Sampling for the Collection of Agricultural Statistics in England and Wales

The agricultural statistics for England and Wales collected by the Ministry of Agriculture, Fisheries and Food may be conveniently classified under two heads ; physical (crops, livestock, labour, machinery, etc.) ; economic and financial (wages, prices, rents, profit and loss accounts, etc.). This article deals only with the collection of physical statistics, principally from the agricultural censuses in March, June, September and December every year, describing some of the sampling methods used or planned to be used for this purpose.

The Agricultural returns

The primary sources of a great deal of the agricultural statistics are the returns made by the 370,000 occupiers of agricultural holdings of over one acre. These occupiers are clearly in the best position to furnish accurate information as to the uses to which the land is being put, the number of livestock and machines on the land, and the numbers of workers employed. Since 1866 farmers have rendered an annual return in June of the acreage of their crops and grass and the number of their livestock. The number of questions on which separate information is required has slightly increased since pre-war days, mainly because of additional horticultural questions. The June 1957 form contains 147 questions, but even at June 1937 there were as many as 132. The items themselves, and their definitions have changed very little over the years. The tables which follow, which give the results of the June 1957 Census, illustrate the number and variety of the questions asked.

The total numbers in the country of the various types of cattle, sheep, pigs and poultry vary over the years. Some of this variation is purely seasonal—for example, there are many more hens at the beginning of the winter than there are in June, when a number have already been culled or have died. But other variations are caused by farmers' changes in production policy, which in some cases can occur quite suddenly. Thus the numbers of sows and gilts put to the boar may show an upward or a downward swing at any time. For these reasons, it has been found that a single annual return does not give sufficient information for the proper determination of agricultural policy and in December 1939 a regular series of quarterly returns was instituted, for March, September and December, in addition to the normal June return. These three extra returns are shorter than the June return as they are mainly confined to the livestock and labour items, although since March 1956 they have been lengthened by the addition of machinery questions, a procedure which has replaced the special Machinery Censuses which had been held biennially since 1942.

Other Censuses are confined to certain groups of farms. Thus the biannual Glasshouse return is confined to farms with more than a certain area under glass, and the September and December vegetable form is sent only to occupiers who at the previous June returned some area as under vegetables. This process saves troubling the large number of farmers who are not concerned with this item.

Provisional results

Up to June 1951, the provisional results for each June census were obtained about mid-August from nearly all the returns, allowance being made for the missing returns by grossing up the results proportionately to the known areas under crops and grass. For the March, September and December censuses about 85 per cent. of the forms were customarily used for the provisional results, " grossing up " being done from a simple " balanced sample " *i.e.*, after making sure that for each county the average size of holding in the 85 per cent. sample was approximately equal to the known average size of all holdings in that county.

Because of the large proportion of all holdings included, the results were of high accuracy. The differences between provisional and final results were almost entirely due to corrections of errors on individual returns which late enquiries had settled.

Random samples

The possibility of using smaller random samples had been considered for several years but it was not until just after the war that a thorough investigation of the variances of the different items could be calculated, for a couple of counties (using the 1944 Returns). These analyses, from punch card machines, were supplemented by an extended series of calculations on the 1948 returns, for one county. From these calculations it was possible to estimate the expected sampling errors for each item for samples of various sizes and designs. In particular, measures were obtained of the gain in efficiency through using stratified samples with variable sampling fractions or by using ratio methods of estimation (e.g. linking for each item with the same farms' results for a previous census). The Ministry had benefited from the research work on sampling for agricultural statistics, conducted by the staff at Rothamsted Experimental Station.

