LOCAL DEVELOPMENT II — PROVINCIAL

مشروع التنمية المحلية

LD II-P

FINANCED BY U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT

مصدر التمويل : الوكالة الامريكية للتنمية الدولية

Sample Design for Subproject Field Visitation System Field Test in Sharqiya

USAID Contract No.: 263-0182-C-00-8041-00
PROJECT No.: 263-0182-3-60054

October 1991
ME 4-03

Chemonics / 8 El Fawakeh Street / Mohandiseen, Cairo / Tel.: 3602545, 3482247 / Fax: 718986 / Telex: 93037

٨ شارع الفواكه / مهندسين . القاهرة / تليفون : ٣٤٨٢٢٤٧ . ٣٦٠٢٥٤٥ / فاكس : ٧١٨٩٨٦ / تلكس : ٩٣٠٣٧

2000 M St., N.W. / Washington, D.C. 20036 / Tel.: (202) 466-5340 / Fax: (202) 331-8202 / Telex: 440381 CHNC UI

Sample Design for Subproject Field Visitation System Field Test in Sharqiya

October 1991
ME 4-03

Table of Contents

	Foreword	iii
I.	Introduction	1
II.	Field Test Objectives and Constraints	2
	A. Objectives	2
	B. Constraints	3
III.	Sample Design Considerations	4
IV.	Sample Design	11
V.	Sample Selection	12
	A. Sample Selection Methodology	12
	B. Sample Selection Process	12
VI.	Sample	16

Foreword

This document was prepared by Chemonics International Consulting Division for the US Agency for International Development, under Contract No. 263-0182-C-00-8041-00, Project No. 263-0182-3-60054, "Local Development II - Provincial Project." The author of the report was Dr. J. George Caldwell.

I. Introduction

This report describes the sample design used for a field test of instrumentation and collection of certain statistical information about subprojects funded under the Local Development II- Provincial (LDII-P) project. The field test was conducted in the governorate of Sharqiya during August, 1991. The report describes the objectives of the field test, considerations involved in determining the sample design, the sample selection process, and the sample.

The objectives of this report are (1) to document work performed under the contract; (2) to provide information about the field test sample design and sample selection process needed to determine the proper formulas to analyze the sample data; and (3) to describe in detail the sample selection procedures used in the field test in the event that it is desired to employ this type of sample design at a later date.

The field test design is a "complex" sample design. It was necessary to employ a complex design in the field test in order to achieve the field test objectives. Although the sample design was specified in such a way ("self-weighting") that estimates of proportions may read be computed using standard statistical analysis programs (e.g., SPSS), special formulas are required to determine the precision of the sample estimates (standard errors, confidence intervals). This report contains all of the information about the sample design and the sample selection process that is required to measure the precision of the sample estimates and to estimate other quantities of interest (e.g., the intravillage correlation coefficient).

This report does not present or analyze the data collected in the field test, nor does it discuss instrumentation modifications resulting from the field test. A reference is cited for the correct procedures to use to analyze the pilot test data in accordance with the sample design.

II. Field Test Objectives and Constraints

A. OBJECTIVES

This report describes a sampling plan used to collect information in a field test conducted as part of the process of developing a Subproject Field Visitation System. The design of the sampling plan depends on the objectives of the field test and the constraints associated with it. This section identifies the objectives of the field test.

The goal of the field test was to provide information that would be useful in designing the Subproject Field Visitation System. To this end, there were two primary objectives. First, it was desired to collect information that would enable the "closing" of various open-ended response categories in the Subproject Status Assessment Form. Second, it was desired to obtain information about the magnitude of the intravillage correlation coefficient for the variables being measured. This information about the intravillage correlation coefficient would be of value in determining the sample design for the final system.

As a secondary objective, it was desired to identify and document field procedures for conducting a subproject sample survey in a governorate. As an incidental objective, it was desired that the information collected on subproject status during the survey could be used as feedback to improve the operational status of the visited subprojects and to improve the management systems used by the pilot-test governorate to monitor subprojects. Achievement of this objective would involve production of estimates of the proportions of subprojects possessing certain characteristics (e.g., the proportion of subprojects of nonoperational status), not of the monetary value associated with certain types of subprojects (e.g, the estimated total monetary value of all nonoperational subprojects).

The field test sample design was a probability sample design, with a modest subproject sample size. It would not have been necessary to use a probability sample for the field test sample if the field test objectives had been solely to obtain information that would be helpful in closing response categories, or identifying field procedures, or even for improving subprojects or systems in a governorate. The reason for using a probability sample was to obtain approximate estimates of the intravillage correlation coefficients, of reasonable accuracy and measurable levels of precision.

With respect to analysis of the field test data, it is planned to conduct a "components of variance" analysis, to determine the various sources of variation in the variables being measured. This analysis will estimate the variance of the markaz means, the within-markaz variance of the village means, and the within-village variance of the subprojects. These variance estimates will then be used to estimate the intravillage correlation coefficient. All of these quantities will be of value in determining what type of sample

design is appropriate for the final system (e.g., cluster, multistage, stratified), and the allocation of sampling effort (sample sizes) among the different stages or strata.

B. CONSTRAINTS

The constraints placed on the field test were the following.

1. The field test was to be conducted during the month of August, in order to enable presentation of results to USAID by the end of September.

2. Available field data collection personnel resources would allow visits to approximately 125 subprojects and approximately five marakez.

