

## Peru Enterprise Informal Survey 2010 Data Set

### 1. Introduction

This document provides information on the World Bank's Informal Enterprise Survey (IFS) collected in Peru from June 10, 2010 to July 20, 2010.

- 1.1. The World Bank's Informal Enterprise Surveys (IFS) collect data on non-registered business activities in every region of the world. The IFS are implemented in parallel to the World Bank's Enterprise Surveys (ES), which interview formal, private, non-agricultural firms in countries around the world ([www.enterprisesurveys.org](http://www.enterprisesurveys.org)).

The IFS use a standardized survey instrument designed to assess the business environment for non-registered businesses within a well-defined universe of activities, which have been identified using information from previous iterations of the studies. The IFS cover business environment topics including: general business characteristics, infrastructure, crime, sales & supplies, finance, labor, registration, business environment, and assets.

The IFS are conducted using a uniform sampling methodology in order to minimize measurement error and yield data that are comparable across the world's economies.

- 1.2. The Informal Surveys aim to accomplish the following objectives:
  - To provide information about the state of the private sector for informal businesses in client countries;
  - To generate information about the reasons of said informality;
  - To collect useful data for the research agenda on informality; and
  - To provide information on the level of activity in the informal sector of selected urban centers in each country

This report describes the survey population, the sampling methodology and the sample structure of the Peru informal survey; additionally, this report provides information that may be useful when using the data, including information on non-response rates and country-specific factors that may have affected survey implementation.

## 2. Survey Target Population

- 2.1. The primary sampling units of the IFS are non-registered business entities.<sup>1</sup> For consistency, “registration” is defined according to the established convention for the Enterprise Surveys in each country. In these surveys, the requirements for registration are defined on a country-by-country basis consulting information collected by Doing Business and information from the in-country contractors.
- 2.2. For Peru, informal firms were defined as those not registered with the Superintendencia Nacional de Administración Tributaria (SUNAT).
- 2.3. At the beginning of each survey, a screening procedure is conducted in order to identify eligible interviewees. At this point, a full description of all the activities of the business owner or manager is taken; based on its principal activity, a business is then classified in the manufacturing or services stratum using a list of activities developed from previous iterations of the survey. Certain activities are excluded such as strictly illegal activities (e.g., prostitution or drug trafficking) as well as individual activities that are forms of selling labor like domestic servants or windshield washers.

## 3. Sampling Methodology

- 3.1. In each country, the IFS are conducted in two selected urban centers, which are intended to coincide with the locations for the implementation of the main Enterprise Surveys. The overall number of interviews is pre-determined, and these interviews are distributed between the two urban centers, according to criteria such as the level of business activity and each urban center’s population, etc.
- 3.2. In Peru, the urban centers identified were Lima and Arequipa. The target sample for both urban centers was 240 interviews.
- 3.3. Sampling in the IFS is conducted within clearly delineated **sampling areas**, which are geographically determined divisions within each urban center. **Sampling areas** are defined at the beginning of fieldwork, and are delineated according to the concentration and geographical dispersion of informal business activity.
- 3.4. The number of **sampling areas**, and the geographical area they contain, is determined with the goal that each **sampling area** will yield four effective interviews.

---

<sup>1</sup> Within the economics literature the concept of the informal firm is defined in many different ways. At times, combinations of these definitions are used in academic and policy work. In order to avoid confusion and also facilitate the standardization of methodology and data comparability, the IFS equate informality with non-registration.

- 3.5. In Peru, each **sampling area** was designed to contain a physical area, on average, of no less than the equivalent of eight city blocks. These **sampling areas** may or may not correspond to the administrative districts of the urban center.
- 3.6. In both Lima and Arequipa, for a total of 240 interviews in each city, 60 **sampling areas** were identified ( $240/4 = 60$  sampling areas), respectively.
- 3.7. The placement of **sampling areas** in each city was determined, consulting local knowledge, using the criteria of the concentration of informal business activity and territorial dispersion.
- 3.8. In order to provide information on diverse aspects of the informal economy, the sample is designed to have equal proportions of services and manufacturing (50:50). These sectors are defined by responses provided by each informal business to a question on the business's main activity included in the screener portion of the questionnaire.

