PD-ABL-363
95442

AGRICULTURAL DATA COLLECTION PROJECT
ASSP/USAID

SURVEY DESIGN FOR BALOCHISTAN PROVINCE
MAY 31 - JUNE 20, 1991

By

Paul W. Cook
Mathematical Statistician
Remote Sensing Section
National Agricultural Statistics Services
USDA, WASHINGTON, D.C.

June 20, 1991

TRIP REPORT : Pakistan Agricultural Data Collection Project

DATES: May 31 - June 20, 1991

TRAVELER: Paul W. Cook, Mathematical Statistician
Remote Sensing Section
Survey Research Branch
Applications and Research Division

SUBJECT: Survey Design for Baluchistan Province, Pakistan

Purpose of the Trip:

To recommend a sample size for Baluchistan Province consistent with the sample needed for all of Pakistan and to examine the area sampling frame developed for Baluchistan.

Preparations for setting the sample allocation:

Before setting the sample size for Baluchistan Province, information on the other three provinces (Punjab, Sindh, NWFP) was necessary to obtain estimates for the population totals (N_h segments) within the strata on a nationwide level to determine if the proposed sample size for Baluchistan was consistent with that for the nation. Although current year data were not available, data from the agricultural census for 1987-1988 provided helpful information on the amount of cultivated land within the districts for each province. Development of estimated population totals for each province proceeded with the use of these data.

Evidence from earlier stratification efforts (using aerial photos) in four districts of the Punjab and three districts of the Sindh Provinces showed that a small proportion of non-cultivated land would become a part of strata 10, 20, and 31 (see Training Document for Area Frame Construction using SPOT Imagery for strata definitions, segment sizes, and procedures followed). Therefore, to determine an average amount of cultivatable waste on a province level that might be added to the three agriculture strata, paired t-tests between total cultivated land and total cultivated land including cultivatable waste for all districts in the Punjab and Sindh Provinces were run individually by province. Since the earlier stratification had added very little cultivatable waste to each district, a 95% left sided confidence interval calculated from the t-test mean and standard deviation provided an estimated value for the non-cultivatable waste that might be excluded during the

BEST AVAILABLE COPY

stratification process for the remaining districts in each of the two provinces.

These estimates of percentage of land area with each stratum, from the earlier stratification of three districts for the Sindh and four districts for the Punjab Provinces, made possible an estimated allocation of sample segments for these two provinces. NWFP Province was an unknown because stratification work had not yet started there. As an approximation to what might be found during the stratification work, the breakdown for Baluchistan was adjusted to provide an estimate for the strata in NWFP. That permitted the development of an estimated sample allocation for the national level. Since much of the needed (real) data will not be available until the stratification of Pakistan is complete and survey data are available, the sample allocation at the national level is recognized as not being precise. However, the sample allocation at the national level did allow an evaluation of the approximate degree of precision provided by a sample size of 4,000 segments as proposed earlier by the FBS. The attached Appendix A provides these preliminary sample sizes on a stratum level.

Review of the Area Sampling Frame at the FBS ADC Cell:

My evaluation of the area sampling frame construction started on Monday, June 3, with an overview of the operations and methods used by the ASF Section in stratifying the land. Their principles of operation closely followed those rules and procedures presented in the paper AREA FRAME DESIGN FOR AGRICULTURAL SURVEYS by Jim Cotter and Jack Nealon of the National Agricultural Statistics Service (NASS). My evaluation of the SPOT imagery corresponding to the area frame maps and the digitized primary sampling units confirmed that they had closely adhered to those guidelines as given in pages 8-24 of the above referenced report. The only deviation in these procedures was that the count units for each district were ordered as a serpentine across strata on each topographical map rather than the usual ordering within district (county in the US)[1]. This change in procedure required the ordering of the topographical maps within the districts in a serpentine manner to create the geographic sampling units. Khalid proposed this method of ordering the count units.

---

1. The ASF Section explained that the reason for this change in the procedures usually used in area frame construction was that district boundaries were often difficult to determine on the topographical maps. Additionally, separate work teams often needed to work on more than one topographical map at one time. This change made these uses of the maps possible.

Mark Harris and Jack Nealon with NASS agreed that this method was a reasonable way to do the ordering in phone conversations on Sunday, June 16. The procedures section that concludes this report gives more detail on their comments.

Khalid Mahmud, Chief of the ADC Cell Area Sampling Frame (ASF) Section, instructed Iftikhar Cheema, who is in charge of the stratification of Baluchistan Province, to assist me in an analysis of the stratification work. My request was to review at least one SPOT scene for each of the districts in Baluchistan.

My work on examination of ASF construction for Baluchistan Province continued on Tuesday and Wednesday, June 4 and 5. During this time, I examined SPOT scenes with their corresponding strata maps from each of 12 districts (Quetta, Pishin, Nasirabad, Turbat, Chagai, Kalat, Lorali, Panjgur, Kachhi, Khuzdar, Gwadar, and Zhob). Because of the nearly 200 SPOT scenes and 600 topographical maps that cover all of Baluchistan, the limited time of the TDY was sufficient to examine only a representative sample of the scenes used. By evaluating the stratifications for these 12 images, I gained a better understanding of the enormous range of terrain features that occur in Baluchistan. This understanding helped me reach a better appreciation of the work that had been done and the limitations imposed by the great variation in topography within the Province.

My overall impression was that the stratification work was of very high quality. The six teams of Statistical Assistants, each headed by a Statistical Officer, had taken great care in delineating the strata. The rugged terrain of much of Baluchistan Province had at times simplified the operations (mountains at high altitudes were quite clearly Stratum 41) while in other places, where orchards or other agriculture appeared in mountain valleys, delineation of the strata was much more difficult and time consuming. Some areas that had appeared to have agriculture according to an examination of the SPOT scenes, required field evaluations that showed many areas to have only wild grasses or shrubs with no current agricultural use. Whether such areas could be used in the future for raising livestock was not clear. Bob Addison and members of the FBS spent nearly 20 days in the fall of 1990 doing field observations to evaluate the accuracy of observations from the SPOT imagery.

Although the stratification work was quite well done, my examination of the imagery did suggest that there were some areas of Stratum 41 that had potential for agriculture whereas other regions, particularly the high mountains and sandy deserts, would never support agriculture. Since Stratum 41 in Baluchistan consists of an estimated 66 % of the country's entire wasteland (non-agricultural) area and 88 % of all land area in the entire Province, a further breakdown of this stratum would be helpful (see Appendix B). This observation does not mean that stratum 41 was

stratified incorrectly, because, by definition, land is not supposed to be put in stratum 41 unless a portion of it is cultivated at some time during the year. Screening procedures should be helpful while doing this future breakdown (further description of these procedures follow in the next section).

Sample Allocation Method:

The method of obtaining the sample sizes shown in Appendix B is outlined on pages 21-24 of AREA FRAME PROBABILITIES SAMPLES, PIADIC Information Management Series published by Instituto InterAmericano De Cooperation Para La Agricultura. The approach suggested by Deming (1960) for estimating variances based on the sample unit distribution curve was the technique employed in estimating the sample segment variances. Mark Harris of NASS, in his TDY Trip Report, dated October 21, 1987, explained in some detail the justification for his use of selected values of these constants when he selected the sample size for Sheikhupura District of the Punjab Province. Target segment sizes for strata 10 and 20 have changed to 0.20 sq km because the sample segment sizes have been reduced to 0.20 sq km from 0.25 sq km to avoid fractions.

Because of the large number of segments in stratum 41 for the Baluchistan and NWFP Provinces, the largest value for crops expected to be found in those segments was taken to be 0.01 sq km. instead of 0.43 sq km as proposed by Mark Harris. However, even this small value might introduce errors in data expansion, because of the large expansion factor for stratum 41. Therefore, with the cooperation of the ADC Cell, the current plan is to allocate only 12 samples to Stratum 41 (at least for Baluchistan) to permit a greater sampling rate for the other strata. They will examine the SPOT scenes to locate areas of potential agriculture within Baluchistan's stratum 41 and will choose the segments according to these areas. The enumerators will use screening questionnaires for areas chosen as segments to assess the presence of agriculture and report this information back to the FBS. This method will permit the breakout of any areas where agriculture may exist within Stratum 41. A sample of segments should later be allocated to some areas of stratum 41 (where agriculture is possible), but not to the truly nonagricultural areas of the stratum. These procedures should result from my suggestion for the creation of new strata from Stratum 41 (see Appendix B) along with the remarks of Ron Fecso, Senior Research Statistician for the NASS.

