Household survey on the equipment and usage of information and communication technologies (CIT-H 2002)

Final results (may 2003)

Index

1 Methodology of the survey

1.1 Antecedents

Statistical research on the use made by households or individuals of new Information and Communication Technologies (TIC) has increased enormously in recent years as these tools were being developed and implanted. It cannot be denied that the official statistical agencies have been sluggish in satisfying the demand of information for what is known as *information society* or *new economy*. Only the *Observatorios de la Sociedad de la Información* as well as private statistical and market research consultancies and some regional governments have carried out specific surveys. However, when this statistical work was launched, no harmonised statistical methodology was used that would have allowed a comparison of the results in European Union countries and, in Spain, to have comparable results for the autonomous communities (CCAA).

Due to these drawbacks, the available data were not as good as the users expected. Among these shortcomings are the variety of sources (both public and private) that hampered the coherence of data, gaps in the information and the lack of an activities and products classification that would have led to a better knowledge and comparability of the data.

That is why at the beginning of 2001, upon the initiative of the *Comisión del Mercado de las Telecomunicaciones* (CMT), contacts were established between the latter and the INE to envisage a household survey on the use of information and communication technologies (TIC).

On principle, the requested information was to be concerned with the availability of personal computers, Internet and electronic commerce in households and in how far their members used them. However, on the proposal of the CMT, information was also requested regarding fixed and mobile telephony, satisfaction with telephone companies and providers of Internet services, reasons why a change of company is envisaged and opinions on telephone companies and providers of Internet services. The CMT also supplied technical advice on the above mentioned variables.

To give legal form to this statistical co-operation, the INE and the CMT signed at the beginning of 2002 a co-operation agreement for a research on said variables. This *Encuesta de Hogares sobre Equipamiento y Uso de Tecnologías de Información y Comunicación* (TIC-H 2002) was mainly financed by the CMT. It became official with its inclusion in the 2002 Statistical Programme, in the 2001-2004 National Statistical Plan.

On the other hand, the European Union statistical office. EUROSTAT, recently granted priority to this kind of research, which is confirmed by the creation in March 2002 of a new unit for statistics on the information society. As a result, their availability greatly improved, an example being the advance results of the 2002 survey on TIC usage, referred to 2001 while a second survey will be carried out in 2003.

In 2002, EUROSTAT started for the first time an ad-hoc survey on TIC usage in households. Although participation was voluntary, most member countries co-

operated. In 2003 a new, also voluntary, survey and the draft of a Regulation are being elaborated. At the same time, steps are taken toward the statistical harmonisation of definitions and nomenclatures as well as of data collection tools.

Through the channel of the INE, Spain partook of this initiative seizing on the above mentioned statistics that were to be carried out jointly with the CMT (TIC-H 2002 Survey), to which end an EUROSTAT subsidy was granted. Spain thus started a new stage in statistical research on *Information Society* in which comparability among CCAA will be possible for the first time. In parallel, statistical convergence was sought with those CCAA Statistical Institutes that elaborate their own studies in this field (Cataluña, Navarra, País Vasco). This allows a better usage of public resources and avoids unnecessary duplication of data collection.

Thanks to these fresh statistical and also social aspects, users will be provided with quality information that makes comparability among the CCAA and the EU member states more fluent.

1.2 Objectives and scope of the survey

The objectives of the survey under consideration, are:

− A more thorough knowledge of the information and communication (TIC) technologies facilities in Spanish households (television, radio, fixed and mobile telephony, computer equipment, Internet…)

− A more thorough knowledge of how the Spanish population uses Internet and mobile telephony. For internet, frequency, payment formulas, access places, access obstacles, electronic commerce activities and telework, are the most researched upon aspects.

− Usage for international comparability and compliance with the information requirements by international organisms on the equipments and implementation of communication and information technologies.

− Attainment of information comparable among autonomous communities on the equipment and implementation of communication and information technologies in households in their respective territories.

