ESTIMACIÓN DE ÁREAS PEQUEÑAS EN LA ENCUESTA INDUSTRIAL DE LA C.A. DE EUSKADI

EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADISTICA

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

Elaboración:
EUSTAT
Euskal Estatistika Erakundea
Instituto Vasco de Estadística

Edición:
EUSTAT
Euskal Estatistika Erakundea
Instituto Vasco de Estadística
Donosita-San Sebastián, 1 – 01010 – Vitoria-Gasteiz

Tirada:
500 ejemplares
I/2006

Impresión y encuadernación:
Estudios Gráficos Zure S.A.
Carretera Lutxana-Asua, 24-A
Erandio-Goikoa (Bizkaia)

ISBN: 84-7749-425-8

Depósito Legal:

PRESENTACION

Eustat, consciente de la creciente demanda de estadísticas de calidad cada vez más desagregadas, constituyó hace dos años un equipo de investigación compuesto por miembros de distintas áreas de Eustat y miembros 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 de la estadística oficial.

El trabajo que aquí se presenta, hace referencia a la primera encuesta de Eustat en la que se ha definido un sistema de estimación en áreas pequeñas. Esta operación ha sido la Encuesta Industrial Anual, principalmente debido a su especial relevancia dentro de las encuestas económicas de Eustat y a ser un sector clave en la economía vasca.

En esta publicación se presentan las estimaciones comarcales para la Encuesta Industrial de la C.A. de Euskadi correspondientes a los años 2002 y 2003. Debido a la novedad y complejidad metodológica que suponen este tipo de trabajos, este documento tiene como objetivo fundamental describir de forma pormenorizada los estimadores y modelos utilizados, tanto para los valores totales como para los errores cuadráticos medios correspondientes.

El contenido de esta publicación pretende ser el punto de partida de un proyecto de aplicación de estas técnicas en otras encuestas de Eustat-, así como aportar un material útil a todos los usuarios interesados en el conocimiento y utilización de métodos en áreas pequeñas.

Vitoria-Gasteiz, enero 2006

JOSU IRADI ARRIETA

Director General

Indice

Capítulo

1

Encuesta Industrial de la C.A. de Euskadi

1.1.- Descripción de la Encuesta Industrial de la C.A. de Euskadi

· Antecedentes

Esta operación se puso en marcha en 1981, teniendo desde su creación como objetivo fundamental el conocimiento pormenorizado del entramado industrial vasco, dada su importancia tanto en términos de valor añadido como de empleo. La información básica se obtiene a partir de las principales partidas de la cuenta de pérdidas y ganancias, y la consiguiente estimación, a partir de ellas, de las principales macromagnitudes.

Esta operación estadística se realiza en colaboración con el Órgano Estadístico Específico adscrito al Servicio de Estadística y Análisis Sectorial del Departamento de Agricultura, Pesca y Alimentación.

· Características Técnicas

Universo: El ámbito poblacional se circunscribe a aquellos establecimientos cuya actividad principal, medida en términos de valor añadido generado, sea industrial.

Incluye, según la Clasificación Nacional de Actividades Económicas de 1993 (en adelante CNAE-93), las siguientes secciones:

Sección C:	Industrias extractivas;
Sección D:	Industria manufacturera;
Sección E:	Producción y distribución de energía eléctrica, gas y agua.

Geográfico. Las unidades estadísticas que están ubicadas en el ámbito geográfico de la C.A. de Euskadi, aun cuando su sede social o gerencia se encuentre fuera de ella.

Temporal. El período de referencia es el ejercicio económico del año natural. Excepcionalmente, de presentarse establecimientos cuya contabilidad vaya referida a períodos de tiempo que no correspondan al año natural, se referirá la información a los ejercicios que finalizan dentro de los años correspondientes.

· Marco de la encuesta

El marco de la encuesta es el Directorio de Actividades Económicas de Eustat. Su utilización permite la elaboración de un muestreo probabilístico que acote los errores muestrales.

· Unidad Estadística

La unidad estadística es el establecimiento definido como una unidad que ejerce, exclusiva o principalmente, una o varias actividades situada en un mismo emplazamiento geográfico.