Only two crops, wheat and the combined category of fruits (consisting of apples, dates, and citrus fruits) in Baluchistan appear to have the potential for being estimated at sufficiently precise levels to provide CV's near 5%. Although fruit is grown in all strata, the orchard stratum 62 contains most of the fruit grown in the province. Similarly, wheat is primarily in strata 10, 20, and 31, but very little is in the other strata. To optimize the

sample size for each of the two most easily estimated crops, two optimal allocations were run using zero ranges in strata where the crop under consideration was not prevalent. Thus, sample allocation for wheat used zeros for all strata except 10, 20, and 31. The sample allocation for fruit used zeros in all but the 51, 52, and 62 strata to make an allocation of sample segments to estimate fruit. The sum of the two sample sizes was 416 (see Appendix A, Baluchistan wheat and fruit sample allocations for details of the number of segments in each stratum).

Another change for the fruit stratum 62 in Baluchistan was to set the range from Deming's formula for stratum 62 at 0.60 instead of 0.45 as given by Mark Harris. This change in values came about from examination of the SPOT scenes that suggested that for many of the fruit segments a much higher proportion of the fruit segments would contain agriculture. This change meant that a smaller sample size of segments for stratum 62 should provide lower CV's than if the range of 0.45 were assumed. Actual segment data will allow a more accurate determination of the population standard deviation and show if this assessment has been realistic.

The ADC Cell will choose a sample of size 12 for stratum 41 after locating areas of potential agriculture within stratum 41 through examination of the SPOT imagery and any ancillary information that might be available. This small sample of segments will provide a means of monitoring any changes in agricultural use for these areas of stratum 41. The purpose of the sample is to monitor changes in agricultural use, not to make estimates of cropland in this stratum. The expansion factors for stratum 41 are so large that unrealistic estimates of cropland present would be most likely if these data were expanded.

The suggested sample size from the allocation formula allotted only four samples to each of strata 51 and 52. Khalid Mahmud suggested a small increase in the sample size for these two strata with small compensating decreases in the other strata. Because these changes in sample sizes were small and would provide better estimation for strata 51 and 52 with little corresponding decrease in precision for the two crops of interest, his suggested sample size was adopted (see Table 1).

Procedures for Sample Selection and Substrata Development:

The only change in the procedures of stratification, mentioned above, was that of ordering each topographical map (containing count units numbered in a serpentine within the map), individually, in a serpentine fashion within the district. This procedure would take the place of carrying the ordering of count units across the maps within a district. Although no documentation was available, Mark Harris, during a phone conversion on June 16, agreed with this solution: for each district, order the topo maps in a serpentine

fashion.

Although the effect on the final estimates from the samples drawn from the two possible orderings of the count units was not completely clear, a further conversation that day with Jack Nealon brought the suggestion of increasing the sample size by 5 % or so to allow for possible non-optimum allocation of the sample in creation of the resulting substrata. Unfortunately, costs for enumeration in Baluchistan are quite high and a larger sample size larger (from 400 to 416 pls 12 segments in stratum 41) than the ADC Cell orginally proposed had already been accepted. Therefore, the recommendation was to make no more increases in the sample size.

There was limited time available to study the possible orderings of the count units. Therefore, adoption of the serpentine ordering of maps was necessary to develop the substrata within the three days left to do the procedures.

A SAS program developed by Mark Harris during an earlier trip (Trip Report dated October 27, 1987), and subsequently updated, permitted development of the substrata and random selection of the samples within those substrata. Khalid Mahmud provided the serpentine ordering of the maps by district. The districts were ordered according to the cluster analysis by crops (see Appendix C for a description of the clustering procedure used). These orderings permitted creation of the substrata as described above and thereby the selection of the count unit in which each segment in the selected sample would be found. This information permitted the ADC Cell to select each segment in the sample by procedures described in the AREA FRAME document written by Jim Cotter and Jack Nealon as mentioned previously. (See Appendix D for the details of the segment allocation for Baluchistan Province).

TABLE I

FINAL SAMPLE SIZE FOR BALUCHISTAN PROVINCE
CHANGE IN ALLOCATION SUGGESTED BY
KHALID MAHMUD

STRATUM	NEYMAN ALLOCATION	MODIFIED SAMPLE PLAN
10	144	136
20	104	100
31	96	88
41	0	0
51	4	16
52	4	12
61	64	64
TOTAL	416	416

NOTE: AN ADDITIONAL 12 SEGMENTS IN STRATUM 41 WILL BE ADDED AT A LATER DATE.

PRELIMINARY SAMPLE SIZES

This Appendix presents the tables showing the Neyman Segment Allocation for the national level and for each of the four provinces (Baluchistan, NWFP, Punjab, and Sindh). Each table consists of two pages as follows:

1. The first page gives the following information:

    a. Stratum
    b. Stratum Definition
    c. Target Segment Size (Defined for that stratum
    d. Segments (The population N_h for that stratum)
    e. Range (The Deming's range value used for that stratum)
    f. Distribution (The assumed type of distribution using Deming's theory)
    g. Constant (The constant value for the given distribution to multiply times the range to generate the estimate of the population standard deviation);

2. The second page gives the additional information as follows:

    a. Stratum (as on the first page)
    b. Segments (as on the first page)
    c. S_h (The population standard deviation)
    d. Three columns of calculated values needed for Neyman Allocation
    d. The calculated sample sizes in shaded columns for the crops examined in that province or on the national area.

The assumptions made for estimation of the Deming range values are the same as in Mark Harris's Trip Report of October 27, 1987 except for Baluchistan. In Baluchistan, the fruit stratum, 62, appeared to have a higher concentration of fruit trees than Mark had assumed, so the value for Baluchistan Province was 0.60 square km as the range. This value made possible the selection of a smaller fruit sample concentrated on estimation of the fruit acreage and a separate sample for the estimation of wheat. The two sample sizes are added together to obtain the full sample size for Baluchistan.

NATIONAL SAMPLE

Area Under Wheat (1987-1988) = 7,308,400 Hac.
Area Under Rice (1987-88) = 1,963,000 Hac.
Area Under Cotton (1987-88) = 2,567,800 Hac.
Area Under Maize (1987-88) = 853,900 Hac.

(Source: Agricultural Statisitics of Pakistan,-1987-88).

Stratum	Definition	Target Size Km2.	Frame Units	Segments	Range(h)		Dist	Const.
10	Cultivated 60-100%	0.2		1,037,763	0.2	:	EQI	0.2
20	Cultivated 30-59%	0.2		96,005	0.2		EQI	0.2
31	Cultivated 1-29%	0.75		65,448	0.4	:	TRI	0.24
41	Non-Ag	0.75		668,241	0.01		TRI	0.24
51	Urban	0.1		46,310	0.03	:	UNI	0.29
52	Villages	0.1		68,803	0.03	:	UNI	0.29
61	Forest	0.75		10,893	0.01		TRI	0.24
62	Orchards	0.75		1,598	0.45		TRI	0.24

C.V. W (%) = 1.17
C.V. R (%) = 4.35
C.V. C (%) = 3.34
C.V. M (%) = 10

S = CV * TOTAL AREA

Neyman Segment Allocation Procedure

NATIONAL SAMPLE

W =	73084	Area under WHEAT in Sq. Km.	Sw	355.0829	Vw =	73[illegible]56.595	n(w) =	4,040
R =	19630	Area under RICE in Sq. Km.	Sr =	[illegible]353.905	Vr =	729[illegible]53.749	n(r) =	4,051
C =	25678	Area under COTTON in Sq. Km.	Sc	357.3452	Vc =	735555.299	n(c) =	4,016
M =	8539	Area under MAIZE in Sq. Km.	Sz	353.9	Vm =	729145.21	n(m) =	4,051

