Journal of the
Indian Society of Agricultural Statistics

XVII] 1965 [No. 2

ESTIMATION OF STALK-BORER INCIDENCE IN SUGARCANE CROPS

By R. S. Srivastava

Indian Institute of Sugarcane Research, Lucknow

Introduction

The borer incidence in sugarcane crop is defined as the proportion of the infested stalks in the total population. Major attempts to estimate it by suitable sampling methods have been made by Khanna and Bandyopadhyay (1951), Rojas (1953) and Srivastava (1957). Khanna and Bandyopadhyay have considered the incidence values on the primary units belonging to a Poisson distribution and accordingly applied the Square-root transformation. Rojas has assumed these to belong to a Normal distribution while Srivastava has taken this to be a Binomial distribution and has converted all these values by inverse sine transformation. These transformed values were then analysed and the sampling scheme which gave the lowest coefficient of variation or within unit variance was recommended for general purposes. These recommendations are not similar and experience has shown that the estimates obtained from these were considerably different from the actual values. This, therefore, calls for further research and presently a probe into the assumed existence of a specified population and the transformation applicable thereto has been made.

Problem

If 'X' and 'Y' be the number of infested and the total stalks in the population and 'x' and 'y' the corresponding numbers in a random sample, then the incidence is defined as the proportion 'X/Y' and for a given precision, the problem is to estimate 'X/Y' with the help of the sample values.

It is well known that for the application of any statistical theory the main consideration is that for the random variable under study there should exist a probability distribution having a known mathematical form. According to Deming (1950) this shape is reproduced hour after hour, day after day, so long as the process remains in statistical control, that is, exhibits properties of randomness. Unless this holds, the statistician can never make meaningful and useful predictions about the future samples. Thus the distribution is either known to exist or assumed to exist from the past experience. As stated, the previous workers have assumed the distribution of the incidence values on the primary units to the Poisson, Normal or Binomial distribution. Let this postulate be examined.

In any sampling scheme of sugarcane, it is physically impossible to take the stalks or the clumps, as primary units because this demands at least an ordering of all the stalks or the clumps in the population so that the selection of a random sample may be possible. On the other hand, if portions of row lengths or small area units are considered then it is almost sure that units cannot be so framed as to have equal number of stalks. Thus in any primary unit both the infested (x) and the total stalk (y) will be subject to error, that is, will have variations. If the variations be small in magnitude, then the above assumptions might hold; but if they are not 'x' and 'y' will have to be considered separately. Khanna and Bandyopadhyay have observed that 'there is a considerable amount of non-homogeneity in the incidence data'; the scrutiny about the distributions of 'x' and 'y', therefore, becomes all the more essential. If both 'x' and 'y' are subject to error, then the incidence values 'x/y', $0 \leqslant x \leqslant y$, obtained from the primary units do not form a consistent estimate for 'X/Y', and then a suitable statistic to estimate the incidence in the population will have to be found out. It is on these lines that the investigations have been carried out and are presented in this paper.

For this, the primary units of different row length cuts have been considered and from the change in the mean and variance of 'x' and 'y' separately the shape of the distribution of 'x' and 'y' has been investigated and the statistic based on a valid transformation has been suggested. Lastly, with the help of a random sample, the relative efficiencies of the statistic based on the above transformation in relation to other ones have been calculated.

MATERIAL

Two plots of $55' \times 27'$ of sugarcane having nine rows of $55'$ row length each at the farm of the Indian Institute of Sugarcane Research

Lucknow, were taken for these studies. The number of stalks in each clump was noted and the location of clumps in the plot was marked to scale on a graph paper. Complete enumeration of the stalk-borer (*Chilotraea auricilia* Ddgn.) infested stalks was done by splitting each stalk. Such enumerations were made in the Entomology Section and the data of two years—1954–55 and 1957–58—were utilized. These two years were deliberately chosen because of the low infestation in one and the high in the other.

