STANDARDISATION AND SYSTEMATISATION OF RESPONSE RATE CALCULATION.

EUSKAL ESTATISTIKA ERAKUNDEA
BASQUE STATISTICS OFFICE

Donostia-San Sebastián, 1
01010 VITORIA-GASTEIZ
Tel.: 945 01 75 00
Fax.: 945 01 75 01
E-mail: eustat@eustat.es
www.eustat.es

Presentation

This document sets out the research conducted by Eustat in 2006 into non-response in household surveys. This is part of a project that continues in 2007 and where the most general objective is to implement actions to deal with non-response.

The work carried out comes under one of the works of the 2005-2008 Basque Statistics Plan, regarding statistical methods R&D&I, and which is aimed at researching and applying new mathematical-statistical methodologies in statistical surveys. Excellence management is the spirit underlying these studies and is defined as a continuous management process.

The inclusion of the non-response study in the Plan works reflects the concern that EUSTAT shares with other international statistical offices for this issue. The non-response is a phenomenon that occurs to a greater or lesser extent in many statistical surveys, that may sometimes detract from the reliability and always makes the result estimates less accurate.

EUSTAT has always monitored the non-response of household surveys. However, the need to establish common measurement criteria or standards has been raised, based on the European methodology published for that purpose by the Institute for Social & Economic Research, at the University of Essex (United Kingdom).

I hope that this report is useful for all those parties interested in this sphere of statistics.

Vitoria-Gasteiz, May 2007
Josu Iradi Arrieta
General Director

SUMMARY

This document[1] is divided into the following chapters:

The first chapter sets out the objectives and the methodology that is going to be applied in the work document. The methodology refers to the definitions and the data processing that is going to be used.

The second chapter considers how each of the outcomes have been categorised and establishes the standard definitions to be used to calculate the rates.

The definitions of the different response rates are set out and explained in the following point: response rate, co-operation rate, contact rate, refusal rate and eligibility rate. Each of the formulas are reflected with their relevant explanation.

Chapter Four sets out the results obtained for the Population in Relation to Activity Survey (PRA), using the codification and definition of the response rates proposed by the ISER.

Chapter Five contains the conclusions reached and the last point sets out the bibliography used.

[1] EUSTAT wishes to thank Susana Sanz for her excellent research work within the framework of the EUSTAT Mathematical-Statistical Methodology Research Grant , specifically in the field of Non-Response

Contents

Tables

Graphics

Chapter
1

1. Introduction

Introduction and goals

This document aims to set out a systemisation method to measure the non-response in individual and family statistical surveys.

This document also considers an application of this method for a specific household survey, the Population in Relation to Activity Survey (PRA).

First of all, the response rates need to be highlighted as one of the fundamental quality indicators of the surveys. And after analysing them, its purpose is to adopt the relevant measures to improve the survey.

The importance of knowing the response rate can fundamentally be specified in the following aspects:

- insofar as the non-response may introduce a measurement error or bias, due to the fact that the non-respondents may differ from the respondents in significant characteristics. And therefore, great accuracy does not guarantee error absence .

- insofar as non-response reduces the accuracy of the estimations, as there will be less cases available for the analysis.

This document proposes a method to measure the non-response, the one put forward by the Institute for Social & Economic Research (ISER) in 2001. There are specifically two advantages to this method:

- first of all, it proposes a full codification of all outcomes, or situations where a survey may finish at the end of the collection task. This is particularly useful to classify outcomes from different surveys in the same way: refusals, non-contacts, non-surveyable, etc.

- secondly, it puts forwards the formulas, in accordance with the above codification, for the response, co-operation, contact, refusal and eligibility rates.

The approach set out in this paper comprises adopting a non-response rate calculation model that can be used for the various household surveys conducted by EUSTAT, systematize it and obtain comparable results in the different surveys and in time. The fundamental objective would therefore be to establish the quality of the surveys in this sense, report on it and act to improve the survey.

The known standards to calculate the response rates are those of the aforementioned ones of the ISER, together with those of the American Association for Public Opinion Research (AAPOR) in 2000.

The AAPOR, in the same way as the ISER, sets out a classification method for outcomes (different collection methods, in this case) and rate obtaining formulas.

There are very small differences between both models. According to the ISER authors, they attempt to extend the model put forward by the AAPOR for face-to-face surveys, to United Kingdom standards, and in particular to public entity surveys. They have promulgated the use of this standard to facilitate comparisons between surveys, understanding of trends and response rate guidelines.