As a general rule, services must constitute an ongoing business enterprise and so exclude the sale of manual labor, as noted above. Manufacturing activity in the informal sector includes business activity requiring inputs and/or intermediate goods. Thus, for example, the processing of coffee, sugar, oil, dried fruit, or other processed foods is considered manufacturing, while the simple selling of these goods falls under services. If an informal business conducts a mixture of these activities, the business is considered under the manufacturing stratum.

- 3.9. Thus, each **sampling area** was designed with the goal of obtaining two interviews in services and two interviews in manufacturing.
- 3.10. In order to ensure a degree of geographical dispersion within each **sampling area**, two **starting points** were identified.
- 3.11. Each **starting point** was designed to correspond to four city blocks, which were numbered sequentially (see below). The first **starting point** was identified as Starting Point A and the second as Starting Point B.



3.12. Proceeding from each **starting point**, interviewers were instructed to begin on block 1, defining the starting block and corner. Each interviewer was instructed to attempt to achieve two interviews from each starting point, ideally one interview in manufacturing and one in services.

3.13. Interviewers were instructed to proceed clockwise around block 1 from Starting Point A; if the target interviews were not achieved, interviewers proceeded to block 2, Starting Point A, and so forth until completing a circuit of block 4.

After achieving two interviews from Starting Point A, interviewers were instructed to cease work in the blocks assigned to that given Sampling Point and repeat the same procedure in Starting Point B, beginning with block 5.

3.14. Using local knowledge, within each block all houses and shops were checked for unregistered businesses, following the pre-fixed route described above, until the allotted quota of interviews for the sampling area was reached.

3.15. Due to the sampling procedures described above as well as the geographical dispersion of the targeted informal businesses, detailed maps using GPS coordinates and Google maps were produced.

3.15.1. Each **sampling area**, including its two **starting points**, were delineated using Google maps, with the GPS coordinates of the starting points being systematically recorded.

3.15.2. Additionally, when obtaining a complete interview, the exact address of the informal business (or where the interview took place) was registered by the interviewer. Once in the office, this address was searched in Google maps, and its GPS coordinates were registered in a fieldwork report.

3.15.3. If no address was immediately available, using local knowledge, the GPS coordinates were determined using imaging via Google maps.

3.15.4. Documentation of this mapping procedure (if available) can be found in the accompanying appendices. In order to preserve confidentiality, the exact coordinates of businesses are not published.

#### 4. Sampling Structure

4.1. In **Peru**, 480 interviews were realized. The breakdown of the final sample was:

City	Interviews
AREQUIPA	239
LIMA	241
TOTAL	480

4.2. Moreover, the overall distribution by identified sectors is given in the following table (manufacturing sectors highlighted in grey):

<b>Sample distribution by activity sector</b>	Lima	Arequipa	Freq.
Manufacturing of clothes or shoes	43	61	104
Manufacturing of baked food (Baker)	3	10	13
Manufacturing of coffee, sugar, oil, dry fruits and other processed foods(exclude restaurants, street food-sellers,)	18	0	18
Manufacturing of handcrafts	4	11	15
Manufacturing of furniture	2	11	13
Manufacturing of metal products	16	11	27
Manufacturing of household items	5	1	6
Manufacturing of tools and instruments	1	0	1
Other manufacturing (not included above)	29	12	41
Selling food or Groceries (street food sellers, restaurants)	23	59	82
Selling of clothes or household items	12	2	14
Selling of business/computer/phone services	0	7	7
Selling of other goods	60	26	86
Transport Services	1	0	1
Cleaning and washing services	0	1	1
Hairdressers and barber shops	7	12	19
Professional Services (including internet services)	2	5	7
Repairing services	10	6	16
Other services (not included above)	5	4	9
<b>Total</b>	<b>241</b>	<b>239</b>	480

- 4.3. Due to issues of non-response (see below), in the process of fieldwork, the implementing contractor was unable to obtain the targeted four interviews in each of the originally delineated **sampling areas**.
- 4.4. As a result, replacement **sampling areas** were delineated, *ex post*.
- 4.5. In sum, there were 70 **sampling areas** (60 original, 10 replacement) in Arequipa and 72 zones in Lima (60 original, 12 replacement).