3. All large subprojects (i.e., of USAID funding greater than or equal to LE 200,000) were to be included in the field test sample.

4. Fourteen subprojects suspected or known to be problematic would be included in the field test sample. To keep the field-test costs low, the subprojects in this "purposive" sample would be selected from marakez to be visited in the probability sample of small subprojects or the certainty sample of large subprojects (i.e., no new marakez would be added to the field test for the purpose of visiting problematic subprojects).

5. The sample design was originally restricted to subprojects in three sectors -- potable water, roads, and buildings. During the course of discussions about the field test, it was decided that the field test also include environmental subprojects.

The estimation of the intravillage correlation coefficient and other population characteristics of interest (means, proportions) would be based on the probability sample, not on the subprojects included in the field test for other reasons.

It was suggested to perform the field test in the governorate of Sharqiya, since that governorate had a large allocation of funds in each LDII-P funding cycle, and therefore has a substantial number of subprojects. The existence of a large population of subprojects would enable estimation of the population characteristics of interest, and would increase the applicability of the field-test experience to other governorates. That governorate agreed to participate in the field test.

III. Sample Design Considerations

This section presents the considerations that led to the field test design.

Three-Stage Sample Design. To estimate the intravillage correlation coefficient, it was necessary to have reasonable-sized samples of subprojects within villages. The average number of subprojects per village is about eight. To reduce survey costs, it was desirable to use a three-stage sample design, in which a first-stage sample of marakez would be selected, a sample of villages would be selected from each sampled markaz, and a sample of subprojects would be selected from each sampled village.

As will be seen from the discussion that follows, it was possible to develop a very efficient sample design based on three-stage sampling. The design enables estimation of the intravillage correlation coefficient ***and*** estimation of proportions of subprojects having certain attributes (characteristics). It is emphasized, however, that although the developed sample design is ***efficient*** for the purpose of estimating proportions (i.e., it returns a relatively high level of precision for the level of sampling effort expended), the level of precision associated with the design is probably not very high.

The reason for this situation is that although it is less costly to sample subprojects in the same village than in different villages, subprojects in the same village are probably more similar to each other with respect to the variables being measured than they are to subprojects in different villages. This means that the amount of information that several subprojects in the same village yield about subprojects in the whole population is not expected to be as great as the amount of information that the same number of subprojects in different villages would yield about the whole population. This reduction in the level of precision for a multistage or cluster sample of a specified size compared to the level of precision for a simple random sample of the same size is often referred to as the "cluster effect," or the "effect of intracluster correlation."

If the desire to produce estimates of the degree of similarity of subprojects in the same village (as measured by the intravillage correlation coefficient) were not a factor in this survey, and if it were necessary to achieve a high level of precision from the sample estimates of proportions, the proposed three-stage sample design would probably not be recommended. Instead, we would probably propose a design based on simple random sampling of subprojects. The problem with using a simple-random-sampling design for the field test, of course, is that although it would produce better estimates of proportions it would not enable the estimation of the intravillage correlation coefficient and other related quantities of interest (variance components).

This situation, in which a design may be good for some purposes but poor for others, emphasizes the importance of taking into account all of the survey objectives in designing a sample. The proposed sample design is appropriate for the field test with its specific objectives, but it is probably ***not*** appropriate

for other applications since the estimates of proportions derived from the sample will probably be of low precision (because of the cluster effect). Although a multistage sample design (of size about 100 subprojects) may not produce an adequate level of precision for a single governorate, a design of this type may be quite desirable (making efficient use of sampling resources and producing a satisfactory level of precision) for producing national-level estimates in a full-scale implementation of the Subproject Field Visitation System.

Markaz, Village, and Subproject Sample Sizes. Since the markaz chief has a high level of influence on subproject-related matters in each village, it was expected that the variation among villages within the same markaz would not be very great. For this reason, from the viewpoint of making estimates of population means and proportions it was desirable to sample a large number of marakez and a small number of villages per markaz, since each additional village in the same markaz probably provides relatively little additional information. We hence restricted the within-markaz village sample size to two. (A minimum of two villages is selected per markaz, in order to permit estimation of variances.) In view of the constraint that the markaz sample size should be approximately five, it was decided to select a sample of six marakez, and to sample two villages from each of the sampled marakez. Hence a total of 12 villages would be visited.

With respect to the sampling of subprojects within villages, it was decided ***not*** to subsample subprojects within villages, but to include all of the subprojects in a sample village in the sample. The proposed design, therefore, is more properly referred to as a ***two-stage cluster sample design*** rather than a ***three-stage sample design.*** The reasons for the decision to use cluster sampling are the following.

First, the average number of subprojects per village nationwide is about eight. If all were included in the sample, the expected sample size would be 96 subprojects, which is within the total sample size constraint. Second, from the point of view of estimating the intravillage correlation coefficient, it is desirable to have at least several subprojects per village. The sample size of eight is not unreasonably large. Third, from the viewpoint of validity it was considered desirable to sample all subprojects in a village. The reason for this is that some difficulty was anticipated in identifying and locating subprojects, and it was considered that inclusion of ***all*** of the subprojects in a village in the sample might reduce the chance that similar subprojects might be misidentified.

Assuming an average of 8 subprojects per village, the expected sample size is, as noted above, 96 subprojects. Note that with cluster sampling, the subproject sample size is not controlled. Instead, the exact sample size depends on which marakez and villages fall in the sample.

