ESTIMATION OF STEM BORER DAMAGE IN RICE FIELDS*

BURTON T. OÑATE**

Stem borer is considered as a major insect pest of rice. Considerable losses are incurred annually from the attack of this insect. However, precise methods of estimating the incidence of stem borer attack are not available. This paper will describe a simple but precise method of estimating the damage caused by this major insect pest of rice. These estimates with productivity data also can be used to explain the state and nature of yield loss.

Stem borer incidence in a rice field is usually measured as the number of dead hearts (X_i) per hill at various stages of vegetative growth or the number of white heads (X_i^*) per hill at maturity. If we assume a finite universe of hills in a rice field, then the parameters may be designated as follows:

$$\bar{X} = \sum_{i=1}^{N} (X_i/N)$$ is the population mean of dead hearts per hill,

or

$$\bar{X}^* = \sum_{i=1}^{N} (X_i^*/N)$$ is the population mean of white heads per hill.

*Results given in this paper were presented to the Symposium on the Major Insect Pests of Rice held last September, 1964, at the International Rice Research Institute, Los Baños, Laguna, Philippines.

**Statistician. The International Rice Research Institute. Los Baños, Laguna, Philippines.

where

N is the size of the universe or the total number of hills.

The population variance is defined as

$$S^2(X_i) = \sum_{i=1}^{N} (X_i - \bar{X})^2/(N-1)$$

and

$$\sigma^2(X_i) = [(N-1)/N]\ S^2$$

for dead hearts. Similar equations are derived for X_i^*. In general, N is usually large so that numerically S^2 is equal to σ^2.

Incidence may also be expressed as the ratio of X_i or X_i^* to the number of tillers in a hill (Y) or to the number of bearing panicles (Y_i^*) at harvest time, respectively. These ratios are expressed as

$r_i = (X_i/Y_i)$ for number of dead hearts to total number of tillers in a hill.

and

$r_i^* = (X_i^*/Y_i^*)$ for number of white heads to total number of bearing panicles in a hill.

The parameters in the population of dead hearts (X_i) and total tillers (Y_i) are

$\bar{R} = \sum_{i=1}^{N} r_i/N$ as the population mean of ratios

and

$\bar{Q} = (X./Y.)$ as the ratio of population totals X. and Y. (or ratio of population means $\bar{X}$ and $\bar{Y}$).

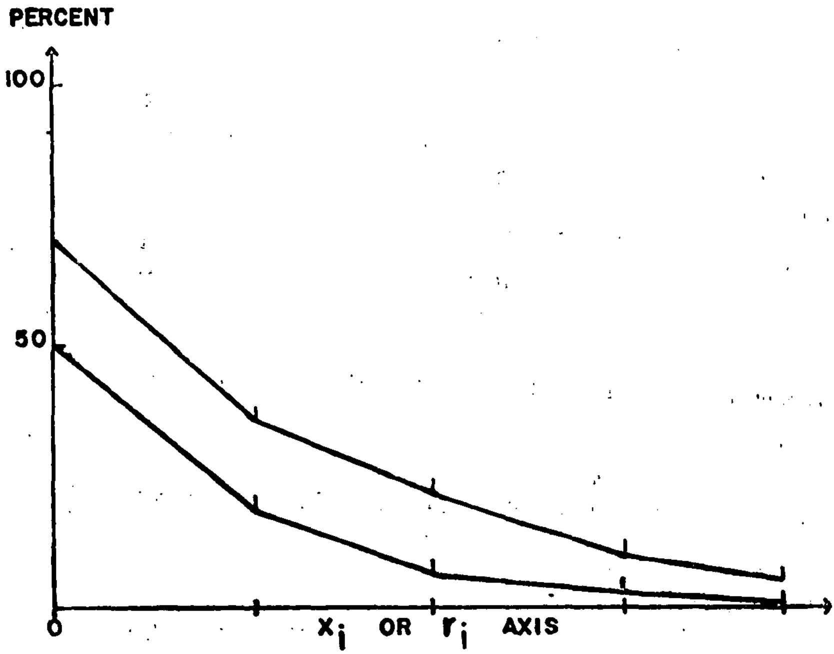
Similar formulas may be given for the X_i^*'s and the Y_i^*'s.

It is important to indicate which of the parameters $\bar{X}$, $\bar{R}$ or $\bar{Q}$ is being estimated. The variance of the estimator also can be derived.