	Nh	Sh=Const*Range(h)	Sh^2	Nh*Sh	Nh*(Sh^2)	Wheat nh	Rice nh	Cotton nh	Maize nh
10	1,037,763	0.04	0.0016	41510.5200	1660.4208	3,089	3,089	3,062	3,089
20	96,005	0.04	0.0016	3840.2000	153.6080	286	286	283	286
31	65,448	0.096	0.009216	6283.0080	603.1688	486	492	463	486
41	668,241	0.0024	5.76E-06	1603.7784	3.8491	119	119	118	119
51	46,310	0.0087	7.569E-05	402.6970	3.5052	30	30	30	30
52	68,903	0.0087	7.569E-05	598.6861	5.2077	44	45	44	45
61	10,893	0.0024	5.76E-06	26.1432	0.0627	2	2	2	2
62	1,598	0.108	0.011664	172.5840	18.6391	13	13	13	13
	1,995,061			54437.7167	2448.4614	4,040	4,051	4,016	4,051

PUNJAB SAMPLE

Area Under Wheat (1987-88) = 5.343,800 Hac.
Area Under Rice (1987-88) = 1,085,400 Hac.
Area Under Cotton (1987-88) = 1.935,600 Hac.
Area Under Maize (1987-88) = 336,700 Hac.

(Source: Agricultural Statistics of Pakiatan, 1987-88).

Stratum	Definition	Target Size - Km2.	Frame Units	Segments	Range(h)		Dist	Const.
10	Cultivated 60-100%	0.2		654024	0.2	:	EQI	0.2
20	Cultivated 30-59%	0.2			0.2		EQI	0.2
31	Cultivated 1-29%	0.75		38304	0.4	:	TRI	0.24
41	Non-Ag	0.75		55443	0.01		TRI	0.24
51	Urban	0.1		8830	0.03	:	UNI	0.29
52	Villages	0.1		29980	0.03	:	UNI	0.29
61	Forest	0.75		465	0.01		TRI	0.24
62	Orchards	0.75			0.45		TRI	0.24

C.V. W (%) = 1.25
C.V. R (%) = 6
C.V. C (%) = 3.5
C.V. M (%) = 20

S = CV * TOTAL AREA

Neyman Segment Allocation Procedure

PUNJAB SAMPLE

W =	53438	Area under WHEAT in Sq. Km.	Sw	667.975	Vw =	446190.601	n(w) =	2,053
R =	10854	Area under RICE in Sq. Km.	Sr =	651.24	Vr =	424113.538	n(r) =	2,159
C =	19356	Area under COTTON in Sq. Km.	Sc	677.46	Vc =	458952.052	n(c) =	1,996
M =	3367	Area under MAIZE in Sq. Km.	Sz	673.4	Vm =	453467.56	n(m) =	2,020

Stratum	Nh	Sh=Const*Range(h)	Sh^2	Nh*Sh	Nh*(Sh^2)	Wheat nh	Rice nh	Cotton nh	Maize nh
10	654024	0.04	0.0016	26160.9600	1046.4384	1,772	1,863	1,722	1,743
20		0.04	0.0016	0.0000	0.0000	0	0	0	0
31	38304	0.096	0.009216	3677.1840	353.0097	249	262	242	245
41	55443	0.0024	5.76E-06	133.0632	0.3194	9	9	9	9
51	8830	0.0087	7.569E-05	76.8210	0.6683	5	5	5	5
52	29980	0.0087	7.569E-05	260.8260	2.2692	18	19	17	17
61	465	0.0024	5.76E-06	1.1160	0.0027	0	0	0	0
62		0.108	0.011664	0.0000	0.0000	0	0	0	0
	787,046			30309.9702	1402.7076	2,053	2,159	1,996	2,020

Neyman Segment Allocation Procedure

SINDH SAMPLE

Area Under Wheat (1987-88) = 1,024,800 Hac.
Area Under Rice (1987-88) = 721,800 Hac.
Area Under Cotton (1987-88) = 629,600 Hac.
Area Under Sugarcane (1987-88) = 207,100 Hac.

(Source: Agricultural Statistics of Pakistan, 1987-88).

Stratum	Definition	Target Size - Km2.	Frame Units	Segments	Range(h)		Dist	Const.
10	Cultivated 60-100%	0.2		330021	0.2	:	EQI	0.2
20	Cultivated 30-59%	0.2		57423	0.2		EQI	0.2
31	Cultivated 1-29%	0.75		11762	0.4	:	TRI	0.24
41	Non-Ag	0.75		54825	0.01		TRI	0.24
51	Urban	0.1		35933	0.03	:	UNI	0.29
52	Villages	0.1		37060	0.03	:	UNI	0.29
61	Forest	0.75		8248	0.01		TRI	0.24
62	Orchards	0.75			0.45		TRI	0.24

C.V. W (%) = 5
C.V. R (%) = 6.9
C.V. C (%) = 8
C.V. S (%) = 24.5

S = CV * TOTAL AREA

Neyman Segment Allocation Procedure

13

SINDH SAMPLE

W =	10248	Area under WHEAT in Sq. Km.	Sw	512.4	Vw =	262553.76	n(w) = 1,152
R =	7218	Area under RICE in Sq. Km.	Sr =	498.042	Vr =	248045.834	n(r) = 1,219
C =	6296	Area under COTTON in Sq. Km.	Sc	503.68	Vc =	253693.542	n(c) = 1,192
S =	2071	Area under SUGARCANE in Sq. Km.	Ss	507.395	Vm =	257449.686	n(s) = 1,174

Stratum	Nh	Sh=Const*Range(h)	Sh^2	Nh*Sh	Nh*(Sh^2)		Wheat nh	Rice nh	Cotton nh	Sugarcane nh
10	330021	0.04	0.0016	13200.8400	528.0336		873	924	903	890
20	57423	0.04	0.0016	2296.9200	91.8768		152	161	157	155
31	11762	0.096	0.009216	1129.1520	108.3986		75	79	77	76
41	54825	0.0024	5.76E-06	131.5800	0.3158		9	9	9	9
51	35933	0.0087	7.569E-05	312.6171	2.7198		21	22	21	21
52	37060	0.0087	7.569E-05	322.4220	2.8051		21	23	22	22
61	8248	0.0024	5.76E-06	19.7952	0.0475		1	1	1	1
62		0.108	0.011664	0.0000	0.0000		0	0	0	0
	535,272			17413.3263	734.1971		1,152	1,219	1,192	1,174
						Cost	305,299	323,104	315,931	311,335

	Cost/Segmnt (Rs.)
Punjab	265
Sindh	265
Baluchistan	795
NWFP	530

Neyman Segment Allocation Procedure

FRUIT SAMPLE USING FBS ESTIMATESS OF ACREAGE

BALUCHISTAN SAMPLE

Area Under Wheat	277,000 Hctares
Area Under Rice	115,000 hectares
Area Under Jowar	46,000 Hectares
Area Under Fruits	46,000 Hectares

(Source: Agricultural Statistics of Pakiatan, 1987-88).