In addition to the expected sample size of 96, there were nine large subprojects in Sharqiya (four large potable water subprojects and five large roads subprojects).

Summing the expected sample size of 96, the nine large subprojects, and the 14 problematic subprojects, the expected total field test sample size was 119.

Markaz and Village Sample Selection Procedures. It was decided to select the markaz sample without replacement, using probabilities proportional to the number of villages in each markaz, and to select the two villages per sampled markaz also without replacement, using equal probabilities. The reason for this approach is a little complicated, and will be described in the paragraphs that follow.

In general, the use of cluster sampling is efficient. It returns a high level of precision for the sampling effort expended because of reductions in the amount of time spent in visiting markaz and village officials, and in the travel time between subprojects (since a smaller number of marakez and villages is sampled than if simple random sampling is used). Usually, however, the precision of a cluster sample of a specific size is substantially less than the precision of a simple random sample of the same size. The reason for this is that, in most socioeconomic surveys, the intracluster correlation coefficient is positive for most variables. This means that (as discussed above), with respect to the variables being measured, the sample units (subprojects) within the same cluster (village) are more similar to each other than to subprojects in other clusters. Hence, they provide relatively less information than do the sample units of a simple random sample.

To limit the loss in precision due to clustering, it is generally advisable to employ cluster sampling only when the cluster size is not very large (as is the case in the present application). Otherwise, it is advisable to subsample from the clusters, i.e., to use multistage sampling. Since the cluster size is small in the present application, there was no need to consider subsampling (i.e., selecting a sample of subprojects from each sampled village). Hence, each subproject in a sampled village was included in the sample with probability one.

The way in which the markaz and village samples are selected has a substantial effect on the precision of the estimates. In order to return a high level of precision for the level of effort expended, it is desirable that each subproject in the population have a comparable probability of selection (i.e., that the probabilities of selection of the subprojects be similar, for all subprojects in the population). If the subprojects in a sampled village are included with probability one, this objective implies that the probabilities of inclusion of the sample villages should be approximately equal.

Since the marakez vary somewhat relative to the number of subprojects they contain, the precision of the sample estimates would be reduced if the marakez were selected with equal probabilities. The precision is increased if the marakez are selected with probabilities approximately proportional to the number of subprojects in each markaz. Furthermore, since the markaz population is small (15 marakez in all in Sharqiya), a substantial reduction in the variance is realized if the marakez sample is selected without replacement. The problem that arises is how to do this and at the same time keep the probabilities of selection of the villages approximately equal.

A solution to this problem is as follows. Select the marakez without replacement with probabilities proportional to the number of ***villages*** in each markaz, and select the two villages from each sampled markaz without

replacement with equal probabilities. Since the number of subprojects per village does not vary tremendously, this procedure is similar to selecting the marakez with probabilities proportional to numbers of subprojects. Furthermore, the probability of selection of a village is:

Probability of selection of a village

= (Prob of selection of markaz) x (Prob of selection of village conditional on selection of markaz)

= constant x (no. of villages in markaz) x (2/(no. of villages in markaz)

= constant,

as was desired. It follows directly that the probabilities of selection of the subprojects are equal, since:

Probability of selection of a subproject

= (Prob of selection of a village) x (Prob of selection of a subproject conditional on selection of village)

= 1.0 x constant

= constant.

So we see that if the marakez are selected with probabilities proportional to the number of villages, and if an equal number of villages is selected from each sampled markaz with equal probabilities, then the probabilities of selection of the villages are uniform (i.e., the same for every village in the population). Furthermore, under these conditions if the probabilities of selection of subprojects from sampled villages are constant (i.e., the same proportion of subprojects is selected from each village), then the probabilities of selection of the subprojects is also uniform. In the present application, the probability of selection of each subproject from a sampled village is 1.0.

As noted above, it is proposed to select the sample of two villages within a sample markaz using nonreplacement sampling. The reason for using nonreplacement sampling is that the number of villages per markaz is small, and a significant reduction in the variance is realized by using nonreplacement sampling instead of replacement sampling (because of the finite population correction).

In summary, the proposed sample design will provide a high return of precision for effort expended, because of the following desirable features:

- Uniform probabilities of selection of subprojects
- Selection of marakez with probabilities approximately proportional to size

- Small village sample sizes within sampled markaz
- Small cluster sizes
- Nonreplacement sampling of marakez
- Nonreplacement sampling of villages

Since the probabilities of selection of the subprojects is constant, the design is called "self-weighting." This means that, even though the design is a "complex" sample design, the estimates of means and proportions can be computed as simple averages. The practical advantage of this is that standard statistical program packages (e.g., SPSS) may be used to compute the sample estimates (although they cannot be used to compute the variances of the sample estimates).

The only potential difficulty with the proposed design is that, if care is not taken in the implementation of probability-proportional-to-a-measure-of-size (PPMS) sampling without replacement (i.e., the procedure proposed for selection of the markaz sample), no known formulas may exist for computation of the variances of the sample estimates. To avoid this problem, special methods are used to implement the PPMS sample selection. For the present application, we used one such method, the Rao-Hartley-Cochran method. This method is described in William G. Cochran, Sampling Techniques, 3rd edition (Wiley, 1977).