The research covered all those living in main family dwellings. The survey was carried out in the 52 provinces of Spain, Ceuta and Melilla.

There are several reference periods for a number of variables and characteristics defined in their description. These periods are:

− Reference period for the survey's results. Second quarter, 2002.

− Information reference period, which is usually that of the interview. Some variables, however, referred to different periods, as explained below:

• Last month for questions on mobile telephone consumption.

• Last three months for the use of internet by the household members, place and frequency of the use of computers, place and frequency of the use of Internet, time of sporadic use of Internet, services used in Internet and electronic commerce.

• Last 12 months for expenditure on mobile phone facilities

• Next 6 months for the intention to change fixed telephony, mobile telephony and Internet operators.

• Three months and sometime before for purchases on Internet (electronic commerce).

There are also two questions whose reference periods remain open, according to the information provided by the interviewed person: expenditure and time spent on fixed telephony, expenditure on mobile telephony with pre-paid card (without contract).

1.3 Sampling design

1.3.1 TYPE OF SAMPLING. STRATIFICATION

Its general characteristics are those of the Encuesta de Población Activa (Labour Force Survey EPA), since the sample comprised the dwellings included in this survey for the fourth, fifth and sixth time during the first quarter of 2002. Those interviewed for the fourth time were only used to increase the sample of those Autonomous Communities made up of a single province.

This EPA based sampling framework made it possible to know in advance all the socio-demographic variables of the household members. As will be explained later, this allowed a random sampling selection of respondents regarding mobile telephony and the crossing of results with the above mentioned socio-demographic variables. For each household, only members aged 16 or over were selected since this is the EPA age threshold. The sampling type was two-staged and stratified.

The first stage units were the census sections; those of the second stage, the main family dwellings. Since the research addressed all people aged 16 and over, no sub-sample was used. Only for data regarding mobile telephony, a person older than 16 was randomly selected to provide the information.

The framework for the selection of the sample included the areas of the census sections referred to 1 January 2001. For the second stage units, use was made of the main family dwellings in all the sections selected for the sample.

The strata were defined for each autonomous community and province, according to the size of the municipality to which the section belongs. An independent sample was designed for each autonomous community, since this is the level of breakdown aimed at by the survey.

1.3.2 SAMPLING SIZE. ALLOCATION

To determine the sampling size, some knowledge is needed on the basic characteristics of the target population and on the level of breakdown of the data to be analysed. Information stemming from outside sources and from the pilot study, helped to determine the sampling size necessary to produce estimates with a sampling error below 10% for at least the most important results of the autonomous community survey. The final sample included 20.001 dwellings distributed over 1.177 census sections.

The distribution of the sample among autonomous communities was made according a compromise between the uniform and proportional allocation. In every CCAA, the sample was distributed among strata in a strictly proportional way.

Table 1: Distribution of the theoretic sample

Autonomous communities	Nr. of Sections	Nr. of dwellings
01. Andalucía	150	2.526
02. Aragón	56	989
03. Asturias (Principado de)	58	885
04. Baleares (Illes)	45	724
05. Canarias	60	1.062
06. Cantabria	34	539
07. Castilla-León	83	1.282
08. Castilla-La Mancha	67	1.087
09. Cataluña	126	2.221
10. Comunidad Valenciana	94	1.707
11. Extremadura	52	815
12. Galicia	80	1.323
13 Madrid (Comunidad de)	78	1.602
14 Murcia (Región de)	46	830
15. Navarra (Comunidad Foral de)	39	651
16. País Vasco	69	1.085
17.Rioja (La)	27	447
18. Ceuta y Melilla	13	226
TOTAL	1.177	20.001

1.3.3 SELECTION OF THE SAMPLE

Within each stratum, the sections were selected with a probability proportional to their size. Within each section, the dwellings were selected with the same probability through systematic random start sampling.

This procedure led to self-weighting samples for each stratum.

For each dwelling, a household questionnaire was completed. Besides general data on the household and fixed telephony, it collected information on the usage

of mobile telephony supplied by a previously and randomly selected member of the household. **All the Internet users had to complete a questionnaire on the subject**.

