

Survey Solutions Training Projects

Sergiy Radyakin, Senior Economist, The World Bank

May 2018

Tasks:

1. Describe in a document (text or presentation format) your strategy for collecting this data.
2. How does the strategy affect the design of the questionnaire? What parts of the questionnaire are responsible for holding which information? Etc.
3. Which assumptions have you made if the outline of the task was ambiguous? Do you think these assumptions are realistic?
4. Decide on the structure of the questionnaire, which sections and sub-sections will it consist of? Which questions will you put and in what sequence? What are the conditions you need to add to establish a certain logic of the questionnaire?
5. Design the actual questionnaire. Test it.
6. Share the questionnaire with a colleague along with the description of your strategy and this description of the project. Get feedback, convince your colleague that your design corresponds to the strategy, or make necessary changes to address the feedback.
7. Prepare instructions for the person managing the input information for your survey.
8. Simulate the survey with some hypothetical data to obtain some observations.
9. Analyse the data you've collected in Stata, SPSS, R, or any other statistical package you are comfortable with to obtain the indicators mentioned in the assignment.
10. Compile a presentation showing the results of your work:
 - a. outline any of the assumptions you've made and how this affects the quality of data you collect or the process of data collection;
 - b. describe how the data was collected, which roles were practiced, etc;
 - c. show examples of input data and how the data is exported from Survey Solutions;
 - d. show your analytical results: any tables, charts that you've produced.
 - e. Outline any challenges or problems you've encountered in the process, and how you've solved them.
 - f. Draw conclusions from this exercise. How will this affect your work or research.

Project: Library Inventory

Background

In 2015 the Government of Utopia has implemented an initiative in which 6,000 community libraries were set up based on the initial endowment of books. You are now tasked to design a questionnaire and prepare a description of a survey for checking in 2018 the status of these libraries and books originally supplied to them.

Different libraries have received different number (from 120 to 180) and varieties of books. You were told that the implementing ministry has retained an Excel file with two sheets, one listing the libraries with their attributes and another outlining the titles of the books, along with some information about them, which were sent to the libraries. While you don't have access to this Excel file at this moment, the following description was communicated:

These are the attributes of a **library**:

- Identification number
- Province
 - North
 - South
 - East
 - West
 - Center
 - Highlands
- Address
- Contact phone number

These are the attributes of a **book**:

- Title
- Author(s)
- Issue year
- Number of pages
- Book condition in 2015:
 - mint (as new)
 - used
 - heavily used
 - damaged.
- Kind of damage if a book is damaged
 - pages missing,
 - writing on pages,
 - water damage,
 - binding broken, etc

During the time between 2015 and 2018 the following changes could occur:

1. A community library has closed;
2. A book in the library has gone missing / is lost;
3. A book in the library has changed it's condition, for example from 'used' to 'damaged'.
4. The community library may have acquired additional books since it's opening.

Due to the rush to implement the program not all of the attributes were accurately recorded in 2015, but you know that the libraries didn't receive identical copies of books, and that the title and the issue year were always accurately recorded (uniquely identify a book in the library). You want to fill in the gaps by acquiring the missing information and correcting inaccurate records.

For the books that were acquired by the library during its operation, you want to collect all of the book's attributes, and in addition the year and the month when it was acquired and the method of acquisition (purchased or donated), and price paid (only if the book was purchased).

Optional Tasks:

- I. The personnel of the library can restore books (for example, change from 'damaged' to 'used'), but can't make its condition 'mint (new)'. Permit improvements in the condition of the book, but don't permit 'mint (new)' condition unless the book was in this condition in 2015.
- II. you are concerned that the enumerators will not have attention to details and may confirm the book is still in the library even though it is missing. How can you address this concern in the questionnaire design and survey strategy?
- III. You received additional information that all the books in mint (new) condition in 2015 were accurately recorded in 2015 based on the purchase records. How will you use this information?
- IV. Consider that each book that is in the library can be either physically there or on loan to a user, in which case you want to collect the date when the book was last loaned to the user. The loan date may not be in the future and may not be more than 90 days earlier than the date when the library is surveyed. (Library rules stipulate that books not returned within 90 days of the loan date should be considered lost). Note, that different libraries will be surveyed on different dates.