The much earlier attempt by another US organisation, the Council of American Survey Research Organisations (CASRO), on which the AAPOR method was based, is usually also highlighted in the bibliography about non-response and standards. That council appointed a work group in 1979 to make recommendations about response measurements as quality indicators in the surveys. In 1982, they published a report with formulas to calculate the response rate, some definitions for the appropriate application of the formula and recommendations in general.

Some entities have published response rate calculation methods. Special mention should be made of the review by the US Federal Committee on Statistical Methodology of the different calculation possibilities, associated to the surveys that use them.

Yet it should also be pointed out that the formulas that are applied in the response statistical operations are often not explicitly published. This rather complicates the comparison of results between surveys.

Some authors have expressed concern or have openly criticized this situation where there is no agreement when using the terms. In the European sphere, the ISER has conducted a general review of the surveys and, based on various studies, concluded that the formulas and definitions are not sufficiently precise, which is particularly true in the case of Labour Force Surveys.

Criticism of this type is also to be found in America. Tom W. Smith, in his article *Developing Nonresponse Standards*, reviews the non-response definitions and standards in various fields of social research, in the United States, preferably: the academic world, commercial research and federal offices. He concludes that emphasis has been placed on standardisation, in the form of definitions and recommendations, but the organisations in general do not put them into practice, that the response rates are not usually published, and when they are published, they are not often defined or the same definitions are not used.

Chapter
2

2. Categorisation of the results (Outcomes)

In this chapter, we firstly explain how each of the outcomes have been categorised and then set out the standard definitions to be used to calculate the response rates.

The following categories are applied to the household face-to-face household surveys. By household surveys, we refer to surveys where the household is the sampling unit, as in the PRA.

The proposed categorisation of the results is set out as a 3-level hierarchical framework. The first two must be used as described herein for all the surveys. The use of the third level is optional. In other words, the codes shall consist of a standard two-level code and a third level digit. The categories are numbered with three digit codes, with each digit representing a hierarchical level.

The first level categories would be:

1. Complete survey

2. Partial survey

3. No contact

4. Refusal

5. Other non-response type

6. Unknown eligibility

7. Not eligible

ISER notation used

Once the outcomes are codified using the ISER codification, the following standard definitions, which will be used to calculate the rates, are applied. The number in brackets refers to the categories defined above.

I = Complete interview (1)

P = Partial interview (2)

NC = No contact (3)

R = Refusal (4)

O = Other non-response type (5)

UC = Unknown eligibility, contact (641, 651, 661 and part of 67). (Reference: Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys)

UN = Unknown eligibility, no contact (61, 62, 63, 642, 652, 662, 68 and part of 67). (Reference**:** Recommended Standard Final Outcome Categories and Standard Definitions of Response Rate for Social Surveys)

NE = Not eligible (7)

Ec = Estimated proportion of the contacted cases of unknown eligibility that are eligible

En = Estimated proportion of the non-contacted cases of unknown eligibility that are eligible

RR = Response Rate

COOP = Co-operation Rate

CON = Contact Rate

REF = Refusal Rate

ELIG = Eligibility Rate

Chapter
3

3. Definition of the response rates

The definitions of the different response rates are set out below:

Response rate: The response rate indicates the proportion of achieved interviews in all the eligible cases. It is calculated by dividing the sum of all the complete and partial interviews by the sum of all the complete and partial interviews, plus the refusals, non-contacts, other response type and unknown eligibility.

Numerator = Achieved Interviews

Denominator = Eligible households

$$RR_O = \frac{I + P}{(I + P) + (R + NC + O) + e_C UC + e_N UN)}$$

When estimating *en*, i.e., the estimated proportion of non-contacted cases of unknown eligibility that are eligible, one has to be guided by the best available objective information and a proportion must not be selected to boost the response rate.

It would be appropriate for some surveys to assume that the proportion of eligible cases among those cases where the eligibility is unknown is the same as among the cases where the eligibility has been established. For other surveys, it will be appropriate to assume e=1. In the PRA, e=1 is taken, which would be the most unfavourable case.

Co-operation rate The co-operation rate indicates the proportion of the number of achieved interviews of those cases that were contacted during the field work. It is calculated by dividing the sum of all the complete and partial interviews by the sum of all the complete and partial interviews, plus the refusals, other non-response type and contacted cases of unknown eligibility.

Numerator = Achieved Interviews

Denominator = Contacted households

$$COOP = \frac{I + P}{(I + P) + R + O + e_C(UC)}$$

Contact rate The contact rate measures the proportion of all the cases where a member of the household was contacted by the interviewer, even though they subsequent refused to answer or they were unable to give any type of information. It is calculated by dividing the sum of all the complete and partial interviews, plus the refusals, other non-response type and contacted cases of unknown eligibility by the sum of all the complete and partial interviews, plus the refusals, non contacts, other non-response type and unknown eligibility units.