Additional Remarks. The proposed design is an interesting one. Often, in two-stage sampling, the first-stage sample is selected with probabilities proportional to size, and a fixed-size sample is selected from each sampled second-stage unit. This results in constant selection probabilities for the second-stage units (self-weighting, efficient). Alternatively, the first-stage units may be selected with equal probabilities (e.g., a certainty stratum of the largest units), and second-stage samples selected whose sizes are a fixed proportion of the unit sizes. This also results in constant selection probabilities for the second-stage units. A less-efficient approach is to select the first-stage units with probabilities proportional to size, and then to select the same proportion of subprojects from each sampled first-stage unit.

If there were only two stages of sampling in the present application, it would not be possible to use PPMS sampling for the first-stage sample units (marakez) and to select the final-stage units (subprojects) with certainty, and also satisfy the condition that each subproject have the same (unconditional) probability of selection. This follows since:

Prob of selection of a subproject

= (Prob of selection of markaz) x (Prob of selection of subproject conditional on selection of markaz)

= constant x (number of subprojects in markaz) x constant,

which is not a constant.

With three stages of sampling, however, it *is* possible both to use PPMS sampling for the first stage (marakez) and to select the final-stage units (subprojects) with certainty, and also satisfy the condition that each subproject have the same (unconditional) probability of selection. This is made possible by the presence of the second stage of sampling (villages). We have:

Prob of selection of a subproject

= (Prob of selection of markaz) x (Prob of selection of village conditional on selection of markaz) x (Prob of selection of subproject conditional on selection of village)

= constant x (number of villages in markaz) x constant / (number of villages in markaz) x constant

= constant.

Hence the presence of a sampling stage between the first and final stages of sampling enables the development of a particularly efficient sample design. The field test sample design has the unusual characteristic of achieving all of the objectives and satisfying all of the constraints placed on the field test survey in an efficient manner. Usually, sample survey designs involving multiple objectives or constraints and complex populations represent compromises. They achieve increased precision for some estimates of interest at the expense of decreased precision for other estimates of interest (e.g., stratum means vs. total-population means), or they obtain increased precision by selecting units at one stage in a certain way at the cost of decreased precision at another stage.

Note that the sample design involves the use of sampling proportional to measures of size related to numbers of ***subprojects***, but ***not*** proportional to measures of size related to the ***monetary value*** of subprojects (pounds or dollars). The reason for this is that it was desired to use the survey to produce estimates of proportions (or possibly numbers) of subprojects possessing certain attributes, ***not*** to produce estimates of monetary amounts (e.g., an estimate of the total dollar volume associated with nonoperational subprojects). A separate stratum of large subprojects was included in the field test because of the political visibility of such subprojects, ***not*** because monetary estimates were desired.

In the final Subproject Field Visitation System, it may be desired to produce estimates by sector (e.g., separate estimates for potable water subprojects, buildings subprojects, and road subprojects). For the field test, the sample sizes were too small to enable the production of estimates by sector at a useful level of precision. Note that if the final system involves multi-stage sampling nationwide it may be desirable to use subsampling of subprojects within villages (to reduce the within-village sample size and thereby reduce the effect of intravillage correlation on reducing the precision of the estimates). The total number of subprojects varies substantially by sector, and

so the sampling fractions would likely vary by sector. Since the average number of subprojects in all sectors per village is only eight, however, there would be little advantage in employing subsampling within villages, from the point of view of producing sector-specific estimates. Since the sector sample sizes would be much smaller than the total (all-sector) sample size, and the cluster effect could render the estimates of unusable precision.

These and other considerations will be taken into account in the process of developing a sampling plan for the final system. In the design of the sample for the final system, it will be important to identify all estimation objectives, just as this was important for the field test. The nature of the sample design will be determined by these objectives (as well as constraints). The final design may or may not involve the use of a multi-stage sample design or clustering. In any event, the information obtained in the field test about the intravillage (and possibly intramarkaz) correlation, will assist the development of a good sampling plan for the final system.

IV. Sample Design

This section summarizes the probability sample design proposed and used for the Subproject Field Visitation field test. (Recall that the probability sample is in addition to all nine large subprojects in the sectors of interest and the judgment sample of 15 "problematic" subprojects.)

1 *Sharqiya.* The field test was conducted in Sharqiya.

2 *Two-Stage Cluster Sampling Plan.* A two-stage cluster sampling plan was proposed. The first-stage sample units were marakez, selected without replacement with probabilities proportional to the numbers of villages, using the Rao-Hartley-Cochran sample selection procedure. The second-stage sample units were villages, selected without replacement with equal probabilities. The third-stage sample units were subprojects; all of the subprojects (in the potable water, roads, buildings, and environmental sectors) in a village were included in the sample, i.e., the proposed sample design is a "cluster" sample design. The sample sizes were: three marakez, two villages per sampled markaz, and all of the subprojects (in the relevant sectors) in each sampled village. Since the average number of subprojects (in the sectors of interest) per village is approximately eight, the expected subproject sample size was 96.

V. Sample Selection

A. SAMPLE SELECTION METHODOLOGY

As mentioned earlier, the sample design calls for use of the Rao-Hartley-Cochran (RHC) technique for selecting the markaz sample. This technique is one of a number of techniques for selecting a probability-proportional-to-a-measure-of-size (PPMS) sample with unequal probabilities without replacement, for which the variances of the sample estimates are known. We selected the RHC method because it is relatively simple and general. It is appropriate for situations in which the sample size exceeds two, as is the case in the present application (markaz sample size equal to six). (Large nationwide sample surveys often involve the selection of samples of two from a large number of small strata. Many of the methods for nonreplacement PPMS sampling are designed for sample sizes of two, and are not appropriate (inapplicable or too complex) for larger sample sizes.)

