

**ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA SOBRE LA  
SOCIEDAD DE LA INFORMACION - FAMILIAS DE C.A. DE EUSKADI**



**EUSKAL ESTATISTIKA ERAKUNDEA  
INSTITUTO VASCO DE ESTADISTICA**

Donostia-San Sebastián, 1  
01010 VITORIA-GASTEIZ  
Tel.: 945 01 75 00  
Fax.: 945 01 75 01  
E-mail: [eustat@eustat.es](mailto:eustat@eustat.es)  
[www.eustat.es](http://www.eustat.es)

---

# Presentación

Eustat, consciente de la creciente demanda de estadísticas de calidad cada vez más desagregadas, constituyó en 2003 un equipo de investigación compuesto por miembros de Eustat y de la Universidad. El objetivo era trabajar en la mejora de las técnicas de estimación en diferentes operaciones estadísticas, e introducir técnicas de estimación en áreas pequeñas basadas en modelos en la producción estadística. Los resultados de estos trabajos han sido la aplicación del sistema de estimación en áreas pequeñas a la Estadística Industrial de periodicidad anual, editado por Eustat en un Cuaderno Técnico en 2005 y a la Encuesta de Población con Relación a la actividad, editado por el Eustat en un Cuaderno Técnico en el 2008.

Esta metodología de estimación se ha aplicado a otra operación estadística, igualmente relevante dentro de la producción de Eustat, la Encuesta sobre la Sociedad de la Información - familias, que facilita a los usuarios resultados anuales sobre el acceso y el uso de Internet, así como otras tecnologías de la información en la C.A. de Euskadi a nivel de Territorio Histórico. Las estimaciones basadas en métodos de áreas pequeñas dan información de las 20 comarcas estadísticas en las que está dividida la C.A. de Euskadi.

El objetivo de esta publicación es aportar material útil a todos los usuarios interesados en el conocimiento y utilización de métodos en áreas pequeñas.

Este documento tiene dos partes diferenciadas. La primera, la metodología que se ha utilizado, con algunos aspectos propios sobre los estimadores y la información auxiliar empleada; y la segunda, la presentación de los resultados comarcales correspondientes a los años 2005, 2006, 2007 y 2008.

Vitoria-Gasteiz, mayo de 2009

JOSU IRADI ARRIETA

Director General

---

# Índice

PRESENTACIÓN .....	2
ÍNDICE .....	3
INTRODUCCIÓN.....	4
LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS (ESIF) .....	6
2.1 DESCRIPCIÓN DE LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS DE LA C.A. DE EUSKADI .....	6
2.2 ESTIMADORES UTILIZADOS EN LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS (ESIF) DE LA C.A. DE EUSKADI .....	7
SISTEMA DE ESTIMACIÓN EN ÁREAS PEQUEÑAS EN LA ESIF .....	10
3.1 ESTUDIO DE ESTIMADORES .....	10
3.2 ESTIMACIÓN DEL ERROR CUADRÁTICO MEDIO .....	14
3.3 SOFTWARE EMPLEADO .....	19
ESTIMACIONES COMARCALES 2005-2008 .....	20
4.1 DEFINICIONES .....	20
4.2 RESULTADOS .....	22
CONCLUSIONES.....	27
BIBLIOGRAFÍA.....	28
ANEXO .....	30

## Introducción

Actualmente la estadística oficial tiene que responder a una demanda de información de calidad, cada vez más desagregada sobre los principales indicadores sociales y económicos.

Un modo de afrontar esa demanda de desagregación es aumentar los tamaños de muestra, con todos los costes que esto conlleva, y seguir aplicando los estimadores basados en el diseño utilizados actualmente en la estadística oficial.

Otra alternativa, en vías de investigación, es utilizar técnicas de estimación más complejas, asistidas y basadas en modelos.

El presente documento tiene como finalidad difundir los resultados de la tercera operación abordada según esta metodología en EUSTAT, la Encuesta sobre la Sociedad de la Información - familias (en adelante ESIf).

Un antecedente en el ámbito europeo del uso de estimación basada en modelos en estadística oficial se encuentra en la Oficina Nacional de Estadística de Reino Unido (Office for National Statistics, ONS). Allí las estimaciones basadas en modelos de datos de desempleo en "local authority" o municipalidades, a partir de su encuesta de fuerza de trabajo (Labour Force Survey, LFS), han sido aceptadas recientemente como estadística nacional. Es la primera vez que se otorga este rango en Reino Unido a estimaciones basadas en modelos (CLARKE y otros, 2007).

Otras estimaciones en áreas pequeñas obtenidas en la ONS aún tienen el rango de estadística experimental, lo que significa que aún están sujetas a posibles mejoras metodológicas.

En general se están dando pasos en el ámbito internacional en la aceptación de las estimaciones en áreas pequeñas como estadística oficial, considerada como aquella que cumple todos los requisitos del Código de Buenas Prácticas de las estadísticas oficiales. Por un lado, esto implica nuevos retos para la investigación en estos métodos, y por otro, la presentación y explicación adecuada de estos resultados a los usuarios.

En este documento se van a presentar varios aspectos. En la parte teórica, hay dos partes: en primer lugar, se van a exponer las principales características de la Encuesta sobre la Sociedad de la Información, con los estimadores de resultados y de errores que se utilizan (capítulo 2); a continuación, se va a tratar el sistema de estimación de las áreas pequeñas que se ha aplicado a la ESIf (capítulo 3).

En la parte aplicada se van a comentar los resultados obtenidos a partir de la mencionada encuesta, con esta metodología, para las comarcas de la C.A. de Euskadi. Se muestran resultados para las siguientes magnitudes: Usuarios de Internet, siempre para el colectivo de 15 y más años (capítulo 4) y dividido por sexos. Finalmente, se

extraen las conclusiones del trabajo (capítulo 5) y se muestra la Bibliografía. En Anexo, se detalla la división de municipios en comarcas de la C.A. de Euskadi.

## La Encuesta sobre la Sociedad de la Información - familias (ESIf)

### 2.1 Descripción de la Encuesta sobre la Sociedad de la Información - familias de la C.A. de Euskadi

La Encuesta sobre la Sociedad de la Información - familias (ESIf) se puso en marcha en el 2000, con el objetivo de disponer de información rica y detallada de la demanda creciente por parte de la ciudadanía de mayores cotas de información. El auge de los diversos canales de acceso, en los que Internet ocupa un lugar preferente, con amplias opciones de desarrollo futuro, motiva la implicación de Eustat en este ámbito de trabajo.

Más concretamente, la finalidad de la operación es producir información estadística continua sobre el acceso y el uso de Internet, así como otras tecnologías de la información. El acceso a la información esta unido al concepto mismo de bienestar y antes o después será considerado un Derecho Fundamental.

Esta información se obtiene para las principales características a nivel de Territorio Histórico, tal y como corresponde a su diseño muestral, que se comentará más adelante.

La población de referencia de la ESIf es la residente en viviendas familiares en la C.A. de Euskadi. El marco de la encuesta es el Directorio de Viviendas y el Registro Estadístico de Población de la C.A. de Euskadi. El primer año para el cual hay datos completos de la ESIf es 2000. Desde entonces la encuesta ha sufrido algún cambio en el tamaño y diseño muestral, así como en el marco de muestreo y los tratamientos de elevación. Los datos de la ESIf se obtienen a partir de un cuestionario específico y se realiza por muestreo sobre la población de la C. A. de Euskadi. Se trata de una encuesta ligada a la muestra y a los procesos de recogida de la Encuesta de Eustat sobre la Población en relación con la Actividad (PRA), que se viene realizando desde 1985. Actualmente se hacen dos tomas anuales: una en el 2º trimestre y otra en el 4º trimestre.

Este reparto de viviendas se realiza de modo proporcional a la raíz cuadrada del número de viviendas de los Territorios Históricos, para reducir las diferencias en el tamaño de población entre ellos. Para garantizar este reparto, los Territorios Históricos forman los estratos de la muestra.

En cada vivienda de la muestra la selección de la primera persona se realiza de forma aleatoria mediante tablas de Kish (en base a la semana de encuestación y el número de individuos, estudiantes y ocupados respectivamente de la vivienda); además, en el supuesto de que en la vivienda residan estudiantes y/o personas ocupadas, una de cada grupo será seleccionada por el mismo procedimiento. Desde el año 2003 se

completa la muestra con todas las niñas y niños entre 6 y 14 años residentes en la unidad familiar.

El paso de los datos muestrales a las estimaciones se realiza después de un proceso de elevación o calibración. En este proceso se calculan unos elevadores o pesos para individuos y familias, en función de las proyecciones de ambos tipos.

## 2.2 Estimadores utilizados en la Encuesta sobre la Sociedad de la Información - familias (ESIf) de la C.A. de Euskadi

### 2.2.1 Definición de los estimadores y fórmulas de elevación

Para estimar las características de la encuesta se considera el siguiente estimador y se detallan las formulas utilizadas en el cálculo:

#### **2.2.1.1 Estimador para personas mayores de 15 años**

En cada estrato  $h$  (Territorio Histórico), se obtienen también estimadores de calibración en dos fases:

Fase 1: Estimador de Horvitz Thompson, a partir del cálculo del factor de diseño ó inverso de la probabilidad de selección de cada persona. No hay falta de respuesta parcial, esto es, de personas dentro de la vivienda, así que no hay corrección por falta de respuesta.

$$\hat{X}_h = \sum_{i=1}^{v_h} \sum_{j=1}^{n_{v_i}} w_{hij} X_{hij}$$

donde:

$n_{v_i}$  es el número de personas muestreadas en la vivienda  $i$ .