Numerator = Contacted households

Denominator = Eligible households

$$CON = \frac{(I + P) + R + O + e_C(UC)}{(I + P) + (R + NC + O) + e_C UC + e_N UN)}$$

Refusal rate: The refusal rate indicates the proportion of all the estimated eligible cases that refused to answer. It is calculated by dividing the refusals by the sum of all the complete and partial interviews, plus the refusals, non-contacts, other response type and unknown eligibility.

Numerator = Household that refuse to answer

Denominator = Eligible households

$$REF = \frac{R}{(I+P)+(R+NC+O)+e_C UC + e_N UN)}$$

Eligibility rate The eligibility rate is defined as the number of eligible cases among the total cases.

Numerator = Eligible households

Denominator = Total households (eligible + non-eligible)

$$ELIG = \frac{(I+P)+(R+NC+O)+e_C UC + e_N UN}{(I+P)+(R+NC+O)+(UC+UN)+NE}$$

Chapter
4

4. Methodology Applications

The above definitions have been applied to the sample of the Population in Relation to Activity Survey (PRA).

The PRA is a longitudinal survey that is conducted every three months. The sample consists of 5088 dwellings in each of the different surveying quarters. Each dwelling remains in the panel for 8 rotation waves, which is equivalent to 2 years, after which the dwelling is removed from the panel. The sample is therefore renewed by an 1/8 each quarter.

1. Population in Relation to Activity Survey (PRA)

As is explained in Chapter 2, each one of the outcomes needs to be categorised before calculating the response rates. Once the outcomes have been codified, the standard definitions are applied (Complete interview, No contact...) which we will then used to calculate the rates.

In the case of the PRA, as it is a longitudinal survey and in order to be able to study the evolution of the rates over the different quarters, different types of response rates have been calculated. First of all, the response rates are analyse per quarter, then by wave, and finally, by quarter and wave simultaneously.

By applying the ISER-proposed categorisation, the left-column of the following table sets out the contact form outcome codes that are used for the PRA and these codes with the three-digit ISER codification in the right hand column.

Contact form Codes		Three-digit ISER codification	
00	Affirmative	**13**	Complete survey
01	Not located.	**63**	Unknown eligibility. Not located.
02	Not surveyable	**76**	Not eligible. Collective Institution/Establishment
03	Disappeared	**7**	Not eligible. Disappeared
04	Alterations	**7**	Not eligible. Alterations
05	In ruins	**72**	Not eligible. In ruins
06	Construction	**71**	Not eligible. Construction
07	Inaccessible	**62**	Unknown eligibility. Inaccessible
08	Temporary dwelling	**75**	Not eligible. Temporary dwelling
09	Other purposes	**74**	Not eligible. Non-residential dwelling/for other purposes
10	Uninhabited, empty	**73**	Not eligible. Uninhabited, empty
11	Refusal	**432**	Refusal
12	Holidays	**52**	Other non-response type. Holidays
13	Prolonged absence (other reason)	**31**	No contact. Extended absence, no contact with nobody in the household
14	Already surveyed	**563**	Other non-response type.
15	Recovered (and lost)		
19	Other outcomes	**563**	Other non-response type.

The tables of the different response rates obtained quarterly, by wave, and quarterly and by wave simultaneously are set out and analysed below. These are automatically calculated using the new programme in SAS. Each of the tables are analysed in the following points.

Quarterly response rates

The first table sets out the different response rates for the 8 first quarters when the PRA has been performed so far, along with the mean of the 8 quarters.

T.1 QUARTERLY RESPONSE RATES

QUARTER	20044	20051	20052	20053	20054	20061	20062	20063	AVERAGE
RESPONSE RATE	83,9	82,1	82,6	82,8	83,8	85,1	86,0	86,9	84,2
COOP. RATE	90,0	88,9	89,2	89,4	89,9	90,9	91,6	92,2	90,3
CONTACT RATE	93,2	92,4	92,6	92,7	93,2	93,7	93,9	94,2	93,2
REFUSAL RATE	7,8	8,5	8,2	7,8	7,2	6,4	5,7	4,7	6,9
ELIG. RATE	92,3	91,6	91,5	91,0	90,9	90,6	90,6	91,1	91,0