Procedures for implementing the RHC method are described in detail Cochran's book, Sampling Techniques. The technique involves randomly splitting the population of sample units (marakaz) into a number of groups equal to the desired sample size, and selecting one unit (markaz) from each group according to the probabilities specified for each unit in the group. In the present application, the selection of the unit (markaz) from a group is done using PPMS selection where the measure of size is the number of villages in the markaz. Since the units are assigned randomly to groups, this method does not keep the probabilities of selection exactly proportional to the measure of size. Because of this, the method may lose some precision; its advantages are simplicity and generality.

B. SAMPLE SELECTION PROCESS

1. Purpose of Section

This section describes the sample selection process in some detail. The detailed description is presented both to enable determination of correct procedures for analysis of the field test data and to facilitate use of this method by others at a future time.

The sample selection procedure involves the use of a table of pseudorandom numbers. Such a table was generated using a GW-BASIC program, and is presented in Table 1. Random numbers are selected from the table in sequence, starting from the beginning of the table and proceeding row by row, moving across each row to the right. Whenever a random number is required, the next number in the table is selected. The Table begins with the sequence .49839, .72405, .22986,.... Hence the first number selected from the Table is .49839, the second is .72405, the third is .22986, and so on.

2. Selection of the First-Stage (Markaz) Sample

We now describe the procedure used to select the markaz sample, in accordance with the RHC technique. Table 2 lists all of the marakez in Sharqiya, and the number of villages in each. The first step of the RHC method involves the assignment of the marakez to random groups -- six random groups, since a sample of six marakez is desired. A random number is selected for each marakaz, as shown in Table 2 (column 3). The ranks of these random numbers is indicated in column 4 of the Table. The marakez are reordered by rank (thereby placing them in random order), and listed in Table 3.

We shall now combine the marakez into six random groups, using the randomly ordered list of Table 3. Since there are 14 marakez and 6 groups, each group will be formed to have two or three marakez. With the RHC procedure, the number of marakez in each random group may be specified in advance. The variance of the estimates is minimized if the sizes (numbers of villages) of the groups is as uniform as possible. In this application, we shall use four groups of size two (villages) and two groups of size three (villages).

The usual method for organizing the marakez into random groups would be to proceed down the randomized list of marakez presented in Table 3, forming the six random groups as the first two marakez of the randomized list, the next two, the next two, the next two, the next three, and the last three. If this is done, however, it is clear (from Table 3) that the total numbers of villages in the random groups would vary substantially. This substantial variation occurs because of the small size of the village sample in each sample markaz (i.e., 2 or 3). Such variation would have the effect of introducing substantial variations in the probabilities of selection for the subprojects, i.e., the design would not be approximately self weighting. If the design is not self weighting, the variance of the sample estimates will increase, and it will no longer be appropriate to use the unweighted sample means as estimates. Instead, the availability of correct estimates would be somewhat delayed until the correct (weighted) estimates could be computed using special-purpose statistical software.

To avoid this complication, it was decided to slightly modify the sample selection from the standard RHC procedure, in such a way that the total number of villages in each random group were quite close (to 86/6 = 14, the average number). Use of this modification would have the advantage that the design would be approximately self weighting, but it would have the disadvantage that the standard RHC formulas for estimating the variances would not be quite correct. Since the proposed modification would have the effect of *reducing* the variances, however, the RHC formulas would still be applicable as approximate upper bounds for the variances. In any event, since the modification was considered to have only a slight effect on the variance, and would produce an approximately self-weighting sample, it was decided to implement it.

The proposed modification is as follows. Since there is a total of 84 villages and six groups, it is desirable (as noted above) that each group contain approximately 84/6 = 14 villages. The first markaz (of the randomly ordered list) is assigned to the first group. This markaz contains 4 villages. The list is

scanned to see if there is a markaz that contains 14-4 = 10 villages. There is – the third one in the list. That markaz is hence combined with the first one in the list to form the first random group. The first random group is now complete – it contains two marakez and a total of 14 villages.

Continuing, the second markaz in the randomly ordered list is assigned to the second random group. That markaz contains 3 villages. There is no other markaz containing 14-3 = 11 villages, so we simply assign the next markaz of the list having fewer than 11 villages. That markaz is the fourth one of the list, containing 6 villages. The group now contains two marakez and a total of 3+6 = 9 villages. We now scan the list for a markaz containing 14-9 = 5 villages, and hence select the ninth one in the list for inclusion in the second random group. The second random group is now complete – it contains three marakez and a total of 14 villages.

This process continues until all six random groups are formed. The groups to which the various marakez belong are identified in the last column of Table 3. The numbers of villages in the six random groups are 14, 14, 14, 15, 14, and 13.