These data when reproduced on a graph paper could be utilized for proper subdivisions into primary units. The entire row lengths were partitioned into cuts of 1′–11′ and the number of infested and total stalks were counted in each cut. Those units which had no stalks were left out of the count. Thus the number of primary units, their

TABLE I

The means and the variances of the infested and the total stalks in primary units of varying lengths

Length of the primary unit in feet	1954–1955					1957–58				
	Total No. of units	Infested stalks		Total stalks		Total No. of units	Infested stalks		Total stalks	
		Mean	Variance	Mean	Variance		Mean	Variance	Mean	Variance
1	2	3	4	5	6	7	8	9	10	11
1	371	1·09	1·299	3·79	5·543	262	1·90	4·027	3·84	6·254
2	236	1·71	2·567	5·96	11·398	205	2·43	5·786	4·90	10·394
3	162	2·49	4·022	8·62	15·196	153	3·08	7·212	6·31	13·968
4	123	3·31	6·116	11·33	22·043	117	4·02	10·455	8·21	20·716
5	99	4·08	7·044	14·24	21·800	96	4·76	12·686	9·53	23·294
6	82	4·94	9·311	17·11	29·073	79	5·63	14·214	11·63	24·022
7	70	5·77	12·405	20·10	35·861	69	6·72	19·358	13·58	35·838
8	61	6·62	18·268	22·67	45·139	61	7·72	17·120	16·18	38·540
9	55	7·35	16·953	25·40	53·876	55	9·05	29·910	17·83	45·610
10	49	8·20	18·122	28·31	53·151	49	9·88	29·290	19·65	40·635
11	45	8·98	21·355	31·27	48·862	45	10·53	36·855	22·31	55·330

location, the number of infested and total stalks in each of them were recorded.

For different lengths of the primary units, the number of units in the plots, the means, variances for the stalk-borer infested and total stalks for the years 1954–55 and 1957–58 have been given in Table I.

It is seen that both the means and variances have steadily increased with the size of the primary units. As such it appears that even when 'x' and 'y' are both considered separately the increase in the size of primary unit does not decrease the within unit variation so that specific unit might be recommended for sampling. Further, the variances for all the series are always higher than the mean, which indicates that there might be a functional relationship between the variance and the mean quite different from Poisson for which the variance and mean are equal or Binomial for which the variance is less than the mean. The investigations for this relationship were, therefore, carried out by fitting distributions on the observed series of infested and total stalk counts.

FITTING THE DISTRIBUTION ON THE DATA

The observed number of infested and total stalks in the primary units from 1′ to 6′ units, the mean and variance for each series have been shown in Table II. The series for units higher than 6′ have been left out because for applying x^2 test too many cell frequencies had to be grouped together making the fit too artificial.

The analysis regarding the shape of the population is invariably made with one of two aims (*a*) to fit in a biological process for which a distribution is known and (*b*) to interpret the data by a suitable transformation which may be valid with the knowledge of the distribution. From Table II it is evident that all the series have an over-dispersion because the variances are significantly greater than the mean. When the over-dispersion is present, a number of distributions depending on the increasing skewness and tail length have been examined by Anscombe (1950). These are the Thomas, Fisher's logarithmic, Neyman's contagious, Polya Aeppli and Negative Binomial (*NB*) distributions. Of these Thomas and Neyman's contagions admit of a number of modes, Polya Aeppli two and *NB* only one. The logarithmic distribution, though having one mode, is more skew. Thus considering the number of modes and the tail-length the *NB* distribution was preferred for investigation if this might be useful for the interpretation of the data.

Following Anscombe for *NB* distribution, the probability of 'r' counts is given by

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

This has two parameters 'm' and 'k'. The best estimate of 'm' is '$\bar{r}$' the sample mean and is fully efficient. The 'k' is estimated by $\{(\bar{r})^2/(S^2-\bar{r})\}$ where S^2 is the sample variance, and for a large sample the errors of estimation of 'm' and 'k' are independent.