1. Distribution of Incidence

The pattern of the distribution of dead hearts (X_i), white head (X^*_i) counts, r_i or r^*_i is shown in Figure 1 where a large proportion of the observations X_i, X^*_i r_i or r^*_i is zero.

FIGURE I. PATTERN OF DISTRIBUTION OF INCIDENCE



This situation gives rise to large sampling variability of the original X_i or r_i.

2. Pattern of Variability

As measured by the cv($\bar{x}$) in percent, the variability of x is high. The variability is exhibited by the results given in Table 1, Figures 2 and 3. Even with high mean incidence, the variability is still very high. Note the marked linear relationship between S* and $\bar{X}$* in Table 1 and Figure 1 and also the relationship between s and $\bar{x}$ in Figure 2. This was used by the author in devising a simple method of approximating the needed sample size for a given level of $\bar{x}$ (Oñate, 1964).[a] Also, from Table 1, the size of sample needed to reduce the cv(x) to 10 percent will exceed 500 random hills. This situation calls for a simple but precise method of sampling for stem borer incidence.

TABLE 1

MEAN (X*) AND STANDARD DEVIATION (S*) OF WHITE HEADS (X_i^*) AND THE COEFFICIENT OF VARIABILITY OF x*, CV(x*), IN PERCENTAGE FOR VARYING VALUES OF n . FOUR UNIFORMITY EXPERIMENTS. IRRI. 1964.

	Insecticide treatment and variety	$\bar{X}$* (mean)	S* Standard deviation	CV($\bar{x}$*) in percent for different sizes of sample n			
				1	100	200	500
I	Lindane-Chianung 242	0.0512	0.2636	510	51	37	23
II	Lindane-Taichung Native 1	.0810	.3656	420	42	30	19
III	Endrin-Chianung 242	.1425	.4544	320	32	23	14
IV	Endrin-Taichung Native 1	.2784	.6691	240	24	17	11

a B. T. Oñate. Statistics in Rice Research. Bound manuscript. Part II. The International Rice Research Institute. 1964.

FIGURE 2. THE RELATIONSHIP BETWEEN $\bar{X}^*$ AND S^* AND THE COEFFICIENT OF VARIABILITY, $CV(X_i^*)$ IN PERCENTAGE FOR WHITE HEADS COUNT. FOUR UNIFORMITY TRIALS. IRRI. 1964.

