INVESTIGATIONS ON SAMPLING FOR ESTIMATION OF CROP ACREAGES—I

By V. K. Mokashi
Institute of Plant Industry, Indore

The method of sampling as a means of collecting agricultural statistics was almost unknown in India some years ago. But now random sample surveys for estimation of crop yields have become a routine feature. The first yield survey on cotton by random sampling was conducted in 1942–43 in Akola District of Madhya Pradesh (Panse and Kalamkar, 1944, 1945). But the total production of any crop depends upon the yield per acre of the crop and the total area under the crop. The area figures for calculating the cotton production from these surveys was provided by the patwaris. During the course of these surveys, the need for verifying the accuracy of acreage data was increasingly felt. The present study had its origin in this need. The method of complete enumeration of area as followed by patwaris, works satisfactorily in practice, provided there is an adequate and effective supervision over his work. But it appears worth investigation whether crop areas can be estimated with enough accuracy by resorting to random samping method. Apart from its possible use in giving rapid estimates of crop areas in special inquiries, random sampling is considered by some to be the only feasible procedure for estimation of areas in places where there is no patwari agency. Secondly, even where the area is completely enumerated, the development of a suitable sampling method will provide a scientific check in the hands of the supervising staff over the work of patwaris. Reliable preharvest forecasts can also be made by the use of random sampling technique. Sampling theory for the purpose of estimation of population characters and the precision of sample estimates is now well established. Its application for estimation of crop acreages and their precisions has been considered here and illustrated in this article.

As in yield surveys now in progress in most of the states in India under the auspices of Indian Council of Agricultural Research, so also in sampling for crop acreages, sampling is considered in two stages. The village with its well-defined boundary constitutes a natural and convenient primary sampling unit and a field or block of fields within a village a sampling unit. Part I deals with the single stage sampling of primary units (villages) without any sub-sampling within.

Several aspects of this problem of single stage sampling have been studied and are discussed below.

MATERIAL

The material for the present investigations is furnished by the data for complete enumeration of the agricultural area for Akola District in Madhya Pradesh for the year 1945–46. The district is typical of the temporarily settled areas with a long established system of complete enumeration of agricultural land. The district has an area of about 4,092 sq. miles and consists of six tahsils and 1,731 villages. The patwari records were available for 1,723 villages in the whole district. Data on two crops were studied. Cotton is the main crop of the district and was, therefore, examined in considerable detail. Since, however, cotton occupies a very large acreage, *viz.*, 6·3 lacs of acres, another crop, wheat, with an area of 1·2 lacs of acres, was included for study as representating a crop occupying much less area. This was done to verify to what extent sampling is feasible for a crop with a relatively smaller area. The village crop area was treated as a variate in these investigations.

The distribution of village cotton areas was found to be skew and showed that there was a small number of villages with large cotton areas and a large number of them with smaller cotton areas. Further, village cotton areas were found to be highly correlated with respective sizes of the village, the size being represented by the total geographical area of the village. Such correlation is seen from the typical diagram (Fig. 1) shown for one tahsil. Other tahsils exhibited similar relationship. The villages varied greatly in their sizes and were grouped into four size strata. The strata and the variability of the village cotton areas within each stratum are shown in Table I.

TABLE I

Stratification by size of the village

Size of stratum	Standard deviation of cotton area per village	Total No. of villages in the stratum
Below 600 acres ..	64	457
601 to 1,500 do ..	119	789
1,501 to 2,400 do ..	186	286
Above 2,400 do ..	460	191

The table shows that the population to be sampled consists of a small number of large-sized villages with high variability and a large number of small-sized villages subject to a considerably lesser variability. Therefore, for a given amount of sampling, considerable gains could be accomplished by including in the sample a greater portion of the larger villages.

Optimum Allocation

According to the principle of optimum allocation of sampling units (Neyman, 1934), the total number of villages required for providing an estimate of the total cotton area for the district, with a given precision, should be distributed among different strata in proportion to $N_i\sigma_i$, where N_i is the total number of villages in the i-th stratum and σ_i is the standard deviation of the i-th stratum. From the knowledge of the average variance within strata, a total sample of 108 villages was found to give approximately the five per cent. standard error of the mean. This was the sample used for the study. It was distributed among different size strata in an optimum manner.

