{ax + bx^2}}\,dx = c'\sqrt{\frac{a}{b}}\ \sinh^{-1}\sqrt{\frac{bx}{a}}$$

analogous to that given by Bartlett.[3] The multiple correlation coefficient was calculated for finding out if the approximation was appropriate. Its value came to 0·9608 which indicates that even for small samples the transformation may work well. The value of k in the present case comes to 0·370. In fact the experimenter should himself decide about the value of k either from a population study or from some previous knowledge. The values found out in the population study of Table II varied from 3·10 to 17·78. It appears that the minimum value of k for sugarcane clump population would be near about one, and therefore the data collected at Section 3 (*c*) were analysed by transforming with value of $k = 1$. Further, Bartlett[5] has suggested that when the observed values r are small then $(r + 0 \cdot 5)$ should be used instead of r. The analysis of variance with the observed values and both the transformed values is presented in Table IV.

TABLE IV

Analysis of variance

Source	d.f.	Untransformed			$\sinh^{-1}\sqrt{r}$			$\sinh^{-1}\sqrt{r+0\cdot5}$		
		S.S.	M.S.	F	S.S.	M.S.	F	S.S.	M.S.	F
Between varieties	5	336·34	67·27	12·96*	8·36	1·67	7·66*	4·28	0·86	9·07*
Within varieties	24	219·70	..	..	5·85	..	..	2·65	..	..
Residual	210	1090·13	5·19	..	45·81	0·22	..	19·82	0·09	..
Total ..	239	1646·17			60·02			26·75		

* This denotes significance at 1 per cent. level.

TABLE V

Summary of Results

	Co. 1058	Co. 1062	Co. 975	Co. 1105	Co. 1096	Co. 453	C.D.
Untransformed data	5·73	4·05	3·08	2·95	2·43	2·25	1·0284
Transformed data ($\sinh^{-1}\sqrt{r+0\cdot5}$)	1·56	1·44	1·32	1·30	1·23	1·15	0·1347

The relative efficiencies because of transformations come out to 23·79 and 54·99%. It appears that when the numbers are small, the transformation due to Bartlett doubled the efficiency of the analysis of the variance. The summary of results given in Table V from the analysis of variance of the untransformed and the transformed data are different from one another. The differences which were not previously visible became apparent. Thus the $\sinh^{-1}$ transformation besides being valid to the tiller data is also necessary for discriminating between the varieties with regard to tillers in clumps.

6. SUMMARY

Theoretically it has been found that the distribution of sugarcane clumps with regard to the number of tillers is a Negative Binomial. The observations made from twelve large samples of different varieties have been found to agree with this model to a significant degree.

The analysis of variance on the number of tillers r in a clump requires modification. The observed values should be first transformed to $\sinh^{-1}\sqrt{r}$ (or to $\sinh^{-1}\sqrt{r+0\cdot5}$ if r be small) and then analysed. This is not only valid for small samples, but also necessary to bring out the varietal differences more precisely.

9. ACKNOWLEDGMENT

I am very much indebted to Dr. A.R. Roy, Professor of Statistics, Lucknow University, who guided me in preparation of this paper. I am also highly grateful to Shri R. R. Panje, Director, Indian Institute of Sugarcane Research, Lucknow, and Dr. K. Kishen, Chief Statistician, Department of Agriculture, Uttar Pradesh, whose discussions helped me very much.

8. REFERENCES

1. Anscombe, F. J. .. "The statistical analysis of insect counts based on negative binomial distribution," *Biometrics*, 1949, **5**, 165–73.
2. Barber, C. A. .. "Studies in Indian Sugarcane No. 4. Tillering or underground branching," *Mem. Dept. Agric. Ind. Bot. Ser.*, 1919 **10** (2), 39–153 (Original not seen).
3. Beall, G. .. "The transformation of data from entomological field experiments so that analysis of variance becomes applicable," *Biometrika*, 1942, **32**, 243.
4. Bartlett, M. S. .. "The use of transformation," *Biometrics*, 1947, **3**, 39–52.
5. Polya, G. .. *An Introduction to Probability Theory and Its Applications*, 1950, **1**, 128 (by W. Feller, publ. by John Wiley and Sons).
6. Quenouille, M. H. .. "A relation between the Logarithmic, Poisson and Negative Binomial series," *Biometrics*, 1949, **5**, 162–64.