With the mean and variances as obtained for each series of Table II, 'm' and 'k' were estimated and the expected frequencies and the values of χ^2 as obtained have been shown in Table III.

The low values of χ^2 show that the *NB* distribution can be profitably used to interpret not only the distribution of the infested stalks, but for the total stalks also. It now becomes clear that the incidence values cannot be considered as they are, but a suitable statistic based on *NB* distribution should be employed to estimate incidence on a given precision.

The Statistic to Estimate Incidence

Let 'r' be the observed number of counts from a *NB* distribution. The variance 'V' of r is then functionally related to the mean 'm' by $V = m + m^2/k$. A new variable 't' a function of 'r' but independent of 'm' will, therefore, be given by

$$t = \int \frac{1}{\sqrt{r+\frac{r^2}{k}}}\,dr = k^{-\frac{1}{2}}\log\left(\sqrt{1+kr}+\sqrt{kr}\right).$$

This variate 't' is a normal variate and Beall (1954) has used this transformation on actual data and mentions that 'in particular the third moment of the transformed variable seems to approach zero'. Thus if a sample of 'n' units is chosen at random and the counts of the infested and total stalks be $(x_1, y_1), (x_2, y_2), \cdots, (x_n, y_n)$, then the incidence is to be estimated either by

$$T_1 = \frac{1}{n}\sum\frac{x_i}{y_i} \quad \text{or} \quad T_2 = \frac{\sum\frac{x_i}{n}}{\sum\frac{y_i}{n}}.$$

TABLE

Observed number of primary units having different

Number of infested stalks	1954–55 Size of primary unit in feet						1957–58 Size of primary unit in feet					
	1	2	3	4	5	6	1	2	3	4	5	6
0	135	52	23	8	2	1	69	51	25	16	9	6
1	130	78	39	22	13	6	70	41	31	17	13	4
2	65	54	27	27	19	11	46	30	19	11	8	11
3	25	22	32	22	18	14	39	31	23	17	11	8
4	13	12	22	14	11	13	13	15	13	10	11	6
5	2	8	3	8	10	9	8	14	15	13	8	6
6	0	6	4	5	5	7	6	6	10	6	8	3
7	1	4	9	8	6	4	4	8	7	6	6	9
8	..	..	2	3	8	6	4	4	2	9	6	9
9	..	..	1	6	2	2	0	1	2	5	5	4
10	..	..	..	..	5	4	2	3	4	7	2	4
11	..	..	..	..	..	5	1	1	2	..	3	3
12	..	..	..	..	..	..	..	..	..	..	3	3
13	..	..	..	..	..	..	..	..	..	..	3	2
14	..	..	..	..	..	..	..	..	..	..	..	1
15	..	..	..	..	..	..	..	..	..	..	..	..
N	371	236	162	123	99	82	262	205	153	117	96	79
Mean	1·09	1·71	2·49	3·31	4·08	4·94	1·90	2·43	3·08	4·02	4·76	5·63
Variance	1·299	2·567	4·022	6·116	7·044	9·311	4·027	5·786	7·212	10·455	12·686	14·214

II

numbers of stalk-borer infested and the total stalks

Number of total stalks	1954–55 Size of primary unit in feet						1957–58 Size of primary unit in feet					
	1	2	3	4	5	6	1	2	3	4	5	6
1	61	20	2	..	..	..	41	..	8	3	2	1
2	72	22	6	1	..	..	44	24	8	3	2	6
3	59	20	6	3	..	..	63	25	20	12	6	2
4	54	27	10	2	..	..	32	36	23	10	7	4
5	45	27	13	5	..	..	29	32	17	11	4	5
6	39	23	14	9	..	..	18	20	15	11	7	4
7	17	28	15	8	5	..	12	16	13	11	6	6
8	7	15	21	13	4	..	9	11	9	8	14	3
9	8	14	10	9	5	..	4	9	6	8	5	10
10	2	6	12	9	10	8	3	8	12	6	8	5
11	3	7	15	8	8	7	4	8	9	5	3	5
12	2	4	11	8	9	2	3	6	4	8	5	6
13	2	8	11	6	8	4	..	10	2	2	4	7
14	..	4	7	4	5	6	..	..	3	4	7	1
15	..	1	3	12	9	3	..	..	2	3	3	4
16	..	..	6	9	5	5	..	..	2	12	..	2
17	..	..	..	8	5	11	..	..	..	..	5	2
18	..	..	..	1	3	11	..	..	..	..	5	3
19	..	..	..	1	6	4	..	..	..	..	1	2
20	..	..	..	7	6	2	..	..	..	..	2	1
21	..	..	..	..	11	1	..	..	..	..	.	..
22	..	..	..	..	..	4	..	..	..	..	..	..
23	..	..	..	..	..	3	..	..	..	..	..	..
24	..	..	..	..	..	3	..	..	..	..	..	..
25	..	..	..	..	..	8	..	..	..	..	..	..
N	371	236	162	123	99	82	262	205	153	117	96	79
Mean	3·79	5·96	8·62	11·33	14·24	17·11	3·84	4·90	6·31	8·21	9·53	11·63
Variance	5·543	11·398	15·196	22·043	21·800	29·073	6·254	10·394	13·968	20·716	23·294	24·022

TABLE

Expected number of primary units having different

No. of infested stalks	1954-55 Size of primary unit in feet						1957-58 Size of primary unit in feet					
	1	2	3	4	5	6	1	2	3	4	5	6
0	137	59	23	9	5	2	73	45	22	11	6	6
1	125	67	36	19	11	6	66	46	29	16	10	8
2	67	50	35	23	15	10	47	36	27	18	13	9
3	28	30	27	22	16	11	30	26	22	16	13	9
4	14	16	18	18	15	11	19	18	17	14	12	9
5	..	8	11	13	12	10	11	12	12	11	10	8
6	..	..	6	8	9	9	7	8	8	9	8	6
7	..	6	..	5	6	6	4	5	6	6	6	5
8	..	..	6	..	..	6	2	..	..	..	3	4
9	..	..	..	16	10	5	..	9	..	..	..	3
10	..	..	..	..	..	..	3	..	10	16	15	..
11	..	..	..	..	..	5	..	..	..		..	..
12	..	..	..	..	..	..	..	..	..	..	..	9
13	..	..	..	..	..		..	..	..	..	..	..
14	..	..	..		..	..	..	..	..	..	..	..
15	..	..	..	..	..	..	..	..	..	..	..	..
N	371	236	162	123	99	82	262	205	153	117	96	79
M	1·09	1·71	2·49	3·31	4·08	4·94	1·90	2·43	3·08	4·02	4·76	5·63
K	5·685	3·410	4·047	6·067	5·616	5·583	1·699	1·759	2·296	2·511	2·859	3·692
χ^2	0·896	8·757	4·590	9·518	7·057	2·209	6·485	6·438	5·320	9·188	5·310	7·083
**d.f.*	3	5	4	7	6	7	6	7	7	7	7	6
Significance	85%	10%	30%	20%	30%	90%	30%	45%	60%	20%	60%	30%