$w_{hij} = w_{hi} p_{hij}^*$ , factor de diseño de la persona  $j$  de la vivienda  $i$ , siendo  $w_{hi}$  el factor del diseño de la vivienda  $i$

$p_{hij}^* = e_{hi}$  si la persona  $j$  es estudiante.

$= o_{hi}$  si la persona  $j$  es ocupado.

$= p_{hi}$  en otro caso.

donde:

$e_{hi}$  número de estudiantes en la vivienda  $i$ .

$o_{hi}$  número de ocupados en la vivienda  $i$ .

$p_{hi}$  número de personas muestreables en la vivienda  $i$ .

$X_{hij}$  es el valor de la característica a estimar de la persona  $j$  de la vivienda  $i$  del estrato  $h$ .

Fase 2: Ajuste por calibración.

$$\hat{X}_h^* = \sum_{i=1}^{v_i} \sum_{j=1}^{n_{ij}} w_{hij}^* X_{hij}$$

con  $w_{hij}^*$  obtenidos a partir de los factores iniciales  $w_{hij}$  aplicando post-estratificación según la variable cruce del territorio, grupo de edad (6 grupos: 15-24, 25-34, 35-44, 45-54, 55-64, >=65) y sexo. Se trata de población proyectada por TH, grupo de edad y sexo con referencia al día 15 del mes central del trimestre.

### 2.2.2 Método de estimación de los errores de muestreo

El método utilizado es el Método de Expansión de Taylor. Permite calcular estimaciones del error muestral para totales, medias y ratios en muestras con estratificación, clústers y probabilidades desiguales. El método obtiene aproximaciones lineales del estimador y calcula su varianza utilizando ésta como estimación de la varianza muestral.

La expresión para el cálculo de la varianza estimada para la media poblacional es la siguiente:

$$\bar{V}(\hat{Y}) = \sum_{h=1}^H \frac{n_h(1-f_h)}{n_h-1} \sum_{i=1}^{n_h} (e_{hi\cdot} - \bar{e}_{h\cdot\cdot})^2 \quad (2)$$

Donde:

$$e_{hi\cdot} = \left( \sum_{j=1}^{m_{hi}} w_{hij} (y_{hij} - \hat{Y}) \right) / w_{\dots}$$



$$\bar{e}_{h..} = \left( \sum_{i=1}^{n_h} e_{hi.} \right) / n_h$$

y

$$w_{...} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} w_{hij}$$

Notación:

$h = 1, 2, \dots, H$  indica el estrato con un total de  $H$  estratos.

$i = 1, 2, \dots, n_h$  indica el número de clusters en el estrato  $h$ , con un total de  $n_h$  clusters.

$j = 1, 2, \dots, m_{hi}$  indica el número de unidad dentro del cluster  $i$  del estrato  $h$ , con un total de  $m_{hi}$  unidades

$$n = \sum_{h=1}^H \sum_{i=1}^{n_h} m_{hi}$$

es el número total de observaciones en la muestra.

$w_{hij}$  indica el elevador de la observación  $j$  en el cluster  $i$  del estrato  $h$

$Y_{hij} = ( Y_{hij}(1), Y_{hij}(2), \dots, Y_{hij}(P) )$  son los valores observados de la variable  $Y$  en la observación  $j$  del cluster  $i$  del estrato  $h$ . (variables numéricas y categóricas).

Este cálculo se realiza con el procedimiento PROC SURVEYMEANS del paquete estadístico SAS (Sas Institute Inc. 2004).

## Sistema de Estimación en Áreas Pequeñas en la ESIf

### 3.1 Estudio de estimadores

La metodología de estimación se ha establecido a partir del análisis de diversos estimadores tanto clásicos como asistidos y basados en modelos para estimar el porcentaje de usuarios de Internet mayores de 15 años, por sexo, en 20 comarcas de la C. A. de Euskadi y en las 3 capitales de los Territorios Históricos.

Para ello se ha evaluado el error cuadrático medio de los estimadores calculados.

Se han evaluado:

- Estimadores basados en el diseño
- Estimadores basados en modelos

#### 3.1.1 Estimadores basados en el diseño.

##### 3.1.1.1 Directo:

$$\hat{y}_d^{directo} = \frac{\sum_{j=1}^{n_d} \tilde{w}_j y_j}{\sum_{j=1}^{n_d} \tilde{w}_j} N_d$$

donde  $y_j = 1$  (usuario de internet)  $y_j = 0$  (no usuario)

$N_d$  número de personas con edad > 15 en la zona d.

$n_d$  tamaño muestral en la zona d

d es el área pequeña (zona)

$\tilde{w}_j$  peso calibrado a partir del peso del diseño

##### 3.1.1.2 Posestratificado

$$\hat{y}_d^{post} = \sum_g \hat{y}_{dg} N_{dg}$$

donde

$N_{dg}$  número de personas con edad > 15 en la zona d y en el grupo g.

$\hat{y}_{dg}$  es la media calculada con el estimador directo anterior,  $\hat{y}_{dg} = \frac{\sum_{j \in S_{dg}} \tilde{w}_j y_j}{\sum_{j \in S_{dg}} \tilde{w}_j}$

### 3.1.1.3 Sintético

$$\hat{y}_d^{\text{sin } t} = \sum_g \hat{y}_g N_{dg}$$

donde

$N_{dg}$  número de personas con edad > 15 en la zona d y en el grupo g.

$\hat{y}_g$  es la media calculada con el estimador directo anterior

### 3.1.1.4 Compuestos

$\hat{y}_d^{\text{dep}} = \lambda_d \hat{y}_d^{\text{post}} + (1 - \lambda_d) \hat{y}_d^{\text{sin } t}$  donde  $\hat{y}_d^{\text{post}}$  es el estimador postestratificado y  $\hat{y}_d^{\text{sin } t}$  es el estimador sintético calculado con la media por grupo del total en el área calculada con el estimador directo.

y donde  $0 \leq \lambda_d \leq 1$  viene dada por

$$\lambda_d = \begin{cases} 1 & \text{si } \hat{N}_d \geq \alpha N_d \\ \frac{\hat{N}_d}{\alpha N_d} & \text{en otro caso} \end{cases}$$

$\hat{N}_d = \sum_d w_j$  es el total poblacional estimado en cada área d y  $\alpha$  es un parámetro. Evaluamos el estimador compuesto para distintos valores de  $\alpha = \frac{2}{3}, 1, 1.5$  y  $2$ . En adelante compuesto1 ( $\alpha = \frac{2}{3}$ ), compuesto2 ( $\alpha = 1$ ), compuesto3 ( $\alpha = 1.5$ ) y compuesto4 ( $\alpha = 2$ ).

Se calculan los estimadores anteriores para los siguientes grupos g:

- Grupo de edad y sexo. (12 categorías, 6 por sexo: 15-24, 25-34, 35-44, 45-54, 55-64 y  $\geq 65$ )
- Grupo de edad y sexo (12 categorías, 6 por sexo: 15-24, 25-34, 35-44, 45-54, 55-64 y  $\geq 65$ ) \* Nivel educativo reagrupado (2 categorías: "Medios superiores y Superiores" y "Secundarios y profesionales y Primarios o menos").

### 3.1.2 Estimadores basados en modelos.

#### 3.1.2.1 Estimadores basados en un modelo lineal mixto

Cuando se plantea un modelo lineal no se controla que la predicción obtenida esté en el intervalo  $[0,1]$ . Se está modelizando la probabilidad de ser usuario de Internet, por ello los valores obtenidos deberían estar entre cero y uno. En caso de que se obtenga una predicción negativa se transforma en 0 y en el caso de que sea mayor que 1 se transforma en 1.

Se ha estudiado un modelo lineal mixto desagregado por sexo, es decir un modelo para mujeres y otro para hombres, con efecto aleatorio de área. La información auxiliar empleada fueron los grupos de edad y sexo (6 categorías por sexo).

$$y_{dj} = \beta_1 x_{dj1} + \beta_2 x_{dj2} + \beta_3 x_{dj3} + \beta_4 x_{dj4} + \beta_5 x_{dj5} + \beta_6 x_{dj6} + v_d + e_{dj}$$

con  $d = 1, \dots, D_s$ , y  $j = 1, \dots, n_d$ ,

Donde:

$y_{dj}$  es la variable de interés para el individuo  $j$  de la comarca  $d$ ,

$x_{dj1}, \dots, x_{dj6}$  valores de la variable auxiliar (categorías de edad-sexo),

$\beta_1, \dots, \beta_6$  son los efectos fijos del modelo,

$v_d$  efecto aleatorio común para todos los individuos de la comarca  $d$ ,

$e_{dj}$  son los errores aleatorios específicos de cada individuo.