· Diseño muestral y extrapolación

Se realiza un muestreo probabilístico en dos fases: una primera en la que se seleccionan con probabilidad "uno" todas las unidades que tengan más de 19 empleados; en la segunda fase se realiza un muestreo aleatorio estratificado donde las variables de estratificación son:

a) Territorio Histórico: Álava, Bizkaia y Gipuzkoa.

b) Actividad: Clasificación Nacional de Actividades Económicas (CNAE-93) a nivel de subclase, es decir, a 5 dígitos. Posteriormente para su difusión se utiliza la clasificación normalizada de EUSTAT A84. La clasificación A84 es una desagregación de la A60 (CNAE-93 a 2 dígitos) en función de la estructura económica específica de la C.A. de Euskadi.

El tamaño de la muestra seleccionada es de 3.000 unidades estadísticas, aproximadamente.

Previamente a la extrapolación, se post-estratifican los establecimientos muestrales, según los tres Territorios Históricos (Álava, Bizkaia, Gipuzkoa) , subclase de la CNAE-93 a 5 dígitos y cinco tamaños de establecimientos, que son:

Entre 1 y 19 empleados;

Entre 20 y 49 empleados;

Entre 50 y 99 empleados;

Entre 100 y 499 empleados;

Mayores o iguales a 500 empleados.

El paso de datos muestrales a las estimaciones se realiza a través de una matriz de elevadores por cada estrato. La variable que se utiliza para la obtención de los elevadores es el número de ocupados de los establecimientos industriales. El uso de esta variable está justificado debido a que es la más correlacionada con las principales variables económicas que intenta medir la encuesta.

1.2.- Estimadores utilizados en la Encuesta Industrial de la C.A. de Euskadi

Todas las unidades estadísticas con más de 19 empleados son autoponderadas, por lo que el interés radica principalmente en las estimaciones dentro del estrato de empleo de 1-19 empleados. En lo que sigue se describe la estimación dentro de una subclase de actividad (CNAE-93 a 5 dígitos) el total de una variable cualquiera y, así como de su correspondiente coeficiente de variación.

Actualmente, para la estimación del total de la variable y se utiliza el estimador indirecto de razón o estimador sintético, utilizando como información auxiliar el número de empleados de los establecimientos. Esta decisión se tomó tras realizar un estudio descriptivo exhaustivo en el que se comprobó que existe una correlación positiva fuerte entre el número de empleados de los establecimientos y la magnitud de las principales variables de la Encuesta Industrial.

El estimador indirecto de razón de una variable de interés cualquiera y, cuando se dispone de una variable auxiliar x, está, en el caso de la Encuesta Industrial, asistido por el modelo de regresión lineal simple heterocedástico del tipo:

$$y_{hj} = x_{hj}\beta + \varepsilon_{hj} \qquad \text{con} \quad \mathrm{var}(\varepsilon_{hj}) = \sigma^2 x_{hj}, \qquad (1)$$

donde h hace referencia al estrato y j a la unidad estadística. Los estratos son los Territorios Históricos (Álava, Bizkaia y Gipuzkoa) dado que en todo lo que sigue se supone el interés radica en una única subclase. El modelo lineal simple (1) contiene perturbaciones heterocedásticas siendo la varianza función lineal creciente del número de empleados de los establecimientos industriales.

El estimador del total de la variable y en una subclase dada en el Territorio Histórico h viene dado por:

$$\hat{t}_{yh.SYN} = X_h \hat{\beta} = X_h \frac{\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj} y_{hj}}{\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj} x_{hj}},$$

donde $X_h = \sum_{j=1}^{N_h} x_{hj}$, w_{hj} es el peso de muestreo de la unidad j en el Territorio Histórico h, x_{hj} recoge el empleo del establecimiento j del Territorio Histórico h y n_h es el tamaño de la muestra en el Territorio Histórico h.

El estimador de la varianza del estimador indirecto de razón se puede aproximar mediante:

$$\hat{\mathrm{var}}(\hat{t}_{yh.SYN}) \approx N^2\left(1-\frac{n}{N}\right)\frac{1}{n}\left(\frac{X_h}{\sum_{h=1}^{H}\sum_{j=1}^{n_h} w_{hj}x_{hj}}\right)^2 \hat{\mathrm{var}}(\varepsilon)$$

donde $\hat{\mathrm{var}}(\varepsilon)$ es la varianza muestral de los residuos del modelo heterocedástico (1) con todos los datos muestrales (es decir, se calculan los residuos en toda la CA de Euskadi, no sólo en el Territorio Histórico h) y el resto de la notación es la habitual.