By June 1951 it was considered that experiments had proceeded far enough to take the first steps in applying the theory to the practice of presenting the results of the current agricultural censuses. For that June census, preliminary results for England and Wales were estimated by using one-half of the 12,000 or so parishes in the country. In this first year of the scheme, various methods of selecting and grossing up such a sample of parishes were tried, and as a result of these experiments it was decided that the best combination of reliability, speed and economy would be achieved in future by selecting parishes at random from each sub-district in a county, after first dividing up the district into pairs of adjacent and similarly

sized parishes to ensure an even distribution over the district. The results would be grossed up separately for each item within a county, by comparing the current results with those for the same parishes at the previous June. For a sample of this size (50 per cent. of each county) this grossing up system is more efficient than any simpler method, except for some of the rarer items which are grown or kept on one set of farms in one year and on another in the next year. Such items will have a relatively high sampling error under any grossing up method, even with a large sample of efficient design, because there is no way of pinning them down into fixed " specialist " strata, and they are not usually more prevalent in larger farms. Most of these items are horticultural ones, and they would be analysed separately, and later, on a " 100 per cent. provisional " basis—*i.e.* a complete count of all returns, subject only to adjustment in subsequent months through settling of doubtful entries. Since only one-sixth of all returns has any horticultural entries, it is possible to obtain these 100 per cent. provisional results for horticulture quite quickly by picking out these relatively few returns from the main bulk, and then sorting them back for later parish summaries of all the items on the return.

The differences between the 50 per cent. provisional results and the 100 per cent. final figures (produced later at the end of the year) are not due to sampling errors alone, because of the later adjustment to queried forms. Even so, the differences of the six Junes (1951 to 1956) for which the system has been used, are small and unbiassed. About three-quarters of all the 80–90 items have errors of less than one per cent. for England and Wales totals, and most of the main items' errors are less than one-half of one per cent. In fact, results are sufficiently reliable for county totals to be issued for most items or combination of items for over 1,000 (acres or numbers).

The results of this relatively modest incursion into the territory of random samples has been to encourage further steps. One of these is that for June 1957, provisional results have been obtained from a one-third sample of parishes, which gave rise to a slight increase in sampling errors. There was, however, a saving in clerical labour. In recent years, the completed census forms have been sent by farmers to the County Officers, who booked them in and scrutinised them. To effect substantial overall staff economies, this process has been centralised, the forms all now being sent direct to the Ministry headquarters at Guildford. This change of procedure has of course involved headquarters in a new high peak of work, for booking in and scrutinising (although this is far less than the economy effected by abolishing the load at county level) and the reduction in the size of sample required for the provisional results has reduced the casual staff needed at that time.

For the censuses at March, September and December, the change to a one-third sample system had taken place as from March 1953. For these three censuses, only one-third of the farmers are sent forms each time, so that the results grossed up from the one-third samples are the only ones obtained for these dates, and the load of form-filling for farmers was almost halved, *i.e.* each farmer received a June form, and another for only one of the three census dates March, September or December, as compared with the previous practice of every farmer receiving all four census forms every year.

The selection of these one-third samples was made from address lists. These lists are printed from over 370,000 metal plates, *i.e.* one for every agricultural holding of more than one acre in England and Wales. The plates are kept in county order, and, within each county, in order of parish number, and of holding number within parishes, the latter numbering showing some slight (inverse) correlation with holding size for holdings of under 100 acres. For the one-third samples, these plates were simply marked X, Y or Z, in turn, the X holdings being in the March samples, the Y's in the September and the Z's in the December sample. Under this system, farmers receive forms for the same dates each year. In March, 1958, the system has been slightly modified. The sampling unit is now the parish, instead of the holding. This allows the grossing up factors described below to be more up-to-date and also to be calculated by fewer staff.

Consistency of results

The danger of such a rigid system, namely perpetuating a bias at a particular date, was expected to be negligible since, for each item, a ratio method of grossing up was used, so that if for example the March sample by chance included too many cattle, the bias would only show itself if these cattle holdings had in fact changed their numbers at a different rate (or direction) from " cattle holdings " in the remaining two-thirds of the population. Tests on the variances for this grossing up method indicated that with such large samples this type of bias was likely to be negligible, and indeed results for England and Wales during the period of nearly five years since the samples were initiated have been satisfactorily consistent, both as between successive samples and by comparison with the annual full censuses at June.

This consistency is measured for the livestock items by using statistics of fatstock marketing and slaughter to help to link the results at successive census periods. For acreage items, farmers are asked in March to forecast the acreage of main crops which they hope to have sown by June. The relations between these forecasts and the actual June acreage are still of the same order of magnitude as they were in the years of full censuses at both dates.