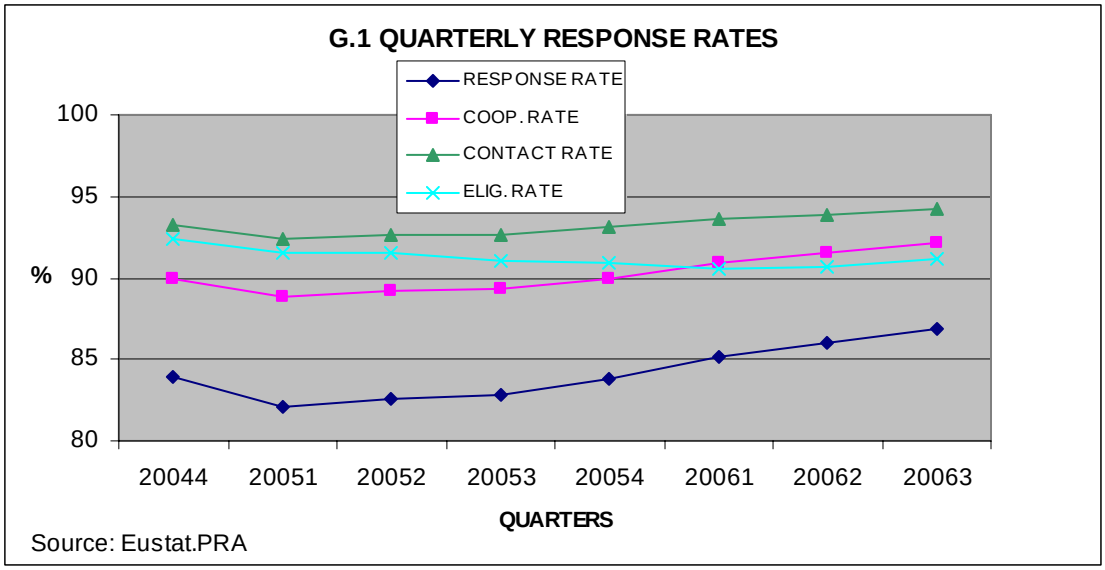
Source: Eustat. PRA

As can be seen in Table 1, the response rate is nearly 84% in the first quarter and then falls 2 points in the second quarter. From then onwards, a slight improvement is noted in the third and fourth quarter, when it reaches nearly 83%. From the fifth surveying quarter onwards, the response rate begins to increase more notably, on the basis of one point or more in the following quarters. The rate for the last quarter is 86.9%, which is proof of the upwards evolution of the rate over the quarters.

It should be pointed out that in the first quarter, as the whole sample is new, is when more non-response outcomes occur and these outcomes are accumulated to those of the renewal housing in the second quarter, which is why the response rate drops.

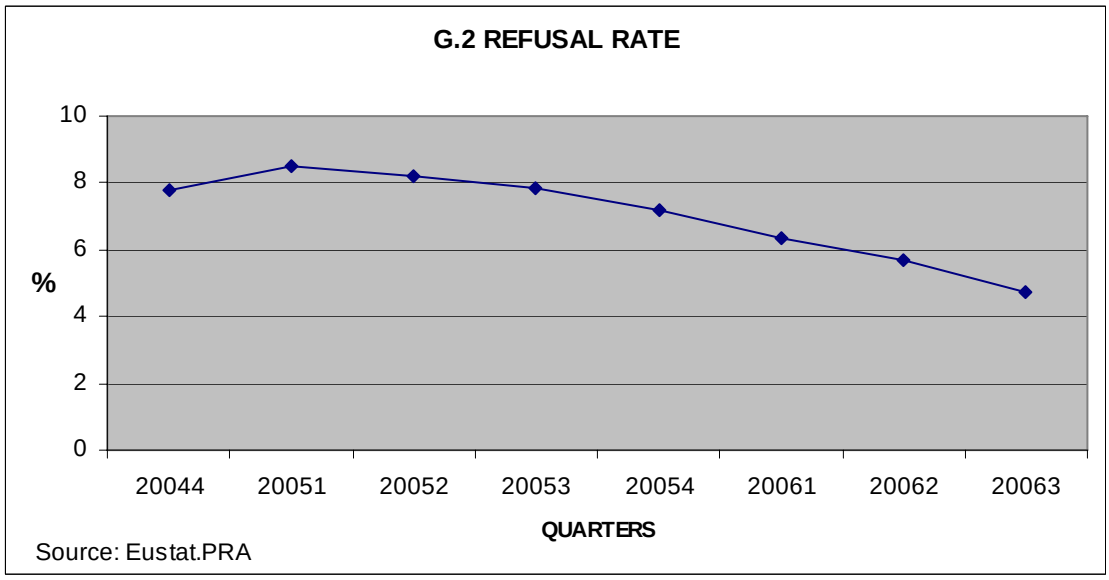
The impact of the outcome number of the first quarter will decrease as those dwellings are removed from the panel throughout the 8 waves. In fact, as can be observed, the rate is increased from the second quarter onwards, which demonstrates, on the one hand, that the first-quarter dwellings that presented a large number of outcomes are gradually removed from the panel, and on the other hand, that the new dwellings entering the panel show a lower number of outcomes.