El modelo superpoblacional en forma matricial sería el siguiente:

$$Y = X\beta + Zv + \varepsilon, \quad v \sim N(0, \sigma_v^2 I_D), \quad \varepsilon \sim N(0, \sigma_e^2 I_N)$$

Una vez desarrollada la versión proyectiva del modelo con la parte muestreada y la no muestreada, los predictores del modelo quedarían de la siguiente forma:

La media:  $\hat{y}_d^* = \bar{X}'_{d(p)} \hat{\beta} + \hat{\gamma}_d (\bar{y}_d - \bar{x}'_{d(p)} \hat{\beta})$  con  $d=1, \dots, D$  y donde  $\hat{\beta} = \tilde{\beta}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$  ha sido evaluado con las estimaciones de los componentes de varianza

$\bar{X}'_{d(p)} = (1, \bar{x}_{d(p)})$  y  $\bar{x}_{d(p)} = \frac{\sum_{j \in N_d} x_{dj}}{N_d}$  es la media poblacional de la variable auxiliar

para una comarca  $d$  (muestreados o no).

El total:  $\hat{t}_d^* = X'_{d(p)} \hat{\beta} + N_d \hat{\gamma}_d (\bar{y}_d - \bar{x}'_d \hat{\beta})$  con  $d=1, \dots, D$  donde  $X'_{d(p)} = (N_d, X_{d(p)})$ ,  $N_d$  es el número total de individuos en la comarca  $d$ . Si se desarrolla esta versión del predictor del total basado en modelos se llega a un estimador compuesto de la siguiente forma:

$$\hat{y}_d = \hat{\gamma}_d [\bar{y}_d + (\bar{X}_{d(p)} - \bar{x}_d)' \hat{\beta}] + (1 - \gamma_d) \bar{X}'_{d(p)} \hat{\beta} \quad \text{con } d=1, \dots, D \quad \hat{\gamma}_d = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \frac{\hat{\sigma}_e^2}{n_d}}$$

Es un estimador compuesto donde la primera parte es la suma ponderada del estimador de regresión generalizado y la segunda parte es un estimador sintético basado en el modelo. En este caso  $\hat{\gamma}_d$  (que es el equivalente de  $\lambda_d$  para el compuesto) depende de la varianza relativa  $\hat{\sigma}_v^2$ , de la varianza total  $\hat{\sigma}_v^2 + \hat{\sigma}_e^2$  y del tamaño de la muestra en el área  $n_d$ .

Este cálculo se realiza con el procedimiento PROC MIXED del paquete estadístico SAS (Sas Institute Inc. 2004). Este procedimiento usa por defecto el método de máxima verosimilitud restringida (REML) para la estimación de las componentes de varianza  $\sigma^2 = (\sigma_e^2, \sigma_v^2)$

### 3.1.3 Conclusiones

Una vez realizado el estudio con todos los estimadores basados en el diseño, el estimador más adecuado por comarcas ha resultado ser el compuesto 4, que incluye como variables auxiliares el sexo y edad. El compuesto 4, con  $\alpha=2$ , es, entre los cuatro compuestos, el que mas peso da a la parte sintética, es decir a la parte prestada del resto de dominios en forma de proporciones de usuarios de internet. Es el estimador con menor error y el más estable en sus estimaciones.

Los resultados obtenidos para 20 comarcas y las 3 capitales no difieren considerablemente entre el Modelo y el Compuesto 4, eran muy similares. En las comarcas más grandes, la estimación basada en el modelo es ligeramente mayor que la del compuesto y en las demás comarcas es ligeramente menor que la del compuesto. Las estimaciones proporcionadas por el modelo son más inestables en la comarcas con menor muestra. Las estimaciones del estimador Compuesto 4 son más estables incluso en las comarcas donde la muestra no está bien repartida entre la población.

Respecto a los errores, los errores de ambos estimadores son bajos pero no son comparables ya que se han calculado por diferentes métodos.

Analizando la estabilidad y los errores de las diferentes estimaciones, se ha decidido emplear el estimador compuesto 4.

## 3.2 Estimación del error cuadrático medio

### 3.2.1 Procedimientos para el cálculo del error cuadrático medio (MSE)

Se han utilizado tres métodos de estimación de los correspondientes errores cuadráticos medios, el método de linealización de la varianza y los métodos de remuestreo siguientes: el método jackknife y el método bootstrap. Se estudia el comportamiento del estimador del MSE utilizando los tres métodos.

Se evaluaron los 3 métodos para los estimadores compuestos y los estimadores postestratificado y sintético.

Los métodos de remuestreo se basan en la evaluación de los estadísticos en remuestras ó submuestras obtenidas a partir de los datos originales, y mediante esos valores se obtienen estimadores de las medidas de exactitud ó de la distribución muestral del estadístico.

En el caso del método jackknife se dispone de tantas submuestras como clusters tenga la muestra, ya que se obtienen por sucesivas eliminaciones de clusters en la muestra original. Para cada submuestra se definen nuevos pesos y se calcula el estimador (postestratificado, sintético ó compuesto). Luego se obtiene la varianza y sesgo de los estimadores como se detalla posteriormente.

En el método bootstrap, las submuestras se obtienen mediante muestreo aleatorio simple, pero se ha de determinar cuántas son necesarias. De forma análoga, para cada submuestra se definen nuevos pesos y se calcula cada estimador. Con dichas estimaciones se obtiene el error cuadrático medio como se detalla posteriormente.

En lo que sigue se utilizan los siguientes índices:

- $h$  es el número de estrato, donde  $h = 1, 2, \dots, H$
- $i$  es el cluster  $i$ -ésimo en el estrato  $h$ , donde  $i = 1, 2, \dots, n_h$
- $j$  es la unidad  $j$ -ésima del cluster  $i$  en el estrato  $h$ , donde  $j = 1, 2, \dots, m_{hi}$

Se ha considerado que el estrato  $h$  es el territorio histórico (th), el cluster  $i$  es la vivienda (que se identifica mediante la variable  $nt1\_idev$ ) y por último  $j$  es la persona encuestada en dicha vivienda.

Los errores del estimador basado en el Modelo Lineal Mixto se calculan en base a la formulación de Prasad y Rao (1990), que proporcionan un estimador del error

cuadrático medio (MSE) del predictor de la media en la versión proyectiva, válido cuando los estimadores de los componentes de varianza se han obtenido por REML o por el método de los momentos.

### 3.2.1.1 Método de linealización de la varianza.

El método de linealización ó método delta consiste en aplicar un desarrollo en serie de Taylor.

Definimos para cada dominio D las siguientes variables indicadoras:

$$I_D(h, i, j) = \begin{cases} 1 & \text{si } (h, i, j) \text{ está en } D \\ 0 & \text{en otro caso} \end{cases}$$

$$z_{hij} = y_{hij} I_D(h, i, j) = \begin{cases} y_{hij} & \text{si } (h, i, j) \text{ está en } D \\ 0 & \text{en otro caso} \end{cases}$$

$$v_{hij} = w_{hij} I_D(h, i, j) = \begin{cases} w_{hij} & \text{si } (h, i, j) \text{ está en } D \\ 0 & \text{en otro caso} \end{cases}$$

El estimador de la media en un dominio D viene dado por la expresión:

$$\hat{Y}_D = \left( \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} v_{hij} z_{hij} \right) / v_{\dots}, \text{ donde } v_{\dots} = \sum_{h=1}^H \sum_{i=1}^{n_h} \sum_{j=1}^{m_{hi}} v_{hij}$$

La varianza linealizada del estimador de la media en un dominio D viene dada por:

$$V\hat{a}r_L(\hat{Y}_D) = \sum_{h=1}^H V\hat{a}r_h(\hat{Y}_D) \text{ donde } V\hat{a}r_h(\hat{Y}_D) = \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} (U_{hi} - \bar{U}_{h..})^2$$

$$U_{hi} = \frac{1}{v_{\dots}} \sum_{j=1}^{m_{hi}} v_{hij} (z_{hij} - \hat{Y}_D) \text{ y } \bar{U}_{h..} = \frac{1}{n_h} \sum_{i=1}^{n_h} U_{hi}$$

### 3.2.1.2 Método jacknife para la estimación de la varianza y del sesgo.

Para aplicar el método jacknife al esquema de muestreo utilizado en la ESIf se ha de eliminar un cluster (vivienda) cada vez. Definimos unos nuevos pesos, dados por:

$$w_{j(hi)} = \begin{cases} w_{hij} & \text{si la unidad } j \text{ no está en el estrato } h \\ 0 & \text{si la unidad } j \text{ está en el cluster } i \text{ del estrato } h \\ \frac{n_h}{n_h - 1} w_{hij} & \text{si la unidad } j \text{ está en el estrato } h \text{ pero no en el cluster } i \end{cases}$$

Sea:

- $\hat{\theta}$  el estimador compuesto 4 obtenido con los datos de una simulación utilizando pesos  $w_{hij}$
- $\hat{\theta}_{(hi)}$  el estimador compuesto 4 obtenido con los datos de la submuestra resultante de eliminar el cluster  $i$  (vivienda) del estrato  $h$  (th) de dicha simulación y utilizando los pesos  $w_{j(hi)}$

El estimador jacknife de la varianza en el estrato  $h$  viene dado por:

$$\hat{V}ar_{JK(h)}(\hat{\theta}) = \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} [\hat{\theta}_{(hi)} - \hat{\theta}_{(h.)}]^2$$

Donde:

$$\hat{\theta}_{(h.)} = \frac{1}{n_h} \sum_{i=1}^{n_h} \hat{\theta}_{(hi)}$$

El estimador jacknife del sesgo de un estimador en el estrato  $h$  viene dado por:

$$Se\hat{s}go_{JK(h)}(\hat{\theta}) = (n_h - 1)(\hat{\theta}_{(h.)} - \hat{\theta})$$

El estimador jacknife del MSE en el estrato  $h$ :

$$M\hat{S}E_{JK(h)}(\hat{\theta}) = \hat{V}ar_{JK(h)}(\hat{\theta}) + Se\hat{s}go_{JK(h)}^2(\hat{\theta})$$

Como los estratos son independientes el MSE del estimador viene dado por:

$$M\hat{S}E_{JK}(\hat{\theta}) = \sum_{h=1}^H \left[ \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} [\hat{\theta}_{(hi)} - \hat{\theta}_{(h.)}]^2 + ((n_h - 1)(\hat{\theta}_{(h.)} - \hat{\theta}))^2 \right]$$

### 3.2.1.3 Estimación bootstrap del error cuadrático medio.



A continuación se describen los pasos a dar para construir la versión del bootstrap reescalado en un muestreo simple estratificado propuesta por Rao y Wu (1988).