1.3.4   ESTIMATORS

To estimate the characteristics of the survey, the following types of estimators were considered:

• Estimator for household data

• Estimator for data on Internet usage and electronic trade

• Estimator for mobile telephony data

In all the cases, use was made of ratio estimators. A final re-weighting was made taking into account information from outside sources. The auxiliary information was:

For household data, the distribution of the total number of households by province and size of the household.

For personal data, the total population distributed by autonomous community, age and sex.

The reference universe of households and people to which the sampling values obtained by the survey have been weighted up, was that of the second quarter of 2002: 13.712.939 main households and 33.932.985 people aged 16 or over.

The mathematical calculation of the estimators is shown below:

A) Estimator for household data

For the characteristics of households, an estimator produced by the following steps is applied:

A.1. Expansion estimator based on the design factor, including correction of non-response at section level.

In each stratum h, the estimator of the total of an characteristic X is obtained through the following expression:

$$\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{v_{ih(e)}} \frac{V_h}{n_h . v_{ih(e)}} . x_{hij}$$

where:

V_h = dwellings in the stratum h

$v_{ih(e)}$: effective sample of dwellings in section i.

n_h : number of sampling sections in stratum h.

x_{hij} : Value of the target characteristic in dwelling j.

A.2. Separate ratio estimator, to fit the survey's data to the population projection in each stratum h.

$$\hat{\hat{X}}_h = \frac{\displaystyle\sum_{i=1}^{n_h} \sum_{j=1}^{v_{ih(e)}} \frac{x_{hij}}{v_{ih(e)}}}{\displaystyle\sum_{i=1}^{n_h} \sum_{j=1}^{v_{ih(e)}} \frac{p_{hij}}{v_{ih(e)}}} \cdot P_h$$

where:

p_{hij} : Total population in the sample (aged 16 or over) in dwelling j.

P_h : Projection of the population over 16 in stratum h.

A.3. The final estimator is obtained by applying re-weighting techniques to the above estimator, for which the CALMAR software is used. EPA information for the size of dwellings was an outside source.

B) Estimator for data on the use of Internet and electronic commerce

Those data are obtained from the individual questionnaire.

Since all those who use Internet in the household must contribute to the survey, the estimator is obtained as for the households but with the following changes:

B.1. Since the characteristics estimated in this paragraph refer to individuals, the estimator that appears in paragraph A.2 facilitates a reweighting by age groups and sex. As an outside source, the population projections by age groups, sex and autonomous communities were considered.

B.2. Correction of non response in individual questionnaires.

If F_{hij} is the final household j factor obtained in the above paragraph, the final factor of a member of said household in the age and sex (es) group is obtained through the formula

$$F'_{hij(es)} = F_{hij} \cdot K_{h(es)}$$

where:

$$K_{h(es)} = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{v_{ih(e)}} F_{hij} \cdot n_{hij(es)}}{\sum_{i=1}^{n_h} \sum_{j=1}^{v_{ih(e)}} F_{hij} \cdot n'_{hij(es)}}$$

where

$n_{hij(es)}$: The total members of household j with age and sex group (es) who use Internet according to the general household questionnaire (and who must consequently complete the individual questionnaire).

$n'_{hij(es)}$: The total members of household j with age and sex group (es) whose individual questionnaire is completed.

C) Estimator for mobile telephony data

In this case the sampling information stems from a person selected among the members of the household. The estimator is similar to the household one, but includes a factor that incorporates the corresponding selection probability.

C.1. Estimator based on the design factor, with correction of non response.

$$\hat{X}_h = \sum_{i=1}^{n_h} \sum_{j=1}^{v^*_{ih(e)}} \frac{V_h \cdot p_{hij}}{n_h \cdot v^*_{ih(e)}} \cdot x_{hij}$$

where:

$v^*_{hi(e)}$ = the sampling dwellings where information on mobile telephony is obtained.

p_{hij} = Total of population (aged 16 and over) in the sample of household j.