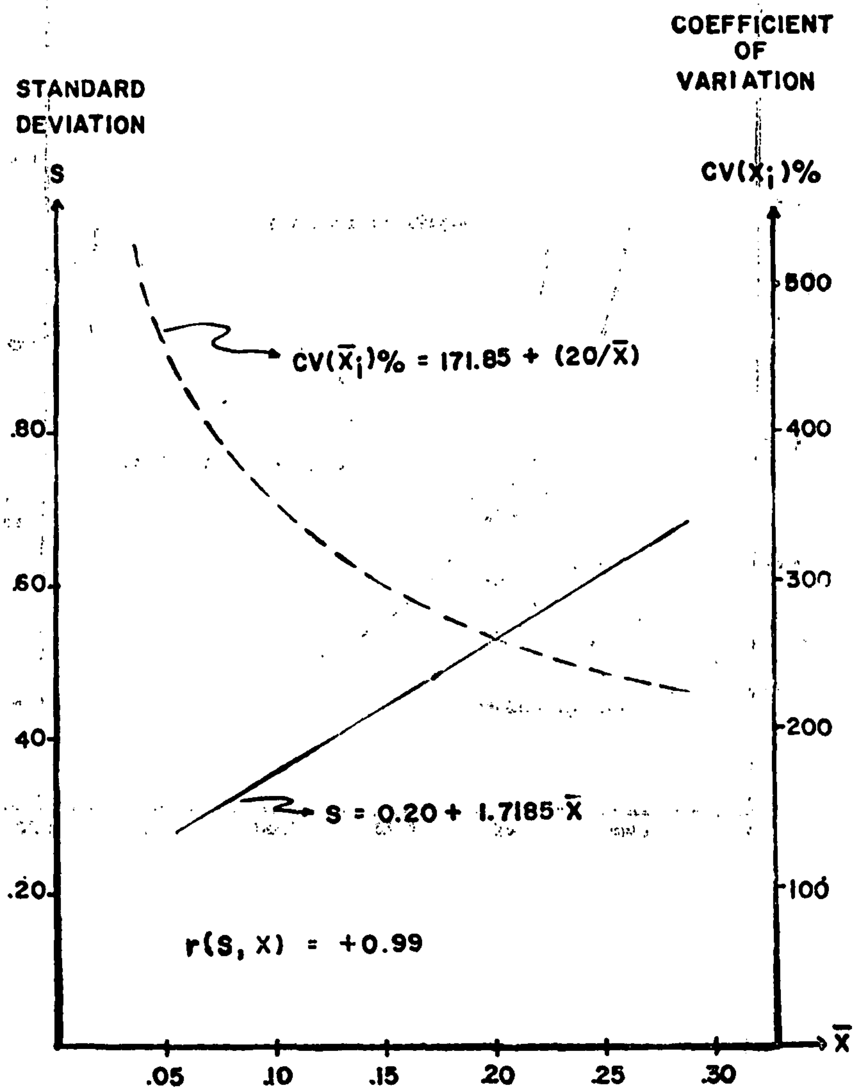
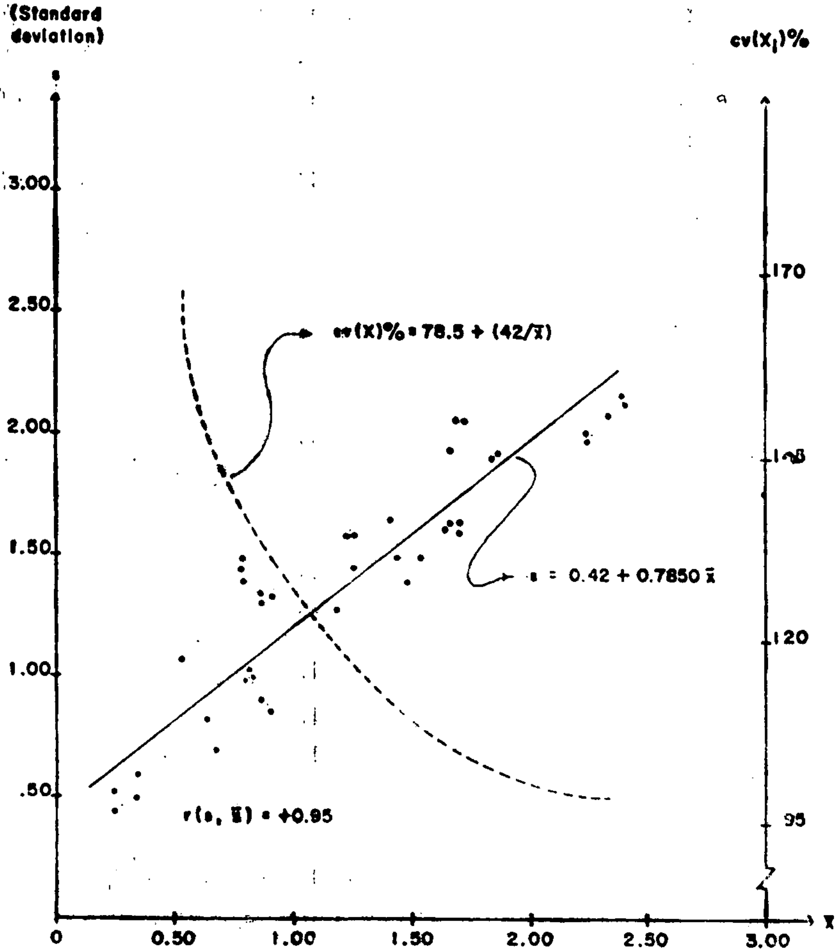


Figure 3. THE RELATIONSHIP OF $\bar{x}$ AND s FOR 40 RANDOM RICE VARIETIES TESTED FOR DEAD HEARTS INCIDENCE. IRRI. 1962.



3. Estimators

3.1 Random sampling

In dealing with incidence of stem borer attack, the parameters under study must be defined concisely, so that there will be no misunderstanding on what is being measured. These parameters where described earlier.

From the theory of a finite universe, our random sampling estimators of X and S^2 are

$$\bar{x} = \sum_{i=1}^{n} (X_i/n)$$

and

$$s^2 = \sum_{i=1}^{n} (X_i - \bar{x})^2/(n-1)$$

respectively, where

$\bar{x}$	is the sample mean,
s^2	is the sample variance,
and n	is the sample size.