1. Fijado un estrato  $h$ , tenemos una muestra con  $n_h$  clusters. Extraemos una submuestra con  $n_h - 1$  clusters mediante muestreo aleatorio simple con reemplazamiento de la muestra del estrato  $h$ . Repetimos este proceso de forma independiente en cada estrato.

2. Para cada submuestra  $r$  ( $r=1,2,..R$ ) construimos un nuevo peso:

$$w_{hij}(r) = w_{hij} \frac{n_h}{n_h - 1} m_i(r)$$

donde  $m_i(r)$  es el número de veces que el cluster

$i$  es seleccionado en la submuestra, y calculamos  $\hat{\theta}_r^*$  utilizando los nuevos pesos  $w_{hij}(r)$ .

3. Repetimos los pasos 1 y 2,  $R$  veces.

4. Para obtener el estimador bootstrap del error cuadrático medio realizamos:

$$M\hat{S}E_B(\hat{\theta}) = \frac{1}{R-1} \sum_{r=1}^R (\hat{\theta}_r^* - \hat{\theta})^2$$

donde:

- $\hat{\theta}$  es el estimador compuesto 4 obtenido con los datos de una simulación utilizando pesos  $w_{hij}$ .
- $\hat{\theta}_r^*$  es el estimador compuesto 4 obtenido con los datos de la submuestra  $r$  utilizando los pesos  $w_{hij}(r)$ .

Una de las cuestiones a decidir es el tamaño de  $R$  para que el método funcione correctamente. Se consideran distintos valores de  $R$ , a partir de un cierto tamaño se estabiliza y a la vista de los resultados se opta por utilizar  $R=200$ .

### 3.2.1.4 Estimación del error cuadrático medio para el Modelo.

El error cuadrático medio se calcula como suma de las aproximaciones de las contribuciones de la estimación de los efectos fijos y aleatorios del modelo.

Prasad y Rao (1990) proporcionan un estimador del error cuadrático medio (MSE) del predictor de la media en la versión proyectiva, válido cuando los estimadores de los componentes de varianza se han obtenido por REML o por el método de los momentos. Viene dado por:

$$M\hat{S}E[\hat{y}_{d(p)}] = g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2) + 2g_{3d}(\hat{\sigma}^2)$$

y como el tercer término es muy pequeño, que representa el error debido a la estimación de las componentes de varianza, se puede despreciar y en la práctica se calcula el siguiente:

$$M\hat{S}E[\hat{y}_{d(p)}] \approx g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2)$$

Donde:

$$g_{2d}(\hat{\sigma}^2) = (\bar{X}_{d(p)} - \hat{\gamma}_d \bar{x}_d)' \hat{\Phi}_s (\bar{X}_{d(p)} - \hat{\gamma}_d \bar{x}_d)$$

$$g_{1d}(\hat{\sigma}^2) = (1 - \hat{\gamma}_{dc}) \hat{\sigma}_v^2$$

$$\hat{\Phi}_s = (X_s' \hat{V}_s^{-1} X_s)^{-1}$$

El MSE de predictor del total por comarca se estima multiplicando el estimador del MSE de la media por el cuadrado del tamaño poblacional de la comarca  $N_d^2$ . En efecto

$$M\hat{S}E[\hat{t}_d] \approx N_d^2 [g_{1d}(\hat{\sigma}^2) + g_{2d}(\hat{\sigma}^2)]$$

$g_{1d}(\hat{\sigma}^2)$  representa el error cometido en la estimación (supuestos conocidos todos los parámetros, los  $\beta$  y las varianzas), es decir el error al estimar los efectos aleatorios.

$g_{2d}(\hat{\sigma}^2)$  representa el error cometido debido a la estimación de las  $\beta$ .

$g_{3d}(\hat{\sigma}^2)$  representa la estimación del resto de los parámetros (estimación de los parámetros de varianza) y disminuye en cuanto aumenta el número de áreas pequeñas (Cuando el número de áreas del estudio es pequeño, no es recomendable usar modelos mixtos de efectos aleatorios porque los errores aumentan).

La mayor contribución al error suele ser debida a la estimación de los efectos fijos  $g_2$ , seguida de la estimación de los efectos aleatorios  $g_1$ .

### 3.2.2 Conclusiones

Los análisis realizados muestran que el estimador Bootstrap es el estimador más adecuado para el cálculo del error cuadrático medio de los estimadores compuestos y los estimadores postestratificado y sintético. El Método de linealización de la varianza y Método jackknife para la estimación de la varianza y del sesgo presentan errores muy parecidos al Bootstrap. Computacionalmente el Bootstrap es más eficiente.

En la estimación del error del modelo se calculo extraordinariamente el coeficiente  $g_{3d}(\hat{\sigma}^2)$  de segundo orden y efectivamente, no aportaba demasiado al error y su calculo complicaba demasiado el proceso.

Tanto los errores del estimador Compuesto 4 y los del Modelo Lineal Mixto con efectos aleatorios son bajos. Estos errores obtenidos con métodos tan distintos no se pueden comparar. Uno esta basado en el diseño y el otro esta basado en el Modelo.

### 3.3 Software empleado

Para el estudio de esta metodología y la aplicación de los estimadores indicados anteriormente se ha utilizado la programación informática basada en SAS. Se han elaborado programas específicos, diseñados a modo de macro informática, que ejecutan las diferentes tareas descritas: elaboración de estimaciones por comarcas y el cálculo de los errores cuadráticos medios para los diferentes métodos.

La macro proporciona estimaciones calculadas mediante el estimador compuesto (el parámetro alpha es un parámetro de entrada) y al ser éste combinación de un estimador postestratificado y un sintético, también proporciona las estimaciones calculadas mediante éstos.

Otros parámetros de entrada de esta macro son: la variable a estimar, las variables auxiliares a utilizar, la opción de calibración de las estimaciones comarcales a las territoriales obtenidas mediante la estimación directa de la encuesta, el método de estimación de los errores cuadráticos medios (y en el caso del bootstrap, el valor de R, numero de submuestras).

El programa también ofrece la posibilidad de obtener medias móviles de varios trimestres consecutivos. Además se calculan los errores cuadráticos medios y sus correspondientes coeficientes de variación (cociente de la raíz cuadrada del error cuadrático medio y la estimación), con los tres métodos mencionados.

Esta macro se aplica trimestralmente a las muestras de la ESIf y se obtienen, con los parámetros determinados como mejores, las estimaciones de las magnitudes mencionadas y sus errores cuadráticos medios.

## Estimaciones comarcales 2005-2008

### 4.1 Definiciones

A continuación se presentan las estimaciones obtenidas utilizando el sistema de estimación antes expuesto en la Encuesta sobre la Sociedad de la Información - familias (ESIf), para los años 2005-2008.

Las estimaciones se refieren al porcentaje de usuarios de Internet de 15 o más años que se han conectado a Internet, ya sea en el hogar, en el centro de trabajo o centro de estudio o en otro lugar, en las 20 comarcas de la C.A. de Euskadi y en las capitales de los Territorios Históricos, por sexos.

Junto con las estimaciones se ofrecen en las tablas los coeficientes de variación (CV) de las mismas.

La división oficial en comarcas de la C.A. de Euskadi es la siguiente:

Alava: Valles Alaveses, Llanada Alaveses, Montaña Alaveses, Rioja Alaveses, Estribaciones del Gorbea y Cantábrica Alaveses:

Bizkaia: Arratia-Nervión, Gran Bilbao, Duranguesado, Encartaciones, Gemika-Bermeo, Markina-Ondarroa y Plentzia-Mungia

Gipuzkoa: Bajo Bidasoa, Bajo Deba, Alto Deba, Donostia-San Sebastián, Goierri, Tolosa y Urola Costa

(Ver Anexo, con la relación de comarcas y municipios)



Seguidamente, se pasa a presentar los resultados más destacables.