Methods of Estimation

The villages were selected randomly in each stratum. For this sample the total cotton areas were calculated for different strata. These totals were expanded to obtain an estimate of district cotton area, by weighting them by the reciprocals of the sampling fraction, *i.e.*,

$$Y = \sum_{i=1}^{k} \frac{s(y_i)}{n_i} N_i \tag{1}$$

where sy_i is the total of y for the sample in the i-th stratum based upon n_i number of villages in the i-th stratum and k is the number of strata. The sampling variance of Y is given by

$$V(Y) = \sum_{i=1}^{k} N_i^2 \frac{\sigma_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i} \tag{2}$$

The estimate of σ_i^2 is provided by the sampling variance of the villages in the i-th stratum. In view of the high correlation between the village cotton area and its size, two alternative methods, *i.e.*, ratio and the regression methods of estimation, were employed for estimation of the district cotton area. The linear regression estimate

$Y_{reg.}$ for the population total of Y, the cotton area, is given by (Cochran, 1942),

$$Y_{reg.} = \sum_{i=1}^{k} N_i \left[\bar{y}_{is} + b_i \left(\bar{x}_{ip} - \bar{x}_{is}\right)\right] \qquad (3)$$

where p and s refer to population and sample values respectively. The variance of $Y_{reg.}$ is given by

$$V(Y_{reg.}) = \sum_{i=1}^{k} \frac{N_i - n_i}{N_i} \cdot N_i^2 \sigma^2 y_i \left(1 - \rho_i^2\right)$$

$$\left[\frac{1}{n_i} + \frac{(\bar{x}_{pi} - \bar{x}_{si})^2}{s\,(x_i - \bar{x}_{si})^2}\right] \qquad (4)$$

The mean ratio is determined for each stratum from the sample and was multiplied by the corresponding total agricultural area of the stratum and pooled over the whole district. The total cotton area is given by

$$Y_r = \sum_{i=1}^{k} X_i \bar{r}_i \qquad (5)$$

where $\bar{r}_i = 1/n_i \overset{ni}{\Sigma} y/x$ and X_i is the population value of x in the i-th stratum. The variance of this estimate is given by Yates (1949) as

$$V(Y_r) = \sum_{i=1}^{k} X_i^2 \, V(\bar{r}_i) \, \frac{N_i - n_i}{N_i} \qquad (6)$$

Table II shows the estimates of the total cotton area by the three methods and their standard errors. It was observed that the unweighted mean of the ratios for the individual villages does not show any

TABLE II

Estimated cotton area and percentage standard error with size-stratification

Method of estimation	Estimated cotton area in 1,000 acres	Percentage standard error
Direct ..	633	3·6
Regression ..	628	2·8
Ratio ..	630	2·6

trend over varying values of x and therefore, it is unlikely, that it will be biased (Yates, 1946).

Geographic stratification is a common mode of stratification adopted in many surveys. Apart from its effect on the accuracy of the sample estimate convenience for administrative purposes is an advantage in favour of this kind of stratification. In case of area enumeration, the work is done by the departmental staff as part of their normal duties and any plan of sampling for area enumeration should, therefore, fit in with the administrative machinery. With this point in view, geographic stratification was studied in combination with size stratification. The tahsil was adopted as a geographical stratum and each tahsil was further divided into the same four-size strata for combining the two modes of stratification. Another sample of 108 villages was selected for study and was distributed among strata in an optimum manner. From this sample the estimates of the district cotton area along with their percentage standard errors were calculated by the three methods and are given in Table III.

TABLE III

Estimated cotton area and its percentage standard error with size-cum-*geographic stratification*

Method of Estimation	Estimated cotton area in 1,000 acres	Percentage standard error
Direct ..	648	3·0
Regression ..	608	2·4
Ratio ..	631	2·0

The comparison between Tables II and III indicates that the errors for each method are reduced by a small extent, by employing geographic stratification in combination with size stratification.

The above methods were employed for the estimation of area of another crop namely, wheat, occupying a relatively smaller area. The same sample was used for this purpose. Table IV gives the standard errors in percentages.

For a given size of the sample, standard errors are much higher for district wheat area than that for cotton. It was to be expected as both size stratification and methods of estimation would be less efficient for wheat owing to the poor correlation observed between

TABLE IV

Percentage standard errors of estimated wheat area

Method of estimation	Size stratification	Size-*cum*-geographic stratification
	percentage	standard error
Direct ..	12·1	10·9
Regression ..	11·0	14·3
Ratio ..	11·2	10·6

the size and wheat area of a village. Among the methods of estimation, the ratio and regression methods of estimation are superior to direct method of estimation but the reduction in error is not appreciable. Ratio is superior to regression very slightly.