It is not claimed that a completely self-consistent link-up is obtained between each sample date. From time to time anomalies arise : for example at a September census in one year it might be found that the estimated number of breeding sheep is high in relation to the total of the relevant items recorded at the previous June, the recorded slaughterings between June and September and a factor for death rate obtained by averaging " unrecorded disappearance " between June and September in previous years. Such a high figure for breeding sheep at September

might be due to a combination of abnormal features for the year in question, of which one might well be a high actual sampling error. Thus, from the estimates of sampling error for that item, there is roughly a one in forty chance that the September estimate in a particular year is two per cent. too high and this two per cent. difference involves about 150,000 head. But similar anomalies do arise in the June (100 per cent.) census itself, in June-June comparisons, so that a freak result at a sample date need not be attributed to sampling error ; it merely means that the existence of such potential sampling errors adds another factor to the list of possible explanations, and complicates to some degree the work of interpreting results. It is clear that a sampling system, to be of any practical use in interpreting short-term trends, must be designed to ensure that expected sampling errors for each item are less than deviations over the short period which are considered to be significant enough with 100 per cent. censuses to indicate a new trend, superimposed upon any normal seasonal movement. But this general principle is easier to state than to apply, because for almost every item opinions differ as to the minimum size of a quarterly fluctuation needed to indicate a change of trend, even with full censuses at every date. Each piece of the jig-saw puzzle has its own blurred edges.

Despite these doubts and difficulties, it might be arbitrarily suggested as a working rule, that an effective sampling scheme for a quarterly census should be designed to produce expected sampling errors for the dozen or so main items of not more than one per cent. This criterion is met in the one-third census scheme. For example wheat, barley and oats acreages, and numbers of cows, ewes, pigs or hens all have expected sampling errors of under one per cent.

Other samples

Within the last three years the Department has extended the use of samples to those of smaller size designed for a particular commodity. In each case, the sample was selected systematically within relevant strata, using a variable sampling fraction. The sampling frame was usually a master sample of returns with the main items card-punched (in full or in code form), although in one or two instances it was more convenient to pick direct from the agricultural returns themselves.

Egg yields

The first of these small samples concerned a survey of egg yields under different systems of management, conducted from December 1955 to March 1957. Three thousand holdings were selected from different flock sizes, including a relatively larger proportion of the bigger flocks, and results for each flock-size stratum were combined by using weights, which took account of the variable sampling fractions.

In this survey, the selected farmers were visited by local staff (Field Officers of the County Agricultural Executive Offices—now the Divisional Offices) and asked to record the number of eggs produced each day by their hens, subdividing flocks by systems of management—e.g. those in batteries and those under " free range ". These daily record cards were collected each month by the Field Officers, and each farmer was asked to continue in the scheme for a full year. The service was entirely voluntary and involved some trouble to farmers, but in fact the initial response rate was high and was well maintained throughout the months of the survey. In this, as in other surveys described below, the individual returns are confidential.

This was the first time that the Ministry had conducted a full-scale survey into egg yields, from a random sample, although there had been a survey confined to the East Midlands in 1950, which provided a valuable guide.

Milk yields

In order to reduce the amount of time required of each farmer, an alternative procedure, inspired by a recent survey conducted by the Department of Agriculture for Scotland, was adopted for this Department's next survey of yields—namely of milk yields. For this survey, which covered the year beginning October 1956, 200 holdings per month were chosen (stratification by herd size) for each month separately, and farmers were only asked to record details of the milk production *on the previous day*. The response to this lighter task was excellent, even for months when, because of petrol restrictions, the forms were sent and returned by post instead of being recorded at an interview by a Field Officer.

Stocks and yields of cereals

For some types of survey, however, it is important to maximise the precision of estimates of monthly trend rather than of each month's separate estimate, even though the latter is also of importance. Thus in 1956–7, the Department undertook a survey into the stocks of home grown wheat and barley remaining on farms each month from end-November to end-March. For this purpose 3,000 holdings were chosen (stratified by cereals acreages) and were each sent a form every month.