## 5. Non-response

- 5.1. The overall survey response rate among contacted, eligible businesses for the Peru IFS was estimated at 25%. The implementing contractor noted a number of factors that negatively affected response rates:
  - 5.1.1. Interviewers were occasionally confused with government officials (Sunat, Municipality, etc.); the implementing contractor suspected that this confusion lowered response rates.
  - 5.1.2. Some respondents were reluctant to give personal and/or sensitive business information.
  - 5.1.3. Locating the business owner was occasionally difficult, requiring several visits.

## 6. Database Structure

- 6.1. One version of the questionnaire was used for all interviews; variable *a0* indicates the module of the questionnaire. There are two levels of distinction in the informals data set: the data are distinguished by both the urban center/region and as a manufacturing or services business or activity, as detailed above.
- 6.2. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.
- 6.3. There are 2 establishment/business identifiers, *idstd* and *id*. The first is a global-unique identifier. The second is a country-unique identifier.
- 6.4. As noted above, there are two levels of distinction: industry (manufacturing or services) and region. Variable *a3a* denotes the region<sup>2</sup> (here: urban center) where the interviewed business is located. The variable *sect* indicates whether a business or activity was treated as manufacturer or service for the purposes of

---

<sup>2</sup> Term “region” used in order to maintain comparability with the formal Enterprise Surveys.

the interview. It is important to note that this variable is based upon variable *Sc2b*, which codes in detail the activity of the business or activity for the purposes of the questionnaire. *Sc2a* codes in detail the **main** product of the business or activity (in terms of sales value). As noted above, if an informal business or service has a main activity that is a service, yet has a manufacturing product, then the business is treated as a manufacturer. This treatment allows certain questions to be asked that only pertain to manufacturers (i.e., regarding inputs).

- 6.5. The screener information section (section Sc.) contains questions about the business activity and basic physical location of informal businesses; section B provides general information on the business and its ownership; section C discusses location and infrastructure; section I contains questions on crime; section D information on sales and supplies; section K is on finance; section L poses questions on labor; section R contains questions on registration; section M the business environment; and section N includes questions on business productivity.
- 6.6. All data are made available using Stata software. Due to restrictions of this software, variable labels are often shortened in the full data set. ***Please consult the full questionnaire for the complete text of questions and additional information.***

## **7. Country-specific implementation comments**

- 7.1. The implementing contractor noted the following relevant issues in Peru during fieldwork:
  - 7.1.1. Question Sc2 caused occasional confusion.
  - 7.1.2. Questions related to total business sales, presented difficulties for respondents; in many cases these questions had to be repeated multiple times.
  - 7.1.3. Respondents did not answer with precise information; they provided averages or were unwilling to answer certain questions.
  - 7.1.4. In instances where a business owner worked with a spouse, neither the owner nor his/her spouse was included on the payroll or was paid a fixed salary. In these cases a second visit was often necessary to obtain more precise answers.
  - 7.1.5. Respondents divided their attention between customers and the interviewer. In addition, sometimes respondents were too busy to answer questions and the interviewer had to set an appointment to complete the

interview. Often, re-visits were necessary for recovering parts of information.

- 7.1.6. The implementing contractor noted a fear of crime in various areas where informal businesses were located.
- 7.1.7. To avoid confusion with Control Entity Staff (Sunat, Municipality, etc.), interviewers identified themselves as staff from the implementing contractor with a visible photo ID and badge.