

Sample Design and Estimation in the Island Context

Peter Wingfield-Digby
Statistical consultant
P.O. Box 308 Chiang Mai
Chiang Mai 50000, Thailand
pwdigby@loxinfo.co.th

1. Introduction

Experience on the recent 2003 Labour Force Survey in the Kingdom of Tonga provides a good illustration of some of the sample design and estimation problems that are likely to be experienced in the Pacific and in similar island nations elsewhere. Selecting an efficient sample is a particularly challenging exercise in this context.¹

Tonga consists of 170 islands, spread over a vast expanse of the Pacific Ocean, between latitudes 15°S and 24°S and between longitudes 173°W and 179°W. The total land area, including about 8,000 hectares of inland water, is only about 75,000 hectares. Tonga has a population of about 100,000, living in some 16,000 households on 48 inhabited islands. Administratively, the country is divided into five divisions: the main division of Tongatapu, with about 70,000 people, and the much smaller divisions (in terms of population size) of Vava'u, Ha'apai, 'Eua and the Niuaus.

The Tonga Statistics Department has had some previous experience of labour force surveys. The first survey had been conducted in 1990, and had covered about 1900 households. For the second survey, which was conducted in 1993/94, the sample size was increased to more than 3300 households. That second survey was carried out in two phases, six months apart, so as to take some account of seasonality, but in the analysis the results of the two phases were pooled, and no separate results were given for each phase. Survey reports were successfully completed and published for both surveys.

In all sample survey work, certain general principles should guide one's work. First of all, one should try to go for quality not quantity. Secondly, one should use one's field resources as efficiently as possible. In the case of the Pacific, this may well mean leaving out some of the more remote islands, particularly if they are very small. Thirdly, one should design a user-friendly questionnaire that is easy to administer. Fourthly, one should train and supervise one's interviewers as carefully as possible. And finally, one should exercise tight control over all stages of data entry, editing and processing, so that one finishes up with a reliable data set.

¹ The author was involved in this survey as a consultant for the Asian Development Bank under project number TA No. 3432: TON/Poverty Assessment. The Tonga Statistics Department has placed a full copy of the survey report on the PRISM website: <http://www.spc.int/prism/country/to/stats/pdfs/lfs/lfs2003.pdf>

2. Sample design for the 2003 Tonga LFS

In preparing the sample design, there are a number of key issues that must be addressed. These are:

- General structure and timing of the survey
- Domains for analysis
- Survey population vs. Target population
- Method of sample selection
- Fine tuning of the sample design

Structure and timing

The survey planned to use a questionnaire that was broadly similar to that used for the earlier LFSs, but some small changes were made, to improve the flow of the questions.

The Statistics Department requested that the third LFS should be carried out in similar fashion to the second LFS, with fieldwork being spread over two phases six months apart, so as to capture some seasonal variation. The second LFS had run up against a financial shortfall in its second phase, but a major part of the funding for the third LFS was coming from the Asian Development Bank, and it was expected that resources would be sufficient to complete the survey this time. The two phases of fieldwork commenced in February and August 2003 respectively.

Domains for analysis

Given the relatively small size of the Tongan population and its vast geographic distribution, great care is required to ensure that an efficient sample is selected, which provides adequate representation of the population but which is at the same time cost-effective. The population of Tonga is heavily weighted towards the main division of Tongatapu. Of the 16,000 households in Tonga, almost 11,000 are to be found in Tongatapu, nearly 3,000 in Vava'u, about 1,500 in Ha'apai, only 800 in 'Eua and 400 in the Niuas.

The two earlier LFSs had produced results only at the national level, but this time there was a demand for more disaggregated data. The consultant's initial idea was that it might be sufficient to provide results for just two domains: the main division of Tongatapu, consisting of the main island of Tongatapu and ten other small islands nearby, would form one domain, and the other four divisions the other domain. Another option might have been to produce results separately for urban and rural areas, but one problem was how to define exactly what was meant by 'urban' and 'rural'. In any case, one might question whether the urban/rural dichotomy has much meaning in the context of small Pacific islands.

Meetings were held within the Statistics Department, and a user-producer workshop held with representatives of the four key ministries with a direct interest in the survey: the Ministry of Finance; the Central Planning Department; the Ministry of Labour, Commerce and Industry; and the Ministry of Education.

From these discussions it became clear that a greater level of disaggregation was in fact required. The interest was in having separate results for each of the five divisions. This stated need from the clients provided a challenge to those responsible for sample design, in view of the very small population size of 'Eua and the Niuas. It was agreed that the survey would be conducted in two phases, six months apart, so as to get some measure of seasonal variation. The survey report would contain some simple analysis of the key aggregates for each phase, presented at the national level. The main analysis, however, would be done on the pooled survey results from the two phases, with the analysis being presented in terms of island divisions.