The variance of $\bar{x}$ is

$$\sigma^2(\bar{x}) = [(N-n)/N] \; S^2/n$$

which simplifies into

$$\sigma^2(\bar{x}) \doteq S^2/n$$

if the finite population correction (fpc) is

$$[(N-n)/N] \doteq 1 .$$

This variance $\sigma^2(\bar{x})$ is estimated by

$$s^2(\bar{x}) = [\ (N-n)/N\ ]\ s^2/n$$

or

$$s^2(\bar{x}) \doteq s^2/n$$

if $[n/N]$ is very small.

The results given in Table 1, Figure 2 and Figure 3 show that random sampling will result in estimators with very high sampling variability even for a large n.

3.2. Screening techniques

This technique was utilized by Oñate (Part III, 1964, pp. 94-95) in sampling for stem borer incidence. It is assumed in the application of this technique that the units, U_i with $X_i = 0$ can easily be distinguished in the field. If so, these units are screened and ignored in the samplnig procedure cedure (Cochran, 1953). The mean of the Xi = 0 is zero and the variance of the mean also is zero. By definition,

variance fos the whole population σ^2 is larger than σ^2_{nz}, the variance for the non-zero (nz) population. This relationship is described below:

$$\sigma^2_{nz} = \sum_{i=1}^{PN} (X_i - \bar{X}_{nz})^2 / PN$$

$$= (1/P)\ [\ \sigma^2 - (Q/P)\ \bar{X}^2\ ]$$

$$= (1/P)\ [\ \sigma^2 - (PQ)\ \bar{X}^2_{nz}\ ]$$

$\bar{X}_{nz} = \sum_{i}^{PN} X_i/PN$ is the population mean of the non-zeros.

P is the proportion of N that is non-zero.

$Q=(1-P)$ is the proportion of N that is zero.

and

$$\bar{x} = P\bar{X}_{nz}$$

Our estimator of $\bar{\bar{X}}$, the population mean of hills attacked in the whole population is

$$\bar{\bar{x}} = P\bar{x}_{nz} + Q\cdot 0$$

where

$$\bar{x}_{nz} = \left(\sum_{i=1}^{n^*} X_i/n^*\right) .$$

n^* is the sample size in the non-zero population.

and

P and Q are as defined before.

The variance of $\bar{\bar{x}}$ is

$$\sigma^2(\bar{\bar{x}}=P\bar{x}_{nz}) = P^2\sigma^2(\bar{x}_{nz})$$

$$= (P^2\sigma^2_{nz})/n^*$$

This form of $\sigma^2(\bar{\bar{x}})$ indicates that there are three sources which are responsible for the reduction of the variance of $\bar{\bar{x}}$ with the screening of $X_i=0$.

These sources are as follows:

a) σ^2_{nz} will be smaller than σ^2 as indicated by the conditions given above.

b) P^2 appears in the numerator and $0 < P < 1$,

and

c) The finite population correction of $\sigma^2(\bar{\bar{x}})$ will be smaller than in $\sigma^2(\bar{x})$. This relationship is given by

$$[(N_{nz}-n^*)/N_{nz}] < [(N-n^*)/N] .$$

Note that the ordinary sample mean which is obtained without screening is $\bar{x}$ and $\sigma^2(\bar{x})=\sigma^2/n^*$ where we have ignored the fpc. In this formula, we can express the variance in terms of either S^2 or σ^2 from the relationship

$$\sigma^2 = [(N-1)/N] S^2$$
$$= ks^2 .$$

The comparison will be in terms of $\sigma^2(\bar{x})$ and $\sigma^2(\bar{\bar{x}})$. Thus

$$(\sigma^2/n^*) - (P^2\sigma^2_{nz}/\sigma^*) \geq 0$$

implies a gain in the screening method. From this relationship, the relative efficiency can be expressed as

$$[\sigma^2/P^2\sigma^2_{nz}] \; 100\% = (1/P) \; [1+Q/cv(X_{inz})^2] \; 100\%$$

which is a function of P and $cv(X_{inz})$. The results are given in Table 2 for some values of P and $cv(X_{inz})$. Figure 4 illustrates the relationship. We can estimate σ^2 by ks^2 and σ^2_{nz} by $k' s^2_{nz}$. Our example estimate will be derived by the ratio

$$(ks^2/k's^2_{nz}) \geq P^2$$

which indicates the relationship necessary to attain efficiency in the use of the screening method over the purely random case.

