AGRICULTURAL STATISTICS

Chairman : Mr. K. RAMIAH

STATISTICAL DEFINITION OF STANDARD YIELD OF CROPS*

By P. C. MAHALANOBIS

The object of crop-cutting experiments for revenue purposes is to determine the yield of one or more standard crops grown on different classes of land in different years. Let us consider a hypothetical experiment which admittedly is impossible to carry out in practice. Let us suppose that we gather the whole of the crop on each individual plot throughout the area under consideration (which may be a *taluk*, or a sub-division, or a district, or the whole province). Let us also suppose that this work is repeated from year to year for a long period. If this can be done it would be possible to prepare a two-way yield table shown schematically in the following diagram.

Nature of Season	Class of Land					
	α	β	γ	δ	ε	etc.
(1)	(2)	(3)	(4)	(5)	(6)	(7)
α'	$y(\alpha\alpha')$	$y(\beta\alpha')$	$y(\gamma\Psi')$	$y(\delta\alpha')$	$y(\varepsilon\alpha')$	
β'	$y(\alpha\beta')$	$y(\beta\beta')$	$y(\gamma\beta')$	$y(\delta\beta')$	$y(\varepsilon\beta')$	
γ'	$y(\alpha\gamma')$	$y(\beta\gamma')$	$y(\gamma\gamma')$	$y(\delta\gamma')$	$y(\varepsilon\gamma')$	
δ'	$y(\alpha\delta')$	$y(\beta\delta')$	$y(\gamma\delta')$	$y(\delta\delta')$	$y(\varepsilon\delta')$	
ε'	$y(\alpha\varepsilon')$	$y(\beta\varepsilon')$	$y(\gamma\varepsilon')$	$y(\delta\varepsilon')$	$y(\varepsilon\varepsilon')$	
etc.						

We may classify the land into a convenient number of grades such as α, β, γ etc. The actual number of such classes, of course, is a matter to be settled from practical considerations, and may be 4, or 5, or 16, or 20, or any other convenient number. These are shown as columnar headings in the diagram. In the same way we may also classify the seasons (on the basis of amount and distribution of rainfall together with such other factors as may be considered appropriate from the agricultural point of view) into a number of types α', β', γ' etc. Here again the actual number may be 4 or 5 or 15 or any other number. These are shown schatically in column (1).

If the field survey is repeated for a very long time we shall get records for each class of land (α, β, γ etc.) for a large number of years of each type (α', β', γ'....etc.) Taking the average of these yields we can then get the average or expected yield for each type of land in each type of season ; and we can enter these expected or normal yields in the appropriate place in the diagram. Thus $y(\alpha\alpha')$ will represent the expected yield in land of class α in a season of type α' ; $y(\gamma\beta')$ will represent the yield in land of class γ in season of type β' and so on.

If once such a table could be formed then there would be no difficulty in deciding the yield of any particular class of land in any particular season. We cannot, however, construct such a table by a complete survey of all plots continued from year to year as this is altogether impracticable from the point of view of both expenditure and administrative arrangements. It is, therefore, necessary to proceed by the method of random samples, that is, we have to construct an approximate table by conducting crop-cutting experiments on a *limited number*

* From a note dated 23rd September, 1940, to the Director of Agriculture, Bombay.

of holdings distributed among a *limited number of classes of land* in each year. In doing this we have no right to assume that the proportional yield of two classes of land would remain the same under different seasonal conditions, thus the yield of land of class δ may be half of the yield of land of class α in an exceptionally good season of type α'; and yet in a bad season of type γ' the yield of land of class δ may be reduced to only ten percent of the yield in land of class α. In other words, it is not possible to assume that the expected yield in different columns would be in a constant proportion which means that it may be necessary to use different proportions for estimating the yield in the same pair of classes of land in different years.

In crop-cutting experiments we have to settle what should be the area of the ultimate sampling units or individual cuts in the field ; we have also to decide how many classes of land should be included in the sampling programme ; and how many sampling units should be allotted to each class of land. Finally we have also to decide how the sampling units assigned to any particular class of land should be distributed over the geographical area. As the time or money available for crop-cutting experiments is not unlimited it is necessary to settle these details in such a way that the work may be done with minimum expenditure.

This then is the ultimate statistical problem. The total cost being given, we have to find out what is the area of each sampling unit ; how many such sampling units should be used altogether ; how these should be allotted to different classes of land ; and how the sampling units assigned to each class of land should be distributed geographically in order that the margin of error in the final result should be reduced to a minimum. Alternatively, we may define the object in a slightly different way. The order of accuracy of the final results being given, we have to find out what would be the best size and distribution of sampling units in order to do the work at a minimum cost.

I may explain here that by 'total cost' in this connexion I mean essentially the total amount of time spent by workers of different grades plus the expenditure for usual appliances and contingencies. It is, however, convenient to convert the cost of human labour into money value as this enables rough comparisons being made for different regions.

One thing should be noticed in this connexion ; it is not possible to give a complete sampling programme all at once. The technique has to be developed on experimental lines. The method of the exploratory survey is of great advantage in this matter. We start with whatever knowledge we possess, and conduct a small scale experiment keeping detailed record of costs. The information gathered in this way is then used to prepare the plan for a survey on a larger scale in the next season. In this way the work is expanded from year to year by the method of trial and success until an efficient technique is developed.

In the schematic diagram we have intentionally used neutral phrases like α, β, γ or α', β', γ'..... It is possible, however, to attach anna labels. That is, we may, if we like, call land of class α as 16-anna land ; land of class β as 15-anna land etc. Similarly we may attach anna labels to the seasonal types. Thus α' may denote 16-anna season, β' 15-anna season and so on, proceeding by intervals of one anna. (Or we may condense the classification by using intervals of two annas or four annas). Even if we use the anna labels it must however be emphasized that these are mere labels. Once this is understood, and it is remembered that the yields in different columns cannot in general be assumed to be proportional, it is easy to see that 8-anna land in one year may give 50 percent of the yield of 16-anna land, and in another year may give only 10 percent.

[*Paper received : 23 September, 1940*].