A similar survey is being held in the current production year (1957–58), but this time it is being linked to an experimental survey on yields of wheat, barley and oats. This combined survey, which at the time of writing has not been completed, is of a complex design. The main features are :—

(*a*) Stratification by the relevant cereals acreage.

(*b*) Four periods in July–September with questionnaires on forecasts of yield per acre per field. In these periods, some farms are asked at each date—for trends—and the remainder once only, partly for a series of independent estimates and partly to avoid troubling farmers more than absolutely necessary at the busy harvest time.

(*c*) Survey conducted by post, except for one farm in each National Agricultural Advisory Service (N.A.A.S.) district, which is visited by the N.A.A.S. District Advisory Officer each time. From these latter farms a link can be obtained between the weighted sample yields and those used in the official estimate of yields which are obtained from reports supplied by the same N.A.A.S. officers, who make their own local enquiries.

(*d*) A questionnaire on stocks on farm and total production per farm (linked to June 1957 cereal acreage) to be sent out each month from end-October to end-May, to the same farms and also to 2,000 extra wheat and barley farms.

Pig population

There is one small sample survey which is held every month. This is a sample, the series for which began in January 1956, to enable a close and frequent check to be made on the trends in the pig population, which are susceptible to up or down swings which may occur, almost without warning, in between the normal quarterly census dates.

The information for the enquiry is collected by visits to about 1,600 farms per month, the main stratification factor in this case being the number of sows. It is, however, necessary to include several farms without any pigs returned at the censuses used as the sample base (currently, March 1956). This procedure allows for the fact that several occupiers who kept no pigs at March 1956, may be keeping some in the month in which they are sampled.

Results from the pig samples are grossed up by reference to the numbers returned on the same holdings at a recent census date. This ratio method is not applied to every item separately: thus the correlation between sows pregnant for the first time (" gilts in pig ") at one month and the numbers a few months later may even be negative, because of seasonality of replacements. Sampling errors for each of the items, as grossed up, were calculated early in the series and these calculations are repeated from time to time to make sure that variances and co-variances have not changed significantly. As an example of the efficiency of a small sample selected for a specific purpose, as contrasted with large samples selected for a multitude of items, and without optimum stratification, it may be of interest to record that the 1,600 " pig holdings " have a sampling error for total sows of some 2½ per cent., whereas for the 125,000 holdings in the one-third sample census the sampling error for the same item, if raised simply by the reciprocal of the sampling fractions (i.e. multiplying by three) would be 1¼ per cent., a reduction of one-half for a sample nearly 80 times as large.

Present developments

Futher experiments in sampling designs are still proceeding within the Department. Since the summer of 1956, the Branch concerned with agricultural censuses and sample surveys has been housed at Guildford, where the punch card installations are available. With the aid of these machines experiments are under way on the selection of samples of a size intermediate between the small samples and the one-third samples. These experiments are testing the efficiency of such medium sized samples, stratified by size of farm and certain other factors, and picked with variable sampling fractions, for the purpose of obtaining quicker but still reliable first results from the returns. It seems likely that the use of sampling methods will continue to be extended over the next few years.