TABLE 2

RELATIVE EFFICIENCY IN PERCENT OF SCREENING TO NON-SCREENING OF ZEROS BY PROPORTION OF ATTACKED HILLS (P) AND COEFFICIENT OF VARIATION OF NON-ZEROS [CV(X_{inz})%].

	Coefficient of variation [CV(X_{inz})%]						
Proportion (P)	5	10	20	30	40	50	60
.05	1,905,000	480,000	125,000	35,000	20,000	15,000	10,000
.10	361,000	91,000	23,000	7,000	4,000	2,000	2,000
.20	160,500	40,500	10,500	3,000	1,500	1,000	1,000
.30	93,573	23,643	6,327	1,665	999	666	666
.40	60,250	15,250	4,000	1,250	500	500	500
.50	40,200	10,200	2,800	800	400	400	400
.60	26,726	6.560	1,760	640	320	320	160
.70	16,940	4,433	1,287	429	286	286	143
.80	10,125	2,625	750	250	250	125	125
.90	4,551	1,221	444	222	111	111	111
1.00	100	100	100	100	100	100	100

With the data in Table 1, we can find out the reduction in the variance $\sigma^2(\bar{\bar{x}})$ as compared to $\sigma^2(\bar{x})$ The comparison between $\sigma^2(\bar{\bar{x}})$ and $\sigma^2(\bar{x})$ is given in Table 3 for 44 experiments. The gain in statistical precision ranges from 171 percent to 13,900 percent or an average of about 1250 percent [IRRI Annual Report, 1964]. Lower P values will result in higher relative efficiencies.

FIGURE 3. RELATIVE EFFICIENCY OF THE SCREENING METHOD AS COMPARED TO THE NON-SCREENING METHOD IN THE SAMPLING FOR STEM BORER INCIDENCE FOR VARYING VALUES OF P AND CV(X_{inz})