We shall now illustrate the selection of a single markaz from the first random group, using probability sampling with the probabilities proportional to the number of villages in each markaz. Refer to Table 4. The marakez in the first group are marakez numbers 6 and 7 (of the original list, numbered from 0 to 13). The numbers of villages in these two marakez are 4 and 10 respectively. We cumulatively sum the numbers of villages of the group, obtaining 4 and 14. We shall now select a markaz with probability proportional to the number of villages by randomly selecting one of the integers between 1 and 14. If the selected integer is between 1 and 4, then the first markaz (number 6) is selected. If the selected integer is between 5 and 14, then the second markaz (number 7) is selected. Clearly, the probability of selection is 4/14 for the first markaz and 10/14 for the second markaz (i.e., the probabilities of selection are proportional to the numbers of villages, 4 and 10, of the marakez).

To determine a random integer between 1 and 14, we simply select the next random number of Table 1 (.978), multiply it by 14, and round up. This process produces .978 x 14 = 13.69, which is rounded up to 14. Hence the random integer is 14. The integer 14 falls in the interval 5-14 associated with the cumulative-sum range for the second markaz (number 7), and so that markaz is selected.

This process is repeated for each of the other 5 random groups. As is seen in Table 3, the following six marakez fall in the sample: 7, 8, 11, 12, 3, and 2 (or, in order, 2, 3, 7, 8, 11, and 12).

3. SELECTION OF THE SECOND-STAGE (VILLAGE) SAMPLE

The procedure for selecting the village sample is illustrated using Table 5. For each sample markaz, two villages are to be selected with equal probabilities, without replacement. The procedure for doing this will be described for the first sample markaz, markaz number 7. That markaz contains 10 villages. A random number is selected, multiplied by the number of villages (10), and rounded up. The resulting integer is the number of the first sample village. This process is repeated to obtain the number of the second sample village. If the same integer is obtained as for the first sample village, a new random number is selected and the process is repeated until a village number different from that of the first sample village is obtained.

In the case of markaz number 7, the random number .721 was selected. Multiplying by the number of villages (10) and rounding up, we obtain the integer 8. Hence village number 8 falls in the sample. A second random number is selected -- .932. Multiplying by 10 and rounding up we obtain the integer 10. Hence village number 10 falls in the sample. The village sample for markaz number 7 hence consists of villages number 8 and 10.

This process for selecting two sample villages from each markaz is repeated for the other five sample marakez, as depicted in Table 5.

The complete sample is listed in Table 6. The numbers of subprojects in each village were obtained from the Quarterly Progress Report data base. Since all subprojects in a sample village are to be surveyed, the total subproject sample size is obtained by summing the subprojects for all sample villages. The total number of sample subprojects is hence seen to be 87. The Table also presents a breakdown of the number of subprojects in the potable water, roads, and buildings sectors. (The Table was prepared before the decision to add the environmental subprojects to the field test was made, hence those subprojects are not included in the Table.)

In addition to the 87 small subprojects, nine large subprojects were included in the field test. One of these large subprojects is located in one of the sample villages and the others are located in seven other villages.

VI. Sample

Table 7 presents a combined list of the 87 small and 9 large subprojects. This Table also presents the names of the marakez and villages in which the subprojects are located. The names, as well as other information on the subprojects, were extracted from the Quarterly Project Report data base.