Survey population vs. target population

The data contained in Table 1 highlight the problem facing those concerned with sample design. In the case of Tonga, there is one large island that accounts for two-thirds of the total population, a further ten islands with populations of over 100 each, and then 37 other islands that each has a population of less than 100. Table 2 shows the 48 islands that were inhabited at the time of the last census, classified according to the division in which they are located.

Table 1: Tonga: Inhabited islands (1996 census) by number of households on each

Number of households on the island	Number of:		Cumulative frequency		Percent cumulative frequency	
	Islands	Households	Islands	Households	Islands	Households
1-4	13	26	48	16194	100.0	100.0
5-9	1	6	35	16168	72.9	99.8
10-19	5	79	34	16162	70.8	99.8
20-29	4	92	29	16083	60.4	99.3
30-39	4	134	25	15991	52.1	98.7
40-49	3	125	21	15857	43.8	97.9
50-99	7	469	18	15732	37.5	97.1
100-499	7	1112	11	15263	22.9	94.3
500-999	2	1321	4	14151	8.3	87.4
1,000-9,999	1	2108	2	12830	4.2	79.2
10,000+	1	10722	1	10722	2.1	66.2
Total	48	16194				

Source: Tonga Statistics Department, *Population Census 1996: Administrative Report and General Tables*, 1999.

After careful examination of the data shown in Table 1, it was decided to limit the survey population to those 18 islands containing at least 50 persons each. Although this meant that the 30 smallest islands were not covered, the survey would still be covering 97 percent of the total population of Tonga. It is recognised that the employment characteristics of people on these very small islands, some of them very remote, may be rather different from those in the rest of the country. Their exclusion from the sampling frame used for the survey, and the resulting slight loss in accuracy of the survey results, was considered a price worth paying, since it was going to be counterbalanced by a substantial reduction in field costs. Table 3 shows only those islands with more than 50 inhabitants.

Table 2: Distribution of Tonga's 48 inhabited islands by division

<u>Tongatapu</u>	<u>Vava'u</u>	<u>Ha'apai</u>	<u>'Eua</u>	<u>Niuas</u>
11	17	16	1	3
'Ataa	Foeata	Foa	'Eua	Niuafu'ou
'Atataa	Hunga	Fonoifua		Niuatoputapu
'Eueiki	Kapa	Fotuha'a		Tafahi
Fafaa	Kenutu	Ha'afeva		
Nukunukumotu	Koloa	Ha'ano		
'Oneata	Lape	Kotu		
'Onevai	Nuapapu	Lifuka		
Pangaimotu	Ofu	Lofanga		
Tongatapu	Okoa	Mango		
Velitua Hahake	Olo'ua	Matuku		
Velitua Hihifo	'Ovaka	Mo'unga'one		
	Pangaimotu	Nomuka		
	Tapana	'O'ua		
	Taunga	Tofua		
	'Utungake	Tungua		
	Vaka'eitu	'Uiha		
	Vava'u M.			
100 %	100 %	100 %	100 %	100 %

Table 3: Distribution of Tonga's 18 islands with population of 50 or more, and proportion of households covered in each division (97.1% of the population)

<u>Tongatapu</u>	<u>Vava'u</u>	<u>Ha'apai</u>	<u>'Eua</u>	<u>Niuas</u>
1	7	7	1	2
Tongatapu	Hunga	Foa	'Eua	Niuafu'ou
	Kapa	Ha'afeva		Niuatoputapu
	Koloa	Ha'ano		
	Nuapapu	Lifuka		
	Pangaimotu	Nomuka		
	'Utungake	Tungua		
	Vava'u M.	'Uiha		
99 %	95 %	85 %	100 %	93 %

Method of sample selection

The decision was made to use PPS (probability proportional to size) sampling, since this would result in fixed workloads. Any changes in the measure of size used for selecting the census blocks would be accommodated by weighting the sample data during the analysis to take account of the difference between the original measure of size for each census block and the size found during the listing operation.

For each round of the survey, it was decided to select 50 census blocks, with a systematic sample of 20 households being taken within each selected block. This gave a total sample of 1,000 households in each round, which was expected to provide two samples, each with about 4,000 persons aged 10 and over. In order to allow for analysis by island division, each island division was treated as a separate stratum, and the 50 census blocks in each round were distributed as follows: 14 to Tongatapu, 12 to Vava'u, 10 to Ha'apai, 8 to 'Eua, and 6 to the Niuas.

The sample design took account of the fact that the survey would be carried out in two rounds, about six months apart, so as to obtain some measure of seasonal variation in employment. It was also designed to ensure that, when data from the two rounds was combined, there would be sufficient households to allow separate analysis for each island division.