AGRICULTURAL CENSUS RESULTS

England and Wales

June 1939 and June 1957

Acreage of Crops and Grass

Thousand acres

Description	June, 1939	June, 1957 (Provisional)
Wheat	1,683	2,033
Barley	910	2,386
Oats	1,358	1,323
Mixed Corn, for threshing(1)	83	329
Rye, for threshing	(2)	24
Total corn	4,034	6,095
Beans, for stockfeeding	133	86
Peas, for stockfeeding	37	14
Potatoes, first earlies	56	96
Potatoes, main crop and second earlies	398	470
Total potatoes	454	566
Turnips and swedes for stockfeeding	396(3)	244
Sugar beet (for sugar)	337	415
Fodder beet (all types of high dry matter content)	(2)	14
Mangolds	210	148
Rape	53	118
Cabbage, kale, savoys and kohl rabi, for stockfeeding	94	345
Vetches or tares	49	16
Mustard, for seed, fodder or ploughing in	48	32
Linseed	} 4	(4)
Flax, for fibre		(4)
Hops	19	20
Orchard with crops, fallow or grass below the trees	236	241
Orchards with small fruit below the trees	18	8
Small fruit, not under orchard trees	29	30
Vegetables for human consumption (excluding potatoes) and hardy nursery stock, flowers and crops under glass	275	419
Fruit and vegetables, not grown primarily for sale	(2)	8
All other crops	48	47
Bare fallow	355	302
Total of crops and fallow (tillage)	6,830	9,170
Lucerne	32	102
Temporary grass (including clover and sainfoin)		
for mowing	1,304	2,375
for grazing	768	1,799
Total temporary grass	2,072	4,174
Total Arable Land	8,935	13,446
Permanent grass for mowing	4,612	2,897
Permanent grass for grazing	11,097	8,129
Total permanent grass	15,709	11,026
Arable and permanent grass temporarily out of use through flooding on the East Coast January/February 1953	—	5
Total Acreage of Crops and Grass(5)	24,643	24,476
Rough grazing—Sole right	4,179	3,618
Common Rough Grazings	1,361	1,483(6)
Total Rough Grazings	5,541	5,101

(1) In 1939 included areas cut green.
(2) Not separately returned.
(3) Includes turnips and swedes for human consumption.
(4) Included with " all other crops ".
(5) Excludes rough grazing.
(6) Provisional.

Numbers of Livestock

Thousands

Description	June, 1939	June, 1957 (Provisional)
Cattle and Calves		
Cows and heifers in milk :—		
For producing milk or calves for the dairy herd	2,255	2,142
Mainly for rearing calves for beef		467
Cows in calf but not in milk :—		
Intended for producing milk or calves for the dairy herd	392	361
Intended mainly for rearing calves for beef		85
Heifers in calf (first calf)	459	659
Bulls being used for service	91	60
Bulls (including bull calves) being reared for service	43	23
Other cattle :—		
2 years old and over — Male (steers)	(1)	530
2 years old and over — Female	(1)	513
Total	944	1,043
1 year old and under 2 — Male (steers)	(1)	587
1 year old and under 2 — Female	(1)	1,010
Total	1,346	1,597
Under 1 year old (excluding bull calves being reared for service) — Male (steers)	(1)	624
Under 1 year old (excluding bull calves being reared for service) — Female	(1)	1,073
Total	1,242	1,696
Total	6,770	8,134
Calvings		
Number of Calvings (calvings which occurred during the three months preceding date of census)		
Heifers that calved for the first time during March, April and May	(2)	172
All other cows that calved during March, April and May		601
Total		773
Sheep and Lambs		
Sheep one year old and over :—		
Ewes kept for breeding	7,160	6,211
Two-tooth (shearling) ewes or gimmers	1,477	1,359
Rams kept for service	205	174
Draft and cast ewes	1,021	231
Wethers and other sheep		515
Total one year old and over	9,863	8,491
Sheep and lambs under 1 year old :—		
Ram lambs intended for service	156	(1)
Other sheep and lambs	7,967	(1)
Total sheep and lambs under 1 year old	8,123	7,513
Total	17,986	16,004
Pigs		
Sows in pig	(1)	312
Gilts in pig	(1)	129
Other sows kept for breeding	(1)	166
Total sows for breeding	449	607
Barren sows for fattening	(3)	16
Boars being used for service	30	37
All other pigs (not entered above) :—		
5 months old and over	633(3)	858
2–5 months old	1,516	2,022
Under 2 months old	888	1,214
Total all other pigs	3,036	4,094
Total	3,515	4,754

(1) Not collected separately.
(2) Not returned.
(3) Barren sows for fattening were not collected separately in 1939 but were included among other pigs, 5 months old and over,

Numbers of Livestock (*continued*)

Thousands

Description	June, 1939	June, 1957 (Provisional)
Poultry		
Fowls 6 months and over	23,154	31,889
Fowls under 6 months old :—		
Male	(1)	5,307
Female		31,093
Sex not known		2,343
Total fowls under 6 months old	29,758	38,743
Total	52,912	70,632
Ducks of all ages	2,237	1,101
Geese of all ages	584	429
Turkeys of all ages	693	1,783
Total	56,426(2)	73,945
Horses		
Horses (including Mares kept for breeding) used for agricultural purposes or by market gardeners	549	87
All other horses and ponies (not entered above)	297	95
Total	846	182