Särndal and Hidiroglou (1989) proporcionan una aproximación del sesgo del estimador sintético según la cual $E\left(\hat{t}_{yh.SYN}\right) - t_{yh.SYN} \approx -\sum_{j=1}^{N}\hat{\varepsilon}_j$ donde $\hat{\varepsilon}_j = y_j - x_j\hat{\beta}$. Por consiguiente, el estimador será aproximadamente insesgado si se verifica que $\sum_{j=1}^{N}\hat{\varepsilon}_k = 0$. Esta condición no se satisface normalmente. Si el modelo no ajusta bien en el dominio de interés, la suma de residuales puede estar lejos de cero, indicando un sesgo considerable. En caso contrario, podemos esperar un sesgo limitado. Por ello, es deseable estimar el error cuadrático medio como medida de precisión del estimador. Viene dado por:

$$MSE\left(\hat{t}_{yh.SYN}\right) = \mathrm{var}\left(\hat{t}_{yh.SYN}\right) + \left(sesgo_{h.SYN}\right)^2,$$

y se estima mediante la expresión:

$$\hat{MSE}\left(\hat{t}_{yh.SYN}\right) = \hat{\mathrm{var}}\left(\hat{t}_{yh.SYN}\right) + \left(\sum_{j=1}^{n_h}\hat{\varepsilon}_j\right)^2,$$

donde $\hat{\varepsilon}_j = y_j - x_j^{|}\hat{\beta}$, para $j = 1,\ldots,n$ son los residuos obtenidos a partir del modelo estimado (1) con todos los datos muestrales, aunque en cada Territorio Histórico solamente se suman los específicos de ese Territorio Histórico. El coeficiente de variación se define como

$$\hat{cv}\left(\hat{t}_{yh.SYN}\right) = \frac{\hat{rmse}\left(\hat{t}_{yh.SYN}\right)}{\hat{t}_{yh.SYN}}, \qquad \text{donde} \quad \hat{rmse}\left(\hat{t}_{yh.SYN}\right) = \sqrt{\hat{MSE}\left(\hat{t}_{yh.SYN}\right)}.$$

Capítulo
2

Sistema de estimación en áreas pequeñas en la Encuesta Industrial

2.1.- Introducción

La C.A. de Euskadi se divide en las siguientes 20 comarcas:

- Álava: Valles Alaveses, Llanada Alavesa, Montaña Alavesa, Rioja Alavesa, Estribaciones del Gorbea y Cantábrica Alavesa.
- Bizkaia: Arratia-Nervión, Gran Bilbao, Duranguesado, Encartaciones, Gernika-Bermeo, Markina-Ondarroa y Plentzia-Mungia.
- Gipuzkoa: Bajo Bidasoa, Bajo Deba, Alto Deba, Donostia-San Sebastián, Goierri, Tolosa y Urola Costa.

La actividad industrial de la C.A. de Euskadi no está uniformemente repartida en las 20 comarcas estadísticas y, tanto la importancia del sector industrial como su tamaño varía enormemente entre comarcas. De hecho, hay comarcas en las que la actividad industrial es realmente reducida, por lo que la tarea de estimación comarcal pasa forzosamente por aplicar técnicas de estimación de áreas pequeñas basadas en modelos. El aumento del tamaño muestral requerido para obtener estimaciones comarcales de calidad sería ciertamente costoso.

Los modelos de áreas pequeñas suponen la existencia de un modelo subyacente que siguen todos los datos de la población, pero que se estima con los datos de la muestra (Rao, 2003). Eustat utiliza para la obtención de estimaciones comarcales en la Encuesta Industrial dos tipos de modelos: el modelo de regresión lineal de efectos fijos y el modelo de regresión lineal con efectos fijos y aleatorios, llamado también modelo mixto.

En el modelo mixto el predictor consta de un término común de efectos fijos y otro diferenciado para los elementos de cada comarca d $(d = 1,...,t)$. Este término diferenciado está formado por los efectos aleatorios (v_d), de modo que todos los datos de la misma comarca comparten el mismo efecto aleatorio. En el caso del modelo de efectos fijos no existen términos diferenciados para cada comarca ya que la parte sistemática es común para todas las comarcas. Sin embargo, la especificidad se consigue al proyectar el coeficiente estimado común a la información auxiliar específica de cada comarca.

