

11-15-91
7-15-91

TRIP REPORT

(August 4-15, 1991)

Livestock and Poultry Summary

Agricultural Data Collection Component
Agricultural Sector Support Program

Prepared by:

Fred. B. Warren
Mathematical Statistician
USDA, NASS, Washington, D.C.

I. Scope of work:

The primary objective of this TDY was to develop formulations for summarizing the livestock data collected as a pilot survey with the January 1991 Area Frame Survey in seven districts of the Punjab and Sind.

Secondary objectives were to:

1. review the questionnaire used in the January 1991 ASF survey and recommend improvement, if any
2. review the interviewing procedures being used and recommend improvement, if any
3. review the edit and data entry programs already written by the ADC Project office, and recommend improvements, if any
4. develop formulations for summarizing and expanding segment data to district, province, and country levels
5. document 1), 2), 3), and 4) in a written report.

II. Formulations for summarizing livestock data:

I recommend that the 'weighted segment' approach be used in summarizing livestock data in the Punjab and in Sind. I also recommend that the 'open segment' estimator not be used in preparing estimates of production and area of harvested crops in the future. My reasons for this recommendation are based upon the particular conditions in these two provinces and may not be appropriate to Baluchistan and the NorthWest.

Three alternative methods of summarizing data from area surveys are the 'closed segment', 'open segment' and 'weighted segment' approaches. The 'closed segment' approach requires the farmer to provide reliable information for defined areas within the sample segment. Then because the area to be enumerated is bounded and the probability of selection for the sample segment is known, the closed segment estimates are (aside from non-sampling errors) unbiased. The closed segment approach is, appropriately, used by ADC to summarize reports of planted areas for the rabi crops in January and the kharif crops in July.

The 'open' and 'weighted' segment methods have been developed to handle situations where it is more appropriate to collect data for the entire farm instead of only for land in the sample segments. For example, the farmer may be more able to report production of rice for his entire farm instead of for individual fields. Also, livestock may be anywhere. Therefore, data for the entire farm is collected from all persons with land in the

sample segment. (This is different from the U.S. where data for the entire farm is collected only if the farmer lives in the sample segment.) Because a farmer may have land both in and outside the sample segment, some adjustments in the reported data are needed if the survey results are not to contain gross overestimates.

The 'open segment' method developed by ADC requires that the enumerator determine (for each person having land in the sample segment) if the farmer has any fields which are closer to the farmer's residence than the closest field in the sample segment. If the closest field is in the sample segment (Item 11 of the survey form is marked 'Near'), then the summary program accepts the reported data for the entire farm. Otherwise, if the field closest to the farmer's residence is not in the sample segment (Item 11 is marked 'Far'), the summary program replaces the reported entire farm data with zeroes. Because the report for the entire farm is assigned only to the area segment which contains the field closest to the farmer's residence, this procedure is (theoretically) unbiased. However, because Pakistani farmers generally live in villages, area segments closer to the village will have a higher proportion of 'nearest fields' than area segments which are farther from the village. Therefore, variances of estimates produced by this procedure will be much larger than variances from either the closed or weighted segment approaches.

The 'weighted segment' method requires that the survey collect some measure of the size of the farmer's holdings both in and outside the sample segment. This information is used to allocate, mathematically, a portion of the values reported for the entire farm to the sample segment. This proportion is computed as $W = A_s/A_e$ where

A_s is some measure of the total size of the farmer's fields in the sample segment, and

A_e is the same measure of total size of all fields, both inside and outside the segment.

For the January 1991 Survey, A_e is the entire area of the rabi crop, both inside and outside the segment (Item 13, code 100), and A_s is the total area of the rabi crops inside the sample segment.

Because the total farm area can be uniquely divided between all area segments in which the farmer has any fields, the estimates produced by the weighted segment procedure are also (theoretically) unbiased. Also, because the ratioing procedure tends to remove some of the natural variation in the proportion of individual crops in different area segment, estimates produced by the weighted segment procedure tend to have smaller variances than for estimates produced by the closed segment approach.

Closed segment estimates and coefficients of variation (CV) for rice, cotton, and sugarcane in the four Punjab districts from the July 1990 Survey are listed in the following table with comparable weighted and open survey estimates and CV's from the January 1991 Survey. The weighted survey estimates from the January 1991 survey generally are quite close to the corresponding closed segment estimates from the July 1990 survey. Also, the CV's of the weighted estimates generally are smaller than for the closed segment estimates. This result is not unexpected as the weighted segment method does 'smooth out' some of the variability between segments. However, several of the open segment estimates are quite different from the corresponding closed segment estimates and the CV's often are larger. In particular, the differences between the open and closed segment estimates for cotton in Faisalbab and sugarcane in Jhang are too large to be explained by sampling errors, and the CV's of the total rice and cotton estimates for the Punjab are much larger than for the closed segment estimates. Therefore I have concluded that the open segment approach does not produce estimates of usable accuracy and should no longer be used for the production and harvested acreage estimates from the semi-annual acreage surveys.

The livestock information on the January 1991 Survey was also collected for the entire farm. Therefore, the weighted segment approach should also be suitable for summarizing the livestock data.

I must emphasize that the livestock data was collected only from the sub-population of people who farm some land. Therefore, the survey results will be only for that part of the total universe. Any livestock owned by people who do not farm any land will not be included in the survey results.