Labour

(For 1939 references to age 20 should be read as age 21)

Thousands

Description	June, 1939	June, 1957 (Provisional)
Regular Wholetime Workers		
Male 65 years old and over	375·3	20·2
Male 20 years old and under 65		328·7
Male 18 years old and under 20	44·7	28·3
Male under 18 years old	50·8	35·2
Total	470·8	412·4
Women and girls	40·3	36·1
Total Male and Female	511·1	448·6
Regular Part-Time Workers		
Male 20 years old and over	(1)	42·3
Male under 20 years old		5·1
Total		47·4
Women and girls		31·5
Total Male and Female		78·9
Seasonal and Temporary Workers		
Male 20 years old and over	(1)	42·2
Male under 20 years old		3·7
Total		45·9
Women and girls		33·7
Total Male and Female		79·6
Total Male Workers	534·1	505·8
Total Female Workers	73·0	101·3
Total Workers	607·1	607·1

(1) Not collected separately.
(2) As a result of war-time controls many small-sized holdings were recorded for the first time in 1941

England and Wales

Acreage of Small Fruit

Thousand acres

Description	June, 1939	June, 1957 (Provisional)(1)
Small Fruit		
Strawberries	18·7	16·1
Raspberries	4·1	2·7
Currants, Black	10·4	11·3
Currents, Red and White	2·3	0·8
Gooseberries	9·1	5·7
Loganberries and Cultivated Blackberries	2·5	1·2
Total	47·2	37·9

Acreage of Vegetables for Human Consumption, Hardy Nursery Stock, Flowers and Crops under Glass

Thousand acres

Description	June, 1939	June, 1957 (Provisional)(1)
Vegetables for Human Consumption (excluding potatoes) **Grown in the Open**		
Brussel Sprouts	38·0	42·5
Remaining Spring Cabbage (planted in previous year)		7·8
Summer Cabbage		7·6
Autumn Cabbage		5·2
Winter Cabbage	44·1	12·8
Autumn Savoys		2·6
Winter Savoys		8·8
Kale and Sprouting Broccoli		2·4
Winter Cauliflower or Broccoli (heading) :— :—		
Remaining from previous year's plantings		2·3
Planted in the current year		9·6
Summer and Autumn Cauliflower :—	18·9	
Early Summer sown under glass and planted in the open		6·1
Late Summer and Autumn (open sown)		8·6
Carrots, earlies (grown for bunching only)		2·2
Carrots, main crop	16·1	25·4
Parsnips	(2)	3·7
Turnips and Swedes	(2)	5·1
Beetroot	(2)	8·3
Onions, grown for salad	1·7	1·2
Onions, for harvesting dry		3·0
Beans, Broad		9·9
Beans, Runner	17·8	9·9
Beans, Dwarf or French		2·7
Peas, green for market	60·6	30·8
Peas, green for canning or quick freezing		52·3
Peas, for harvesting dry :—	28·0	
Marrowfats		56·4
Blues		22·0
Asparagus	2·6	1·5
Celery	6·7	4·5
Lettuce	5·9	7·1
Rhubarb	7·2	5·0
Tomatoes (growing in the open)	0·2	0·7
Other vegetables and mixed areas	(2)	17·7
Total	247·7	385·7
Hardy Nursery Stock		
Fruit trees, fruit bushes and other fruit stock		4·1
Ornamental trees and shrubs	10·5	5·1
Other nursery stock (herbaceous plants, alpines, etc.)		3·9
Bulbs and Flowers in the Open		
Bulbs grown for flowers :—		
Daffodils (Narcissi)		3·7
Tulips		1·8
Other bulb flowers	7·7	0·8
Bulbs grown for sale as bulbs :—		
Daffodils (Narcissi)		1·5
Tulips		1·7
Other bulbs		0·2
Other flowers, not under glass	5·8	6·5
Total	13·5	16·2
All Crops grown under Glass	3·3	4·5

(1) These figures are provisional but are based on returns which account for 100 per cent. of the total area returned.
(2) Not returned.