The following graph shows the evolution of the response rates over 8 quarters.



Source: Eustat.PRA

The following graph shows the evolution of the refusal rate.



Source: Eustat.PRA

As can be observed, practically the same occurs with the refusal, co-operation and contact rates as with the response rate, with the difference that the refusal rate drops quarter by quarter (except in the second quarter).

As far as the eligibility rate is concerned, no great variation is observed over the quarters. It ranges between 92% and 90% in all quarters and no great conclusions may be reached, as the extraction procedure of the dwellings from the directory is always performed in the same way and therefore, the differences are not conclusives.

Response rates by wave

The response rates by wave are analysed below. By wave we refer to the order number of the interview, i.e., to the number of visits made to that dwelling.

The following tables set out these rates for the 8 waves up to the time of the report.

T.2 RESPONSE RATES BY WAVE

WAVE	TOTAL DWELLINGS	RRO	COOP	CON	REF	ELIG
1	9534	86,8	92,1	94,3	5,5	92,6
2	8262	85,3	91,1	93,6	6,4	92,0
3	6990	84,6	90,6	93,4	6,8	91,2
4	5717	83,0	89,5	92,8	7,6	90,7
5	4450	82,2	88,7	92,7	8,2	90,0
6	3180	80,9	87,8	92,2	9,1	89,4
7	1908	79,3	86,6	91,5	10,2	88,3
8	637	77,1	86,6	89,0	10,0	89,8

Source: Eustat. PRA

RRO=Response Rate
COOP=Co-operation rate
CON=Contact rate
REF=Refusal rate
ELIG=Eligibility rate

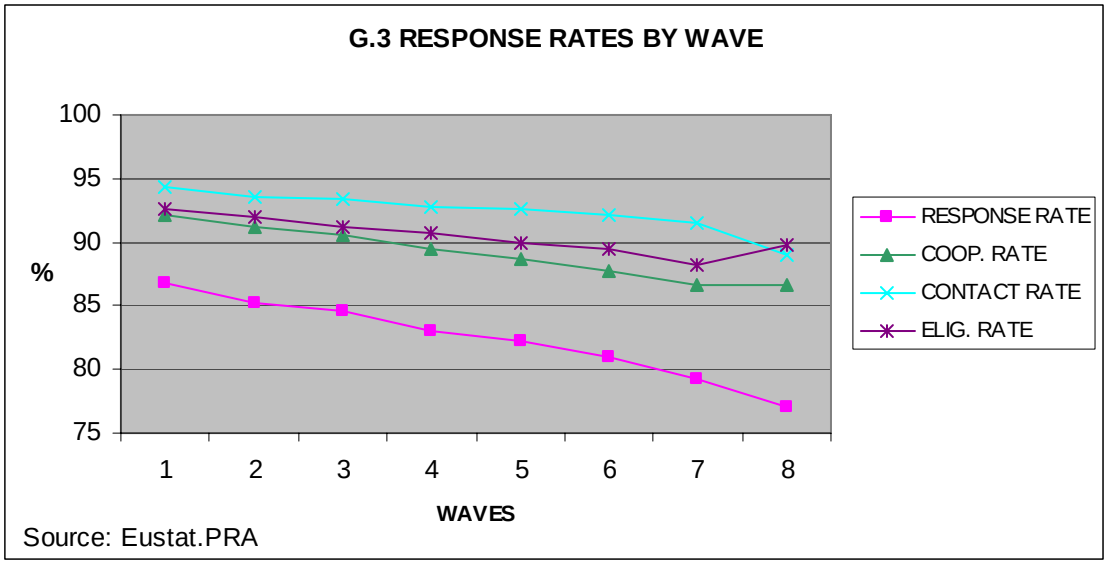
As can be seen in Table 2, the response rate decreases from one wave to another. The response rate for the first wave is 86.8% and decreases over the waves, until it reaches 77.15 in the last wave.

On the other hand, it should be pointed out that the number of dwellings always decreases from one wave to another as an eighth of the dwellings are removed from the panel in each wave.