Stratum	Definition	Target Size Km2.	Frame Units	Segments	Range(h)		Dist	Const.
10	Cultivated 60-100%	0.2		26859	0	:	EQI	0.2
20	Cultivated 30-59%	0.2		19291	0		EQI	0.2
31	Cultivated 1-29%	0.75		7691	0	:	TRI	0.24
41	Non-Ag	0.75		439348	0		TRI	0.24
51	Urban	0.1		1190	0.03	:	UNI	0.29
52	Villages	0.1		1356	0.03	:	UNI	0.29
61	Forest	0.75		1677	0		TRI	0.24
62	Orchards	0.75		1229	0.6		TRI	0.24

C.V. Wheat (%) =	7.3
C.V. Rice (%) =	14
C.V. Jowar (%) =	100
C.V. Fruit (%) =	5

S = CV * TOTAL AREA

Neyman Segment Allocation Procedure

FRUIT SAMPLE USING FBS ESTIMATESS OF ACREAGE

BALUCHISTAN SAMPLE

Wheat =	2770	Area under WHEAT in Sq. Km.	Sw	202.21	Vw =	40868.8841	n(w) =	1
Rice =	1150	Area under RICE in Sq. Km.	Sr =	161	Vr =	25921	n(r) =	2
Cotton =	5	Area under COTTON in Sq. Km.	Sc	5	Vc =	25	n(c) =	782
Fruit =	460	Area under FRUIT in Sq. Km.	Sf =	23	Vf =	529	n(f) =	72

Stratum	Nh	Sh=Const*Range(h)	Sh^2	Nh*Sh	Nh*(Sh^2)	Wheat nh	Rice nh	Jowar nh	Fruit nh
10	26859	0	0	0.0000	0.0000	0	0	0	0
20	19291	0	0	0.0000	0.0000	0	0	0	0
31	7691	0	0	0.0000	0.0000	0	0	0	0
41	439348	0	0	0.0000	0.0000	0	0	0	0
51	1190	0.0087	7.569E-05	10.3530	0.0901	0	0	41	4
52	1356	0.0087	7.569E-05	11.7972	0.1026	0	0	46	4
61	1677	0	0	0.0000	0.0000	0	0	0	0
62	1229	0.144	0.020736	176.9760	25.4845	1	1	695	64
	498,641			199.1262	25.6773	1	2	782	72

Neyman Segment Allocation Procedure

WHEAT SAMPLE USING FBS ESTIMATES OF CROP ACREAGE

BALUCHISTAN SAMPLE

Area Under Wheat	277,000 Hectares
Area Under Rice	115,000 Hectares
Area Under Cotton	500 Hectares
Area Under Fruits	46,000 Hectares

(Source: Agricultural Statistics of Pakistan, 1987-88).

Stratum	Definition	Target Size Km2.	Frame Units	Segments	Range(h)		Dist	Const.
10	Cultivated 60-100%	0.2		26859	0.2	:	EQI	0.2
20	Cultivated 30-59%	0.2		19291	0.2		EQI	0.2
31	Cultivated 1-29%	0.75		7691	0.4	:	TRI	0.24
41	Non-Ag	0.75		439348	0		TRI	0.24
51	Urban	0.1		1190	0	:	UNI	0.29
52	Villages	0.1		1356	0	:	UNI	0.29
61	Forest	0.75		1677	0		TRI	0.24
62	Orchards	0.75		1229	0		TRI	0.24

C.V. Wheat (%) =	5
C.V. Rice (%) =	12
C.V. Cotton (%) =	100
C.V. Fruit (%) =	100

S = CV * TOTAL AREA

Neyman Segment Allocation Procedure

17

WHEAT SAMPLE USING FBS ESTIMATES OF CROP ACREAGE

Sw	138.85	Vw =	19279.3225	n(w) =	344
Sr =	138	Vr =	19044	n(r) =	348
Sc	5	Vc =	25	n(c) =	39,352
Sf =	460	Vm =	211600	n(f) =	32

BALUCHISTAN SAMPLE

Wheat =	2770	Area under WHEAT in Sq. Km.
Rice =	1150	Area under Rice in Sq. Km.
Cotton =	5	Area under COTTON in Sq. Km.
Fruit =	460	Area under FRUIT in Sq. Km.

Stratum	Nh	Sh=Const*Range(h)	Sh^2	Nh*Sh	Nh*(Sh^2)	Wheat nh	Rice nh	Jowar nh	Fruit nh
10	26859	0.04	0.0016	1074.3600	42.9744	143	145	16,359	13
20	19291	0.04	0.0016	771.6400	30.8656	103	104	11,750	9
31	7691	0.096	0.009216	738.3360	70.8803	98	99	11,243	9
41	439348	0	0	0.0000	0.0000	0	0	0	0
51	1190	0	0	0.0000	0.0000	0	0	0	0
52	1356	0	0	0.0000	0.0000	0	0	0	0
61	1677	0	0	0.0000	0.0000	0	0	0	0
62	1229	0	0	0.0000	0.0000	0	0	0	0
	498,641			2584.336	144.7203	344	348	39,352	32

Neyman Segment Allocation Procedure

APPENDIX B

June 9, 1991

COMMENTS ON THE NEED FOR RESTRATIFICATION IN BALUCHISTAN

Paul W. Cook
USDA
National Agriculture Statistics Service
Research and Applications Division

Comment 1.

Stratum 41 in Baluchistan Province (and quite probably in NWFP Province) contains a very large proportion of the total land within the province (see Figure 1). However, in allocating a sample of segments on a national level for Pakistan, Figure 2 shows that only about 10% of the sample will go to Baluchistan Province since so little of the agricultural land is there. Figure 3 illustrates that Baluchistan Province contains over 65% of the estimated wasteland in the entire country. Consequently, any agricultural land found within Stratum 41 will have a very large expansion factor. Indeed, for Baluchistan this expansion factor is nearly 5,000. A one square km of agricultural land in Stratum 41 would expand to nearly 5,000 sq km or 500,000 hectares. Such an estimate would clearly be unacceptable since there are only 1,500,000 hectares of cropland in all of Baluchistan.

Comment 2.

Restratification of Stratum 41 into the additional strata suggested in my report dated June 8, 1991 (part of this appendix) would make sampling of the smaller strata that have agricultural potential possible, because then the expansion factors for the province would be much more reasonable and would not cause such overestimates of agriculture or livestock.

Comment 3.

Because Baluchistan consists of so much wasteland with such a very small portion of Pakistan's agriculture, the estimates of agricultural land will have high CV's. This will not be true in the Punjab and Sindh. My recommendation is that before enumerating a sample from Baluchistan Province, that a sample for the Punjab Province (when the frame is ready) be drawn and prepared for enumeration during the same survey. The CV's and estimates for the Punjab Province should be very good and should make the high CV's obtained from Baluchistan Province of little influence. Thereby, the ASF will be given a fair chance to show its value in obtaining crop estimates.