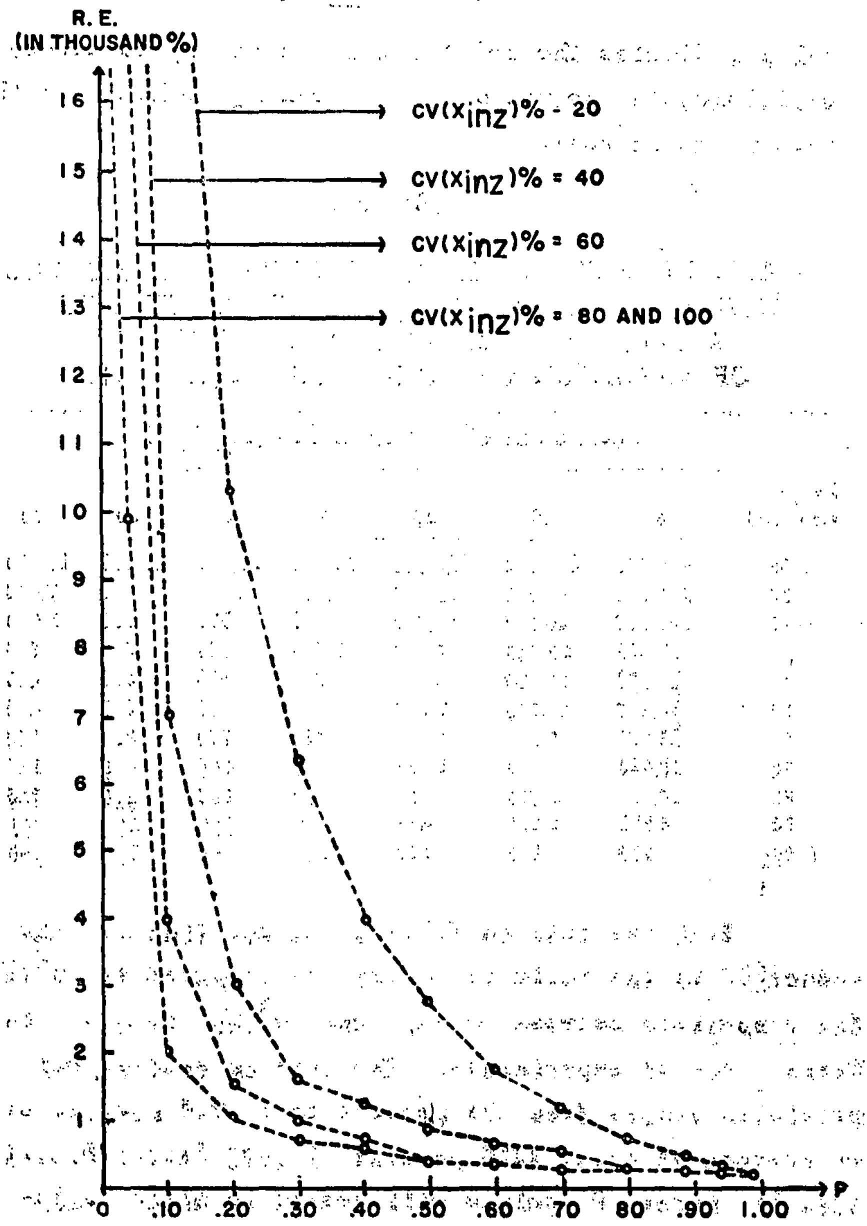


TABLE 3
COMPARISON OF VARIANCE WITHOUT SCREENING AND VARIANCE WITH SCREENING OF ZEROS. IRRI. 1962 AND 1964.*

Experiment number	P or $\hat{P}$	Variance w/o screening	Variance w/screening	Relative efficiency in percent $= [(3)/(4)]100\%$
1	0.0435	0.0695	0.0005	13,900
2	.0688	.1337	.0021	6,367
3	.1113	.2074	.0051	4,067
4	.1879	.4477	.0212	2,112
5	.65	1.71	.62	276
6	.73	2.66	1.22	218
7	.80	3.05	1.72	177
8	.61	4.37	1.59	275
9	.53	2.59	0.67	387
10	.47	1.00	0.15	667
11	.28	0.58	0.04	1,450
12	.80	4.49	2.52	178
13	.40	1.71	0.27	633
14	.55	0.86	0.17	506
15	.83	4.08	2.55	160
16	.60	1.51	0.44	343
17	.52	1.53	0.43	356
18	.30	0.25	0.05	500
19	.17	0.29	0.01	2,900
20	.27	0.37	0.02	1,850
21	.17	0.36	0.02	1,800
22	.30	0.25	0.005	5,000
23	.31	1.17	0.19	616
24	.63	2.83	1.07	264
25	.38	1.98	0.35	566
26	.56	2.44	0.66	370
27	.42	0.73	0.09	811
28	.57	0.81	0.15	540
29	.50	1.61	0.38	424
30	.52	1.03	0.21	490
31	.50	1.05	0.16	656
32	.56	1.56	0.49	318
33	.40	1.76	0.27	652
34	.75	2.61	1.25	209
35	.53	0.51	0.06	850
36	.72	2.29	1.01	227
37	.75	3.69	1.90	194
38	.71	2.29	1.04	220
39	.82	4.71	2.84	166
40	.61	3.93	1.35	291
41	.68	2.10	0.92	228
42	.50	1.03	0.17	608
43	.13	0.17	0.009	1,889
44	.80	2.05	1.20	171
			Average =	1,247

*Experiments 5 to 44 were conducted in the 1962 wet season, while Experiments 1 to 4 during the 1964 dry season.

3.3. Stratified sampling

A field may have sub-areas with different levels of incidence. We can stratify the field in relation to $\bar{X}_i$ where $i=1,2,\ldots,L$ refers to the number of strata or sub-fields. Thus, the overall mean is

$$\bar{\bar{X}} = \sum_{i}^{L} W_i \bar{X}_i$$

where

$W_i = (N_i/N)$ is the weight of the i^{th} sub-area or may represent another weighing pattern which the entomologist may give himself.

and

$\bar{X}_i = \left(\sum_{j=1}^{N_i} X_{ij}/N_i \right)$ is the mean of the i^{th} sub-area.