## 4.2 Resultados

**Tabla 1. Población de 15 y más usuaria de Internet por sexo, Territorio Histórico y comarca (%). 2005-2008**

Fuente: EUSTAT. Encuesta de la Sociedad de la Información. Familias (ESIF)

	2005			2006			2007			2008		
	Estimaciones			Estimaciones			Estimaciones			Estimaciones		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
<b>C.A. de Euskadi</b>	<b>37,8</b>	<b>41,6</b>	<b>34,2</b>	<b>41,3</b>	<b>44,6</b>	<b>38,2</b>	<b>46,7</b>	<b>49,7</b>	<b>43,6</b>	<b>49,9</b>	<b>54,0</b>	<b>46,0</b>
<b>Alava</b>	<b>38,3</b>	<b>39,5</b>	<b>37,2</b>	<b>40,7</b>	<b>42,1</b>	<b>39,0</b>	<b>45,6</b>	<b>47,9</b>	<b>43,3</b>	<b>50,6</b>	<b>54,1</b>	<b>47,0</b>
Valles Alaveses	35,2	37,1	32,9	35,0	36,6	33,0	42,1	41,9	42,4	45,0	46,2	43,4
Llanada Alavesa	39,0	40,4	37,7	41,5	43,0	40,0	46,4	48,7	44,0	51,5	55,1	48,0
Montaña Alavesa	29,6	31,9	26,6	30,0	32,3	27,2	35,7	40,2	29,6	34,9	42,3	24,8
Rioja Alavesa	35,9	37,7	33,9	37,4	36,7	38,0	40,4	42,2	38,0	44,6	48,8	39,9
Estribaciones del Gorbea	36,1	37,7	34,4	39,4	41,0	37,6	44,1	44,5	43,6	47,9	51,3	44,3
Cantábrica Alavesa	35,8	35,5	36,2	38,6	39,8	37,4	43,4	47,3	39,6	48,6	52,5	44,9
<b>Bizkaia</b>	<b>36,1</b>	<b>41,1</b>	<b>31,5</b>	<b>40,7</b>	<b>45,1</b>	<b>36,6</b>	<b>47,1</b>	<b>50,4</b>	<b>44,0</b>	<b>49,5</b>	<b>54,1</b>	<b>45,3</b>
Arratia-Nervión	35,2	37,0	33,3	39,2	42,9	35,4	48,5	50,5	46,4	49,4	53,3	45,6
Gran Bilbao	36,0	41,2	31,0	40,6	45,3	36,4	46,9	50,5	43,0	49,3	54,0	45,0
Duranguesado	37,3	41,2	33,5	41,7	44,8	38,6	47,9	50,2	45,6	51,3	55,3	47,4
Encartaciones	36,1	39,6	32,6	39,2	41,4	37,0	45,6	49,6	41,6	47,8	52,7	42,9
Gernika-Bermeo	33,7	38,7	29,0	38,7	42,3	35,2	44,6	45,3	44,0	47,0	49,8	44,2
Markina-Ondarroa	35,1	40,0	30,2	39,9	44,0	35,6	45,5	48,6	42,9	49,5	53,3	45,5
Plentzia-Mungia	39,2	43,7	34,8	44,3	49,9	38,9	52,3	55,5	49,1	54,3	58,7	49,9
<b>Gipuzkoa</b>	<b>40,4</b>	<b>43,3</b>	<b>37,7</b>	<b>42,6</b>	<b>44,9</b>	<b>40,4</b>	<b>46,4</b>	<b>49,3</b>	<b>43,6</b>	<b>50,1</b>	<b>53,7</b>	<b>46,7</b>
Bajo Bidasoa	41,3	44,4	38,0	44,5	47,9	41,0	49,0	52,3	45,6	51,6	55,9	47,4
Bajo Deba	38,0	41,5	34,6	39,7	42,6	36,9	43,6	46,6	40,6	46,9	50,0	43,9
Alto Deba	39,8	43,2	36,6	41,2	43,9	38,4	44,4	46,7	42,0	48,5	51,2	45,7
Donostialdea	40,9	43,9	38,1	43,2	45,4	41,1	47,0	50,3	44,1	50,9	55,2	47,0
Goierri	39,0	41,1	36,6	41,8	43,6	40,0	45,9	49,3	42,4	48,9	52,4	45,4
Tolosa	39,3	42,0	36,6	40,6	43,1	38,1	43,7	47,0	40,4	47,6	50,4	44,7
Urola Costa	42,1	44,3	39,9	43,2	44,5	42,0	46,8	47,6	46,0	52,0	53,8	50,2

**Tabla 2. Población de 15 y más usuaria de Internet por sexo, Territorio Histórico y capital (%). 2005-2008**

Fuente: EUSTAT. Encuesta de la Sociedad de la Información. Familias (ESIF)

	2005			2006			2007			2008		
	Estimaciones			Estimaciones			Estimaciones			Estimaciones		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
<b>C.A. de Euskadi</b>	<b>37,8</b>	<b>41,6</b>	<b>34,2</b>	<b>41,3</b>	<b>44,6</b>	<b>38,2</b>	<b>46,7</b>	<b>49,7</b>	<b>43,8</b>	<b>49,9</b>	<b>54,0</b>	<b>46,0</b>
<b>Alava</b>	<b>38,3</b>	<b>39,5</b>	<b>37,2</b>	<b>40,7</b>	<b>42,1</b>	<b>39,5</b>	<b>45,6</b>	<b>47,9</b>	<b>43,3</b>	<b>50,6</b>	<b>54,1</b>	<b>47,0</b>
Vitoria-Gasteiz	39,0	40,4	37,7	41,5	43,1	39,9	46,3	48,8	44,0	51,4	55,1	47,8
<b>Bizkaia</b>	<b>36,1</b>	<b>41,1</b>	<b>31,5</b>	<b>40,7</b>	<b>45,1</b>	<b>36,6</b>	<b>47,1</b>	<b>50,4</b>	<b>44,0</b>	<b>49,5</b>	<b>54,1</b>	<b>45,3</b>
Bilbao	35,8	41,6	30,6	40,5	45,8	35,9	46,0	50,1	42,4	48,5	54,0	43,6
<b>Gipuzkoa</b>	<b>40,4</b>	<b>43,3</b>	<b>37,7</b>	<b>42,6</b>	<b>44,9</b>	<b>40,4</b>	<b>46,4</b>	<b>49,3</b>	<b>43,6</b>	<b>50,1</b>	<b>53,7</b>	<b>46,7</b>
Donostia-San Sebastián	42,4	45,6	39,6	44,6	47,6	42,0	48,1	52,4	44,5	51,8	56,6	47,8

**Tabla 3. Coeficientes de variación de la población de 15 y más usuaria de Internet por sexo, Territorio Histórico y comarca (%). 2005-2008**

Fuente: EUSTAT. Encuesta de la Sociedad de la Información. Familias (ESIF)