19

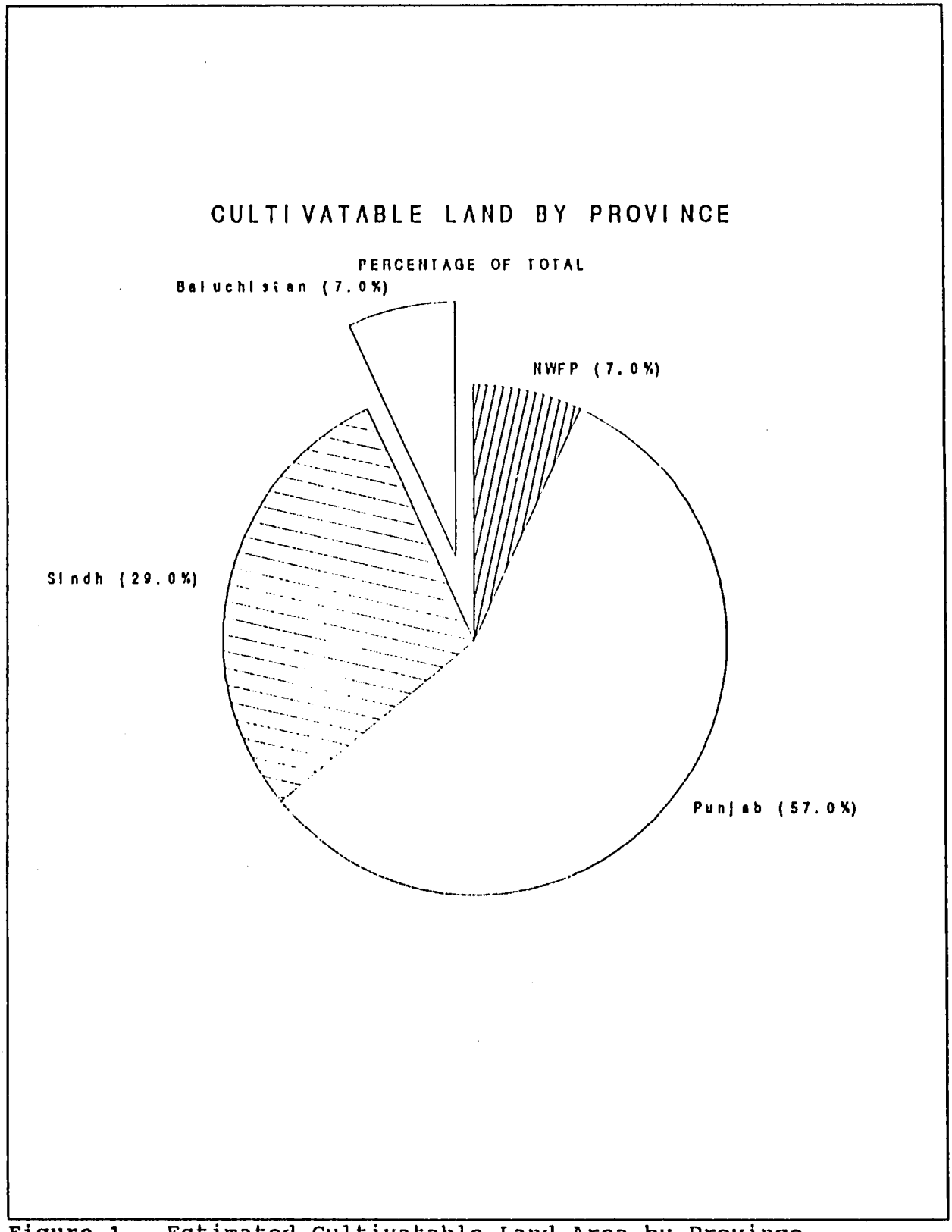
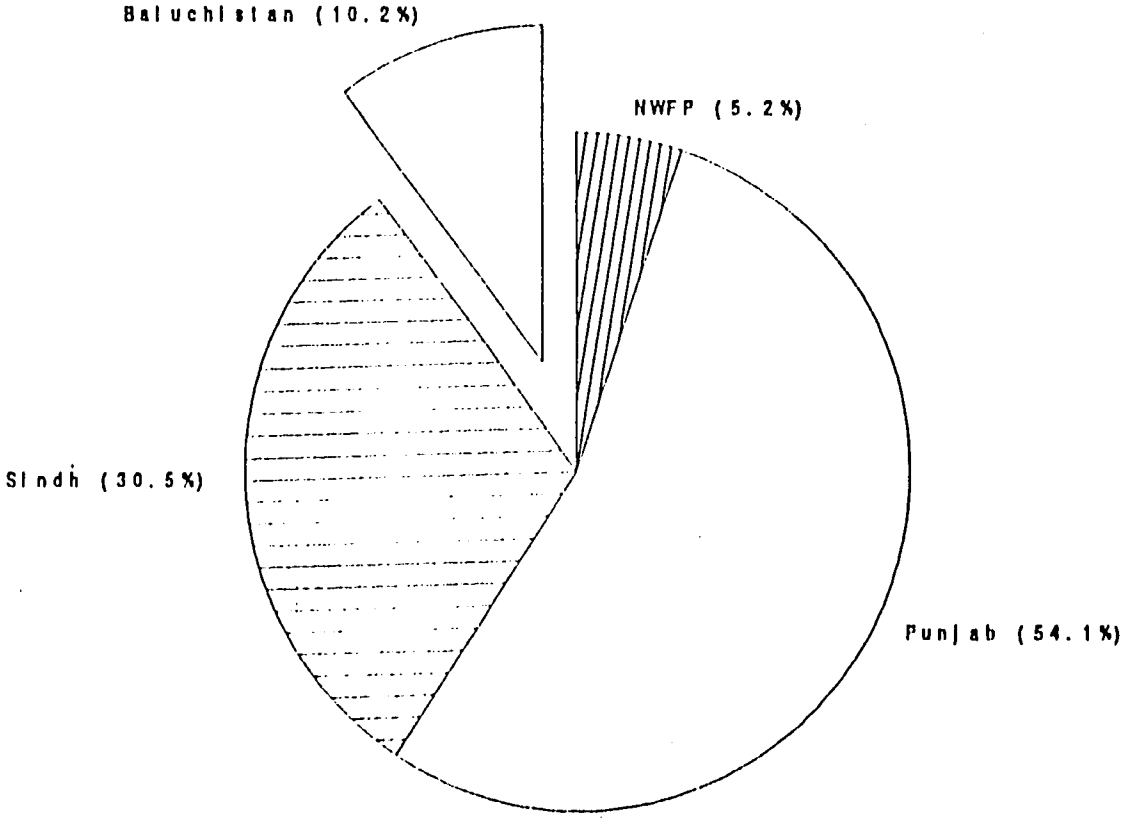


Figure 1. Estimated Cultivatable Land Area by Province

20

SAMPLE SEGMENT DISTRIBUTION BY PROVINCE

PERCENTAGE OF TOTAL



21

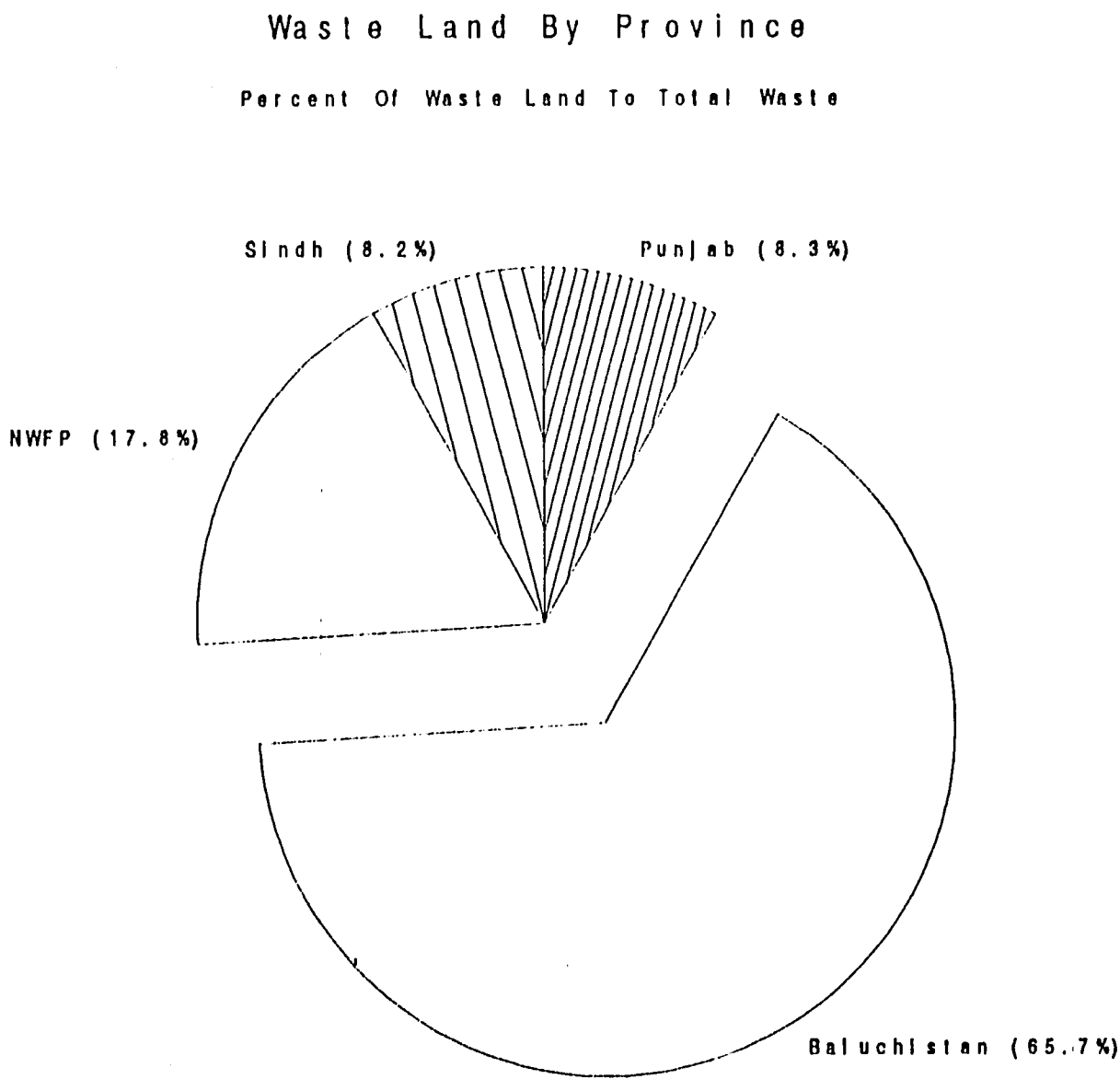


Figure 3 Estimated Wasteland Percentages by Province

22

June 7, 1991

NEW NON-AGRICULTURAL STRATA FORMED FROM STRATUM 41[1]

Paul W. Cook
USDA, NASS, Research and Applications Division

Justification:

An examination of the area sampling frame in Baluchistan showed to me that there are really some chances for agricultural land appearing in Stratum 41. To avoid costly sampling of segments within much of the often desolate land that appears in Stratum 41, I propose that Stratum 41 be divided into new Strata that will permit very much reduced sampling of these areas currently in Stratum 41.