EFFICIENCY OF STRATIFICATION

The effectiveness of stratification appropriate to the design discussed here is now considered. For a stratified sample with a variable sampling fraction, Yates (1949) has discussed the efficiency of stratification. An estimate of the average within strata mean square, s_1^2, and of the overall mean square must be calculated from the proportions $h_i = N_i/N$ of the units of the population in the different strata. If s_i^2 is the mean square within the i-th stratum, s_1^2 is given by $s_1^2 = \sum^{k} h_i s_i^2$ and the overall mean square 's^2' is given by

$$s^2 = s_1^2 + \Sigma\, h_i \bar{y}_i^2 - \bar{y}^2 + \Sigma\, h_i\,(1 - h_i)\,\frac{s_i^2}{n_i}$$

where $\bar{y}$ is the estimate of the population mean derived from the sample and is consequently equal to $\Sigma\, h_i \bar{y}_i$. In case of proportional sampling, the relative precision of a stratified and a simple random sample is given by the ratio s^2/s_1^2. In order to see to what extent stratification is efficient when supplementary information is used, s^2/s_1^2 was also calculated for the ratio method of estimation. Tables V and VI give the relative precision of a stratified and a simple random sample with and without supplementary information.

It will be seen that stratification for village area as a variate has proved highly effective for cotton and to a lesser extent for wheat. The comparison of these figures with those of the ratio method is very interesting and leads to a generally applicable conclusion. The effect

TABLE V

Effects of size stratification

	Cotton	
	Village area	Ratio Y/X
Variance without strata s^2 ..	99254	·0168
,, with ,, s_1^2 ..	34025	·0088
Ratio s^2/s_1^2 ..	291%	190%
	Wheat	
Variance without strata s^2 ..	20069	·0024
,, with ,, s_1^2 ..	16313	·0027
Ratio s^2/s_1^2 ..	123%	89%

TABLE VI

Effect of size-cum-*Tahsil stratification*

	Cotton	
	Village area	ratio Y/X
Variance without strata s^2 ..	123004	·0061
,, with ,, s_1^2 ..	26386	·0043
Ratio s^2/s_1^2 ..	466%	141%
	Wheat	
Variance without strata s^2 ..	18928	·0034
,, with ,, s_1^2 ..	12614	·0036
Ratio s^2/s_1^2 ..	150%	94%

of stratification by size of the village is considerably reduced for the ratio method for cotton and this effect is not visible at all for wheat. The explanation is that stratification by size has very nearly accomplished the same purpose for which the ratio method is employed, *viz.*, to remove from the sampling error a portion of variability resulting from the variation in the size of the sampling unit. Consequently one may either choose a particular factor like size of the sampling unit for stratification or utilize it for the method of estimation.

SIZE OF SAMPLE

The size of the sample required to estimate the population character with a given accuracy, obviously depends upon the variability of the material and the extent to which it is possible to eliminate the different components of variability from the sampling error by proper stratification and other means. The procedure for determining the size of the sample is discussed here.

If n_i sampling units are selected out of N_i units in the i-th stratum, s_i^2 is the mean square within the i-th stratum, the variance of the estimated total, Y, is (Cochran, 1942)

$$V(Y) = \sum_{i=1}^{k} N_i^2 \frac{s_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i} \tag{7}$$

(7) can also be written as

$$V(Y) = \sum_{i=1}^{k} N_i^2 \frac{s_i^2}{n_i} - \sum_{i=1}^{k} N_i s_i^2 \tag{8}$$

the principle of optimum allocation requires that (Neyman, 1934)

$$\frac{n_i}{n} = \frac{N_i s_i}{\Sigma N_i s_i} \tag{9}$$

where $\Sigma n_i = n$ is the total number of units in the sample. From (9) we get

$$n_i = \frac{N_i s_i}{\Sigma N_i s_i} \cdot n \tag{10}$$

A formula for determining n can be derived from (8) and (10). Substituting the value of n_i from (10) in (8), we obtain

$$n = \frac{(\Sigma N_i s_i)^2}{V(Y) + \Sigma N_i s_i^2} \tag{11}$$

The coefficient of variation of Y is

$$c.v.\ Y = \frac{\sqrt{V(Y)}}{Y} \tag{12}$$

and hence,

$$n = \frac{(\Sigma N_i s_i)^2}{Y^2 (c.v.\ Y)^2 + \Sigma N_i s_i^2} \tag{13}$$

The total number of villages required to estimate the distirct cotton area with a given precision is given in Table VII.