Some consideration was given to the possibility of reducing the workload in each census block to 10 households, with a corresponding increase in the number of workloads in each round to 100. This would have had the advantage of reducing the sampling error by spreading the survey over many more areas, but the idea was rejected, mainly because of the cost implications. Another important factor was the relatively small number of census blocks (444) in the country as a whole, and in some of the island divisions.

Fine tuning of the sample design

This sample design was slightly modified before implementation. The small island of 'Atataa was included in the frame, even though it had less than 50 households. This was done because it was close to the main island of Tongatapu and contained more than 40 households. In the event, it was not picked for the survey.

Because the Niuas are so small in terms of population size, it was decided that it would be more efficient to carry out a census of the islands there, rather than attempt to select a sample of six census blocks for each round. There is in fact a total of only 15 census blocks in the Niuas, so 12 of them would have been covered in the survey. Accordingly, the island of Niuafu'ou was covered in the first round of the survey, and the islands of Niuatoputapu and Tafahi (the latter containing only 26 households) were covered in the second round.

One other important activity had to be carried out before the sample could be selected. Some census blocks contain less than 20 households. If one of these census blocks was selected, it could not provide a full workload for an interviewer. Accordingly, efforts were made to group up small census blocks with neighbouring census blocks prior to the selection of the sample, so that this problem was avoided. Some flexibility, however, was allowed in this grouping work. If a census block contained just under 20 households, and could not be conveniently grouped with a neighbour, it was left by itself, but on the understanding that some slight reweighting of the household data would be required at the analysis stage, if that census block was selected and if it was found to still contain less than 20 households.

Once this grouping work had been done, the actual selection of the sample could begin. The census blocks in each division were placed in order according to their census block number, which provides an implicit geographical stratification because of the way the numbering has been organised. A cumulative total of the number of census households was then obtained, and the full sample of census blocks for the two rounds was systematically selected throughout the list, with probability proportional to size. Alternate census blocks were then allocated to the two rounds.

3. Implementation of the sample design

Whilst our focus here is on estimation issues arising from the sample design, it is also helpful to appreciate the context in which the sample design was implemented. Survey designs may well be good in theory, but be derailed by what happens in the field. Luckily in this case it did not happen, though one or two unforeseen events did take place.

The first round of fieldwork commenced during the last week of February 2003, but only on Tongatapu, so as to enable all field staff to become familiar with the survey instrument and with survey procedures. After this first week, those originally assigned to the outer islands moved on to do their work in the census blocks allocated to them. The interviewers travelled by the inter-island boat, while their supervisors took the plane. One supervisor was assigned to each division. In addition, five interviewers were assigned to Vava'u, five to Ha'apai, and three to 'Eua.

In the first round, fieldwork was scheduled to last three weeks in Vava'u, Ha'apai and the Niuaus, and two weeks on 'Eua. Some delays were experienced, particularly because of the poor weather at that time of the year. In the case of Niuafo'ou (in the Niuaus) only one supervisor had been assigned to complete this whole workload, and it proved impossible to complete the workload in time. Accordingly, the supervisor had to spend an extra week there. In the case of 'Eua, the field staff completed their work on time, but were forced to return to Tongatapu by plane, because the wharf at Nafanua Harbour had been damaged by Cyclone Eseta and no ferries were running.

By the time of the second round, field staff were much more familiar with the survey materials. Although four weeks was again allocated for the fieldwork, from 11 August to 9 September, it proved possible to complete almost all the fieldwork within three weeks. In the case of the Niuas, two field staff (a supervisor and an interviewer) were assigned to ensure that both Niuatoputapu island and Tafahi island could be covered completely in the time available. However, there was no boat available at that time to transport them to the Niuas, and the flight was already fully booked for several weeks ahead, so they worked initially on Tongatapu, and only got to the Niuas in September. Visits by boat to two remote islands in the Ha'apai division (Tungua and Nomuka) were also delayed until early September.

Throughout the fieldwork period there were several announcements in the press, and on radio and television, informing the public about the timing of the survey and its objectives. Respondents taking part in the survey, and members of the general public who were encountered by the survey teams, often mentioned that they had already heard about the survey. Advance letters were sent to all Town Officers in the selected areas, informing them of the survey and asking for their cooperation. These letters mentioned the particular date on which they could expect a visit from the field teams. The Town Officers, who are government servants, assisted the field teams in identifying the boundaries of the census blocks and in listing the households. The supervisor then drew the sample of households to be interviewed.