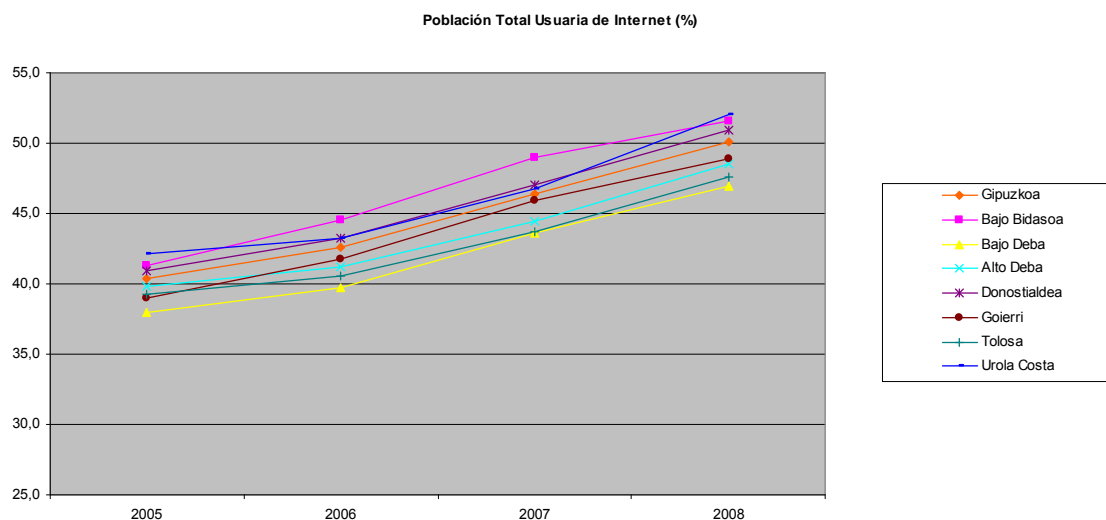
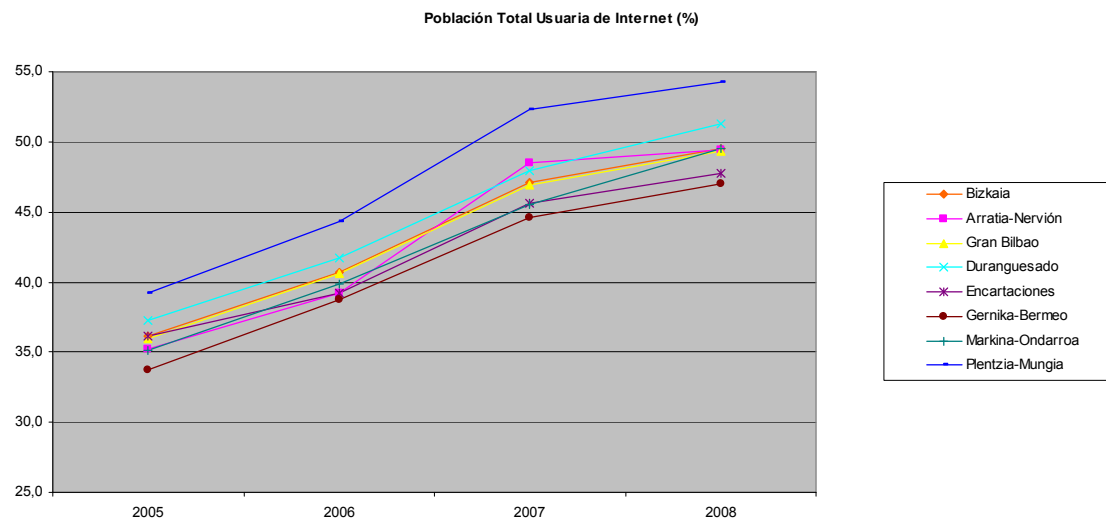
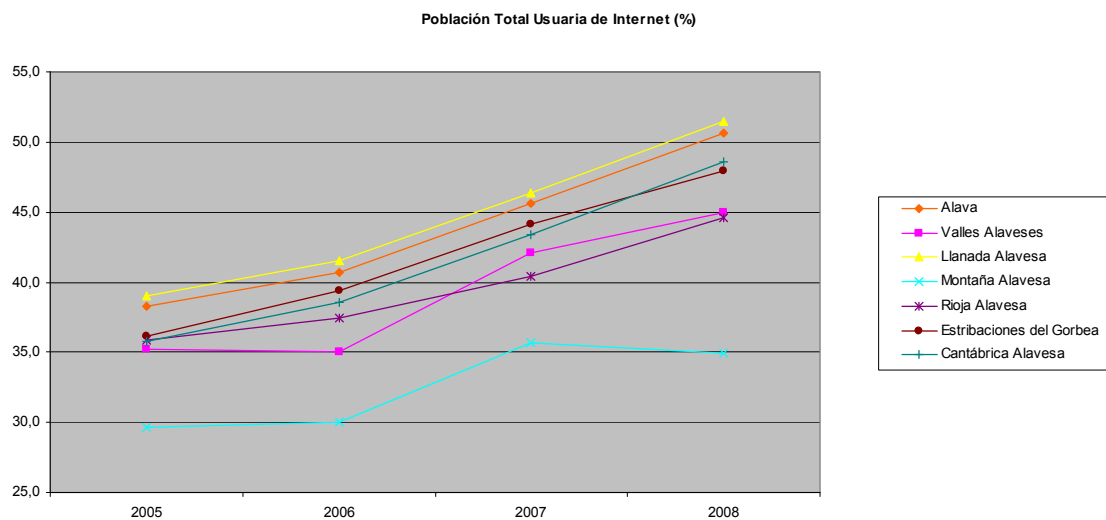
	2005			2006			2007			2008		
	cv			cv			cv			cv		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
<b>C.A. de Euskadi</b>	1,7	2,4	2,5	1,7	2,3	2,4	1,6	2,2	2,2	1,5	2,1	2,2
<b>Alava</b>	3,5	4,7	5,1	3,4	4,7	5,0	3,3	4,6	4,7	3,1	4,4	4,5
Valles Alaveses	3,4	4,4	5,3	4,1	5,4	6,5	3,7	5,2	5,1	3,5	5,2	4,0
Llanada Alavesa	3,3	4,2	5,0	3,1	4,3	4,6	3,0	3,9	4,6	2,9	3,9	4,1
Montaña Alavesa	4,7	5,9	7,7	5,4	7,7	7,4	5,5	7,1	8,9	7,0	7,3	16,1
Rioja Alavesa	4,6	6,7	6,0	5,1	7,4	7,0	5,0	6,7	7,5	4,3	5,6	6,7
Estribaciones del Gorbea	3,7	5,1	5,5	3,6	4,6	5,7	4,5	6,5	6,1	3,9	4,8	6,4
Cantábrica Alavesa	3,6	4,6	5,5	3,5	4,9	5,1	3,3	4,3	5,1	3,0	4,0	4,5
<b>Bizkaia</b>	2,6	3,5	3,9	2,5	3,4	3,7	2,3	3,2	3,3	2,3	3,0	3,4
Arratia-Nervión	4,0	6,0	5,0	4,4	6,5	5,8	3,2	3,5	5,5	3,4	3,7	5,8
Gran Bilbao	2,4	3,1	3,6	2,3	3,0	3,6	2,1	2,8	3,2	2,0	2,7	3,1
Duranguesado	2,9	3,7	4,6	2,8	4,0	4,0	2,3	3,0	3,6	2,2	3,0	3,3
Encartaciones	3,4	4,0	5,8	4,0	4,8	6,4	3,9	5,6	5,3	3,7	5,7	4,3
Gernika-Bermeo	3,4	4,2	5,4	3,1	4,5	4,0	3,1	4,2	4,4	2,6	3,6	3,9
Markina-Ondarroa	3,3	4,0	5,6	3,5	4,2	5,8	3,4	4,2	5,6	2,9	3,5	4,9
Plentzia-Mungia	3,4	4,7	4,7	3,3	4,5	4,8	2,6	3,5	3,8	2,2	3,0	3,3
<b>Gipuzkoa</b>	2,8	3,9	4,1	2,7	3,9	3,9	2,6	3,6	3,6	2,5	3,5	3,6
Bajo Bidasoa	3,1	4,2	4,6	3,1	4,3	4,5	2,7	3,9	3,8	2,8	4,0	3,8
Bajo Deba	3,4	4,5	5,0	3,2	4,2	5,0	2,8	4,0	3,9	2,5	3,7	3,4
Alto Deba	3,2	4,3	4,7	3,1	4,4	4,5	2,8	3,8	4,0	2,6	3,7	3,7
Donostialdea	2,8	3,9	4,0	2,7	3,7	3,9	2,5	3,6	3,4	2,4	3,3	3,4
Goierrri	3,2	4,5	4,0	3,1	4,4	4,3	2,6	3,5	3,9	2,6	3,6	3,6
Tolosa	3,3	4,9	4,3	3,5	4,9	5,1	3,3	4,7	4,6	3,3	4,6	4,7
Urola Costa	3,2	4,5	4,6	3,1	4,3	4,4	2,8	4,0	3,9	2,4	3,6	3,1

**Tabla 4. Coeficientes de variación de la población de 15 y más usuaria de Internet por sexo, Territorio Histórico y capital (%). 2005-2008**

Fuente: EUSTAT. Encuesta de la Sociedad de la Información. Familias (ESIF)

	2005			2006			2007			2008		
	cv			cv			cv			cv		
	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres	Total	Hombres	Mujeres
<b>C.A. de Euskadi</b>	1,7	2,4	2,5	1,7	2,3	2,4	1,6	2,2	2,2	1,5	2,1	2,2
<b>Alava</b>	3,5	4,7	5,2	3,4	4,7	5,0	3,3	4,6	4,7	3,1	4,4	4,5
Vitoria-Gasteiz	3,3	4,3	5,1	3,2	4,3	4,6	3,0	4,0	4,6	2,9	4,0	4,4
<b>Bizkaia</b>	2,6	3,5	3,9	2,5	3,4	3,7	2,3	3,2	3,3	2,3	3,0	3,4
Bilbao	2,7	3,5	4,2	2,6	3,4	3,9	2,3	3,2	3,4	2,3	3,0	3,5
<b>Gipuzkoa</b>	2,8	3,9	4,1	2,7	3,9	3,9	2,6	3,6	3,6	2,5	3,5	3,6
Donostia-San Sebastián	3,1	4,3	4,4	2,8	3,9	4,1	2,6	3,7	3,7	2,6	3,6	3,6

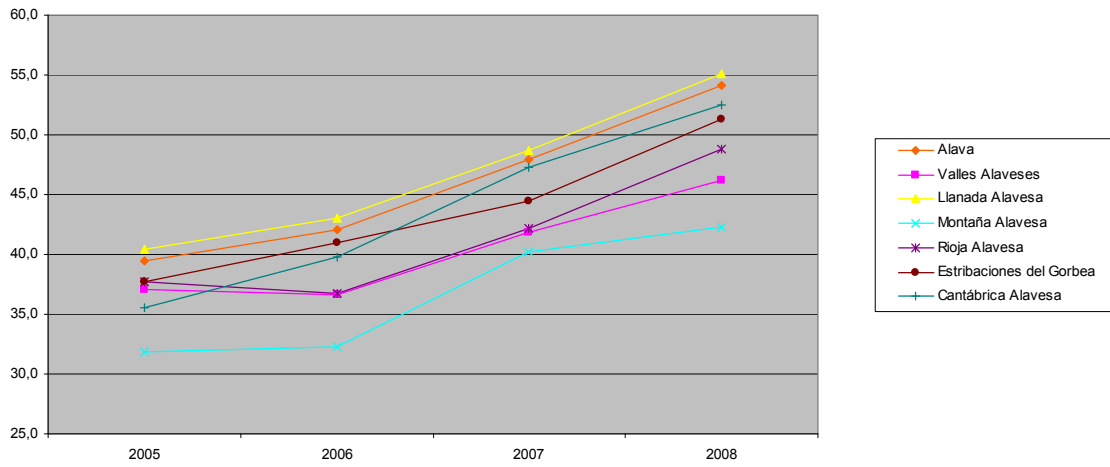
ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS DE LA C.A. DE EUSKADI