TABLE VII

Sample size for a given precision

Level of precision S. E. %	No. of villages required	
	Size stratification	Size-*cum*-geographic stratification
1	633	448
2	282	179
3	148	91
4	97	54
5	61	35

Table VIII gives the number of villages required for a given level of precision for estimation of wheat area of the district.

TABLE VIII

Sample size for a given precision

S. E. %	Size stratification	S. E. %	Size-*cum*-geographic stratification
5	370	3	383
6	303	4	288
7	246	5	224
8	200	6	169
9	169	7	134

These results which are derived for the district estimate indicate that the number of villages required to provide the estimates of district area with a reasonably adequate level of precision is small for a crop with a high intensity of cultivation as in cotton and is much larger for a crop with a lower intensity as in wheat.

Complete Matching

Matched sampling is a device for using auxiliary information available on the sample data. Sukhatme (1950) has considered the possibility of reliably estimating the change in crop acreages from one season to another by enumerating a small number of villages and using the ratio method of estimation on previous year's records of crop areas. In view of the high correlation that usually exists in the two consecutive years' areas, accurate preharvest forecasts of crop acreages can be made possible by using completely matched samples in the successive years. Jessen (1942) has also reported the use of matched samples as an efficient method of measuring year to year changes. He found that in comparison with samples taken independently each year, matched samples are 2·5 to 20 times more accurate for estimating different items of agro-economic interest.

For this study, previous year's records of area for the 108 villages were available. For each of these villages, cotton and wheat area are known for the current year and the previous year. If y represents the current year's village crop area and x represents that for the previous year, the ratio method can be used in the following manner. Let $\bar{y}_{nj}$ be the mean of y in the j-th stratum based on n_j villages and similarly $\bar{x}_{nj}$, the mean of x in the j-th stratum based on the same n_j villages. Then

$$\text{ratio estimate } Y_r = \frac{\sum^{k} N_j \bar{y}_{nj}}{\sum^{k} N_j \bar{x}_{nj}} \cdot N\bar{x}_p$$

and the variance of this estimate is given by

$$V(Y_r) = \bar{y}_p{}^2 \sum \left[N_j \; \frac{(N_j - n_i)}{n_j} \right] \left[\frac{\sigma^2 y_i}{\bar{y}_p{}^2} + \frac{\sigma^2 x_i}{\bar{x}_p{}^2} - \frac{2 \operatorname{cov}(x_j, y_j)}{\bar{y}_p \bar{x}_p} \right]$$

This formula for the variance was used to find out the variance of the ratio estimate from this completely matched sample with size-*cum*-geographic stratification. The standard error per cent. of the ratio estimate for the two crops on the same 108 villages previously selected is given in the following table. The formula for $V(Y_r)$ involves an approximation

$$\frac{\bar{x}_{pj}}{\bar{x}_{sj}} = 1$$

TABLE IX

Percentage standard errors for completely matched sample

Crop	S.E. %
Cotton	0·7
Wheat	3·8

It will be seen that there is a remarkable reduction in the error as compared with the errors previously given. Especially for wheat, which was subject to a considerably greater sampling error, this estimate has proved to be quite reasonably accurate. This is because of the fact that the villages were very variable in regard to either y or x, for wheat area, but the changes from year to year are to a great extent similar. The wheat crop is restricted to certain areas with deep and moisture retentive soil, as stated previously and seasonal fluctuations in this area, in a given tract, would, therefore, be relatively smaller.

INCOMPLETE MATCHING

Although completely matched samples give highly accurate estimates, repeated resurvey of the same units, as is done in complete matching, would be inexpedient and is likely to introduce a slight bias. Sometimes, an incompletely matched sample can be selected in which a part of the previous sample is resurveyed and another part selected anew. Study of the efficiency of an incompletely matched sample has been made by Jessen (1942) and its use has been discussed by Yates (1949) who describes it as sampling on successive occasions with partial replacement of the units.

