SAMPLING METHODS USED IN MANPOWER SURVEYS IN IRAN*

Jamshid Shokrzādeh

Since 1958 various agencies of the Irānian government have carried out studies of manpower problems by means of sample surveys. During this time many different survey designs have been used, and this is a report not only on these past methods but also on the design which has now been adopted by the General Department of Manpower Statistics which is, in our opinion, superior.

The 1337 (1958-1959) Sample Survey[1]

The first survey of this kind conducted in Irān was a study of manpower resources and requirements in urban areas. It covered all the establishments of 135 cities which played some role in the economy except those whose activities were purely agricultural. The survey design divided all the establishments in each city into two groups: large establishments employing 50 workers or more, and small establishments employing 49

*Paper presented at the CENTO Symposium on Manpower Planning and Statistics, Tehrān, November 24th - December 5th 1969.

1. "This survey was nation-wide in scope and covered all segments of the economy: Agriculture (industrial), mining, manufacturing, construction, communication, government and community services, health and sanitation, and education ...."

"Excluded from this survey were: farm workers, domestic servants, household handicraft workers, the armed forces, police, clergy, foreign embassies and other foreign agencies. ..."

"In order to provide the greatest usefulness, the data were tabulated by province, industry, occupation and ownership."

"The ILO International Classification of Occupation (3-digit) was used for job coding and the United Nations Industry Code (3-digit) was used for industry classification.

"The block sampling technique was used in securing a representative sample of small establishments having less than 50 employees in each of the eleven major population centers. The cities in which the sample was applied included the following: Tehrān, Tabrīz, Ahwāz, Mashad, Rasht, Reza'iyeh, Shirāz, Isfahān, Sāri, Kermānshāh, Kermān."

(cont'd xt page)

workers or less. Thus 2,749 large establishments some of which had branch offices all over the country were included and 15,150 small establishments located in 11 sharestān centres were selected by means of simple random sampling. Information on sex, occupation and skill, the number of persons employed, and employment requirements and surplus workers in different occupations,were collected from both the large and the small sample establishments.

The criticisms that can be made of this method are many. Firstly, in 1958 complete information on the type of establishment included in the survey was lacking so that sample selections were made on the basis of an incomplete frame and it was impossible to arrange the sample establishments into homogeneous classes. Moreover, two problems arose in the selection of an establishment as a statistical unit: (i) only regular workers in establishments could be covered, and (ii) some of the workers who were engaged in more than one job were enumerated more than once. In addition, lack of information meant that some of the economic activities carried out in the establishments did not correspond to commonly classified occupations. Secondly, although the survey did give a picture of the kinds of workers employed in urban areas and their distribution among different kinds of economic activities, because it excluded rural workers it threw no light on problems of the geographical mobility of the labour force(including migration) and was therefore of little use in regional planning. Thirdly, a complete study of manpower resources (population and unemployed) is not possible in a survey of the labour force which is confined to establishments.

The 1343 (1964-1965 Sample Survey[2]

This was a household sample survey of the rural and urban population

For further information see: Ministry of Labour and Plan Organisation, *National Manpower Resources and Requirements Survey* (Tehrān:Governmental Affairs Institute, July 1959).

2. This survey was first planned at the beginning of Abān 1342 (October 1963) and was carried out in urban areas from the beginning of Esfand 1342 to Ordibehesht 1343 (February 20th 1964 - April-May 1964) and in rural areas during Xordad and Tir 1343. (May, June, July 1964). For further information see, Ministry of Labour and Social Affairs, General Department of Manpower Statistics, *barresihā-ye masā'el-e niru-ye ensāni* (Studies of Manpower Problems), 4 vols. Tehrān, 1343-45 (1964-1967).

of the country. At this time all cities with a population of 5,000 or more were surveyed. The total number of such cities (186) were divided into 8 groups according to their population; II large cities with a population of 70,000 or more were selected from the first group, and 58 medium and small cities from the other 7 groups were selected by simple random sampling. This meant that 69 out of 186 cities were eventually selected as sample units.

A two-stage sampling design was used for every sample city; blocks were selected by probability proportional to size, (number of doors) and then sample households in the sample blocks were selected by circular systematic random sampling. The sampling design in every large city and also in medium and small cities for the total households under study was self-weighting.