New Strata:

The new strata that I propose are the following:

Stratum 41: Non-Agricultural Mountains:

Mountains having no potential for agriculture because of high altitudes and no source of water.

Stratum 43: Mountainous Areas with Agriculture Potential:

Mountainous Areas having the potential for Agriculture because of lower altitudes and the availability of water sources

Stratum 44: Non_Agricultural Desert Areas:

Desert Areas with no potential for agriculture because of poor soil and lack of water resources.

Stratum 45: Desert Areas with Agricultural Potential:

Desert Areas with good soil having agricultural potential if provided with irrigation canals.

Stratum 46: Mountainous Wild Grass Areas:

Mountainous Areas with wild grass or other natural cover with a potential for future crop.

---

[1] Developed after phone conversion with Ron Fecso on June 6, 1991.

23

Benefits:

The addition of these strata for Baluchistan will permit the division of the current Stratum 41 into those strata that will need no sampling at all, specifically, Strata 41 and 44, and those that will need only a small sample, specifically, Strata 43, 45, and 46.

Development of these strata will help in the sampling plan by permitting more samples to be placed in the agricultural strata. By increasing these sample sizes, the CV's of estimated crops will be improved. These benefits will occur in any province where Stratum 41 could be divided into these additional strata. There is also the possibility that the other provinces might require the creation of additional substrata from Stratum 41 to accomplish this goal. An examination of the satellite imagery in those areas will be needed to determine any such strata.
The process of determining the areas here further stratification work is necessary will not be a difficult one because a very large percentage (perhaps as much as 90%) of Stratum 41 will be in the non-agricultural mountains (Stratum 41) or non-agricultural deserts areas (Stratum 44)strata. The restratification effort should proceed rather quickly and not require a major effort for completion.

Sample Allocation:

To accomplish the flexibility criterion required by the FBS for creating samples at the divisional or district level, the divisional strata sizes are needed to develop number of segments by division. This will permit the allocation of the sample by division within Baluchistan Province. The divisional samples would then be proportionally allocated at the district level at a later time. This level of stratification would not require the creation of geographic sampling units.

Replicated Sampling:

The need for replicated sampling comes primarily from the need for variance reduction, sample rotation, or for adjusting the sample size. The needs of the FBS will determine whether 4, 5, or 10 replicates are needed. These replicate numbers are the ones most easily implemented.

APPENDIX C

CLUSTER ANALYSIS OF DISTRICTS FOR DISTRICT ORDERING IN PAKISTAN

Background:

NASS has developed clustering procedures to order counties in the US according to both crop and livestock data. Jim Cotter in his report on the Area Sampling Frame referred to the clustering methodology, but did not give details on the procedures followed. Robert De Geuss, an ASA Associate, wrote a report titled "Cluster Analysis of County Level Agricultural Data for County Ordering: A Procedure Guide" in which he described the four primary clustering procedures used. Jim Mergerson, a NASS employee in charge of sampling procedures at the Fairfax Area Sampling Frame Office, has established that two of the four methods (Average and Centroid hierarchical clustering) have succeeded in forming good clusters for two states in the US (Iowa and Michigan). Also, these two methods give virtually the same clusters.

The Clustering Procedure:

The SAS clustering procedure for doing the hierarchical clustering is not available in the SAS package in the ADC office. Therefore, Jim Mergerson used data faxed to him by the ADC to do the centroid clustering with the SAS cluster algorithm available to him. The data included district level wheat, rice, jowar, fruit, cotton, maize, sugar cane, sheep, goat, and poultry estimates from the most recent survey data in Baluchistan (see Chart 1). Jim did the clustering on crop only, livestock only, and crop and livestock data.

After careful analysis of the clusters formed in each method, the best organization of the districts appeared to be with the crops only clustering (Chart 2). This clustering was also more logical when considering that the locating of livestock in Baluchistan might be quite difficult because of the nomads who own most of the livestock there. The use of standard sampling techniques from the ASF will not easily locate livestock because the nomads have no fixed area where they keep the livestock.

Chart 3 shows the final clusters developed from the clustering procedure. The ordering of clusters from I through V gives the order in which the districts are to be included in developing the substrata for selecting the final sample of segments.

25

Benefits of Clustering:

The advantages of clustering the districts include the forming of more homogenous substrata for creation of the replicated sampling plan and the determination of how best to order all the districts to form the substrata. There is another consideration of the benefits obtained from doing the clustering procedure. The ADC Cell did make one change in the standard procedure of ordering count units. Instead of ordering the count units within a district, they ordered the count units on each individual map. This procedure could cause the formation of substrata very close together and thereby cause a reduced geographic dispersion of the sample. However, the use of cluster permits the creation of new groups of districts, sometimes separated from each other by another district. The ordering of districts within the clusters should thereby create more homogenous substrata that have sufficient geographic dispersion to generate a representative random sample at the province level. The major benefit, of course, will be a final reduction in the CV's of the acreage estimates.

CHART 1

-Jun-91

LUCHISTAN		CROP ACRES							LIVESTOCK NUMBERS			
VISION	DISTRICT	Wheat	Rice	Jowar	Fruit	Cotton	Maize	Sugarcane	Sheep	Goat	Poultry	Total
etta												
	Quetta	5,251	0	0	8,419	0	0	0	90,272	47,083	279,222	430,247
	Pishin	58,147	0	0	14,483	0	0	0	749,702	560,636	373,708	1,756,677
	Chagai	4,784	0	1,611	1,710	0	0	0	347,263	171,967	114,298	641,633
ob												
	Loralai	53,129	0	5,745	23,789	531	6,215	0	1,136,216	297,450	252,261	1,775,336
	Zhob	17,229	0	373	16,430	0	4,685	0	1,243,474	334,829	150,251	1,767,271
bi												
	Sibi	32,643	499	7,369	7,077	0	245	400	227,862	186,468	109,905	572,468
	Kohlu	22,240	0	309	882	0	0	0	1,402,339	663,921	136,261	2,225,952
	Dera bugti	1,038	72	1,011	27	0	25	0	313,744	232,635	59,035	607,586
sirabad												
	Nasirabad	146,558	256,747	63,994	534	15	0	467	9,739	22,321	17,572	517,947
	Kachhi	36,157	2,088	22,391	536	0	0	0	48,309	165,542	112,321	387,344
lat												
	Kalat	49,316	0	0	9,301	0	0	0	1,219,872	638,692	246,866	2,164,047
	Khuzdar	103,571	1,641	2,829	4,562	0	0	0	2,651,844	2,048,768	378,158	5,191,373
	Kharan	4,572	0	190	3,452	0	0	0	898,290	818,991	233,658	1,959,153
	Lesbala	2,471	146	7,992	1,295	0	0	353	213,301	303,719	336,284	865,561
ekran												
	Turbat	1,473	2,296	0	13,633	0	0	0	77,424	247,500	78,610	420,935
	Panjgur	2,471	198	0	4,942	0	0	0	112,012	89,825	46,359	255,807
	Gwadar	0	0	0	2,446	0	0	0	103,870	162,371	46,265	314,952

aluchistan Province comprises of six Divisions.
ach Division consists of different number of Districts and a District is like a County in US.