Let the sample values in the previous year be denoted by x and those in the current year by y, the values belonging to villages included in both years by x', y' and those included in the previous year by x''. If a fraction λ of all uints included in the previous year is taken for the second year and a fraction $\mu = 1 - \lambda$ is replaced, then,

$$\bar{y} = \bar{y}' + \mu b\,(\bar{x}'' - \bar{x}') \qquad (14)$$

This is the regression estimate of the population mean from the matched sample. In addition to this, another independent estimate can be formed from the villages included in the current year only. There will thus be two independent estimates of the same quantity and the most accurate estimate $\bar{y}_w$ will be provided by the weighted mean of the two. The correct weights are

$$\frac{\lambda}{1 - \mu^2 r^2} \text{ and } \frac{\mu\,(1 - \mu r^2)}{1 - \mu^2 r^2}.$$

where 'r' is the correlation coefficient between x and y derived from the matched portion. These weights are the reciprocals of the variance of the two independent estimates. The variance of this weighted mean is given by Yates (1949)

$$V(\bar{y}_w) = \frac{V(y)(1-\mu r^2)}{n(1-\mu^2 r^2)} \tag{15}$$

where $V(y)$ in the present study is the variance of the individual village crop areas, and n is the total number of villages in the sample. Had the whole sample been completely independent, the estimated mean would have the variance

$$V(\bar{y}) = \frac{V(y)}{n} \tag{16}$$

from (15) and (16) it will be seen that the relative efficiency of incomplete matching depends upon μ and r. Hence

$$\text{efficiency} = \frac{1-\mu^2 r^2}{1-\mu r^2} \tag{17}$$

In the present study, from each tahsil of the Akola District, half the sample of villages was matched, with that of the previous year and half was selected independently for the current year. The value of μ is, therefore, $\frac{1}{2}$. Efficiencies of half matched samples resulting from the data for Akola District by using (17) are given in Table X for cotton and wheat.

TABLE X

Relative efficiency of half matched sample over completely independent sample

Tahsil	Cotton		Wheat	
	r	% efficiency	r	% efficiency
Akola ..	·9773	148	·9969	149
Akot ..	·9815	146	·9650	144
Balapur ..	·9953	149	·9915	148
Maugrulpir ..	·9965	150	·7494	120
Murtizapur ..	·9940	145	·9693	144
Basim ..	·9657	144	·9836	147

From the results of half matched sample, it is possible to generalize the conclusions on efficiencies to be expected for varying values of μ. The percentage efficiencies to be expected from incompletely matched samples for values of $\mu = \frac{1}{4}, \frac{1}{3}$ and $\frac{3}{4}$ are given for the two crops in the following table:

TABLE XI

Percentage efficiency of incomplete matching for varying values of μ

Tahsil	Cotton			Wheat		
	$\mu = \frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$	$\frac{1}{4}$	$\frac{1}{3}$	$\frac{3}{4}$
Akola ..	124	131	163	125	133	173
Akot ..	124	131	165	123	130	158
Balāpur ..	125	132	172	124	132	170
Mangrulpir ..	125	133	173	112	115	118
Murtizapur ..	124	132	172	123	130	160
Basim ..	123	130	158	124	131	166

It will be seen from the above tables that the precision of the estimates of crop acreages in the second year can be substantially improved by resorting to the method of incomplete matching. The correlations are of the order of ·96 to ·99. This high correlation is a very promising feature for adopting this technique.

It is necessary to find out the optimum fraction that should be matched. Assuming that the cost of obtaining information in respect of each of the matched and unmatched units is the same, the problem can be solved by minimising the expression for the variance of the weighted mean as given by (15). We know that $V(\bar{y}_w)$ is

$$V(\bar{y}_w) = \frac{V(y)}{n} \cdot \frac{(1 - \mu r^2)}{(1 - \mu^2 r^2)} = v \qquad (18)$$

then the optimum value of μ, a fraction unmatched can be obtained by putting $\delta v / \delta \mu = 0$, this gives

$$\mu = \frac{1}{r^2} (1 - \sqrt{1 - r^2}) \qquad (19)$$

It corresponds to Jessen's expression (1942) given by

$$\frac{m}{n'} = \sqrt{\frac{1}{1 - r^2}} = \frac{\mu}{\lambda} \tag{20}$$

where m is the number of units unmatched and n' is the number that is matched. We know the value of 'r' for both the crops and we can, therefore, calculate the ratio m/n' given above for these two crops. Table XII gives these values.