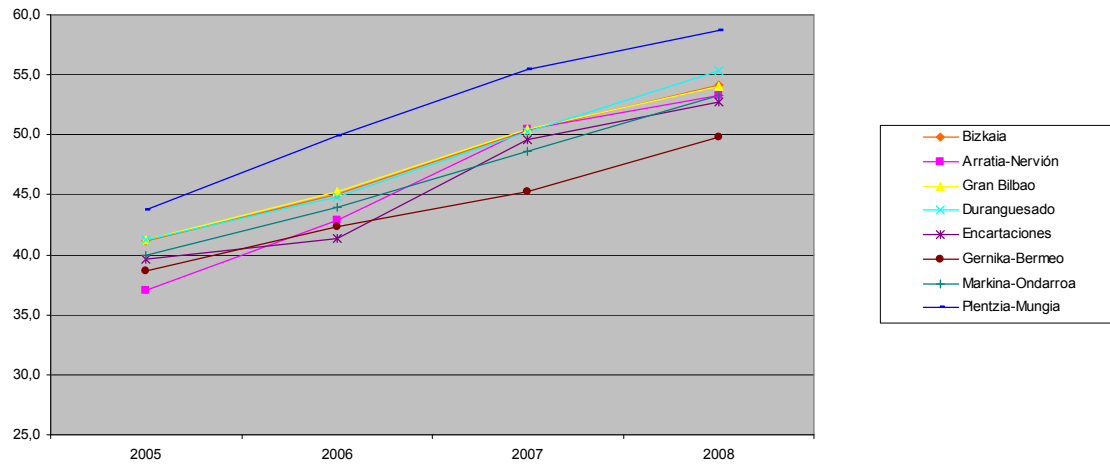


ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS DE LA C.A. DE EUSKADI

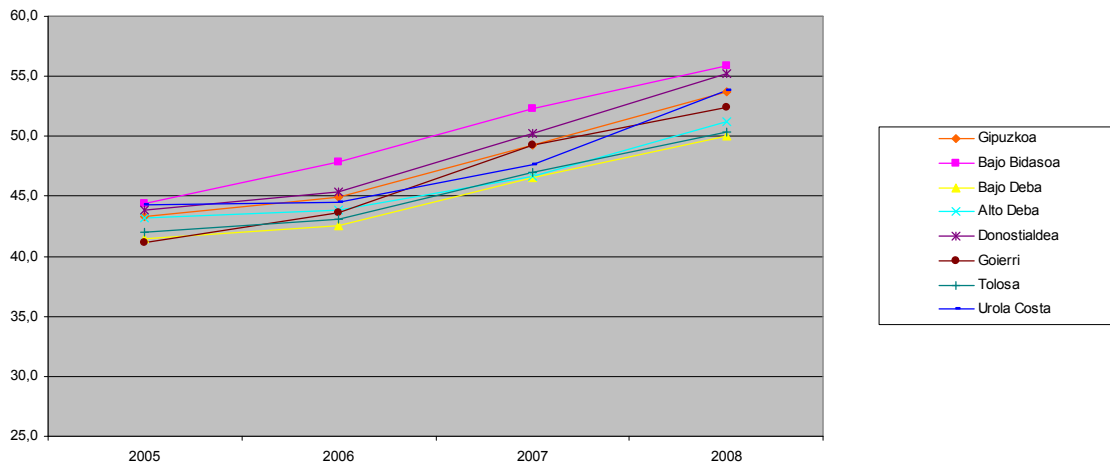
Población Hombre Usuarios de Internet (%)



Población Hombres Usuarios de Internet (%)

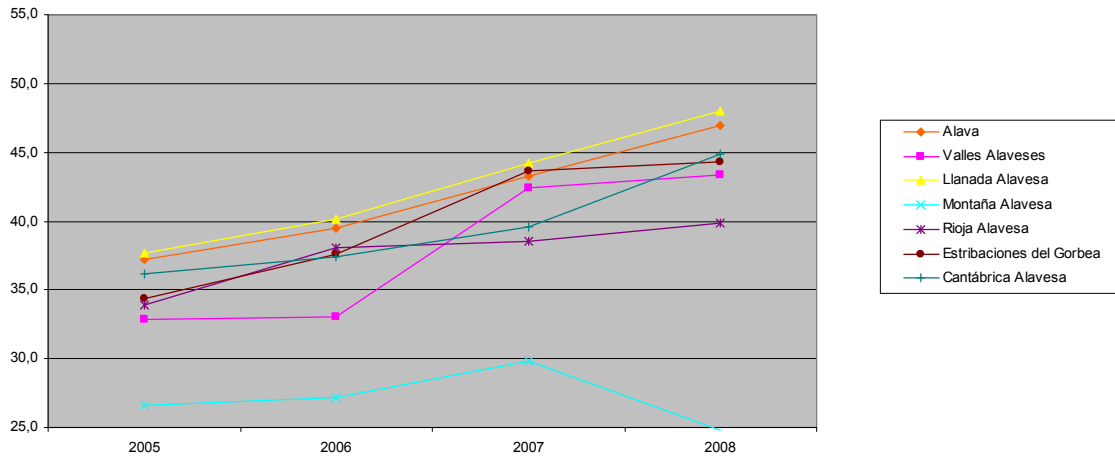


Población Hombres Usuarios de Internet (%)

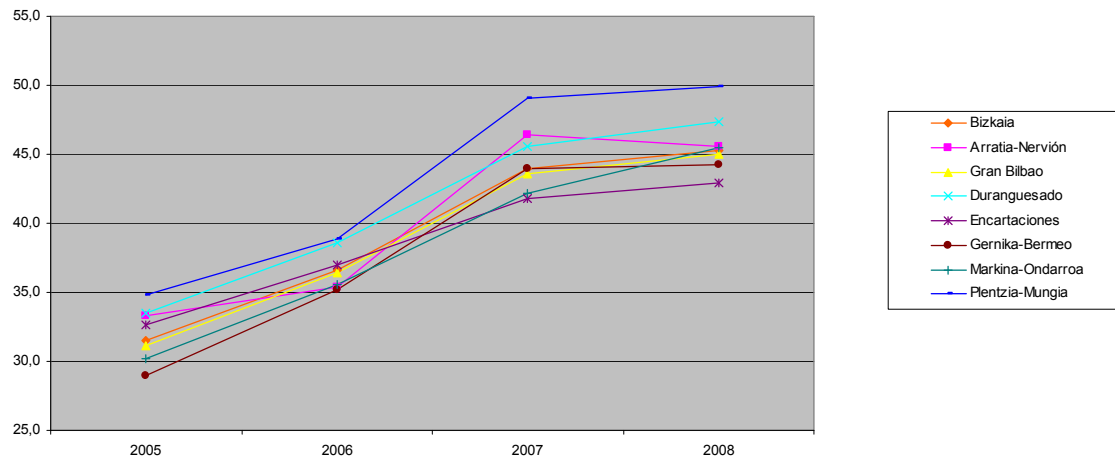


ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA SOBRE LA SOCIEDAD DE LA INFORMACIÓN - FAMILIAS DE LA C.A. DE EUSKADI

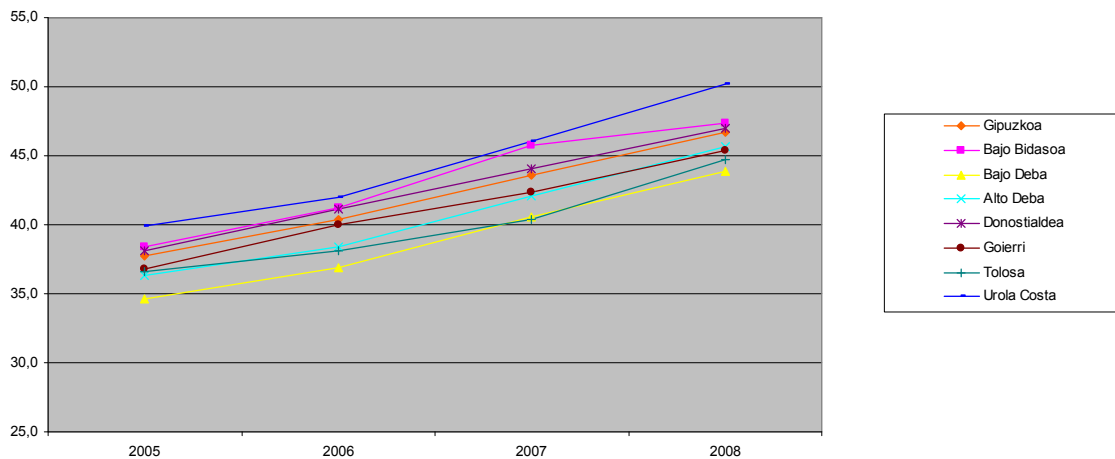
Población Mujeres Usuarias de Internet (%)



Población Mujeres Usuarias de Internet (%)



Población Mujeres Usuarias de Internet (%)



## Conclusiones

La cada vez mayor demanda de información desagregada y la necesidad de no recargar a los informantes hacen que los métodos de estimación basados en modelos estén progresivamente siendo más utilizados en la estadística oficial.