Although every city was known to be divided into several blocks, the number of households in each block was unknown. However, since the blocks had been previously enumerated, 200 sample blocks in every large city (400 blocks for Tehrān), and 318 blocks in all medium and small cities were selected by probability proportional to size (number of doors). All the households in each sample block were listed and an average of 7 sample households per block were selected. The number of sample households in a block were calculated by the formula below:

$$m_i = 7 \frac{M_i}{M'_i} \cdot \frac{n}{\sum_1^n \frac{M_i}{M_i}}$$

m_i = Number of sample households in block i
M'_i = Total number of doors in the city
M_i = Number of households in block i
M'_i = Number of doors in block i
n = Number of sample blocks in the city

Information on household composition, sex, age, marital status, births and deaths, literacy status, migration, occupation, skill, occupational status, wages and earnings, hours of work, and unemployment were collected from the sample households.

The rural section of the survey covered all residential centres with a population of less than 5,000 and the 1335 (1956) census listed 52,953 such villages.[3] These were divided into 3 groups on the basis of the

3. See Abasqoli Khajenouri, Abbās Jāme'i, *tarh-e āmārgiri-ye niru-ye ensāni*,

number of households as follows: (i) villages with 1-200 households, (ii) villages with 201-950 households, and (iii) villages with 951 or more households. Three-hundred of the villages falling into the first and second group were surveyed but, since the third group included only 16 villages, all were included in the sample.

A two-stage sampling design was used: first, sample villages were selected by simple random sampling from every group, and then households in each sample village were selected by circular systematic random sampling. The sampling design in every group for the households under study was self-weighting. A complete list of the households in each village was prepared and divided (according to occupation) into 7 groups; 10 households were then selected by circular systematic sampling. If the number of households in the village were less than 10 then all were surveyed.The data collected for the rural areas was the same as that collected for the urban areas.

This was the first design to follow a scientific statistical method: the sampling method used was based on probability proportional to size (P.P.S.), a method which was itself based on the variance in population, and this meant that the estimates obtained were fairly precise. Thus,despite the limited budget allocated to the survey, and the shortage of trained staff available to carry out the enumeration, it yielded results that were accurate enough to apply to many of the rural and urban manpower problems which beset Irān. However, at the same time, there are still a number of points which are open to criticism. Firstly, in the urban survey the numbers of doors in a block was used for determining probability of block selection (no data on households was available). Thus the weights used for estimation were rather unsatisfactory because generally the correlation between the number of doors and the population is not very high. Secondly, the average number of small households to be selected from a sample block was calculated on the basis of the frame of only one city (Tehrān) but the results were also used for other cities. Calculation of the total number of households to be selected in the urban areas and their proportional distribution among all sample cities would have been more satisfactory. Thirdly, a stratified sampling design was

(Manpower Statistics Project), in Ministry of Labour and Social Affairs, General Department of Manpower Statistics, *barresihā-ye masā'el-e niru-ye ensāni* (Surveys of Manpower Problems), Tehrān, 1345 vol. 1, p.308.

used for the villages although, in view of the great variation in the number of households in Irān's villages, the probability proportional to size method would probably have raised the precision of the estimate. Fourthly, the selection of sample households in the sample villages was done by circular random sampling and households were arranged by occupation for this purpose. However, an arrangement of households by size and characteristics such as unemployment may have given more representative samples because village households differ from one another in terms of the number of members, the number of employed and so on. Finally, data were collected from urban and rural areas all over the country only once a year, whereas the study of manpower problems in different regions requires data for all seasons if it is to be of any use to socio - economic development planning.

The 1344 (1965-1966) Sample Survey[4]

The 1344 survey conducted to collect data on manpower problems was confined to households in urban areas (residential centres with a population of over 5,000). In 1344 (1964-5) residential centres defined as urban areas amounted to 186 and all these were included in the survey.

A two stage sample was employed: each of the 186 cities were divided into blocks and one in every 12 blocks was selected by the probability proportional to size method (number of doors). Sample households in the blocks were then selected by systematic random sampling and the number of households was calculated by the same method as that used in the 1343 (1964-1965) survey. First all the households in the sample block were described, and then an average of 5 per block were selected as samples. Sampling design for all the households in the cities under study was self weighting. The number of sample blocks were divided into four sub-samples and information on the sample households of each part was collected each season. Thus the total sample for the year consisted of four quarterly sub-samples.