TABLE XII

Optimum ratio of unmatched to matched sample units

Tahsil	ratio $\mu/\lambda = 1/\sqrt{1 - r^2}$	
	Cotton	Wheat
Akola ..	4·8	12·8
Akot ..	5·3	3·8
Balapur ..	10·4	7·7
Mangrulpir ..	11·9	1·5
Murtizapur ..	9·2	4·2
Basim ..	3·8	5·6

Matching as a special case of double sampling has been considered here. If correlations were perfect, the gain in efficiency would be 50 per cent. when μ is half. With increasing values of μ, the efficiency goes on increasing up to a point and then decreases. For this particular form of sampling, the smaller the portion of matched sample, the greater the efficiency of the incompletely matched sample. This has been observed by Jessen (1942) also. This is obvious from (14), which is a regression estimate of the mean from the matched portion. Here if μ is zero, the second term on the right-hand side is zero and x' and x'' become equal and we get instead a straight mean. If μ is equal to one, *i.e.*, there is no matching at all, we obviously get a straight mean again. Thus for extreme values of μ, the estimate is similar and consequently there is no gain over that of an independent sample.

SUMMARY

The present investigations were undertaken with the object of developing the area method of sampling for estimation of crop

acreages. The sampling studies were made in conjunction with the statistics of complete enumeration of crop areas in the Akola District of Madhya Pradesh for the year 1945–46, taking cotton and wheat areas as examples of crops with high and relatively lower intensities of cultivation respectively. A number of aspects of single stage sampling of villages is considered. The crop area of a village was treated as a variate for this purpose. The problems studied were:

(1) The relative efficiency of the three methods of estimation, namely, (*a*) simple mean, (*b*) regression estimate and (*c*) ratio estimate, the geographical area of the sampled villages being used as supplementary information for the last two methods of estimation.

(2) Determination of the size of the sample for achieving a given precision of the estimate.

(3) Efficiency of stratification based on (*a*) size and (*b*) size-*cum*-geographic location.

(4) Complete and incomplete matching with previous year's area records.

With reference to the actual data for Akola district that were employed and analysed in connection with the above sampling studies, the results may be summarised as follows:

(1) The ratio and the regression estimates were distinctly superior in precision to the simple mean, the unweighted mean ratio being slightly superior to regression.

(2) The number of villages required to be sampled for estimation of district cotton area with five per cent. standard error was about 61 and 35 for size and size-*cum*-geographic stratification respectively for the simple estimate of the mean. The corresponding numbers of wheat were about 370 and 224 villages respectively.

(3) The efficiency of stratification based on size of the sampling units was observed to be 291 per cent. and 123 per cent. for cotton and wheat respectively. The efficiency of size-*cum*-geographic stratification was seen to be 466 per cent. and 150 per cent. for cotton and wheat respectively. It was also found that stratification by size accomplished more or less the same purpose as the method of estimation based on information regarding size without resorting to stratification.

(4) The gain in precision by complete matching appeared to be 16·3 and 8·7 times for cotton and wheat respectively in comparison to the estimate of the simple mean. Half matching with previous year's records of areas brought about 150 per cent. efficiency in the

estimate. The optimum fraction of the units to be matched is also worked out.

ACKNOWLEDGMENTS

This work was carried out under the guidance of Dr. V. G. Panse, Statistical Adviser, Indian Council of Agricultural Research, New Delhi. I am thankful to him for the keen interest he took in this work and for making useful suggestions during the course of these investigations.

I am also indebted to the Director of Land Records, Madhya Pradesh, and to the Deputy Commissioner, Akola, for kindly making available the patwari registers and village maps for this study.

REFERENCES

Cochran, W. G. .. "Sampling theory when the sampling units are of unequal sizes," *Jour. Amer. Stat. Assoc.*, 1942, **37**, 199.

Jessen, R. J. .. "Statistical Investigation of a sample survey for obtaining Farm Facts," *Iowa Agric. Exp. Sta. Res. Bull.*, **304**, 1942.

Neyman, J. .. "On the two different aspects of the representative method: the method of Stratified Sampling and the method of Stratified Sampling and the method of Purposive Selection," *Jour. Roy. Stat. Soc.*, 1934, **97**, 558.

Panse, V. G. and Kalamkar, R. J. "Estimation of crop yields," *Curr. Sci.*, 1944, **9**, 223.

——— and Sukhatme, P. V. "Crop surveys in India — I," *Jour. Ind. Soc. Agric. Stat.*, 1948, **1**, 34.

Sukhatme, P. V. .. "Sample surveys in Agriculture," *Presidential Address, Section of Statistics, Indian Science Congress*, Poona, 1950.

Yates, F. .. "A Review of Recent Statistical Developments in Sampling and Sampling Surveys," *Jour. Roy. Stat. Soc.*, 1946, **109**, 12.

——— .. *Sampling Methods for Censuses and Surveys*, Charles Griffin and Co., London, 1949.