La obtención de estimaciones de las magnitudes relacionadas con la actividad en áreas pequeñas como son las comarcas, que presentamos aquí, es un paso adelante en la aplicación de las nuevas metodologías de estimación basadas en modelos en el Instituto.

Los resultados presentados en este documento ofrecen una calidad aceptable en términos de precisión. La mayoría de los coeficientes de variación (CV) obtenidos en las estimaciones no supera el 10% y muchos no sobrepasan el 5%.

Eustat, a partir de ahora, puede ofrecer estimaciones comarcales a partir de una encuesta coyuntural con lo que ello supone de aumento de la eficiencia de la operación.

Las estimaciones podrán ser mejoradas en la medida en que una mejor información auxiliar esté disponible. La disponibilidad de una información auxiliar adecuada es fundamental en los modelos y, por ello, es importante contar con unos marcos adecuados y tener acceso a la información de los ficheros administrativos.

Eustat pretende seguir avanzando en el estudio y aplicación de la metodología de estimación basada en modelos para poder ofrecer cada vez información más desagregada y de calidad.

## Bibliografía

CLARKE, PHILIP; MCGRATH, KEVIN; HUKUM, CHANDRA AND TZAVIDIS, NIKOS (2007)

*Developments in Small Area Estimation in UK with focus on current research activities.* IASS Satellite Meeting on Small Area Estimation

EUSTAT (2005)

*Informe sobre el Cálculo de Errores de Muestreo. Encuesta sobre la Sociedad de la Información - familias (ESI).*

[http://www.eustat.es/document/datos/Calculo\\_errores\\_ESI\\_c.pdf](http://www.eustat.es/document/datos/Calculo_errores_ESI_c.pdf)

EUSTAT (2008)

*Proyecto Técnico de la Operación Encuesta sobre Sociedad de la Información - familias. (ESI).*

GHOSH, M. AND RAO, J.N.K., (1994)

*Small Area Estimation: An Appraisal.* Statistical Science, 9, 55-93.

GHOSH, N. AND SÄRNDAL, C.E. (2001)

*Lecture Notes for Estimation for Population Domains and Small Areas.* Statistics Finland, vol. 48.

INSEE INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES (1993)

*“La macro Calmar, Redressement d'un échantillon par calage sur marges”,* Document n° F9310 25/11/1993, Olivier Sautory. Série des documents de travail de la Direction des Statistiques Démographiques et Sociales.. Insee - La macro SAS Calmar.

QUENOUILLE, M. (1949)

*Approximate tests of correlation in time series.* J. Roy Statist. Soc. Ser. B, 11, 18-84.

QUENOUILLE, M. (1956)

*Notes on bias in estimation.* Biometrika, 43 pp. 353-360.

RAO, J.N.K. AND WU, C.F.J. (1988)

*Resampling Inference with Complex Survey Data*. Journal of the American Statistical Association , 83, 231-241

SÄRNDAL, C.E. SWENSSON, B. AND WRETMAN J. (1992)

*Model Assisted Survey Sampling*. Springer-Verlag

SAS INSTITUTE INC., "SAS/STAT® 9. (2004)

"*User's Guide*". Copyright © 2004, Cary, NC, USA. ISBN

TUKEY, J. (1958)

*Bias and confidence in not quite large samples*. Abstract, Ann. Math. Statist., 29, 614

WOODRUFF, R.S., (1971)

*A Simple Method for Approximating the Variance of a Complicated Estimate*. Journal of The American Statistical Association. 66(334), 411-414

## Anexo

### ALAVA/ARABA

**Arabako Ibarrak / Valles Alaveses:** Añana, Armiñón, Berantevilla, Kuartango, Lantarón, Ribera Alta, Ribera Baja/Erribera Beitia, Valdegovía/Gaubea, Zambrana

**Arabako Lautada / Llanada Alavesa:** Alegría-Dulantzi, Arrazua-Ubarrundia Asparrena, Barrundia, Elburgo/Burgelu, Iruña Oka/Iruña de Oca, Iruraiz-Gauna, Salvatierra/Agurain, San Millán/Donemiliaga, Vitoria-Gasteiz, Zaldondo

**Arabako Mendialdea / Montaña Alavesa:** Arraia-Maeztu, Bernedo, Campezo/Kanpezu, Harana/Valle de Arana, Lagrán, Peñacerrada-Urizaharra

**Errioxa Arabarra / Rioja Alavesa:** Baños de Ebro/Mañueta, Elciego, Elvillar/Bilar, Kripan, Labastida/Bastida, Laguardia, Lanciego/Lantziego, Lapuebla de Labarca, Leza, Moreda de Álava, Navaridas, Oyón-Oion, Samaniego, Villabuena de Alava/Eskuernaga, Yécora/Iekora

**Gorbeia Inguruak / Etribaciones del Gorbea:** Aramaio, Legutiano, Urkabustaiz, Zigoitia, Zuia

**Kantauri Arabarra / Cantábrica Alavesa:** Amurrio, Artziniega, Ayala/Aiara, Laudio/Llodio, Okondo

### BIZKAIA

**Arratia Nerbioi / Arratia-Nerviión:** Arakaldo, Arantzazu, Areatza, Arrankudiaga, Artea, Dima, Igorre, Orozko, Otxandio, Ubide, Ugao-Miraballes, Urduña-Orduña, Zeanuri, Zeberio

**Bilbo Handia / Gran Bilbao:** Abanto y Ciérvana-Abanto Zierbena, Alonsotegi, Arrigorriaga, Barakaldo, Basauri, Berango, Bilbao, Derio, Erandio, Etxebarri, Galdakao, Getxo, Larrabetzu, Leioa, Lezama, Loiu, Muskiz, Ortuella, Portugalete, Santurtzi, Sestao, Sondika, Valle de Trápaga-Trapagaran, Zamudio, Zaratamo, Zierbena

**Durangaldea / Duranguesado:** Abadiño, Amorebieta-Etxano, Atxondo, Bedia, Berriz, Durango, Elorrio, Ermua, Garai, Iurreta, Izurtza, Lemoa, Mallabia, Mañaria, Zaldibar

**Enkartazioak / Encartaciones:** Artzentales, Balmaseda, Galdames, Gordexola, Güeñes, Karrantza Harana/Valle de Carranza, Lanestosa, Sopuerta, Trucios-Turtzioz, Zalla

**Gernika-Bermeo:** Ajangiz, Arratzu, Bermeo, Busturia, Ea, Elantxobe, Ereño, Errigoiti, Forua, Gaategiz Arteaga, Gernika-Lumo, Ibarrangelu, Kortezubi, Mendata, Morga, Mundaka, Murueta, Muxika, Nabarniz, Sukarrieta

**Markina-Ondarroa:** Amoroto, Aulesti, Berriatua, Etxebarria, Gizaburuaga, Ispaster, Lekeitio, Markina-Xemein, Mendexa, Munitibar-Arbatzegi Gerrickaitz-, Ondarroa, Ziortza-Bolibar

**Plentzia-Mungia:** Arrieta, Bakio, Barrika, Fruiz, Gamiz-Fika, Gatika, Gorliz, Laukiz, Lemoiz, Maruri-Jatabe, Meñaka, Mungia, Plentzia, Sopelana, Urduliz

## **GIPUZKOA**

**Bidasoa Beherea / Bajo Bidasoa:** Hondarribia, Irun

**Deba Beherea / Bajo Deba:** Deba, Eibar, Elgoibar, Mendaro, Mutriku, Sorluze-Placencia de las Armas

**Deba Garaia / Alto Deba:** Antzuola, Aretxabaleta, Arrasate/Mondragón, Bergara, Elgeta, Eskoriatza, Leintz-Gatzaga, Oñati

**Donostialdea / Donostia-San Sebastián:** Andoain, Astigarraga, Donostia-San Sebastián, Errenteria, Hernani, Lasarte-Oria, Lezo, Oiartzun, Pasaia, Urnieta, Usurbil

**Goierri:** Alzaga, Arama, Ataun, Beasain, Ezkio-Itsaso, Gabiria, Gaintza, Idiazabal, Itsasondo, Lazkao, Legazpi, Mutiloa, Olaberria, Ordizia, Ormaiztegi, Segura, Urretxu, Zaldibia, Zegama, Zerain, Zumarraga

**Tolosaldea / Tolosa:** Abaltzisketa, Aduna, Albiztur, Alegia, Alkiza, Altzo, Amezketa, Anoeta, Asteasu, Baliarrain, Belauntza, Berastegi, Berrobi, Bidegoian, Elduain, Gaztelu, Hernialde, Ibarra, Ikaztegieta, Irura, Larraul, Leaburu, Legorreta, Lizartza, Orendain, Orexa, Tolosa, Villabona, Zizurkil

**Urola-Kostaldea / Urola Costa:** Aia, Aizarnazabal, Azkoitia, Azpeitia, Beizama, Errezil, Getaria, Orio, Zarautz, Zestoa, Zumaia