I am indebted to Khadim Hussain (AFS) for his explanation of stratification procedures, to Javed Iqbal and Mansoor A. Sherazi (SP&D) for increasing my understanding of the January Crop Area Survey, and to M. Jamil Rajput, Waqar Gilani, and Suad Saeed for the data summarized in Table 1.

Table 1:
Comparison of Closed (July 1990), Weighted (Jan. 1991), and
Open segment estimates for rice, cotton, and sugarcane.

District	Crop	Closed		Weighted		Open	
		Acres	CV	Acres	CV	Acres	CV
		(000)	%	(000)	%	(000)	%
Faisalbad							
	Rice	50.9	23.2	49.0	25.4	68.6	36.7
	Cotton	121.0	17.1	106.7	12.6	227.4	11.4
	Sugarcane	227.0	9.2	215.9	8.5	267.5	10.6
Jhang							
	Rice	143.1	16.4	135.3	12.2	82.9	8.4
	Cotton	298.1	11.9	316.7	10.7	211.4	18.7
	Sugarcane	153.7	16.3	170.7	14.5	77.3	19.1
Multan							
	Rice	9.8	32.6	5.2	5.9	2.6	49.4
	Cotton	840.9	6.2	830.4	5.1	880.9	27.0
	Sugarcane	11.8	33.6	9.2	26.6	7.3	.3
Sherkhupura							
	Rice	586.5	6.1	621.5	4.5	812.7	14.0
	Cotton	11.1	48.1	7.2	6.7	6.7	6.3
	Sugarcane	56.8	19.0	43.6	18.1	65.0	27.7
Total - 4 districts							
	Rice	790.3	5.6	811.0	4.3	966.8	12.1
	Cotton	1,271.1	5.2	1,261.0	4.4	1,326.4	18.3
	Sugarcane	449.3	7.7	439.4	7.3	417.1	8.8

III. Review of the questionnaire used in the January 1991 ASF survey and recommended improvements

1. If the 'open segment' approach to estimating production and acres harvested is discontinued, then items 11 (Farmer's residence - Near/Far) and 12 (Field # nearest his residence) are no longer needed!

2. Item 13 asks the farmer to report the total area both inside and outside the segment. This is for both the current rabi crops and the just-harvested kharif crops. The total rabi acreage is used with the sum of the rabi areas reported in items 6, 8, and 10 to determine how much of the whole farm kharif acreages are to be assigned to the sample segment. I have two questions about this procedure.

First, I understand the the total acreage of rabi crops is much greater than the total acreage of kharif crops in some districts, and the reverse in other districts. Since we have no assurance that the ratio of total kharif to total rabi is the same for all fields within a holding, I suggest that a ratio based upon total area (possibly excluding waste) of fields within the sample segment to the total area of all fields both in and outside the segment would provide more consistent estimates of the ratio of the total kharif acreage in the sample segment. This may require further research.

Secondly, my experience is that more accurate reports are obtained if the farmer reports the total acreage in each field instead of for the entire farm. Therefore I recommend that the enumerator ask the farmer to list all fields outside the segment, and to report the total acreage in each. This would insure that the total acreage is reported.

3. Part IV - Livestock Enumeration.

Item 26 asks if the farmer or anyone else has livestock on his land (in and outside the segment). This is the same wording as used in the United States but it may not be appropriate for Pakistan. My understanding is that various species are not necessarily confined to the farmer's own land. For example, they may be in the care of the village herd boy, or wandering beside the road, on hillsides or besides the rivers, or they may be on someone else's land. It would be nearly impossible to enumerate livestock according to where they were. The survey asked for livestock owned without regard for where they were, therefore it would be more appropriate if Item 26 asked "Do you have any livestock, chickens, or other poultry?" In any event,

field tests, first asking the question of a number of Pakistani farmers and then asking them what they did not understand about the question. This would be done with the intent of possibly developing alternative wording for the questionnaire.

III. Review the interviewing procedures being used and recommend improvement, if any

I was not able to observe interviewing procedures in the field. The written instructions available to me appear to be comprehensive and complete.

IV. Review the edit and data entry programs already written by the ADC Project office, and recommend improvements, if any

I have reviewed the data entry (ASFDATC.SAS) and data edit (ASFEDIT.SAS) programs and have no reason to recommend any changes in the programs. However, I feel that both programs lack a sufficient number of comment statements, the type of internal documentation which would be of great benefit to future users of these programs. Comment statements should be used to describe the source of input dataset, the destination (purpose) of output data sets, descriptions of any 'included' programs, the purpose of the various procedures, and explanations of what is happening in different parts of the program. Such information may seem unnecessary to people who know the program well, but could be essential to a new person. A well documented program could easily have as many, on the average, as one line of comments for each two lines of code.

V. Develop formulations for summarizing and expanding segment livestock data to district, province, and country levels

The ADC Project Office had modified the data entry, edit, and summary programs used for the January Area Survey for the purpose of producing 'weighted segment' estimates for the livestock items as well. Major problems with these programs as they currently exist are that they:

1. do not provide direct estimates of the numbers of either total cattle or total buffalo,
2. do not provide estimates of the number of holders having the different species of animals, and
3. continue to need better (more) internal documentation.

I have provided a modification of the existing summary programs (ANISUMM1.SAS and ANISUMM2.SAS) which will correct (1 and (2, and will provide computer listings at the strata, district, province, and country level.