Data processing:

Based on the collected data you need to compute the following indicators:

- 1) Percentage of libraries still in operation (in total and by province);
- 2) Number of books (total and by province) :
 - a) originally distributed;
 - b) currently still in libraries; (and on loan if available);
 - c) lost;
 - d) acquired (break down by mode of acquisition);
 - e) amount paid for purchasing books;
- 3) Percentage of books in each condition;
- 4) Percentage of books the condition of which has:
 - a) Improved;
 - b) Worsened;
 - c) Didn't change.
- 5) Number of books for which the information missing originally was collected; number of books for which information originally contained, but was revised.

Project: Hotel feedback

You are hired by a large international hotel chain (includes 330 hotels in 25 countries) to design a survey and a corresponding questionnaire for the feedback collection from guests about their stay.

Since the guests have left the hotels you are tasked to do this survey over the Internet using Survey Solutions' CAWI functionality.

You are promised by the management of the hotels that the following information will be supplied to you:

- Guest name (name and surname);
- Guest email;
- Hotel name, where the guest has stayed last;
- Date of last check in;
- Date of last check out;
- Total number of visits during the year 2017.

The management is interested in finding out:

- Home country of each tourist;
- Whether the last visit was for business, tourism, connection stop, or some other purpose;
- Level of satisfaction (scale 1..5) of the guest with each of the following: hotel amenities, cleanness, responsiveness of staff, price paid, and overall satisfaction;
- Whether the guest has made a reservation himself or used a travel agency, and if so, what was the travel agency;
- If the guest has had more than 3 stays in the hotel in 2017 and was not satisfied (answered 1 or 2 overall satisfaction) with the last stay ask for a phone number, and convenient date and time for a follow up call.
- If the guest has stayed (last stay) for more than one day, inquire if the guest has used the swimming pool.
- If the guest has stayed for business purposes, ask if the shared office was used, and whether the internet speed there was sufficient. Skip these questions if the purpose of stay was other than business.
- Whether the guest would like to recommend the hotel to his friends, relatives, colleagues at work.

Data processing:

Based on the collected data you need to compute the following indicators:

1. What is the average level of satisfaction overall, and with particular points of interest (amenities, cleanness, etc);
2. Pie chart of main reason for staying (business, tourism, ..)
3. Hotel managers hypothesize that people who stay in the hotel longer (more visits) are more satisfied with the hotel. What results can you put on their table to help support or dismantle this belief?
4. Tabulate the frequency and percent of guests by home country, by region (decide on a suitable grouping of countries in regions);
5. For each of the other indicators that the management was interested in decide on the suitable presentation (a number, table, graph or chart) and write the commands to construct them.
6. Combine your code statements into a single script and your output into a single document or presentation.

Project: land ownership survey.

The Government of Utopia tasks you to prepare a survey in 2019 to determine the accuracy of the land ownership records. For that a sample of 6,000 plots are randomly picked by a computer from the existing land cadastre and it has been checked already that no two plots in the sample belong to the same person.

The strategy of the survey is to send the interviewers to the plot owners and verify the record for the plot in the sample, as well as collect all the other plots owned by that same respondent to be checked later in the office.

The following are the attributes of plots:

- Plot ID, an identifier of the fixed width with format: XXX-NNNNNN-YYYY-NN, where XXX are any letters from A to F range (so each X can be either A or B or C or D or E or F), N denotes a digit, and YYYY is a 4 digit combination of digits representing a year. For example: CAF-223319-1997-02;
- Plot registration office (any of the 69 offices across 8 provinces of the country);
- Plot ownership (sole ownership, shared ownership, etc) and owner name;
- Plot area;
- Plot soil type;
- Plot use type (agricultural use, residential settlement, commercial use, recreational use, wildlife reserve, etc):
 - Registered (as recorded in the registration documents);
 - Actual (as being actually used by the owner, which may deviate from the registered).