III

numbers of stalk-borer infested and total stalks

Number of total stalks	1954–55 Size of primary unit in feet						1957–58 Size of primary unit in feet					
	1	2	3	4	5	6	1	2	3	4	5	6
1	59	14	2	..	..	..	44	..	9	3	1	..
2	62	19	3	1	..	..	43	26	12	5	2	..
3	66	26	7	2	..	..	45	26	16	7	4	1
4	59	30	10	3	..	..	39	29	18	9	5	1
5	45	30	14	5	..	..	31	28	18	11	7	2
6	31	27	16	7	..	..	22	25	17	11	8	3
7	20	23	17	9	6	..	15	21	14	11	8	4
8	12	19	17	10	4	..	9	16	12	10	8	5
9	7	14	16	11	5	..	9	12	10	9	8	6
10	..	11	14	11	7	8	..	9	8	8	7	6
11	10	8	12	11	8	4	..	6	6	7	7	6
12	..	5	9	10	8	4	5	..	8	6	6	6
13	..	4	7	9	9	5	..	7	..	..	5	6
14	..	6	5	8	9	6	..	..	..	8	4	5
15	..	..	..	6	8	6	..	..	5	12	3	5
16	..	..	13	5	7	6	..	..	..	..	3	4
17	..	..	..	..	6	6	..	..	..	..	2	..
18	..	..	..	..	5	6	..	..	..	..	2	..
19	..	..	..	..	4	5	..	..	..	..	1	19
20	..	..	..	28	5	5	..	..	..	..	..	..
21	..	..	..	..	..	4	..	..	..	..	5	.
22	..	..	..	..	13	3	..	..	..	..	..	..
23	..	..	..	..	..	3	..	..	..	..	..	..
24	..	..	..	..	..	..	..	..	..	..	..	..
25	..	..	..	..	..	11	..	..	..	..	..	..
N	371	236	162	123	99	82	262	205	153	117	96	79
M	3·79	5·96	8·62	11·33	14·24	17·11	3·84	4·90	6·31	8·21	9·56	11·63
K	8·194	6·532	11·300	11·983	26·822	24·471	6·107	4·364	5·199	5·390	6·599	10·915
χ^2	7·688	11·117	12·087	8·585	5·494	12·365	11·384	19·305	11·345	7·666	6·480	1·644
d.f.	8	11	12	12	10	8	8	9	11	11	7	6
Significance	45%	40%	40%	70%	85%	10%	15%	20%	40%	75%	45%	90%

* The *d.f.* are two less than the number of cell frequencies compared.

Since 'x' and 'y' both are from NB distribution, these may be converted to normal variates $p = k_1^{-\frac{1}{2}} \log(\sqrt{1+k_1x} + \sqrt{k_1x})$ and $q = k_2^{-\frac{1}{2}} \log(\sqrt{1+k_2y} + \sqrt{k_2y})$ and the corresponding statistics

$$T_1 = \frac{1}{n}\sum \frac{p_i}{q_i} \text{ and } T_2 = \frac{\sum \frac{p_i}{n}}{\sum \frac{q_i}{n}}.$$

It will be examined which of these statistics may be employed for the estimation of incidence.

Following Rao (1952) if μ_1, μ_2 and σ_1^2, σ_2^2 be the means and variances of 'p' and 'q' and ρ the correlation between these then with the usual notations,

$$\mu_{20} = \sigma_1^2, \quad \mu_{02} = \sigma_2^2, \quad \mu_{11} = \rho\sigma_1\sigma_2, \quad \mu_{22} = (1+2\rho^2)\sigma_1^2\sigma_2^2$$

$$\mu_{40} = 3\sigma_1^2, \quad \mu_{04} = 3\sigma_2^2, \quad \mu_{13} = 3\rho\sigma_1\sigma_2^3, \quad \mu_{31} = 3\rho\sigma_1^3\sigma_2.$$

If

$$p - \mu_1 = \xi \text{ and } q - \mu_2 = \eta$$

then

$$\frac{p}{q} = \frac{\mu_1}{\mu_2}\left(1 + \frac{\xi}{\mu_1}\right)\left(1 + \frac{\eta}{\mu_2}\right)^{-1}$$

$$= \frac{\mu_1}{\mu_2}\left\{1 - \frac{\eta}{\mu_2} + \frac{\eta^2}{\mu_2^2} - \cdots + \frac{\xi}{\mu_1}\left(1 - \frac{\xi}{\mu_2} + \frac{\xi^2}{\mu_2^2} - \cdots\right)\right\}$$