(A)

3

8

27

CHART 1 (CONT.) LIVSTOCK.WK1

?-Jun-91

ILUCHISTAN RATIO TO THE TOTAL

IVISION	DISTRICT	Wheat	Rice	Jowar	Fruit	Cotton	Maize	Sugarcane	Sheep	Goat	Poultry
uetta											
	Quetta	0.0122	0.0000	0.0000	0.0196	0.0000	0.0000	0.0000	0.2098	0.1094	0.6490
	Pishin	0.0331	0.0000	0.0000	0.0082	0.0000	0.0000	0.0000	0.4268	0.3191	0.2127
	Chagai	0.0075	0.0000	0.0025	0.0027	0.0000	0.0000	0.0000	0.5412	0.2680	0.1781
hob											
	Loralai	0.0299	0.0000	0.0032	0.0134	0.0003	0.0035	0.0000	0.6400	0.1675	0.1421
	Zhob	0.0097	0.0000	0.0002	0.0093	0.0000	0.0027	0.0000	0.7036	0.1895	0.0850
ibi											
	Sibi	0.0570	0.0009	0.0129	0.0124	0.0000	0.0004	0.0007	0.3980	0.3257	0.1920
	Kohlu	0.0100	0.0000	0.0001	0.0004	0.0000	0.0000	0.0000	0.6300	0.2983	0.0612
	Dera bugti	0.0017	0.0001	0.0017	0.0000	0.0000	0.0000	0.0000	0.5164	0.3829	0.0972
asirabad											
	Nasirabad	0.2830	0.4957	0.1236	0.0010	0.0000	0.0000	0.0009	0.0188	0.0431	0.0339
	Kachhi	0.0933	0.0054	0.0578	0.0014	0.0000	0.0000	0.0000	0.1247	0.4274	0.2900
alat											
	Kalat	0.0228	0.0000	0.0000	0.0043	0.0000	0.0000	0.0000	0.5637	0.2951	0.1141
	Khuzdar	0.0200	0.0003	0.0005	0.0009	0.0000	0.0000	0.0000	0.5108	0.3946	0.0728
	Kharan	0.0023	0.0000	0.0001	0.0018	0.0000	0.0000	0.0000	0.4585	0.4180	0.1193
	Lesbala	0.0029	0.0002	0.0092	0.0015	0.0000	0.0000	0.0004	0.2464	0.3509	0.3885
ekran											
	Turbat	0.0035	0.0055	0.0000	0.0324	0.0000	0.0000	0.0000	0.1839	0.5880	0.1868
	Panjgur	0.0097	0.0008	0.0000	0.0193	0.0000	0.0000	0.0000	0.4379	0.3511	0.1812
	Gwadar	0.0000	0.0000	0.0000	0.0078	0.0000	0.0000	0.0000	0.3298	0.5155	0.1469

Baluchistan Province comprises of six Divisions.
Each Division consists of different number of Districts and a District is like a County in US.

CHART 2

Cluster Analysis of Baluchistan Province
Crops Only 20:35 Tuesday, June 11, 1991

Centroid Hierarchical Cluster Analysis

Eigenvalues of the Covariance Matrix

	Eigenvalue	Difference	Proportion	Cumulative
1	5.1344E9	4.4954E9	0.878456	0.87846
2	6.3898E8	5.9331E8	0.109324	0.98778
3	45664314	21086203	0.007813	0.99559
4	24578111	23415546	0.004205	0.99980
5	1162565	1148916	0.000199	1.00000
6	13649.23	7806.413	0.000002	1.00000
7	5842.822		0.000001	1.00000

Root-Mean-Square Total-Sample Standard Deviation = 28895.87
Root-Mean-Square Distance Between Observations = 108118.4

Number of Clusters	Clusters Joined		Frequency of New Cluster	Pseudo F	Pseudo t**2	Normalized Centroid Distance	Tie
16	Chagai	Kharan	2	2445.42	.	0.020885	
15	Derabugt	Gwadar	2	2046.40	.	0.026090	
14	CL16	Panjgur	3	1534.89	2.95	0.031085	
13	CL14	CL15	5	927.82	4.17	0.037310	
12	Quetta	CL13	6	616.75	4.20	0.060186	
11	CL12	Lesbala	7	459.60	3.85	0.072766	
10	Pishin	Kalat	2	416.42	.	0.094703	
9	CL11	Turbat	8	335.13	5.11	0.100775	
8	Sibi	Kohlu	2	317.92	.	0.129789	
7	CL10	Loralai	3	298.52	2.72	0.135244	
6	Zhob	CL8	3	285.93	1.98	0.158114	
5	CL9	CL6	11	158.16	17.80	0.200640	
4	CL7	Kachhi	4	160.12	7.42	0.287251	
3	CL5	CL4	15	80.07	26.66	0.383911	
2	CL3	Khuzdar	16	84.00	12.39	0.779073	
1	CL2	Nasiraba	17		84.00	2.685532	

BALOCHISTAN PROVINCE

Districts Clusters

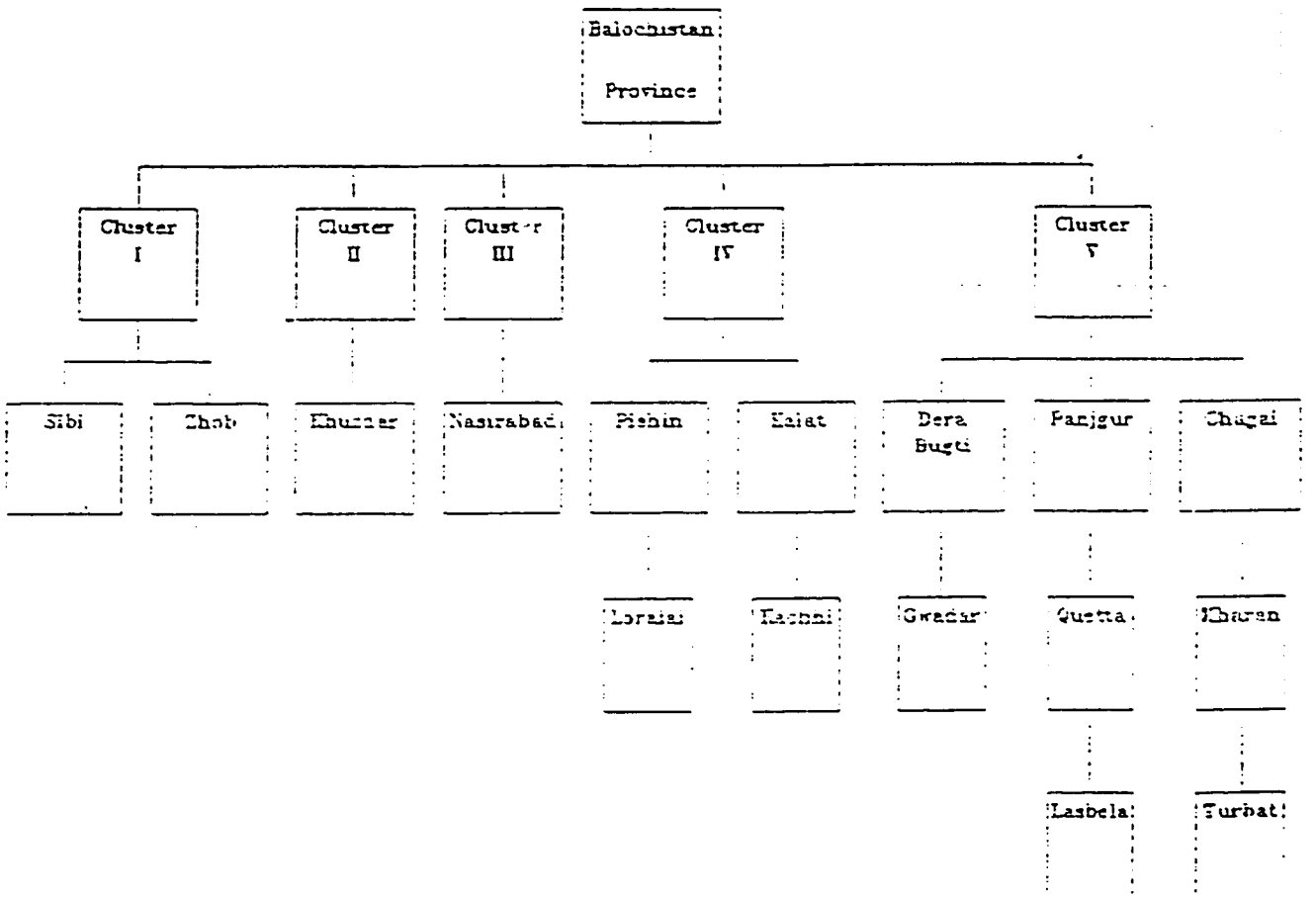


CHART 3

BEST AVAILABLE COPY

APPENDIX D

SEGMENT ALLOCATION : PAPER STRATA CREATION AND SEGMENT SELECTION PROCEDURES

Background:

Jack Nealon has clearly explained replicated sampling, sample allocation using Neyman Optimum Allocation, sample selection, and sample preparation in Chapter 2 of AREA Design for Agricultural Survey, August, 1987. The procedures that he has described were the basis of the segment allocation, paper strata creation, and segment selection procedures used in Pakistan. The following were the specifics of the work in Pakistan:

1. Four sampled paper strata were chosen,
2. A sample of size 416 was chosen as given in Table I of the main section of this paper,
3. The paper strata were developed and segment numbers chosen by a SAS program developed by Mark Harris,
4. The maps for each district ere ordered in serpentine ordering from the Northeast to the Southwest of each district, and
5. A SAS program ordered the maps and the count units on the maps to create a file giving accumulated segment totals to choose the correct count unit in which to find each sampled segment.

Procedures followed:

A SAS program created and described by Mark Harris, with some modifications to use the SAS random number generation procedure, generated the substrata for a replicated sample of size four. Table I displays the printout from this program. This program has selected the specific segments from the paper strata for selection in the count units. Another SAS program was developed to operate on the actual records of each count unit to sort the maps within each district, accumulate the number of segments within the count units by paper stratum and thereby permit the selection of the count unit in which to locate the selected segment number. This procedure is the two-step simple random selection procedure that Jack Nealon described in the Area Frame Design report.

Unfortunately, because of the time necessary to order the maps in a serpentine manner on a district basis, the sample select procedure was not complete at the time of my departure. This appendix will be completed at a later date.

31

Output from Mark Harris's program

STRATA	BIG N	LITTLE N	P STRATA	R1	R2	R3	R4
10	26859	136	1	213	390	279	767
10	26859	136	2	1394	865	1427	1045
10	26859	136	3	1992	2113	2016	1627
10	26859	136	4	2533	2384	2659	2634
10	26859	136	5	3616	3409	3377	3206
10	26859	136	6	4341	4198	4325	4174
10	26859	136	7	5125	5076	5448	5209
10	26859	136	8	5930	6230	5596	6027
10	26859	136	9	6620	6840	6569	7098
10	26859	136	10	7230	7374	7394	7347
10	26859	136	11	8659	7903	8357	8235
10	26859	136	12	8946	9046	8941	9282
10	26859	136	13	9657	9666	10100	10071
10	26859	136	14	11028	10622	10343	10780
10	26859	136	15	11564	11353	11139	11142
10	26859	136	16	12597	12394	12503	12012
10	26859	136	17	13411	13082	13076	13291
10	26859	136	18	14026	13737	13944	13576
10	26859	136	19	14681	14504	14874	14310
10	26859	136	20	15607	15486	15109	15354
10	26859	136	21	16460	15826	16436	16350
10	26859	136	22	16942	17215	17085	16680
10	26859	136	23	17656	17942	17542	18007
10	26859	136	24	18730	18608	18786	18625
10	26859	136	25	19441	19586	19218	19269
10	26859	136	26	20077	20310	20524	19941
10	26859	136	27	20796	21018	21233	21130
10	26859	136	28	22085	21822	22025	21843
10	26859	136	29	22173	22878	22700	22273
10	26859	136	30	23577	23471	22970	23414
10	26859	136	31	23985	24436	24033	23789
10	26859	136	32	24592	24710	25138	24525
10	26859	136	33	25784	25513	25940	25585
10	26859	136	34	26753	26606	26093	26291
20	19291	100	1	319	760	116	294
20	19291	100	2	1105	1213	801	1051
20	19291	100	3	1827	2166	1914	2288
20	19291	100	4	2866	2438	2805	2613
20	19291	100	5	3250	3121	3766	3788
20	19291	100	6	4009	4114	4607	3987
20	19291	100	7	5074	5239	5157	5116
20	19291	100	8	5719	5630	5761	5588
20	19291	100	9	6336	6524	6436	6203
20	19291	100	10	7562	7227	7648	7067
20	19291	100	11	8328	8344	8224	7958
20	19291	100	12	8976	8581	8612	8531
20	19291	100	13	9967	9723	9861	9301
20	19291	100	14	10681	10466	10651	10061
20	19291	100	15	11120	11517	11249	11531
20	19291	100	16	12213	11980	11739	11826
20	19291	100	17	13046	12456	12848	13074
20	19291	100	18	13512	13401	13289	13610
20	19291	100	19	14458	14337	14602	14051
20	19291	100	20	15420	14753	15184	14683
20	19291	100	21	15714	15981	15842	15559
20	19291	100	22	16237	16769	16257	16593
20	19291	100	23	17086	17676	17624	17497
20	19291	100	24	18347	18232	18042	18274
20	19291	100	25	19116	18858	19238	19031

32

STRATA	BIG N	LITTLE N	P_STRATA	R1	R2	R3	R4
31	7691	88	1	195	167	331	276
31	7691	88	2	384	367	443	490
31	7691	88	3	774	1027	842	1004
31	7691	88	4	1127	1154	1365	1202
31	7691	88	5	1694	1669	1568	1664
31	7691	88	6	2061	1870	1911	1765
31	7691	88	7	2330	2152	2262	2331
31	7691	88	8	2557	2694	2744	2746
31	7691	88	9	2811	2884	2924	2829
31	7691	88	10	3303	3285	3308	3152
31	7691	88	11	3714	3713	3754	3566
31	7691	88	12	4174	4122	3846	3971
31	7691	88	13	4408	4532	4465	4321
31	7691	88	14	4798	4783	4724	4695
31	7691	88	15	4958	5062	4913	5168
31	7691	88	16	5491	5371	5453	5399
31	7691	88	17	5886	5846	5855	5615
31	7691	88	18	6036	6064	6220	5975
31	7691	88	19	6603	6572	6299	6399
31	7691	88	20	6716	6864	6722	6859
31	7691	88	21	7304	7127	7081	7212
31	7691	88	22	7565	7576	7519	7653
51	1190	16	1	209	288	221	60
51	1190	16	2	526	454	446	428
51	1190	16	3	613	616	695	826
51	1190	16	4	1110	1024	944	1059
52	1356	12	1	299	221	64	158
52	1356	12	2	684	468	536	880
52	1356	12	3	1292	965	1106	976
62	1229	64	1	62	75	72	58
62	1229	64	2	91	116	109	152
62	1229	64	3	208	174	182	159
62	1229	64	4	299	281	289	253
62	1229	64	5	344	328	332	314
62	1229	64	6	408	460	414	416
62	1229	64	7	496	492	476	475
62	1229	64	8	572	580	558	552
62	1229	64	9	661	669	680	688
62	1229	64	10	691	693	725	724
62	1229	64	11	841	782	820	840
62	1229	64	12	919	863	882	856
62	1229	64	13	935	970	966	969
62	1229	64	14	1073	1068	1013	1030
62	1229	64	15	1146	1106	1079	1095
62	1229	64	16	1216	1193	1209	1159