C.2. Separate ratio estimator to adjust to the population of the stratum.

$$\hat{\hat{X}}_h = \frac{\sum_{i=1}^{n_h} \sum_{j=1}^{v^*_{ih(e)}} \frac{V_h \cdot p_{hij}}{n_h \cdot v^*_{ih(e)}} \cdot x_{hij}}{\sum_{i=1}^{n_h} \sum_{j=1}^{v^*_{ih(e)}} \frac{V_h \cdot p_{hij}}{n_h \cdot v^*_{ih(e)}}} \cdot P_h$$

where:

P_h = : Projection of the population older than 16 in stratum h.

C.3. Application of re-weighting techniques by age groups and sex at autonomous community level (CALMAR) with populations of the second quarter of 2002.

1.3.5 SAMPLING ERRORS

To calculate the sampling errors of the main characteristics, the indirect Jacknife [1]method has been used (method of ultimate clusters).

This method is based on the formation of sub-samples, each of which is obtained by eliminating a primary unit of the total sample. The estimated variance is as follows:

$$\hat{V}ar\left(\hat{X}\right)= \sum_h \frac{(n_h - 1)}{n_h} \sum_{j \in h} \left(\hat{X}_{(hj)} - \hat{X}\right)^2$$

where:

$\hat{X}_{(hj)}$ is the estimate of X when the primary unit j of stratum h is withdrawn from the sample.

$\hat{X}$ is the estimate of X obtained with the whole sample.

n_h is the number of primary units in stratum h.

Sampling errors have been calculated for the six most representative variables tackled by the survey under consideration, i.e.: users of mobile phones, households with computers, households with fixed telephone, households with internet, users of internet and users of electronic trade (latest quarter or sometimes). The sampling error is usually expressed as the variation coefficient (CV), whose formula is as follows:

$$CV \ (\hat{X}) = \frac{D\ (\hat{X})}{\hat{X}} \cdot 100$$

where:

CV is the variation coefficient, *D* the standard deviation and $\hat{X}$ the estimate of the variable (in table number 1, the sampling error appears in percentages).

---

1 The software CALJACK developed by P.Lavallé of Statistics Canada has been used.

Table 1. Variation coefficients in percentages

Autonomous communities	Users of mobil telephone	House-holds with compu-ter	House-holds with fixed tele-phone	Households with access to Internet	Internet users	e-commerce users (in the latest quarter or sometimes)
1 Andalucia	2,92	4,12	1,23	8,87	7,88	12,93
2 Aragon	5,36	6,14	1,66	10,61	10,23	19,38
3 Asturias (Principado de)	4,35	7,16	0,94	11,73	8,01	24,48
4 Balears (Illes)	3,46	5,58	1,30	9,25	8,67	15,68
5 Canarias	2,82	8,19	1,89	11,85	10,85	26,63
6 Cantabria	5,25	6,19	1,24	12,31	7,80	25,61
7 Castilla y León	4,22	5,69	1,10	15,06	13,05	17,86
8 Castilla-La Mancha	3,14	6,71	1,71	14,78	11,96	21,38
9 Cataluña	2,59	4,37	0,61	8,84	7,93	13,08
10 Comunidad Valenciana	2,82	5,40	1,36	10,07	9,71	13,77
11 Extremadura	4,50	9,35	1,92	16,54	11,82	18,35
12 Galicia	3,17	6,33	0,97	12,75	9,07	17,91
13 Madrid (Comunidad de)	2,35	4,04	0,70	6,22	5,96	11,30
14 Murcia (Region de)	4,11	7,36	1,39	11,55	10,66	16,55
15 Navarra (Com. Foral de)	4,07	6,79	0,93	10,80	8,37	15,73
16 Pais Vasco	3,43	5,31	0,60	9,83	9,01	16,19
17 Rioja (La)	7,58	8,56	1,98	16,05	12,61	19,59
18 Ceuta y Melilla	6,69	12,57	4,11	13,24	12,11	15,55
Total	0,93	1,57	0,33	3,09	2,70	4,82

As approximate criteria to give an idea on the reliability of the results, it is considered that cells with fewer than twenty observations should not be deemed reliable and that the weighting results should, therefore, be put aside. When the observations vary between twenty and fifty, the results are reliable but sometimes inconsistent with the context of the data. When there are more than fifty observations, the results may be considered consistent.