Taking expectations of both the sides

$$E\left(\frac{p}{q}\right) = \frac{\mu_1}{\mu_2}\left\{1 + \frac{\mu_{02}}{\mu_2^2} + \frac{\mu_{04}}{\mu_2^4} + \cdots - \frac{\mu_{11}}{\mu_1\mu_2} - \frac{\mu_{13}}{\mu_2^3\mu_1} - \cdots\right\}$$

$$= \frac{\mu_1}{\mu_2}\left\{\sum_0^\infty \frac{(2t)!\,2^{-t}\,v_2^{2t}}{t!} - \rho v_1 \sum_0^\infty \frac{(2t+2)!\,2^{-(t+1)}\,v_2^{(2t+1)}}{(t+1)!}\right\}$$

$$= \frac{\mu_1}{\mu_2}\left\{1 + (v_2 - \rho v_1)\sum_1^\infty \frac{(2t)!\,2^{-t}\,v_2^{2t-1}}{t!}\right\}$$

where v_1 and v_2 are the coefficients of variation of 'p' and 'q' respectively. Thus for T_1 and T_2

$$E(T_1) = \frac{\mu_1}{\mu_2}\left\{1 + (v_2 - \rho v_1)\sum_{1}^{\infty}\frac{(2t)!\,2^{-t}\,v_2^{2t-1}}{t!}\right\}$$

$$E(T_2) = \frac{\mu_1}{\mu_2}\left\{1 + \frac{1}{\sqrt{n}}(v_2 - \rho v_1)\sum\frac{(2t)!\,2^{-t}\,v_2^{2t-1}}{t!}\right\}$$

since the coefficient of variation of $\bar{p}$ is $v_1/\sqrt{n}$ and of $\bar{q}$ is $v_2/\sqrt{n}$. It is seen that $E(T_2) \to \mu_1/\mu_2$ while $E(T_1)$ remains the same for all 'n' and since both have variances of $0\,(1/n)$, T_2 converges Stochastically to μ_1/μ_2 and T_1 to some other value.

T_1 is a biased estimate of μ_1/μ_2 and does not admit a simple correction for bias. Since the bias does not tend to zero as $n \to \infty$, it should be considered inconsistent as an estimate of the parametric function μ_1/μ_2. On the other hand, T_2 is a consistent estimate of the ratio.

Thus the incidence and its precision can be found by converting the observed values to normal variate and calculating T_2. An example is given below where a comparison of the four methods based on proportion, Poisson, Binomial and *NB* variate has been made.

An Example

A random sample of thirty primary units was drawn from a plot to estimate the borer incidence. The records of the infested and total number of stalks have been presented in Table IV.

The incidence values as calculated for proportion and converted on the assumption of Poisson, Binomial and *NB* variate have also been shown in Table IV. The values of k_1 and k_2 required in the conversion of *NB* variate were obtained from the sample. The means and the standard error and the coefficient of variation for each series show that the variation is highest among the incidence values based on proportion followed by transformed Binomial, Poisson and *NB* variate values. The relative efficiencies as compared to proportion are 15·66%, 16·99% and 25·48% for the Binomial, the Poisson and the *NB* variate respectively.

This is as was expected. In the hierarchy of discrete distributions, this is the order for increasing variance in relation to the mean. When p is small and as $n \to \infty$, $np \to m$ Binomial becomes Poisson, while *NB* becomes Poisson as $k \to \infty$. There are biological processes where *NB* fits in well and it is left as a problem to the Sugarcane Entomologists to find if the spread of infestation might conform to any one of them bridging the gap between the mathematics and the natural phenomenon.