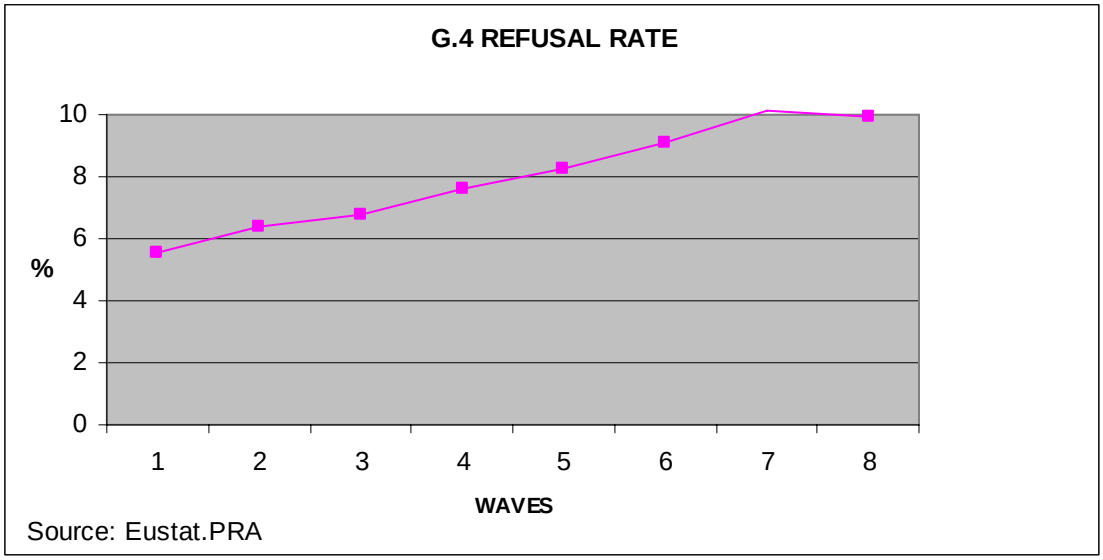
On the other hand, the decrease in the response rate from one wave to another is due to the fact that once the dwellings are surveyed again, the state of the dwellings may change for example, from surveyed to refusal, but not the reverse. Therefore, surveyed dwellings are always "lost" from one wave to another and the response rate therefore drops.

It should be pointed out that the eighth wave only contains the dwellings that were surveyed in the 20044 quarter and more time is therefore necessary for this rate to be more representative.

Graphs 3 and 4 show the evolution of response rate by wave and the refusal rate by wave respectively.



Source: Eustat.PRA



Source: Eustat.PRA

Response rates by quarter and wave

In this last table, we are going to comment response rates by quarter and wave. The following table can be interpreted by columns and diagonally. Diagonally, the evolution of the dwellings per quarter and by wave can be observed. The column interpretation allows us to compare all the first waves of each of the quarters, the second waves of each of the quarters, etc.

T3. RESPONSE RATES BY QUARTER AND WAVE

QUARTER	1	2	3	4	5	6	7	8
20044	83,9							
20051	87,6	81,4						
20052	91,8	86,0	80,4					
20053	90,8	91,0	85,7	79,0				
20054	91,2	89,9	89,9	84,3	78,6			
20061	93,8	90,9	89,6	89,5	83,5	77,6		
20062	88,3	93,1	89,9	88,8	89,0	83,0	77,5	
20063	87,1	88,1	92,6	89,6	88,6	88,7	82,8	77,1

Source: Eustat. PRA

This table is particularly interesting to analyse the behaviour of the renewal dwellings (as has already been stated, 1/8th of the sample is renewed each quarter) in the different quarters.

The first column in yellow shows the response rate for each of the 8 quarters of these renewal dwellings. The second column, in blue, shows the response rate for each of the 7 quarters of the dwellings that answer the survey for a second time, and so on.

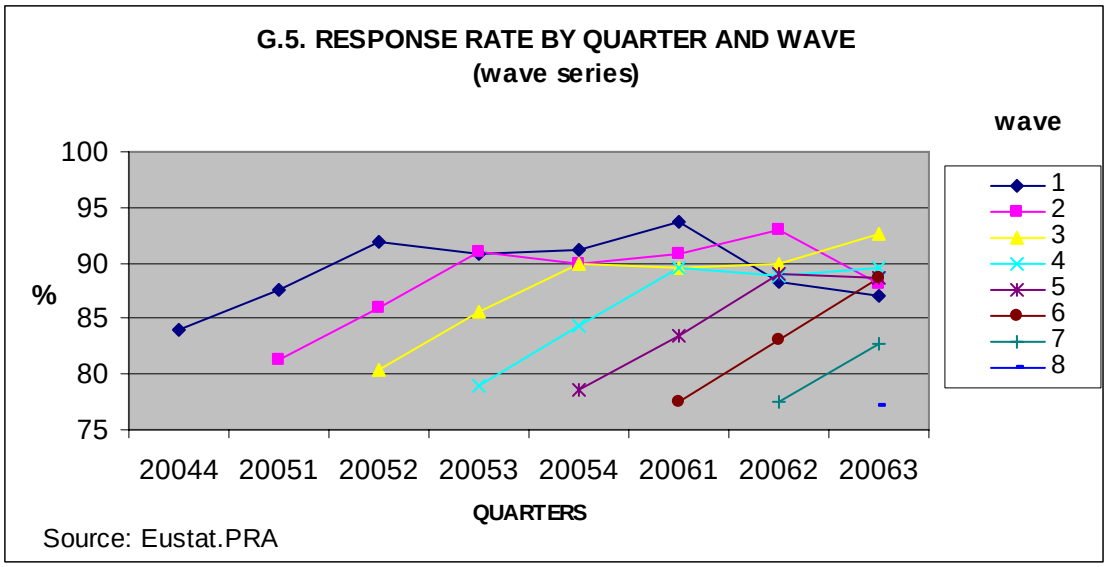
As can be observed, the trend relating to the renewal dwellings increases up to the first quarter of 2006, where it peaks, with a 93.8% response rate. On the other hand, the response rate begins to drop in the two last quarters. In the second quarter of 2006, it drops five and a half points with respect to the first quarter, which is rather significant and the rate fell by one point with respect to the second quarter in the last quarter.