In the plot ID the YYYY year is the year of last change of registration, and last -NN is the number of transaction within that year. So the XXX-NNNNNN part of the identifier uniquely defines the physical plot. When collecting the information on other plots that are owned by the respondent, make sure there are no duplicates in this part of the plot ID.

Actual use is not contained in the registration records, but this status needs to be acquired for a secondary project. In addition, if the actual use type is different from registered use type, collect the explanation of the plot owner why the plot is different.

While the cadastre records contain the plot area always in hectares, the respondents are sometimes reporting the area of plots in ares, or square meters. Allow for this possibility for the collection of plot measurements and standardize each plot size to be expressed in hectares.

The cadastre records may be not accurate in details (individual attributes of the plot, such as size), but also in actualization (the whole plot no longer belongs or never belonged to the person). How do you reflect these possibilities in the questionnaire?

The land cadastre has started its operation in 1973 and was on temporary shutdown from March 1, 1998 to October 17, 2001 due to political instability in the country. No registrations of plots were done during that period.

Optional

It is critical that at least 6,000 respondents are interviewed with at least one plot status collected from each to judge on accuracy of the cadastre. While registration of a plot in the cadastre is mandatory, the participation in the survey is not. Describe the consequences of the refusal of the plot owner to be interviewed in your strategy and the necessary actions of the survey team.

Indicators

The survey you prepare for must enable computing the following indicators:

1. Percent of plots from the original sample, that are still owned by the indicated owner;
2. Percent of plots from the sample of plots still owned by the indicated owner, that have their attributes correctly recorded (all attributes together, and individually by attribute).
3. Percentage of all plots for which the survey data has been obtained, which have exact matching attributes in the cadastre.
4. Average number of plots per land owner, over the whole country and by province.
5. Average land holding per land owner, over the whole country and by province.
6. Pie chart and proportions table for plot use (have side-by-side comparisons for actual use and use declared during registration).
7. Percentage (present as a table and a pie chart) of land owners (all plots) which fall into each of the following groups:
 - a. Tiny (0 ... 0.5 hectare);
 - b. Small (0.5 ... 1 hectare);
 - c. Medium (1 ... 5 hectares);
 - d. Large (5 hectares or more).
8. Show how the plot size has changed over time of registration (it is believed that over time the plots became smaller as there are more transactions both due to volatility of the market and larger population).

Project: Survey of Water Quality

The Ministry of Water Supply (MWS) of Utopia has recently raised concerns about the quality of water consumed by the residents as there has been an increasing number of complains on the quality of water supplied with the pipes and more and more citizens are switching to alternatives (wells, streams, etc).

The MWS has already taken samples of water from a handful of households with inconclusive results. In order to get a more solid base the number of samples must be increased and samples taken at regular intervals (every 4 hours for a period of 72 hours). Unfortunately, the budget of the project can't afford holding an investigator in the household for 72 hours, so you will have to rely on the cooperation of the household members to collect the water samples, and you will plan 2-visits survey. The samples can't be tested in the household or by the household members, and all 18 samples will have to be delivered to the lab for testing.

In addition to resolving the quality of water issue, the Government also wants to understand the water supply profile of the population:

- What are the sources of water utilized?
- Is one or multiple sources utilized by households?
- What sources of water are used for drinking/cooking?
- Do sources vary by season? Depend on some other factors?
- What are the costs associated with water supply? (monetary, non-monetary - such as time to deliver)
- What is the reliability of these sources?
- Who (if anyone) brings the water home (children / adult males / adult females)?
- What are the problems with each water supply used and (if known) for each water supply not used (water taste, smell, color, particles, etc).

In addition, there is also interest in treatment methods utilized by consumers to improve the quality of water and their effectiveness.

Project: Survey of School Dropouts.