2.2.- Modelo lineal mixto

2.2.1.- Versión proyectiva

Se parte de una población formada por los N establecimientos de una subclase de actividad (CNAE-93 a 5 dígitos) concreta. En cada comarca d $(d = 1,...,t)$ hay N_d establecimientos en la población, de modo que $N = \sum_d N_d$. En dicha subclase se muestrean n establecimientos de los que n_d pertenecen a la comarca d. Se propone el siguiente modelo lineal mixto heterocedástico:

$$y_{dj} = \beta_0 + \beta_1 x_{dj} + v_d + e_{dj}, \qquad d = 1,...,t \qquad j = 1,...,n_d, \qquad (2)$$

donde para el establecimiento j de la comarca d, y_{dj} es el valor que toma la variable de interés y x_{dj} es el número de empleados del establecimiento. El número total de establecimientos muestreados en la comarca d es n_d. Los efectos fijos del modelo son β_0 y β_1. El efecto aleatorio común para todos los establecimientos de la comarca d es v_d y e_{dj} son los errores aleatorios específicos de cada establecimiento. Además, se supone que $v_d \subset N\left(0,\sigma_v^2\right)$ y $e_{dj} \subset N\left(0,\sigma_e^2 c_{dj}^{-1}\right)$ son independientes. Para corregir la heterocedasticidad presente en los datos se utilizan los pesos $c_{dj} = 1/x_{dj}$. Cuando $c_{dj} = 1 \ \forall d, j$, este modelo es similar al propuesto por Battese *et al* (1988).

El modelo superpoblacional correspondiente al modelo (2) escrito en forma matricial se expresa como

$$Y = X\beta + Zv + \varepsilon\,, \qquad v \subset N(0,\sigma_v^2 I_t)\,, \qquad \varepsilon \subset N(0,\sigma_e^2 C^{-1})\,, \qquad (3)$$

donde $C = diag(c_{dj})$ ($d = 1,\ldots,t$), es la matriz de pesos del modelo, y j es el establecimiento ($j = 1,\ldots,N_d$). El vector $Y = (Y_1',\ldots,Y_t')'$ es el vector ($N \times 1$) cuyas componentes Y_d' son los valores de la variable de interés para cada comarca, $\beta = (\beta_0,\beta_1)'$ es el vector de coeficientes del modelo, X es la matriz de diseño ($N \times 2$) formada por una columna de unos asociada a la ordenada en el origen y otra columna asociada a la variable auxiliar que es, en este caso, el número de empleados de cada establecimiento. La matriz $Z = diag(1_{N_d})$, $d = 1,\ldots,t$, es la matriz de diseño $N \times t$ diagonal por bloques asociada a los efectos aleatorios. Es decir, para cada comarca d, la matriz Z tiene una columna asociada de unos definida por el vector $1_{N_d} = (1,\ldots,1)'$ de dimensión N_d. Los efectos aleatorios $v = (v_1,\ldots,v_t)'$, son comunes a los N_d elementos de la misma comarca y $\varepsilon = (\varepsilon_1',\ldots,\varepsilon_t')'$ es el vector de errores aleatorios, donde $\varepsilon_d = (\varepsilon_{d_1},\ldots,\varepsilon_{d_{N_d}})'$.

Con objeto de unificar la teoría y presentar las versiones proyectivas y predictivas del modelo (3), procedemos a diferenciarlo en su parte muestreada y no muestreada del siguiente modo:

$$\begin{pmatrix} Y_s \\ Y_r \end{pmatrix} = \begin{pmatrix} X_s \\ X_r \end{pmatrix}\beta + \begin{pmatrix} Z_s \\ Z_r \end{pmatrix}v + \begin{pmatrix} \varepsilon_s \\ \varepsilon_r \end{pmatrix},$$

donde los subíndices s y r denotan los establecimientos muestreados y no muestreados respectivamente. Entonces el modelo muestral puede escribirse como:

$$Y_s = X_s\beta + Z_s v + \varepsilon_s\,, \qquad v \subset N(0,\sigma_v^2 I_t)\,, \qquad \varepsilon_s \subset N(0_s,\sigma_e^2 C_s^{-1})\,,$$

donde $C_s = diag(c_{dj} = 1/x_{dj})$, $d = 1,\ldots,t_s$, $j = 1,\ldots,n_d$ y t_s es el número total de comarcas donde se ha muestreado. La matriz de varianzas y covarianzas de Y_s puede expresarse como $\mathrm{var}(Y_s) = V_s = Z_s\sigma_v^2 Z_s' + \sigma_e^2 C_s^{-1} = diag(V_1,\ldots,V_{t_s})$ donde $V_d = \sigma_e^2 C_d^{-1} + \sigma_v^2 1_{n_d} 1_{n_d}'$ y $C_d = diag(c_{d_1},\ldots,c_{d_{n_d}})_{n_d \times n_d} = diag(c_{n_d})$. Si suponemos conocidos los componentes de varianza $\sigma^2 = (\sigma_e^2,\sigma_v^2)$, el estimador de los efectos fijos así como su matriz de varianzas y covarianzas puede obtenerse por mínimos cuadrados generalizados:

$$\tilde{\beta} = (X_s^{|} V_s^{-1} X_s)^{-1} X_s^{|} V_s^{-1} Y_s, \qquad \text{var}(\tilde{\beta}) = \Phi_s = (X_s^{|} V_s^{-1} X_s)^{-1},$$

donde $V_s^{-1} = diag(V_1^{-1},\ldots,V_d^{-1},\ldots,V_{t_s}^{-1})$, $\quad V_d^{-1} = \dfrac{1}{\sigma_e^2}\left(C_d - \dfrac{\gamma_{dc}}{c_{d.}} c_{n_d} c_{n_d}^{|}\right)$,

$\gamma_{dc} = \dfrac{\sigma_v^2}{\sigma_v^2 + \sigma_e^2 / c_{d.}}$ y $c_{d.} = \sum_{j=1}^{n_d} c_{dj}$.

Sea $1_{n_d}^{|} = (1,\ldots,1)$ de dimensión n_d , entonces la predicción de los efectos aleatorios se obtiene como $\hat{v}_d = \sigma_v^2 1_{n_d}^{|} V_d^{-1}(Y_d - X_d^{|}\hat{\beta}) = \hat{\gamma}_{dc}(\bar{y}_{dc} - \bar{x}_{dc}^{|}\hat{\beta})$, donde $\bar{y}_{dc} = \dfrac{1}{c_{d.}}\sum_{j=1}^{n_d} c_{dj} y_{dj}$, y $\bar{x}_{dc}^{|} = \dfrac{1}{c_{d.}}\sum_{j=1}^{n_d} c_{dj} x_{dj}^{|} = (1, \bar{x}_{dc})$, con $x_{dj}^{|} = (1, x_{dj})$. El predictor de tipo proyectivo de la media de la variable de interés en la comarca d viene dado por:

$$\hat{\bar{y}}_d^* = \bar{X}_{d(p)}^{|}\hat{\beta} + \hat{\gamma}_{dc}(\bar{y}_{dc} - \bar{x}_{dc}^{|}\hat{\beta}), \qquad d = 1,\ldots,t, \tag{4}$$

donde $\hat{\beta} = \tilde{\beta}(\hat{\sigma}_e^2, \hat{\sigma}_v^2)$ ha sido evaluada con las estimaciones de los componentes de varianza, $\bar{X}_{d(p)}^{|} = (1, \bar{x}_{d(p)})$, y $\bar{x}_{d(p)} = \dfrac{\sum_{j \in N_d} x_{dj}}{N_d}$ es la media poblacional del número de empleados en la comarca d para todos los establecimientos de la población: los muestreados y los no muestreados de la comarca d . La correspondiente versión para el total viene dada por:

$$\hat{t}_d^* = X_{d(p)}^{|}\hat{\beta} + N_d \hat{\gamma}_{dc}(\bar{y}_{dc} - \bar{x}_{dc}^{|}\hat{\beta}), \qquad d = 1,\ldots,t, \tag{5}$$

donde $X_{d(p)}^{|} = (N_d, X_{d(p)})$, N_d es el número total de establecimientos y $X_{d(p)}$ es el número total (poblacional) de empleados de la comarca d . El predictor (4) puede expresarse también como suma ponderada de un estimador de regresión generalizado del tipo

$$\bar{y}_{dc} + (\bar{X}_{d(p)} - \bar{x}_{dc})^{|}\hat{\beta},$$

y el estimador de regresión sintético $\bar{X}_{d(p)}^{|}\hat{\beta}$, de modo que,

$$\hat{\bar{y}}_d = \gamma_{dc}(\bar{y}_{dc} + (\bar{X}_{d(p)}^{|} - \bar{x}_{dc})^{|}\hat{\beta} + (1-\gamma_{dc})\bar{X}_{d(p)}^{|}\hat{\beta}, \qquad d = 1,\ldots,t, \tag{6}$$