TABLE IV

A random sample of thirty primary units and the different transformed variates to estimate the incidence

Serial No.	Infested stalks	Total stalks	Incidence Values based on . . .			
			Proportion	Binomial variate	Poisson variate	Negative Binomial variate
1	3	11	·2727	31·50	·5122	1·278
2	2	11	·1818	25·50	·4264	1·153
3	1	21	·0476	12·66	·2182	0·863
4	0	13	·0000	00·00	·0000	·000
5	1	7	·1429	22·22	·3780	1·013
6	1	16	·0625	14·54	·2500	·896
7	1	7	·1429	22·22	·3780	1·013
8	2	11	·1818	25·25	·4264	1·153
9	2	19	·1053	18·91	·3245	1·067
10	2	5	·4000	39·23	·6225	1·304
11	6	13	·4615	42·82	·6793	1·461
12	3	13	·2308	28·73	·4804	1·247
13	6	20	·3000	33·21	·5477	1·376
14	6	14	·4286	40·92	·6547	1·446
15	3	10	·3000	33·21	·5477	1·296
16	4	12	·3333	35·24	·5773	1·355
17	2	15	·1333	21·39	·3651	1·103
18	8	21	·3810	38·12	·6173	1·451
19	11	17	·6471	53·55	·8044	1·589
20	4	16	·2500	30·00	·5000	1·297
21	5	10	·5000	45·00	·7071	1·460
22	1	8	·1250	20·70	·3536	0·992
23	2	22	·0909	17·56	·3015	1·046
24	3	23	·1304	21.47	·3611	1·153
25	4	19	·2105	27·35	·4588	1·267
26	5	10	·5000	45·00	·7071	1·460
27	9	20	·4500	42·13	·6708	1·495
28	3	15	·2000	26·56	·4583	1·222
29	2	20	·1000	18·44	·3162	1·060
30	5	12	·4167	40·22	·6455	1·425
Mean	..	..	·2576	29·113	·4763	1·193
S.D.	..	..	·1625	11·73	·1768	·2966
C.V.	..	..	·6308	·4209	·3712	·2476

SUMMARY

To estimate the borer incidence in sugarcane, the stalks and the clumps do not form manageable primary units, and also any length or area units cannot be framed each of which may contain equal number of stalks. Separate distributions of the infested and the total stalks in the primary units were, therefore, investigated.

It was found that the distributions of the infested and total stalks had variances much higher than the mean and did not belong to Normal, Binomial or Poisson distribution. The distributions for both were similar and conformed to the Negative Binomial distribution.

Since the incidence is the ratio of two random variables, an examination in the validity of assuming these values or belonging to Normal, Poisson and Binomial distribution was made. A consistent statistic for estimating this ratio based on $\sinh^{-1}$ transformation valid for samples from the Negative Binomial distribution is suggested and its efficiency compared with the rest.

The comparative efficiencies of this statistic in relation to the other three based on proportion, Binomial and Poisson distributions were found and the relative efficiency of the suggested statistic was found to be the highest.

ACKNOWLEDGEMENT

I am thankful to late Sri. B. D. Gupta, formerly Entomologist, Indian Institute of Sugarcane Research, Lucknow, for suggesting the problem and providing the necessary data. Thanks are also due to Sri. R. R. Panje, Director, Indian Institute of Sugarcane Research, Lucknow, for the helpful criticism and to Dr. A. R. Roy, Professor of Statistics, Lucknow University, for guidance in the preparation of the paper.

REFERENCES

1. Anscombe, F. J. .. *Biometrika*, 37, Sampling theory of the Negative Binomial and Logarthmic series distributions, 1950, 337.
2. Beall, G. .. *Statistics and Mathematics in Biology*, Iowa State College Press, Ames, Iowa, 1954, 297.
3. Deming, W. E. .. *Some Theory of Sampling*, John Wiley and Sons. Inc., 1950, 502.

4. Khanna, K. L. and Bandyopadhyay — *Proc. Indian Acad. of Sci.*, 1951, **34B**, 210.

5. Rao, C. R. .. *Advanced Statistical Methods in Biometric Research*, John Wiley and Sons, 1952, 154.

6. Rojas, B. .. *Proceedings of the International Soc. Sug. Tech.*, 1953, **8**, 310.

Srivastava, J. N. .. *Annual Report, Indian Institute of Sugarcane Research, Lucknow*, 1957, 56–57, 72.