4. For the results of this survey see: Ahmad Zarandi with the co-operation of the Research Group on Manpower Problems, *barresi-ye niru-ye ensāni-ye jām'e-e shahri-ye Irān bar mabnā-ye āmārgiri nemune'i-ye niru-ye ensāni* (A Study of Urban Manpower Problems in Iran on the Basis of the Manpower Sample Survey) Tehrān, Ministry of Labour and Social Affairs, General Department of Manpower Statistics, 1347 (1968-1969).

The survey was designed to make it possible to obtain data related to population, employment, unemployment, hours of work, and wages and earnings on a seasonal basis for the urban population and yearly for the main regions of the country.

This was the second survey to be conducted over the whole year and on the basis of scientific statistical methods. It yielded the relevant estimates for the purpose of comparing data for the previous year for urban areas, and precision was gained by the adoption of the probability proportional to size method which reduced variance. However, the sampling coefficient used to determine the number of sample blocks in the city was chosen on the basis of past experience, whereas had it been independently calculated the estimates would have been more precise and it would have been possible to indicate confidence intervals.

This survey is also subject to the same criticism as the previous one in that the probability of selection of the sample blocks was determined on the basis of the number of doors per block rather than in terms of the number of households, a much more satisfactory measure.

The 1348 (1969) Sample Survey

The current survey design has taken into account the merits and demerits of the previous designs and is described below.

The sampling design for both the rural and the urban surveys as well as the formula used for estimation will provide measures of changes that take place over different seasons and from one year to another in different statistical regions (Ostāns). Data will be collected by experienced enumerators who have been specially trained for their task.

The rural survey has been designed to cover all residential centres with a population of 5,000 or less; the 1345 (1966) census lists 66,748 such villages. Each statistical centre (Ostān) constitutes one stratum and the sample villages for the whole country are divided into 5 subsamples (n_1, n_2, n_3, n_4, n_5). Each of these sub-samples is distributed in proportion to the population or the number of households among the statistical regions and in each statistical region the sub-samples are separately selected by probability proportional to the population or the number of households of a village. Enumeration from the sub-sample n_1 will be repeated for each of the four seasons of the first year and in addition

the sub-samples n_2, n_3, n_4, n_5 will be enumerated in turn in each of the four seasons. In the second year sub-sample n_2 will be enumerated in each of the four seasons and one of the sub-samples n_1, n_3, n_4, and n_5 will be enumerated each season. This method will continue in the third, fourth, and fifth year. Table 1 shows the samples of villages in different seasons over a period of five years.

Table 1

Samples Over a Period of Five Years

Year	First season	Second season	Third season	Fourth season
1	n_1+n_2	n_1+n_3	n_1+n_4	n_1+n_5
2	n_2+n_1	n_2+n_3	n_2+n_4	n_2+n_5
3	n_3+n_1	n_3+n_2	n_3+n_4	n_3+n_5
4	n_4+n_1	n_4+n_2	n_4+n_3	n_4+n_5
5	n_5+n_1	n_5+n_2	n_5+n_3	n_5+n_4

A complete list of households for every sample village for every season of the year will be prepared and this will show the total number of members, the number of employed, and the number of unemployed for each household. If one village is repeated from one season to another, the household list will be checked and revised. The selection of sample households in the village will be carried out on the basis of the linear systematic method. For this purpose the list of households by total number of members and unemployed members will be arranged as follows:

1-4 persons	1-4 persons	5 persons and more	5 persons and more
with unemployed person	without unemployed person	without unemployed person	with unemployed person

In order to select the sample households, the random start and sampling interval will be previously indicated to the enumerators. The average number of sample households for a village will have been previously calculated and the enumerator will select the required number of households by specified sampling intervals while he is at work in the village.

In this survey design, a two stage sampling method has been used: firstly the sample villages are selected by probability proportional to size and secondly, households in the sample village have been selected

according to linear systematic random sampling. The sampling design in every statistical region for the households under study is self-weighting. Data relating to composition of households, sex, age, literacy, skill, occupational status, wages and earnings, hours of work, unemployment, underemployment, labour force wasted, and births and deaths, are to be collected.

