CROP CUTTING EXPERIMENT ON SUGARCANE IN A FARM CULTIVATION

J. M. SEN GUPTA
Statistical Laboratory, Calcutta.

Introduction

A crop cutting experiment on sugarcane was conducted at the Gopalpur Agricultural Farm in March 1947, under the direct guidance and supervision of statisticians deputed from the Indian Statistical Institute. The farm had been divided into several blocks for purposes of cultivation and the crop in only two of these blocks (M & N) was available for the experimentation, the rest having been already harvested. The object was to devise a sampling technique for forecasting the outturn of sugarcane keeping in mind the following considerations :—

(i) The quantity of cane harvested for sampling is a direct loss to the farm. The fraction sampled must therfore be kept as small as possible.

(ii) The survey should be done much earlier than normal harvesting period to make the forecast useful.

(iii) The sampling procedure must be very simple and straightforward.

(iv) The outturn of canes should be estimated not only for the entire farm, but also for each individual block.

The crop was laid out in furrows, each bigha containing 100 such furrows of an average length of 44 feet. The total area sampled was 36 bighas or 10.91 acres. The sample units consisted of linear cuts of 11 feet, each being taken in four equal sub-cuts of 2 feet 9 inches in length.

A simple device consisting of a horizontal metal rod 11 feet in length resting on two pegs at the two ends provided with five pointers spaced 2 feet 9 inches apart and fixed at right angles to the horizontal piece, was actually used in demarcating the sample cuts. The rod was placed parallel and close to the row of canes and the pointers passing through the canes were the natural borders of the harvested cuts.

Linear cuts were specially adopted to simplify the field work. For one thing, such cuts would avoid the uncertainty of one furrow more or less, which is obviously involved in the use of square or rectangular cuts.

Total outturn is obtained as the product of the estimated mean yield rate per foot of the furrow length and the total running feet (r. ft.) of canes. Total running feet of canes is not directly known but could be estimated from the average furrow length and the total number of furrows. Variability of the individual furrow length, as obtained from the collected records, is only 2.04%. A careful measurement of a number of furrows chosen at random would give us a very precise estimate of the average furrow length.

The effect of size of sample cut on observational bias and coefficient of variation

An overestimating bias in estimated yield rates, usually attributed to the sampler's general tendency[1] of including the bordering plants, is highly pronounced in the case of very small cuts, where the uncertainty at the border is not negligible in comparision to the total yield of the cut. In the present experiment, the sub-cuts into which the sample cut measuring 11 feet was split up, were, under special instructions, harvested strictly in a numerical order, proceeding from one end to the other. The first sub-cut had thus an unconditional freedom at both the borders in the selection of individual canes, while the second, third and fourth had only one of the two ends free for the choice of canes, the first being already defined by the preceding cut. It is, therefore, expected that the first sub-cut would overestimate while the other sub-cuts (through a process of compensation) would lead to estimates almost free from any bias. The estimated yield rates in lbs per foot with standard errors based on cuts of different sizes are shown in Table (1) separately for the Blocks 'M' & 'N'. A glance at the columns (3) and (6) reveals that the first sub-cut overestimates as compared to the other sub-cuts. The bias in case of sub-cut 1, does not, however, come out as statistically significant in relation to the yield rate based on the largest size cut, namely, 11 feet in length. It might have been thought that this may have been due to the size of the sample being too small. But the same pattern repeated in both the Blocks, indicates that the first sub-cut may have a slight overestimating tendency. It is also to be noticed that the yield rate remains fairly steady after a size of 5 feet 6 inches and onwards.

It thus appears that the bias with linear cuts in sugarcane is generally small, and a size of 2 feet 9 inches should be quite adequate. On the other hand a size of 2 feet 9 inches may be rather small compared to the inter plant spacing of 1.5 feet only. Further investigations are obviously needed. To be on the safe side a size of 5 feet 6 inches properly compensated should give satisfactory results.

For effecting an automatic 'compensation' in this sense, a pair of contiguous sub-cuts should be demarcated, while the actual sampling should be confined to the unit, second in order of demarcation. It is important not to give out the real purpose of demarcating the superflous half, so that the enumerator may not get conscious of this dodge. The samplers, for instance, may be asked to count the number of canes for both these units, which would give some useful auxiliary information.

Table (1) Mean yield in lbs per foot and coefficient of variation as estimated from the various sizes

Sub-cuts		Block M (n=144) total field—158310 r.ft.			Block N (n=64) total field—70354 r.ft.		
number (1)	length (2)	mean (3)	S.E. (4)	C.V. (5)	mean (6)	S.E. (7)	C.V. (8)
1	2'-9"	2·591	0·222	102·6	1·347	0·299	177·0
2	2'-9"	2·310	0·194	100·4	1·079	0·274	203·6
3	2'-9"	2·302	0·193	98·2	1·102	0·256	175·8
4	2'-9"	2·472	0·194	94·3	1·137	0·272	191·6
1+2	5'-6"	2·455	0·164	80·3	1·213	0·235	155·5
1+2+3	8'-3"	2·424	0·142	70·1	1·196	0·222	149·1
1+2+3+4	11'-0"	2·430	0·127	62·8	1·181	0·215	145·8

[1] Mahalanobis, P. C. (1944): On large scale sample surveys, *Phil. Trans. Roy. Soc.*, 23(B), 409.

So far as the sampling variation is concerned, the coefficient of variation as shown in cols. (5) and (8) of Table I falls with an increase in size of cut though not in an invers proportion to the increase in cut size. Comparing Block M and N, we note that Block N with a lower yield level has a higher variability.

Study of the concomitant characters

In situations where the direct estimation of any character is a costly affair it would be very helpful if one or more characters highly correlated with the character to be estimated and capable of quick measurement at a very low cost could be found out. The use of such characters for estimation of crop yields would mean that a very large number of sample units could be measured with comparative ease with such characters alone, while actual harvesting being confined to a small fraction of such sampling units.

The present experiment involved examining each individual cane of the second sub-cut (2′-9″) only, with reference to the following characters namely, the length in inches X_1, diameters in inches at foot X_2, at middle X_3, and at top X_4.

The auxiliary variates represent the averages per cut in the case of diameters (x_2, x_3, x_4) and total per cut in the case of length (x_1). The proportionate volume in cubic inches per cut, denoted by x_5 is calculated by the formula $x_5 = x_1(x_3)^2$ for each individual cut. Another estimate of the volume x_6 is calculated as $x_6 = \frac{\pi x_1}{16}\left[x_2^2+2x_3^2+x_4^2\right]$ summed over individual canes in a cut.

The correlation between yield of canes in lbs per cut (y) is considerable with x_1 and significant though lower with each of the three variates x_2, x_3, and x_4. A very high correlation is observed with x_5 and the highest is obtained with x_6. These correlation coefficients are based on 109 cuts excluding the cuts without yield. The parameters 'a' and 'b' of a linear regression on the variates x_1, x_5, and x_6 together with their standard errors are also shown in cols. (5) and (6) of Table (2). All units refer to lbs per 2′-9″.

Table (2) : Correlation between yield rate and the auxiliary variates

auxiliary variates			correlation coefficients	regression coefficients	
symbol	mean	C.V.		$a \pm$S.E.	$b \pm$S.E.
(1)	(2)	(3)	(4)	(5)	(6)
x_1	280·8	75·0	·7096	2·532 ± ·688	·0203 ± ·0015
x_2	1·056	16·6	·5198	—	—
x_3	0·975	15·7	·5613	—	—
x_4	0·855	16·6	·5043	—	—
x_5	216·0	107·0	·8944	1·404 ± ·424	·0242 ± ·0012
x_6	235·8	76·4	·9552	0·615 ± ·285	·0323 ± ·0010

The very high correlation of the yield with the total length per cut, specially with the computed volumes per cut called the x_5 and x_6 variates is encouraging and indicates that the double sampling technique can be applied with advantage. The coefficients 'a' in the fit $y = a + bx$ is significant in all the cases. It may be interesting to compare the coefficient 'b' of the yield rate on x_6, calculated by the least square method with the mean ratio of cane weight to the cane volume which have been obtained as 0.0395 lb per cubic inch with a C.V. of 38.5%.

ALTERNATIVE SCHEMES OF DOUBLE SAMPLING

A comparison of the relative efficiencies of a few alternative schemes, in terms of the cost of operations, is being made here. The field cost components per cut in terms of money have been roughly estimated on the basis of present field experiences as : (i) Rs. 0.062 per auxiliary character (ii) Rs. 0.37 for harvesting etc., and (iii) Rs. 0.0876 for loss of canes. The cost of canes wasted in sample crushing, was valued at Re. 1/- per maund, delivered at the factory gate.

On this basis, four different schemes each using three of the sizes 2'-9", 5'-6" and 11'-0" have been drawn up and shown in Table (3). The first is a single sampling scheme. The other three are double sampling schemes utilising (i) length of canes (ii) length and the mid-diameter of the canes (iii) length and the three diameters of the canes. It has been assumed that the cost increases proportionately with an increase in the size of the cuts and in the number of the auxiliary characters to be read. This assumption may be made for all practical purposes, time spent in making journeys etc. being negligibly small in a compact patch of area sampled. The optimum sizes at two stages of sampling for achieving a precision of 2.5% are shown in cols. (4), (5); (8), (9); (12), (13). Total cost of field operations have been given in cols. (6), (10) and (14) and the fraction of cane sampled from a total block of 100 acres, have been given separately for each size of cut in cols. (7), (11) and (15).

It is evident that a rise in the price of cane or a reduction of cost in taking the auxiliary records would alter the whole set up and the utility of a double sampling scheme will be rendered much more important. As such, it appears that there is enough scope for improving the technique of height and diameter measurements.

TABLE (3) : OPTIMUM SIZES AT TWO STAGES OF SAMPLING FOR A PRECISION OF 2.5%

				2'-9" (C.V. = 100%)				5'-6" (C.V. = 80%)				11'-0" (C.V. = 62·5%)			
scheme	auxiliary variate	correlation coefficient	K	N	n	total cost (Rs.)	fraction sampled in 100 acres.	N	n	total cost (Rs.)	fraction sampled in 100 acres.	N	n	total cost (Rs.)	fraction sampled in 100 acres.
(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
I	nil	—	—	—	1600	714	·33%	—	1024	913	·43%	—	625	1115	·52%
II	x_1	·7090	7·0	2808	1132	982	·24%	1836	724	873	·30%	1033	408	984	·34%
III	x_6	·8944	3·5	2470	640	601	·13%	1581	409	769	·17%	889	239	867	·20%
IV	x_4	·9552	1·75	2000	357	690	·07%	1304	310	928	·13%	745	186	1077	·16%

K = ratio of cost of harvesting to cost of measuring auxiliary variate.

N = no. of sample units for which the auxiliary characters are only to be measured.

n = no. of sample units for which all the characters are to be measured.

From Table (3), it appears that the smallest cut of 2 feet 9 inches in length is in general more efficient than the bigger cuts and the double sampling schemes are in all cases more economical than the single sampling ones. The best results are however obtained with the variate x_6 involving length and the mid-diameter of plants. The smallest sampling fraction is, however, given by the last scheme using length and all the three diameters.

Paper received : 31 October, 1947.