To monitor these calculations, the following figures show the average grossing up factors for each of the three variables included in the survey:

– Household weighting factor: 800

– User of mobile telephone weighting factor: 2.000

– User of internet weighting factor: 900

This approach of data reliability is by no means mathematically accurate and cannot replace the criterion described in the first part of this paragraph for the six previously selected variables.

1.4 Data collection

1.4.1 QUESTIONNAIRES

Two questionnaires were created for data collection: a general questionnaire and an individual questionnaire.

A general household questionnaire requests basic information on members of the household, on information and communication equipment, types of use and expenditure of fixed telephony and general usage (answering yes or no) of Internet by household members aged 16 and over. It also contains specific questions on usage and expenditure on mobile telephony answered **by a single randomly selected member of the household, aged 16 or over**.

An individual questionnaire **for each member of the household aged 16 or over**, who has used Internet anywhere and by any means in the last three months, requests information on computer use anywhere, the use of Internet in or outside the household, its purpose, frequency and the degree of knowledge and attitude toward Internet providers as well as the purchase of goods through this means (electronic commerce).

1.4.2 FIELD WORK: INTERVIEWS

The number of surveyed census sections was 1.177 and the number of households, 7.327 by personal and 12.674 by telephone interviews, totalling 20.001 households. The number of household members included in the surveyed households was 51.486 (referred only to those aged 16 and over). In fact, due to non - response (refusals, families away from their homes), the actual household sample was cut down to 17.148.

After considering several methodologies for the design of data collection tools, two sub-samples were decided upon. One would be carried out entirely by telephone survey (CATI method) and the other by personal interview. This division was even applied to census sections, the general objective being about 60% of each section's households for telephone interviews and the rest for personal ones, which led to the following distribution:

– Number of *face to face* effectively interviewed households: 6.311

– Number of households effectively interviewed by telephone: 10.837

The interviewers which carried out the *face to face* interview, completed a *road map* on which they indicated any incidents and which was afterwards for data treating.

1.4.3   TIME SCHEDULE OF THE DATA COLLECTION

They were the following:

– Overall period for the work: 27 May 2002 to 25 September 2002

– Period of *face to face* interviews: 10 June 2002 to 23 July 2002

– Period of telephone interviews: 20 June 2002 to 6 September 2002

The sample started by announcing the survey in letters sent individually to each household, addressed to the reference person indicated in the EPA sampling base, and meant to explain its purpose and need. In order to reduce the time elapsed between receiving this communication and achieving the survey, the letters were staggered according to the planned weekly personal and telephone interviews so that no more than a week went by between receipt and contact.

On the other hand, a free-call telephone number was available with a view to dispelling the respondents' distrust (as to the survey being indeed required by the INE) and to satisfy their demand of advice; 295 calls were answered.

1.4.4   INCIDENTS DURING DATA COLLECTION

Incidents regarding households

Table 2: Absolute values

	Personal interviews	Telephone interviews	TOTAL
Total of households in the sample	7.327	12.674	20.001
Total of households interviewed	6.382	10.904	17.286
- With complete questionnaires	6.311	10.837	17.148
- With incomplete questionnaires	71	67	138
Total of non response (households)	945	1.770	2.715
- Refusals	220	480	700
- With different human group[1]	141	194	335
- With long absence from home	452	67	519
- Cannot be found	132	999	1.131
- With other incidents	0	30	30

1 With respect to that interviewed in EPA

Table 3: Percentage of households

	Personal interviews	Telephone interviews	TOTAL
Total of households interviewed	87,10	86,03	86,43
- With complete questionnaires	86,13	85,50	85,73
- With incomplete questionnaires	0,97	0,53	0,69
Total of non response (households)	12,90	13,97	13,57
- Refusals	3,00	3,79	3,50
- With different human group	1,92	1,53	1,67
- With long absence from home	6,17	0,53	2,59
- Cannot be found	1,80	7,88	5,65
- With other incidents	0,00	0,24	0,16

Incidents in the variable *usage of mobile telephone*

As mentioned earlier, a member of the household was selected randomly as the official person asked to provide this information; a reserve provider being also available. As a consequence, 16.344 o*fficial* providers and 942 *reserve providers* gave information on mobile telephony, in total 17.286 respondents.

Incidents in the individual questionnaires regarding Internet

Besides the general household questionnaires, individual questionnaires were collected for the use of Internet by all the members of the surveyed households, aged 16 or over, who used it anywhere and for any reason in the last three months. The diagram below summarises this type of surveys and the incidents that not caused questionnaire. The number of theoretic questionnaires refers to those household members who, according to the information of the general questionnaire, used Internet or were likely to have used it. Hence, those known for certain that not having used Internet was not included.

Table 4: Absolute values

	Theoretic Internet questionnaires detected	Complete questionnaires obtained	Questionnaires with incidents	% of incidents over detected
Face to face interview	2.898	2.794	104	3,59
Telephone interview	4.660	4.534	126	2,70
TOTAL	7.558	7.328	230	3,04

Table 5: Percentage over the theoretic total of questionnaires detected

	Face to face interview	Telephone interview	TOTAL
% of Internet questionnaires obtained	96,41	97,30	96,96
% of Internet questionnaires with incidents	3,59	2,70	3,04

1.5 Main results of the survey

The following are the main indicators of the survey:

• Percentage of households with television: 99,5%.

• Average number of television equipment's per household: 1,7%

• Percentage of households with computer: 36%

• Rate of fixed telephony (% of households with this service): 90,2%

• Average monthly expenditure on fixed telephony: 30,86 €

• Percentage of households some of whose members aged 16 or over have a mobile telephone: 65%

• Mobile/user ratio in these households: 1,9

• Percentage of the total population aged 16 or over who use the mobile telephone: 55,5%

• Percentage of households with Internet: 17,4%

• Percentage of people aged 16 or over who used Internet in the last three months: 18,7%

• Percentage of the population aged 16 or over that purchased goods or services through Internet (electronic commerce) for private use in the last three months: 2,1%. (There is moreover 1,1% that made a purchase during the period preceding these 3 months).

• Average expenditure on goods or services in the last three months: 153,30 €.

After the edition of the survey's micro data file, it was crossed with that of the Labour Force Survey that contained the socio-demographic variables of the interviewed households and individuals. This crossing moreover allowed the treatment of a number of inconsistencies not detected in the consistency checks included in the CATI interviews, all of which warrants the robustness and consistency of the final data.

The crossing variables were: size of the household's municipality, number of household members, their sex, age, level of education, labour situation (occupied, unemployed, inactive) and occupation.

Regarding labour situation, emphasis was laid on *students* since this group of the population obviously uses Internet intensively. It therefore includes those classified as *students* in the EPA of the quarter of reference, as well as those who declared they were following some education/training course during the last four weeks and who appeared as non-occupied, provided that the number of hours worked per week is under fifteen.

Regarding the level of studies, a classification into four groups has been considered, according to EUROSTAT guidelines for the breakdown of this variable:

• Illiterate. (Level 0)

• Low. (Levels 1,2). Primary and 1st Stage of Secondary Education

• Medium (Level 3). 2nd Stage of Secondary Education, Vocational Training I, Medium Degree Training Cycles and equivalents

• High. (Levels 5 and 6). University Education, Doctorate, Vocational Training II, High Degree Training Cycles and equivalents.

The levels appear in the INE National Classification of Education (CNED).