Table 1. Table of Pseudorandom Numbers

.94008	.39482	.98002	.56370	.23228	.65518	.03803	.12234	.60336	.62094
.10878	.25603	.41946	.80117	.97750	.67161	.35753	.68253	.49678	.87379
.72097	.93229	.97722	.99323	.93026	.85155	.82808	.86473	.29355	.07831
.44785	.35528	.02235	.54061	.36491	.94223	.49707	.65965	.62815	.54143
.78273	.80292	.44759	.58531	.39451	.98734	.08501	.35102	.31345	.83235
.39458	.53858	.25199	.76305	.21346	.09897	.16968	.46335	.00246	.31229
.03875	.57905	.73664	.11586	.85974	.67102	.04594	.68234	.92714	.41826
.20582	.24788	.34503	.07109	.77785	.19022	.79266	.15422	.50039	.35684
.54012	.41991	.42915	.64306	.71392	.05108	.32788	.82019	.81267	.11445
.30892	.61362	.47266	.26363	.72392	.03099	.48350	.69230	.30437	.94027
.05375	.33728	.55229	.89793	.32652	.33169	.38082	.22702	.14516	.70806
.00488	.34533	.06879	.35147	.08174	.34350	.63801	.20264	.10171	.81313
.89048	.71111	.02871	.92495	.12655	.53830	[illegible]	.75690	.92489	.51070
.42144	.87227	.40813	.82295	.44880	.14783	.75846	.49097	.53775	.85195
.98325	.05512	.37973	.97273	.93072	.53327	.23360	.79613	.55551	.62398
.31622	.73405	.43435	.15957	.44308	.40255	.53166	.80930	.91893	.94986
.21516	.03247	.67840	.63459	.52155	.28151	.67662	.65384	.67208	.45518
.52987	.22127	.83818	.60164	.40619	.89531	.15714	.87157	.16598	.15720
.99767	.72134	.75266	.23921	.47554	.36609	.83328	.20265	.21924	.48290
.68916	.26611	.38556	.16386	.35650	.38349	.92526	.11923	.01698	.07220
.09864	.13058	.63846	.49122	.24122	.75387	.86551	.76929	.77932	.27261
.16012	.58298	.49573	.90097	.19226	.98484	.04529	.23692	.48056	.48152
.72053	.96533	.28289	.36190	.46733	.81111	.73699	.27521	.15049	.74235
.25115	.05916	.06939	.42340	.14475	.81806	.72360	.91806	.53894	.25095
.82850	.64264	.49728	.22970	.58095	.56913	.41629	.62720	.17485	.67831
.66617	.39539	.56680	.66285	.48134	.89348	.49174	.88049	.10113	.46600
.72851	.55712	.16036	.27098	.61558	.23991	.32999	.35803	.81657	.50047
.20774	.99652	.33594	.88792	.45871	.25362	.78458	.13207	.20600	.82253
.89548	.59650	.67136	.10052	.28929	.98171	.06579	.51710	.49001	.27823
.73016	.11211	.19057	.96978	.52587	.36390	.76846	.48635	.47544	.26739
.05139	.90734	.65314	.66225	.33287	.91470	.32565	.81099	.63787	.99603
.34263	.22710	.23842	.89775	.77738	.75317	.41926	.32815	.81747	.88901
.49342	.91060	.90533	.36979	.06790	.38660	.12893	.99417	.55928	.06897
.86234	.70238	.95930	.54492	.65641	.90440	.85056	.62922	.37936	.05805
.67207	.56727	.60743	.49714	.23493	.48830	.26548	.50957	.48794	.10843
.51763	.66546	.53320	.68383	.90790	.39530	.89177	.21382	.75079	.81805
.01929	.89402	.77195	.81926	.07156	.81947	.99879	.37922	.92025	.93773
.90109	.15107	.41842	.25203	.78166	.28892	.91625	.27441	.41691	.52598
.02659	.72881	.94754	.50271	.64061	.50418	.01911	.34488	.19343	.85769
.17270	.49178	.17974	.86776	.48552	.77427	.91942	.13723	.46153	.74295
.57319	.74648	.17209	.24628	.40819	.45657	.64656	.25867	.31364	.16233
.51292	.95865	.56646	.59551	.88843	.05699	.84689	.27791	.52021	.37478
.50411	.72446	.17590	.67161	.07187	.89642	.28428	.72367	.03412	.90445
.82588	.65185	.84063	.96173	[illegible]	.49993	.17261	.58714	.98338	.06265
.45834	.05831	.83476	.96384	.63881	.11481	.12155	.98457	.18331	.91125
.69245	.64126	.35508	.22038	.54166	.01380	.72695	.88627	.44951	.82370
.94692	.52738	.87309	.76217	.75382	.48973	.08689	.06829	.24294	.04998
.57338	.35711	.38161	.86859	.66368	.69778	.77487	.69295	.42816	.94168
.68109	.31046	.81709	.30055	.27758	.75170	.05121	.17459	.17622	.74355
.25244	.23061	.18906	.41228	.38442	.73015	.34402	.23672	.29319	.10759

Table 1 (continued). Table of Pseudorandom Numbers

```
.15043 .25134 .19780 .79829 .11469 .49943 .28093 .31673 .50591 .73617
.58951 .58459 .37151 .38179 .17536 .50892 .80287 .72467 .48581 .81024
.45092 .37451 .00432 .10066 .29180 .96539 .14110 .21205 .64247 .10895
.37376 .77409 .42632 .59748 .63790 .04246 .68883 .85707 .96776 .97694
.36316 .65073 .38685 .56954 .28580 .03148 .04854 .42261 .46219 .54554
.62672 .57699 .58232 .68540 .95641 .59010 .78639 .59296 .92440 .86411
.68042 .01267 .60985 .22403 .60188 .49465 .67492 .74585 .13529 .84777
.00542 .83466 .87789 .84288 .60142 .55278 .45522 .16580 .71785 .19787
.78679 .62057 .25517 .13109 .60165 .58248 .39996 .85517 .20407 .70125
.51556 .34259 .48912 .85415 .28270 .61367 .20834 .34909 .09011 .56472
.48531 .37778 .67516 .04614 .58134 .71886 .01448 .19056 .91102 .69099
.46456 .99096 .30793 .35599 .15242 .12879 .54018 .97198 .61621 .05212
.07610 .69650 .69583 .30373 .09983 .53961 .67357 .09926 .77696 .56705
.66473 .15533 .77015 .35335 .25823 .76763 .00475 .68494 .40727 .13913
.68445 .51329 .56990 .95812 .55684 .15806 .09971 .22635 .42914 .76017
.88641 .33721 .93372 .98496 .24646 .40380 .54378 .97532 .06371 .83705
.49839 .72405 .22986 .35155 .50667 .16820 .74043 .54078 .02614 .59832
.73451 .95526 .84251 .21978 .90845 .56845 .84791 .91864 .91661 .47977
.55659 .71473 .01198 .99274 .87045 .46481 .93680 .94313 .09473 .81566
.06321 .27703 .20361 .40723 .78384 .87904 .35584 .32577 .15248 .65033
.66037 .80694 .77464 .83682 .19198 .18894 .94825 .60726 .25687 .11561
.05422 .88140 .90832 .98961 .19710 .40808 .71232 .67551 .09570 .09336
.76317 .45290 .22373 .57866 .94911 .21956 .20038 .06889 .46795 .94453
.43098 .41052 .59270 .37129 .99764 .41979 .68721 .97084 .79981 .46077
.47173 .14485 .94499 .27352 .53862 .67880 .38930 .05218 .11792 .44505
.82810 .77314 .09299 .35602 .83665 .77312 .26614 .56728 .26072 .57733
Ok
LIST
10 DIM X(10)
20 OPEN"O",#1,"RANDNO1.FIL"
30 RANDOMIZE 5000
40 FOR I=1 TO 100
50 FOR J=1 TO 10
60 X(J)=RND
70 NEXT
80 PRINT USING ".##### ";X(1),X(2),X(3),X(4),X(5),X(6),X(7),X(8),X(9),X(10)
90 PRINT#1,USING ".##### ";X(1),X(2),X(3),X(4),X(5),X(6),X(7),X(8),X(9),X(10)
100 NEXT
Ok
```