Within each i^{th} sub-area, we can screen out $X_{ij}=0$. Our estimator is

$$\bar{\bar{x}} = \sum_{i=1}^{L} W_i \bar{x}_i$$

$$= \sum_{i=1}^{L} W_i \left[P_i \bar{x}_i(nz) + Q_i \cdot 0 \right]$$

$$= \sum_{i=1}^{L} W_i \left[P_i \bar{x}_i(nz) \right]$$

and

$$\sigma^2(\bar{\bar{x}}) = \sum_{i=1}^{L} W_i^2 P_i^2 \sigma^2 \left[\bar{x}_i(nz) \right]$$

$$= \sum_{i=1}^{L} W_i^2 \, (P_i^2/n_i^*) \, [(\sigma_i^2/P_i) - (\bar{x}_i^2 Q_i/P_i^2)]$$

$$= \sum_{i=1}^{L} W_i^2 \, [\, P_i\sigma_i^2 - Q_i P_i^2 \bar{x}_i^2(nz) \,]/n_i^*$$

The finite population correction may be inserted into this variance formula.

In actual sampling work, precise estimate of P_i and Q_i can be obtained from a relatively larger sample ($n_i^{**} >> n_i^*$) while an estimate of $\bar{X}_{i(nz)}$ is given by $\bar{x}_{i(nz)}$ from the smaller sample n_i^* . In fact , the estimate of $\sigma^2_{i(nz)}$ can be obtained from the sample variance formula ,

$$s^2_{i(nz)} = k \sum_{j}^{n_i^*} [\, X_{ij(nz)} - \bar{x}_{i(nz)} - \,]^2 / (n_i^* - 1)$$

and the estimate of $[P_i^2 \sigma^2_{i(nz)}]/n_i^*$ is obtained from

$$[k \, \hat{P}_i^2 s^2_{i(nz)}]/n_i^*$$

where

$\hat{P}_i$ is derived from a larger sample $n_i^{**} >> n_i^*$.

The size of sample for each stratum may be given as

$$n_i^* = n \, [\, N_i S_i / \Sigma N_i S_i \,]$$

where

S_i may be expressed in terms of $\bar{X}_i$.

3.4. Measurement of ratios

Another type of measurement usually employed in stem borer experiments is the ratio

$$r_i = (X_i/Y_i)$$

where

X_i is the count of dead hearts in the i^{th} hill,

and Y_i is the count of tillers in the i^{th} hill.

It is important to note that our U_i's are the hills. If a random sample of size n hills is obtained, then we have the mean of ratios,

$$\bar{r} = (\sum_{i=1}^{n} r_i/n) ,$$

as an unbiased estimate of

$$\bar{R} = \sum_{i=1}^{N} (r_i/N)$$

which is the population mean of ratios (r_i's) on a hill basis.

There is another ratio which is termed as the populaton ratio of means or totals and this is defined as

$$Q = (X/Y)$$
$$= (X./Y.)$$

where

X and Y are population means per hill

and

X. and Y. are population totals of hills respectively.

In the literature, it is not very clear which of R or Q is the parameter which is estimated although in most cases the estimator used is termed the sample ratio of means or totals and this estimator is defined as

$$q = (x/y)$$
$$= x'/y'$$

where

$\bar{x}$ and $\bar{y}$ are sample means,

and

x' and y' are sample totals.

Both $\bar{r}$ and $\bar{q}$ are biased estimates of $\bar{Q}$ but $\bar{q}$ is easier to compute and has a lower upper bound in the relative bias as compared to $\bar{r}$. Note that in this form $\bar{q}$ is identical to the binomial estmiator $\bar{p}$ of $\bar{P}$ = X./Y. since each tiller is observed as either attacked and that

$$\bar{P} = \sum_{ij}^{N,M_i} X_{ij} / \sum_{i}^{N} M_i \, .$$

is estimated by

$$\bar{p} = \sum_{ij}^{n,m_i} X_{ij} / \sum_{i=1}^{n} m_i$$

where $\sum_{i}^{N} M_i$ is the total number of tillers in the universe,

and $\sum_{i=1}^{n} m_i$ is the number of tillers in the sample.

Note that in the binomial, each tiller is assumed to be independent of getting attacked or not. However, we notice that there is a clustering of tillers in a hill. As such there is a tendency for tillers within a given hill to be alike. Also in $\bar{q}$ and $\bar{Q}$, our units, U_i, are the hills or clusters of tillers while in $\bar{p}$ and $\bar{P}$, our units are the tillers. Thus, the universe in $\bar{Q}$ is smaller than the universe in $\bar{P}$. Generally, and in actual sampling work, the hill or some larger unit is the sampling unit and not the tiller within the hill. Thus, the variance

tiller within the h.... Thus, the variance of q . $\sigma^2(\bar{q})$ is the more appropriate variance than $\sigma^2(\bar{p})$. Note that $\bar{p}$ estimates P (population proportion) which is identical to $\bar{Q}$.