Since the necessary budget has been allocated in advance, optimising the number of samples in this design was based on a sampling error of 5 per cent. Thus in some statistical regions, n-50, the primary samples were selected by probability proportional to size (number of households) and were calculated as follows:

Estimate of variance in villages:

$$s^2 = \frac{\sum_1^n z^2{}_i - (\sum_1^n z_i)^2/n}{n-1}$$

Estimate of a characteristic basis of one village:

$$z_i = \frac{x_i}{\pi_i}$$

Number of population = x_i

Probability of selecting sample villages = π_i

Reciprocal of the estimate $c^2 = \frac{s^2}{z^2}$

Maximum of c^2's $= c.^2$

Villages in one statistical region for 5 per cent errors in the estimates: $n. = \frac{c.^2}{(\%5)^2}$

Number of total villages for one season in a year = number of statistical regions, X n.

The estimate of each statistical characteristic is made on the basis of one sub-sample for one season in the statistical region on the basis of formulae below:

Quantity of variable characteristics in household j in village i = x_{ij}

Probability of selecting village i = π_i

Number of villages in the statistical region = N

Number of sample villages in the statistical region = n

Sampling interval in the village i = I_i

Total number of households in the sample village i = M_i

Number of sample households in sample village i = m_i

Estimate of total characteristic in the statistical region:

$$\hat{x} = \frac{1}{n} \sum_{1}^{n} \frac{I_i}{\pi_i} \sum_{1}^{m_i} x_{ij} = \frac{\alpha}{n} \sum_{1}^{n} z_i$$

$$z_i = \sum_{1}^{m_i} x_{ij}$$

$$\hat{V}(\hat{x}) = \alpha^2 \cdot \frac{s^2}{n}, \quad s^2 = \frac{\sum_{1}^{n} z_i^2 - (\sum_{1}^{n} z_i)^2/n}{n-1}$$

If, according to the above formula, $\hat{X}_1$ and $\hat{X}_2$ are respectively the estimates of data for samples repeated and non-repeated, then the estimate of total characteristic and its variance for one season in the statistical region would be:

$$\hat{x} = \frac{\hat{x}_1 + \hat{x}_2}{2}, \quad \hat{V}(\hat{x}) = \frac{1}{4} \{\hat{V}(\hat{x}_1) + \hat{V}(\hat{x}_2)\}$$

The estimate of changes in a characteristic between two seasons in a statistical region is obtained from the formulae below. The difference between repeated sample data for a characteristic in two seasons:

$$z_2 - z_1 = \beta_i$$

Estimate of the change in the characteristic in two seasons:

$$\hat{D} = \frac{\alpha}{n} \sum_{1}^{n} \beta$$

Estimate of variance:

$$\hat{V}(\hat{D}) = \alpha^2 \cdot \frac{s^2}{n}, \qquad s^2 = \frac{\sum_{1}^{n} \beta_i^2 - (\sum \beta_i)^2/n}{n-1}$$

The urban survey included 249 places with a population of 5,000 or more and 9 centres with a population less than 5,000. Thus the total number of urban units included was 258. All the urban centres in every statistical region which had more than 500 households were divided, on the basis of the 1345 (1966) census, into statistical strata containing about 500 households. These strata form the first-stage sampling units (in the same way as villages in the rural areas) and so the method of calculating the number of samples and their distribution into statistical regions in the four seasons of the year, and the method of estimating the data is

exactly the same as that for the rural areas as indicated above. The type of data collected is also the same.

In our opinion this survey design represents and advance on previous ones. Firstly it opens up the possibility of obtaining data necessary to regional socio-economic planning every season of the year and of comparing changes from one season to another in different statistical regions. Secondly, the method used to calculate the number of sampling units, the sampling errors of the estimates, the proportional distribution of samples among statistical regions, and the selection of samples by probability proportional to size, will without any doubt, yield much more precise estimates. Apart from this, since there is a positive correlation between the number of households in a city block and the population of the block, the use of the number of households as a basis for determining the size of the block will probably also raise the precision of the results. The method of estimation used in this survey is easier than that of the preceeding ones but the amount of calculation work involved is greater.

DISCUSSION AND CRITICISM

On the following pages we are printing three short articles we have received in response to Ahmed Kooros' arguments on the relationship between the rate of participation of the rural labour force and economic development in Iran. As is customary, before going to press we have given these comments to Mr. Kooros and asked him for a reply. The articles are as follows: