



Small-area estimation in Official Statistics: ICT survey in Enterprises of the Basque Country

Jorge Aramendi, Jose Miguel Escalada, Elena Goni & Anjeles Iztueta
EUSTAT. Basque Statistics Office

European Establishment Statistics Workshop

Nuremberg – Germany

9 September 2013

Index

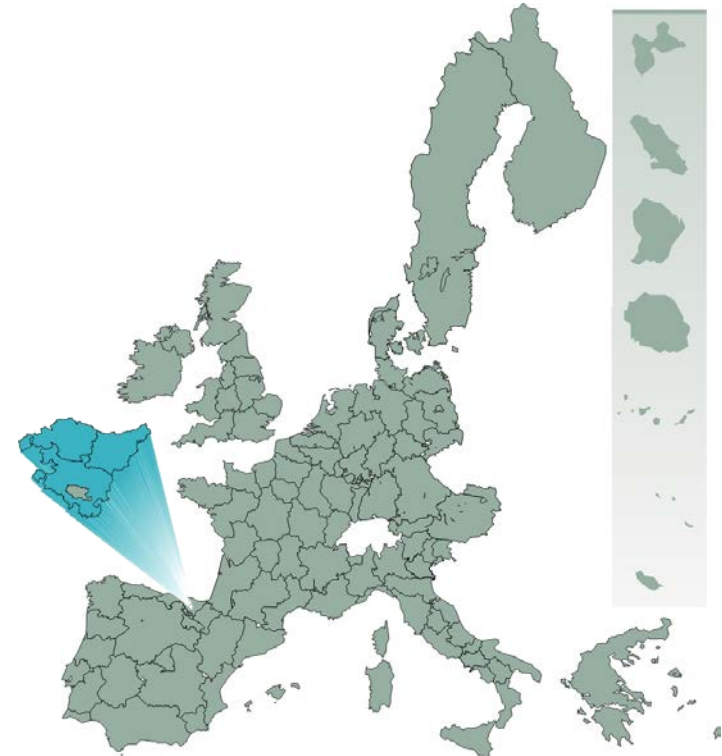
- **Review and Historical Context**
- **ICT Survey Design**
- **Main Objective**
- **Methodology**
 - Variables
 - Auxiliary information
 - Estimators
 - Accuracy
- **Conclusion and Results**

Review and Historical Context

From 2003 onwards, partnership with the UPNA (Navarra Public University) to study small area techniques in EUSTAT surveys

In 2004, work began on studying the first survey

- Industrial Survey (2005)
- Labour Force Survey (2008)
- Information Society Survey in Families (2009)
- Technological Innovation Survey (2010)
- Information and Communications Technologies (ICT) Survey in Enterprises (2012)



ICT Survey design

The ICT (Information and Communications Technologies) Survey in Enterprises was implemented in 2001 to find out the level of **use of new technologies** in the Basque economy

The ICT Survey is a **panel of around 7500 establishments** with an annual renewal between 15% and 20% of the elements.

Stratified design with optimal allocation according to three variables: province, activity and employment stratum.

- **3 provinces:** Araba, Bizkaia and Gipuzkoa
- **65 branches of activity,** based on NACE
- **6 employment strata:** 0-5, 6-9, 10-19, 20-49, 50-99, ≥ 100



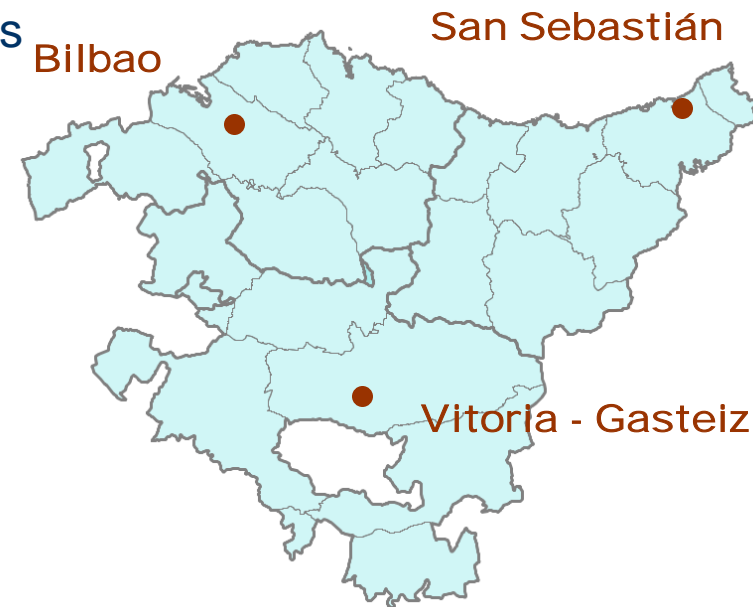
Main Objective

Currently estimates calculated using the **direct Horvitz-Thompson estimator** for 3 provinces, 65 activity sectors and 3 employment stratum.

OBJECTIVE : Obtaining district estimations (20 districts and 3 capital cities) for the main ICT variables.

PROBLEM: Insufficient sample size for direct estimates

So we decided to improve the estimation methodology of this survey and to **introduce small area estimation techniques**.



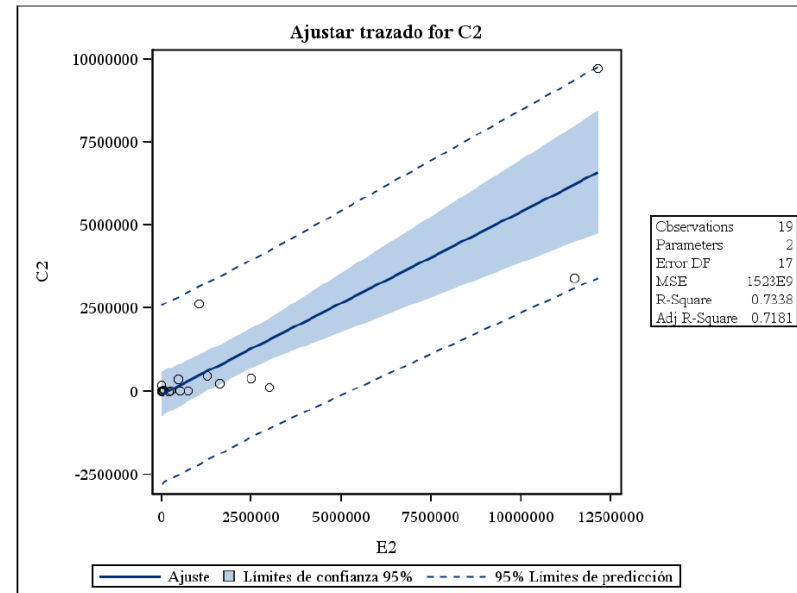
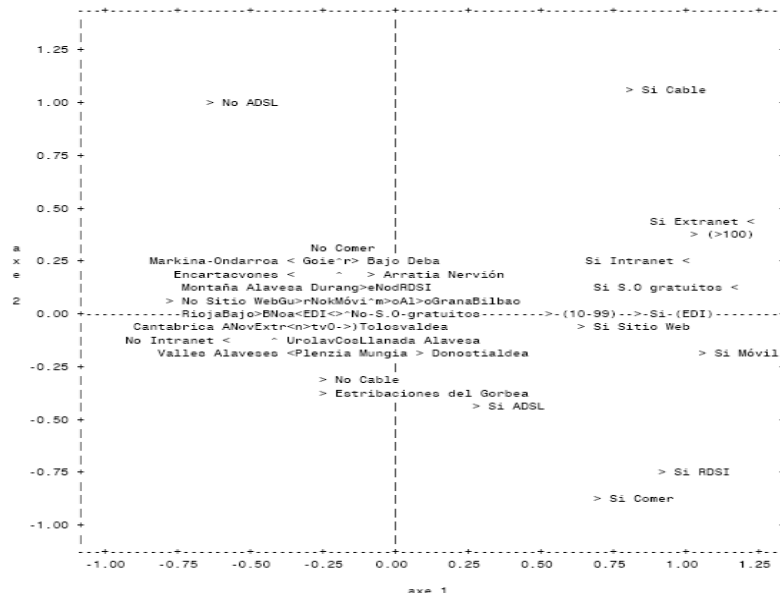
Methodology

- **Variables** : Selection and analysis of the target variables
- **Definition of the small area**: 20 districts and 3 capital cities
- **Auxiliary information**: Availability of sources, quality of the data, ..
- **Estimators**: Select best estimators for the data and asses them

Methodology: Variables

Selection and analysis of the target variables:

- Internet and computer (%)
- e-commerce (sales or purchases) (%)
- web (%)
- freeware operating-systems (%)
- electronic data exchange (%)
- web proceedings with the public administration (%).



Methodology: Auxiliary Information

Analysis of the auxiliary information:

- ✓ DIRAE – Our Directory for economic establishments
- ✓ 184.000 establishments
- ✓ Updated with different **surveys** and **administrative sources**
- ✓ **Employment and Activity** are key variables regularly updated



Methodology: Model selection

Assess several estimators

Design based estimators:

Direct

Postestratified

Synthetic

Composite estimators

Model based estimators:

Logistic regression model, linear mixed model and no-linear mixed model

Analyze the consistency and stability of the aggregated estimates at province level

Methodology: Final Model

First level model:

$$\text{logit}(p^1) = \log \frac{p^1}{1 - p^1} = \beta_0 + \beta_1 x_1 + \dots + \beta_{19} x_{19} + \beta_{20} x_{20} + \beta_{21} x_{21} + \beta_{22} x_{22} + \dots + \beta_{47} x_{47}$$

where

- p^1 is the proportion of establishments that responds affirmatively in a certain variable.
- β_0 is the intercept.
- $\beta_1, \dots, \beta_{19}$ are the coefficients of explanatory variables for the 20 districts.
- β_{20}, β_{21} are the coefficients of explanatory variables for the 3 employment strata.
- $\beta_{22}, \dots, \beta_{47}$ are the coefficients of explanatory variables for the 27 categories in the activity classification.

Methodology: Model

Second level model:

$$\text{logit}(p^2) = \log \frac{p^2}{1 - p^2} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \dots + \beta_{30} x_{30}$$

where

- p^2 is the proportion of establishments that responds affirmatively in a certain variable.
- β_0 is the intercept.
- β_1, β_2 are the coefficients of explanatory variables for the 3 provinces.
- β_3, β_4 are the coefficients of explanatory variables for the 3 employment strata.
- $\beta_5, \dots, \beta_{30}$ are the coefficients of explanatory variables for the 27 categories in the activity classification.

Methodology: Accuracy

Mean Squared Error is calculated using the bootstrap resampling method.

$$e.e.(\hat{\theta}_d) = \sqrt{\frac{1}{R-1} \sum_{i=1}^R (\hat{\theta}_d^i - \bar{\hat{\theta}}_d)^2}$$

where:

θ_d is the total population of a certain variable in region d.

R number of repetitions used in the bootstrap method (R = 200)

i is the repeated number (i = 1, 2, ..., R)

d is the region

$\hat{\theta}_d^i$ the estimator in the i-th bootstrap sample of said total in region d

Conclusion and Results

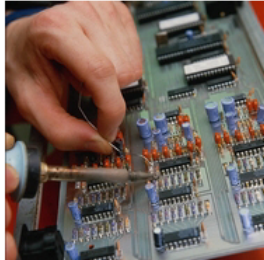
- ✓ Estimates are consistent and stable.
- ✓ The results offer acceptable levels of quality in terms of accuracy. The estimated coefficients of variation (CV) are not excessively high, the majority of the CV-s obtained in the estimations do not exceed 15%. , given the relatively small samples and population in some districts.
- ✓ As a result of the study we have a computer programme based on SAS that is used to analyse this methodology and to apply the mentioned estimators and the calculation of the mean squared errors.
- ✓ We have disseminated district-level information for 3 years 2010, 2011 and 2012

Establishments of 10 or more employees with Internet access that carry out procedures electronically with the administration. Estimation (%) and Variation Coefficient. 2010-2012

	2010		2011		2012	
	Estimation	CV	Estimation	CV	Estimation	CV
Basque Country	84,9	0,01	85,2	0,01	87,8	0,01
Vitoria - Gasteiz	85,4	0,03	85,5	0,02	85,1	0,02
Bilbao	85,4	0,02	84,0	0,02	87,8	0,02
Donostia - San Sebastian	85,7	0,03	87,7	0,02	88,4	0,02
Araba/Álava	84,9	0,02	86,1	0,02	86,0	0,02
Arabako Ibarrek / Valles Alaveses	90,8	0,05	88,7	0,04	84,2	0,07
Arabako Lautada / Llanada Alavesa	85,0	0,03	85,9	0,03	85,1	0,02
Arabako Mendialdea / Montaña Alavesa	84,0	0,08	73,4	0,10	86,5	0,02
Rioja Alavesa	93,0	0,04	91,4	0,03	91,3	0,05
Gorbeia Inguruak / Etribac. del Gorbea	87,0	0,06	87,4	0,05	86,7	0,03
Kantauri Arabarra / Cantabrica Alavesa	77,1	0,05	84,1	0,04	92,2	0,04
Bizkaia	84,6	0,02	83,7	0,02	87,2	0,02
Arratia Nerbioi / Arratia-Nervion	83,8	0,03	88,2	0,02	90,7	0,04
Bilbo Handia / Gran Bilbao	84,4	0,02	83,4	0,02	86,7	0,02
Durungaldea / Duranguesado	85,9	0,04	84,7	0,03	88,3	0,03
Enkartazioak / Encartaciones	77,1	0,07	75,7	0,09	81,5	0,07
Gernika-Bermeo	88,7	0,04	81,4	0,05	93,7	0,03
Markina-Ondarroa	85,9	0,05	92,7	0,03	89,6	0,04
Plentzia-Mungia	87,1	0,03	87,2	0,05	91,5	0,03
Gipuzkoa	85,2	0,02	87,0	0,02	89,5	0,02
Bidasoa Beherea / Bajo Bidasoa	80,2	0,04	80,5	0,05	83,5	0,05
Deba Beherea / Bajo Deba	84,1	0,05	81,8	0,05	89,9	0,04
Debagoiena / Alto Deba	90,8	0,03	95,9	0,02	97,3	0,01
Donostialdea	85,4	0,02	88,0	0,02	89,5	0,02
Goierri	89,0	0,03	89,5	0,04	93,1	0,03
Tolosaldea / Tolosa	87,3	0,03	85,8	0,04	88,1	0,04
UrolaKostaldea / Urola Costa	79,5	0,06	84,2	0,05	85,4	0,05

Establecimientos de 10 o más empleados con sitio web en Euskera. Estimación (%) y Coeficiente de Variación. 2010-2012

	2010		2011		2012	
	Estimación	CV	Estimación	CV	Estimación	CV
C.A. de Euskadi	41,8	0,03	41,0	0,03	41,8	0,03
Vitoria - Gasteiz	34,5	0,07	31,6	0,06	33,0	0,06
Bilbao	42,1	0,06	37,7	0,05	39,4	0,06
Donostia - San Sebastian	58,7	0,05	55,1	0,05	56,9	0,04
Araba/Álava	33,5	0,07	32,2	0,06	32,4	0,06
Arabako Ibarrek / Valles Alaveses	20,0	0,17	18,7	0,11	18,8	0,22
Arabako Lautada / Llanada Alavesa	34,3	0,07	31,6	0,06	32,8	0,06
Arabako Mendialdea / Montaña Alavesa	50,2	0,10	54,0	0,12	54,0	0,06
Errioxa Arabarra / Rioja Alavesa	34,5	0,14	39,4	0,12	32,7	0,13
Gorbeia Inguruak / Estribac. del Gorbea	27,7	0,14	35,4	0,15	28,3	0,15
Kantauri Arabarra / Cantábrica Alavesa	31,7	0,09	32,5	0,10	32,4	0,10
Bizkaia	36,0	0,05	35,2	0,04	35,5	0,06
Arratia Nerbioi / Arratia-Nerviñ	31,6	0,08	34,4	0,09	28,2	0,11
Bilbo Handia / Gran Bilbao	35,7	0,05	34,3	0,05	35,1	0,06
Durungaldea / Duranguesado	33,2	0,09	33,6	0,09	31,3	0,10
Enkartazioak / Encartaciones	43,1	0,15	35,9	0,16	37,2	0,12
Gernika-Bermeo	52,8	0,09	53,0	0,10	51,9	0,08
Markina-Ondarroa	50,0	0,14	60,4	0,13	54,7	0,14
Plentzia-Mungia	32,7	0,07	36,5	0,09	39,6	0,10
Gipuzkoa	54,5	0,05	54,0	0,04	55,5	0,04
Bidasoa Beherea / Bajo Bidasoa	48,9	0,09	45,9	0,10	46,1	0,10
Deba Beherea / Bajo Deba	51,4	0,10	52,7	0,09	52,4	0,09
Deba Garaia / Alto Deba	70,3	0,06	75,6	0,06	75,7	0,05
Donostialdea	52,3	0,05	49,7	0,06	52,7	0,04
Goierri	59,1	0,08	60,0	0,09	60,1	0,08
Tolosaldea / Tolosa	53,8	0,09	57,9	0,08	53,8	0,08
Urola-Kostaldea / Urola Costa	54,4	0,09	55,8	0,08	58,4	0,08



The statistical operation Survey on the Information Society-ESI-Companies, provides regular information on the implementation of New Information and Communication Technology -ICT- in the companies of the Basque Country. Specifically, it records and describes the level of use of the Internet in the different establishments: the systems of Internet access, activities carried out via the Internet, as well as the availability of the website and its main characteristics. It also measures the implementation of E-commerce purchases and sales in economic activity and the means used to carry

Companies

Statistical	Date	
ICT equipment in companies and establishments		
Information technology equipment in establishments in the Basque Country by province, activity branch (A38) and employment strata (%). 2013	07/05/2013	
Information technology equipment in establishments of 10 or more employees in the Basque Country by province, activity branch (A38) and company ownership (%). 2013	07/05/2013	
Network and electronic exchange equipment in establishments in the Basque Country by province, activity branch (A38) and employment strata (%). 2013	07/05/2013	
Network and electronic exchange equipment in establishments of 10 or more employees in the Basque Country by province, activity branch (A38) and company ownership (%). 2013	07/05/2013	
Computer security measures in establishments in the Basque Country according to employment strata and activity branch (%). 2013	07/05/2013	
Information management systems in establishments in the Basque Country by type, according to employment strata and activity branch (%). 2013	07/05/2013	
Use of open-source computer programs in establishments in the Basque Country according to employment strata and activity branch (%). 2013	07/05/2013	
Internet user companies and establishments		
Type of connection in establishments with Internet access in the Basque Country by province, activity branch (A38) and employment strata (%). 2013	07/05/2013	
Type of connection in establishments of 10 or more employees with Internet access in the Basque Country according to province, activity branch (A38) and company ownership (%). 2013	07/05/2013	
Web facilities in establishments with websites in the Basque Country according to employment strata and activity branch (%). 2013	07/05/2013	
Data by capital cities and districts		
Establishments of 10 or more employees by equipment (%). 2012	11/23/2012	
Establishments of 10 or more employees with an Internet connection. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with Internet access that carry out procedures electronically with the administration. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with websites. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with websites in Basque. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with computer networks. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with open-source programs. Estimation (%) and Variation Coefficient. 2010-2012	11/23/2012	
Establishments of 10 or more employees with websites that carry out e-Commerce. Estimation (%) and Variation Coefficient. 2009-2011(*)	11/23/2012	

Related Items

- [Data Bank. Historical series](#)
- [Graphs](#)
- [Press Releases](#)
- [Reports](#)
- [Methodology and Quality](#)
- [Microdata file](#)
- [Definitions](#)
- [Codes and nomenclatures](#)
- [Questionnaires](#)
- [Interactive graphs](#)

Eskerrik asko

Thank you

Danke Schön

Jorge Aramendi
j-aramendi@eustat.es

www.eustat.es