The Government of Utopia is about to introduce incentives for 16-18 year olds to remain in school and complete high school. It is hypothesized that a quarterly monetary subsidy would help as an incentive, but the size of the subsidy is being debated between \$300 and \$800. In order to estimate the effectiveness of this measure you are asked to include a contingency valuation module into the survey to help establish which minimal amount could induce the necessary response.

You have received a sample of 3,200 18-year olds who have dropped out of school within the last school year. You are to prepare a report answering the following questions:

- 1) Whether the decision to drop out was a decision of the pupil or impact of external factors, and if so, which ones?
- 2) Whether the dropout is planning to return back to schooling at some point later or not; if yes, when? If not, why?
- 3) What is the current profile of the dropouts? Are they working at home / hired / doing nothing / etc. ?
- 4) What could be a change that would reverse their decision to drop out and result in them going back to school to finish their education?
- 5) Family profile - looking at the parents and siblings of the dropout, have they attended school? Completed school? And which level? If dropped out, when? And why?

Project: Non-Standard Units of Measurement

The household consumption survey of Utopia will involve a collection of data on the consumption of approximately 300 household consumption items, of which approx 90 are food items, such as apples, bread, corn, dairy products, etc.

Since the large proportion of trade happens through the farmer's markets, you anticipate that the reported amounts of purchase and consumption will be expressed in non-standard units (buckets, heaps, bunches, etc) rather than in standard kilograms and litres.

In preparation for the household consumption survey you are developing a supplementary survey, the purpose of which is to collect the values of the non-standard units prevailing in 5 regions of Utopia: Northern, Southern, Eastern, Western and Central. The data collected as part of the supplementary survey will be utilized to calculate conversion factors from the non-standard units to standard ones during the subsequent household consumption survey.

Your office is seeking answers to the following questions:

- Which non-standard units of measurement are applicable to each consumption item?
- What is the conversion factor of the non-standard units to standard ones?
- How does it vary by region?

Decide on what will be an "assignment" in this survey.

Decide on the protocol of the quality control.

Refer to this publication for a discussion of practical issues related to consumption measurements in non-standard units:

http://siteresources.worldbank.org/INTLSMS/Resources/3358986-1423600559701/LSMS_GB_NSU_Web.pdf

Project: Fuel Price Survey

The Ministry of Transportation of Utopia has hired you to consult on the organization of the High Frequency Fuel Price Survey. The objective is to have an up to date information about the prices of various types of fuels sold around the country.

The following **types of fuel** have been selected for monitoring:

- Diesel fuel;
- Biodiesel fuel;
- Liquefied Natural Gas;
- Super unleaded petrol;
- Premium unleaded petrol.

For biodiesel fuel, it is essential to identify the **source**:

- soybean oil,
- yellow grease,
- corn oil,
- canola oil, or
- animal fats.

The data is to be collected at the refuelling stations, which advertise their prices by posting them on displays. Not all the stations sell all fuel types, and some experience periodic shortages of one, some, or all fuel types. Refuelling stations may be located within cities, in suburban settings, or on major roads and expressways. Some of them belong to major concerns while others work independently. Their sources of fuel also differ, ranging from local refineries to imported fuel. Some stations provide full service while others are self-service.

You expect the price of the biodiesel to be always higher than the price of diesel fuel. In case it turns out to be lower, it is most likely a promotional action, and you need to capture the conditions of that promotion.

Optional:

In a high frequency survey, you can utilize the earlier collected data as a reference. How would you utilize them for the validation purposes?

Indicators:

- Average price in LCU of 1 litre of each fuel type in the country, by region, province, district;
- Average supply (in days) of each fuel type (in the country and by region, province, district) at the stations surveyed;
- Proportion of the stations that experience shortages in fuel supply (by fuel type and geography);
- Average duration of shortage in days for stations that experience shortages.
- Average price difference between independent and concern-owned stations (by fuel type). Fuel sold on stations within 15km distance from the Capital Central Square is subject to an additional 3LCU levy (fixed). Take this into account when calculating the price difference.
- Which stations are more likely to experience shortages, ones that use local suppliers or those that import their fuel?