Implementation of the sample design seems to have worked well. Four cases occurred where a selected census block was a small one that had been linked with a neighbouring one, and in these situations both census blocks had to be listed together and a workload selected. There were three other cases where the original measures of size for selected census blocks were less than 20, and they had not been linked with other census blocks. Since the listing exercise showed that they still had less than 20 households, it was impossible to get full workloads, and adjustments had to be made at the analysis stage. Full workloads were also not possible in another three census blocks, even though their original measures of size were over 20.

Inevitably, one or two mistakes were made in the field. In one case, the interviewer visited the wrong census block. Where this happens, and the error is not corrected at the time, it is important to find out the original measure of size for the wrong census block, so that it can be compared with its present size. In another case, confusion arose when a selected block had been given one name, when in fact that name applied to an island in another location containing only one household. The confusion led to the interviewer going to a census block with a different number, but that workload was accepted. In the third case, the interviewer accidentally switched the order in which two census blocks were visited. One census block that should have been visited in the second round was visited in the first round, and vice versa. The workloads were accepted, despite the errors.

4. Estimation issues

At the processing stage, weights were calculated, to be applied to the selected households in each census block, and to the persons living there. In rough terms, the overall weights ranged from about 2 to 66 for a single round, and from 1 (for the Niuas) to 33 for the pooled data.

These weights took account of four different factors:

- **Omissions from sampling frame.** An allowance was made for the fact that some small islands had been omitted from the sampling frame. Each island division was given its own weight. 'Eua and the Niuas had a weight of 1.
- **Design weight.** This takes account of the different selection probabilities for households in each island division. For each census block there were two design weights, one for use in the analysis of data from a single round, and the other for use in analysis when the data from the two rounds were pooled. For the pooled data, the Niuas had a weight of 1.
- **Change in measure of size.** A weight was calculated to take account of the new information from the listing operation about the current size of the census block, compared with the size used for selecting the census block. This weight was particular to each census block.
- **Small blocks.** A weight was needed to compensate for the six cases where it proved impossible to select 20 households for interview in a selected census block, because the census block was found to contain fewer than 20 households.

Normally on a survey one would also need to apply a weight to adjust for *non-response*, but in the case of the 2003 Tonga LFS this correction factor was not needed, for the simple reason that there was no non-response at the household level. It was reported that every household selected for the survey co-operated in providing information. There was a very small amount of non-response of individuals within selected households, but it was decided it was not worth making special allowance for this, by having separate weights at the household and individual level.

5. General conclusions

The sample design, the selection of the sample, its implementation, and the process of estimation seem to have gone quite well, and no major problems were experienced. The sizes of the samples were sufficient to allow comparison of key employment characteristics at the national level between the two rounds. When the data from the two rounds was pooled, there were sufficient households to enable one to get separate estimates for each of the five administrative divisions of Tonga.

It will be noted that the weighting factors discussed earlier included an element to take account of those islands which had been left out of the sampling frame. These islands form part of the original target population - all households in Tonga - but they were purposely excluded from the survey population. In using a simple weighting factor within each division, we are assuming that the characteristics of people living on those islands that were specifically left out of the sampling frame are similar to the characteristics of those living on the other islands in that division that were included in the frame.

The same kind of argument would have been used, had it been necessary to apply correction factors to take account of non-response. In that case the weighting would be applied at the level of individual census block, and we would assume that the non-responding households in the census block had the same characteristics as those households that did respond.

With the help of a statistics adviser from the Secretariat of the Pacific Community (SPC), sampling errors were calculated for some of the key variables from the survey. This work indicated that, while the precision of most variables was more than adequate, the estimates for unemployment were subject to quite wide margins of error. This was particularly so in those divisions where the unemployed identified during the survey were concentrated in one or two census blocks.

Using the information about the sample design adopted for this survey, along with the new measures of size obtained during the listing exercise and details of household size, it is possible to arrive at estimates of the population for the country as a whole and for each division. The estimates obtained in this way turned out to be considerably lower - by about 10 percent - than the figures one would have expected on the basis of population projections. The reasons for this discrepancy are unclear, but no effort was made to adjust the estimates derived from the sample so that they matched up with the population projections. Instead, the figures were reported 'as is', but a cautionary note was added to the report.

Finally, the 2003 Tonga LFS provides a useful illustration of the way in which both census and survey methods can be used in a single statistical investigation. Where separate survey estimates are required for a number of geographical domains, it is usually efficient to divide the country into separate geographical strata and select samples of approximately equal size in each stratum. But some strata (such as the Niuaus in Tonga) may contain very few households, which would therefore require a very high rate of sampling in order to get a sample of sufficient size. Once the rate of sampling gets above a certain level, so that a high proportion of census blocks will be selected, it may be more efficient to carry out a census in the stratum, rather than a sample survey.