This drop is mainly due to the increase in the number of “refusals” and “non-contacts” in quarters 20062 and 20063 with respect to the other quarters, and particularly to the increase of “other non-response types”, of the “holidays” outcome in quarter 20063.

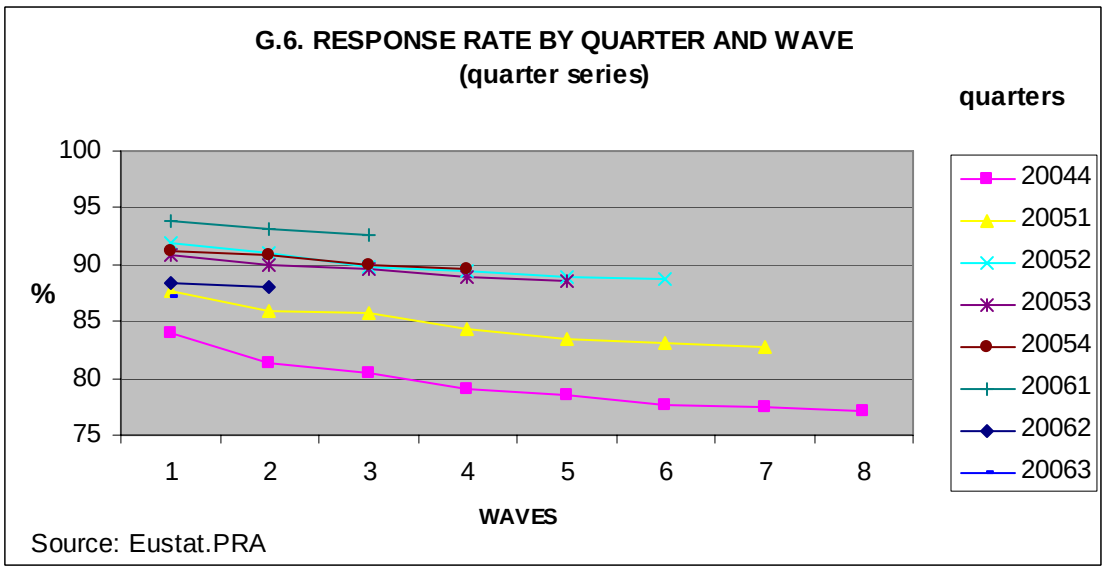
With respect to the rest of the columns, a general upward evolution should be pointed out, although it must be highlighted that as only 8 quarters have been surveyed so far, it is not possible to carry out any comparisons as, only the first sample of quarter 20044 has reached the eighth wave.

This general wave improvement may be due to the increasing experience of the surveyors, the company entrusted with conducting the surveys and the corrective field measures introduced.

The response rates by columns of the above table can be seen in Graph 5. The blue line shows the response rate for the renewal dwellings in each of the 8 quarters, the pink line refers to the response rate of the dwellings that answer the survey for the second time, and so on.



Graph 6 shows response rates taken in a diagonal way on Table 3. The evolution of the rates per quarter and wave can therefore be seen. The pink line, for example, refers to the evolutions of the dwellings from the first surveying quarter, i.e., the dwellings from quarter 20044, across 8 waves. The yellow line refers to the dwellings that are surveyed for the first time in 2005 and their evolutions during the 7 waves so far.



PRA results summary

It should be pointed out in this section that all the results obtained so far are solely based on the 8 first surveying quarters (which is equivalent to 2 years) up to the date of the report. Therefore, only part of the sample surveyed in quarter 2004 have reached the eighth wave. Further time is therefore necessary to be able to extract more final conclusions.