The ratio estimator, $\bar{q}$, has a variance

$$\sigma^2(\bar{q}) = [(N-n)/Nn]\ \bar{Q}^2\ [C_{xx} + C_{yy} - 2C_{xy}]$$

where

C_{xx}, C_{yy} are the square of the coefficient of variation (CV) of X_i and Y_i, respectively,

and

C_{xy} is the similar CV definition for the covariance.

This variance of q is estimated from sample (Σ') by

$$s^2(q) = [(N-n)/Nn\ (n-1)\ \bar{y}^2]\ [\Sigma' x_i^2 + \bar{q}^2\ \Sigma' y_i^2 - 2\bar{q}\ \Sigma' x_i y_i] .$$

Note that numerically $\bar{p} = \bar{q}$, but the variances will differ as shown above. It is important to remember that the units are the hills and not the tillers. Thus, the binomial variance, $\sigma^2(\bar{p})$ is not the appropriate measure.

The relative efficiencies of the screening method for $\bar{x}$ and $\bar{r}$ as the estimators are shown in Tables 3a and 3b, respectively for experiments conducted during the years 1962 to 1964.

TABLE 3a

COMPARISON OF $s^2(x)$ WITHOUT SCREENING AN $s_0^2(\bar{x})$ WITH SCREENING OF ZEROS. COUNTS (X_i) OF DEAD HEARTS. IRRI. 1962-64.[a]

Date	Number of experiments	Relative efficiency with screening of zeros, in percent: Range	Average
July, 1962	44**	166-13,900	1250
August, 1962	40	100- 1,600	328
March, 1963	40	138- 1,800	367
August, 1963	40	198-45,000	4074
March, 1964	40	141- 3,035	543
		Overall average	1312

a Source of basic data: Department of Entomology.
** Includes four uniformity data.

TABLE 3b

COMPARISON OF $s^2(r)$ WITHOUT SCREENING AND s_0^2 WITH SCREENING OF ZEROS. PERCENTAGES (r_i) OF DEAD HEARTS. IRRI. 1962-64.[a]

Date	Number of experiments	Relative efficiency with screening of zeros, in percent: Range	Average
July, 1962	40	163-3,833	727
August, 1962	40	100-1,896	318
March, 1963	40	121-3,161	396
August, 1963	40	206-7,212	1427
March, 1964	40	153-1,384	384
		Overall average	650

a Source of basic data: Department of Entomology, IRRI.

4. Summary and Conclusions

Data from experiments on stem borer incidence from the International Rice Research Institute (IRRI) fields for the years 1962 to 1964 were used to study the problems of estimation of parameters on stem borer incidence.

Results of these analyses and those from the available literature indicate that the assumption of a finite universe (U_f) and finite population (X_i, X_i^*, X_i^{**}) is sound for studies of stem borer incidence. This paper has classified the parameters used in the stem borer incidence and the estimators relevant to each parameter.

The concept of the tiller in the hill as the observational unit (ou) was distinguished from the concept of the hill as the sampling unit (su). Thus, the ratio estimator with the hill as the (su) is the more appropriate than the binomial estimator which uses the tiller or (ou) as the unit.

For counts, the technique of screening out the zeros will result in large relative statistical efficiency averaging about 1300 per cent. The coefficient of variability (cv) of the estimator with screening will be reduced by ($1/\sqrt{13}$). If the cv of the estimator for random sampling is 20 per cent, then the cv of the estimator using screening will be on the average about 6 per cent only. For ratios, the relative efficiency is about 650 per cent. From these results, it is concluded that screening can be used as a precise technique for the estimation of stem borer incidence in experimental fields. This technique was utilized in applied research plots in farmer's paddy fields.

The concepts in the method of sampling with the screening out of zeros were extended to stratified sampling. In conjunction with crop cutting and/or interview survey, the relationship between stem borer incidence and yield may be obtained and yield loss curve can be derived.

6. Literature Cited

Cochran, W. G. Sampling Techniques. John Wiley and Sons, Inc. Chapter III. 1953.

Oñate, B. T. Statistics in Rice Research, Par II, Chapter VI; Part III, Chapter X. 1964.

The International Rice Research Institute. Annual Report. 1964.