Table 2. Illustration of Rao-Hartley-Cochran Method for Selecting Marakaz Sample -- Step 1: Random Ordering of Marakez

Markaz No.	No of Villages	Random No.	Random Order
0	2	.940	13
1	6	.395	6
2	6	.980	14
3	9	.564	8
4	6	.232	4
5	7	.655	11
6	4	.038	1
7	10	.122	3
8	5	.608	9
9	5	.621	10
10	3	.109	2
11	10	.256	5
12	9	.419	7
13	2	.801	12
	84		

Table 3. Illustration of Rao-Hartley-Cochran Method for Selecting Marakaz Sample -- Step 2: Assignment of Randomly Ordered Marakez to Groups

Random Order	Markaz No.	No. of Villages	Random Group No.
1	6	4	1
2	10	3	2
3	7	10	1
4	4	6	2
5	11	10	3
6	1	6	4
7	12	9	4
8	3	9	5
9	8	5	2
10	9	5	5
11	5	7	6
12	13	2	3
13	0	2	3
14	2	6	6

Table 4. Illustration of Rao-Hartley-Cochran Method for Selecting Marakaz Sample -- Step 3: Probability-Proportional-to-Size Selection of a Single Item from Each Random Group

Random Group No.	Markaz No.	No. of Villages	Cumulative Sum	Random No.	Random Integer	Selected Markaz
1	6	4	4	.978 x 14 = 14		
1	7	10	14			* (7)
		14				
2	10	3	3	.672 x 14 = 10		
2	4	6	9			
2	8	5	14			* (8)
		14				
3	11	10	10	.358 x 14 = 6		* (11)
3	13	2	12			
3	0	2	14			
		14				
4	1	6	6	.682 x 15 = 11		
4	12	9	15			* (12)
		15				
5	3	9	9	.497 x 14 = 7		* (3)
5	9	5	14			
		14				
6	5	7	7	.873 x 13 = 12		
6	2	6	13			* (2)
		13				

Table 5. Illustration of the Procedure for Selecting the Village Sample

Sample Markaz	No. of Villages	Random Numbers	Sample Villages
7	10	.721 x 10 = 8; .932 x 10 = 10	8,10
8	5	.977 x 5 = 5; .993 x 5 = 5; .930 x 5 = 5; .852 x 5 = 5; .828 x 5 = 5; .865 x 5 = 5; .294 x 5 = 2	5,2
11	10	.078 x 10 = 1; .447 x 10 = 5	1,5
12	9	.355 x 9 = 4; .022 x 9 = 1	4,1
3	9	.541 x 9 = 5; .365 x 9 = 4	5,4
2	6	.942 x 6 = 6; .497 x 6 = 3	6,3

Table 6. List of Sample Marakez and Villages

Sample Markaz	Sample Village	Total No. of Projects	Water Projects	Roads Projects	Buildings Projects
2	3	7	3	0	4
	6 (#90)	6	2	2	2
3	4	7	5	0	2
	5	8	3	0	5
7	8	7	3	0	4
	10 (#90)	3	2	0	1
8	2	5	2	0	3
	5 (#90)	5	2	2	1
11	1	7	5	0	2
	5	10	2	1	7
12	1	13	3	0	10
	4	9	4	0	5
		87	36	5	46

Table 7. Subproject Monitoring System: Sharqiya Field Test Sample

Markaz	Village	No. of Small Sprjts	No. of Large Sprjts	Total Sample Size	No. Watr Spts	No. Road Spts	No. Bldg Spts
2. Abu Kebeir	3. El Rahmaniya	7	0	7	3	0	4
	90. Abu Keb. Cty	6	0	6	2	2	2
3. Belbeis	4. El Balashoun	7	0	7	5	0	2
	5. El Zawamel	8	0	8	3	0	5
5. El Hesaniya	2. Gezirit Saoud	0	2	2	0	2	0
7. Faqous	4. El Ghazaly	0	1	1	0	1	0
	6. El Samaana	0	1	1	0	1	0
	8. Sawadah	7	0	7	3	0	4
	90. Faqous City	3	0	3	2	0	1
8. Hehiya	2. El Mahdiya	5	0	5	2	0	3
	90. Hehiya City	5	0	5	2	2	1
10. Mashtoul	90. Mashtoul Cty	0	1	1	1	0	0
11. Minya El Qamh	1. Bani Helal	7	1	8	6	0	2
	5. El Telein	10	0	10	2	1	7
12. Zaqazeiq	1. Bani Amer	13	0	13	3	0	10
	4. El Asslogi	9	0	9	4	0	5
	5. El Zankaloun	0	1	1	0	1	0
	90. Zaqazeig Cty	0	1	1	1	0	0
13. Awlad Saqr	90. A. Saqr Cty	0	1	1	1	0	0
Totals		87	9	96	40	10	46