The following should be highlighted with respect to the response rates results obtained so far:

1. The strong ascending evolution of the quarterly response rates so far.

2. The response rates by wave dropped by approximately one point in each wave, which is normal due to the sample's attrition. The evolution of these rates will be seen more clearly when more quarters are available, due to the fact that the eighth wave only includes the dwellings surveyed in the first surveying quarter, i.e. 20044.

3. With respect to the response rates by quarter and wave, it should be recalled that the table can be interpreted diagonally and by columns.

   It should be stressed that the first column of this table refers to renewed dwellings and, in general, show a trend to increase except in the last two quarters. The rest of the columns generally tend to evolve upwards.

   The evolution of the response rates by quarter and wave can be observed diagonally. The evolution of the dwellings from the first wave over the quarters can be seen in each of the diagonals. It can be seen that the rate of the first wave, relating to the renewal dwellings, generally increases over the quarters, which implies that the subsequent diagonal rates also increase.

Chapter
5

5. Conclusions

The following are the outstanding main conclusions from applying the standard proposed by the Institute for Social & Economic Research (ISER):

- First of all, it provides a way of classifying the survey field results or outcomes, more specifically in the case of face-to-face surveys, and which covers a multitude of situations. In addition, using the established level hierarchies, these outcomes are finally classified into “Complete Survey”, “Partial Survey", "No Contact", "Refusal" and “Other non-response types”, along with "Unknown Eligibility” and “Not Elegible”. It thus ensures that the outcomes, sometimes expressed in different ways in different surveys, are always classified in the same way.

- Secondly, this methodology establishes which are the formulas for the response rates and the remaining rates that measure the data collection response components: co-operation rate, contact rate, refusal rate and eligibility rate.

- In this context, according to the panorama that we have described in the introduction, where data collection is not clearly documented, this standard offers the possibility allows these guidelines to be followed for general household surveys, even in the case of personal surveys.

- The use of a standard in a specific survey over time, in a serie that is much longer than the one we have considered in this paper, enables better knowledge of the response mechanisms in the survey and of those factors that can contribute to vary the results that have been obtained in time. A change in the questionnaire, in the collecting method or in the field practices will be reflected in those results, particularly if the components (refusal, contact or co-operation) are also studied, without forgetting the importance of the sample’s frame (eligibility).

- Its use in different surveys allows them to be compared, provided that the designs are similar and the differences in the response rates are not wholly attributable to its different characteristics. In the Labour Force surveys, for example, it is important to compare panel surveys, with waves and collection methods and other similar characteristics.

- Finally, a standard has more value, the greater the number of surveys that use it. And using and reporting the response rates and their components in a uniform and comparable way continues to be the challenge facing statistic offices. In the long term, this will benefit the quality of the surveys and the users in general.

Chapter

6

6. Bibliography

[1] PETER LYNN, ROELAND BEERTEN, JOHANNA LAIHO AND JEAN MARTIN.

RECOMMENDED STANDARD FINAL OUTCOME CATEGORIES AND STANDARD DEFINITIONS OF RESPONSE RATE FOR SOCIAL SURVEYS.

ISER WORKING PAPERS NUMBER 2001-23.

[2] THE AMERICAN ASSOCIATION FOR PUBLIC OPINION RESEARCH.

STANDARD DEFINITIONS. FINAL DISPOSITIONS OF CASE CODES AND OUTCOME RATES FOR SURVEYS . 2004.

[3] STATISTICAL POLICY OFFICE, OFFICE OF INFORMATION AND REGULATORY AFFAIR, OFFICE OF MANAGEMENT AND BUDGET.

STATISTICAL POLICY. WORKING PAPER 31. MEASURING AND REPORTING SOURCES OF ERROR IN SURVEYS. July 2001.

[4] COUNCIL OF AMERICAN SURVEY RESEARCH ORGANIZATIONS (CASRO)

ON THE DEFINITIONS OF RESPONSE RATES

http://www.casro.org/resprates.cfm, (1 de enero de 2006)

[5] TOM W. SMITH

DEVELOPING NONRESPONSE STANDARDS"

SURVEY NONRESPONSE (GROVES Y OTROS ED. 2002).

[6] ATROSTIC, B.K., BATES, N., BURT, G. & SILBERSTEIN, A.

NONRESPONSE IN U.S. GOVERNMENT HOUSEHOLD SURVEYS: CONSISTENT MEASURES, RECENT TRENDS, AND NEW INSIGHTS

JOURNAL OF OFFICIAL STATISTICS, VOL. 17, NO. 2, 2001, PP. 209-226