El peso $0 \leq \gamma_{dc} \leq 1$ mide la proporción de la varianza σ_v^2 relativa a la varianza total $\sigma_v^2 + \sigma_e^2 / c_d$. Si la varianza del modelo es pequeña, los γ_{dc} son pequeños y se concede más peso al componente sintético. Análogamente se concede más peso al estimador de regresión generalizado cuanto mayor sea $c_{d.}$. Cuando $c_{dj} = 1$ el estimador de regresión generalizado es aproximadamente insesgado bajo el diseño si n_d es suficientemente grande. En el caso general, es insesgado bajo el modelo condicionado a la realización de los efectos realizados v_d supuesto que $\tilde{\beta}$ es condicionalmente insesgado para β. Es decir, el estimador BLUP (4) (Best Linear Unbiased Predictor) es condicionalmente sesgado debido a la presencia de este componente sintético $\overline{X}_d^{|}\hat{\beta}$. Cuando el muestreo es aleatorio simple y $c_{dj} = 1$ el estimador BLUP es consistente bajo el diseño para la media de la comarca $\overline{Y}_d$ cuando n_d crece, ya que $\gamma_d \to 1$.

Cuando el modelo es homocedástico y $c_{dj} = 1$ entonces $c_{d.} = n_d$, entonces el predictor proyectivo de la media viene dado por:

$$\hat{\bar{y}}_d^* = \overline{X}_{d(p)}^{|}\hat{\beta} + \hat{\gamma}_d(\bar{y}_{dc} - \bar{x}_{dc}^{|}\hat{\beta}) = \hat{\gamma}_d \bar{y}_{dc} + (\overline{X}_{d(p)}^{|} - \hat{\gamma}_d \bar{x}_{dc}^{|})\hat{\beta},$$

donde $\hat{\gamma}_d$ mide la incertidumbre asociada con la modelización del predictor y toma la forma:

$$\hat{\gamma}_d = \frac{\hat{\text{cov}}(v_d, \bar{u}_{d.})}{\hat{\text{var}}(\bar{u}_{d.})} = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / n_d}.$$

2.2.2.- Versión predictiva

Cuando la fracción de muestreo por comarca $f_d = n_d / N_d$ es significativa, la literatura recomienda utilizar la versión predictiva para obtener la predicción del total o de la media de la comarca d en lugar de la versión proyectiva. Esta versión consiste en diferenciar la parte muestreada de la no muestreada. Así, la predicción de la parte muestreada es la misma muestra, mientras que la no muestreada se predice con el predictor de tipo proyectivo.

Para obtener la versión predictiva se descompone el total $\sum_{j \in N_d} y_{dj} = \sum_{j \in d_r} y_{dj} + \sum_{j \in d_s} y_{dj}$, donde d_s indica la muestra en la comarca d y d_r el resto de los establecimientos no pertenecientes a la muestra de la comarca d. A continuación se descompone la media poblacional:

$$\overline{Y}_d = \frac{\sum_{j\in d_r} y_{dj} + \sum_{j\in d_s} y_{dj}}{N_d} = \frac{(N_d - n_d)\overline{Y}_{dr} + n_d \overline{y}_{ds}}{N_d} = (1 - f_d)\overline{Y}_{dr} + f_d \overline{y}_{ds} . \quad (7)$$

El estimador predictivo de la media del área d , para todo $d = 1,\dots,t$ viene dado por:

$$\hat{\bar{y}}_d = (1 - f_d)\hat{\overline{Y}}_{dr} + f_d \overline{y}_{ds} = (1 - f_d)\hat{\bar{y}}^*_{dr} + f_d \overline{y}_{ds} .$$

Sustituyendo $\hat{\bar{y}}^*_{dr}$ por su expresión $\overline{X}^{|}_{dr}\hat{\beta} + \hat{v}_d$ donde $\overline{X}^{|}_{dr} = (1, \overline{x}_{dr})$ y $\overline{x}_{dr} = \frac{\sum_{j\in d_r} x_{dj}}{N_d - n_d}$, entonces se tiene que:

$$\hat{\bar{y}}_d = \hat{\overline{Y}}_d = (1 - f_d)\left[\overline{X}^{|}_{d(p_r)}\hat{\beta} + \hat{\gamma}_{dc}(\overline{y}_{dc} - \overline{x}^{|}_{dc}\hat{\beta})\right] + f_d \overline{y}_{ds} , \quad (8)$$

que conduce a la versión predictiva del total:

$$\hat{t}_d = X^{|}_{d(p_r)}\hat{\beta} + (N_d - n_d)\hat{\gamma}_{dc}\left(\overline{y}_{dc} - \overline{x}_{dc}\hat{\beta}\right) + \sum_{j=1}^{n_d} y_{dj} , \qquad d = 1,\dots,t \quad (9)$$

donde $\hat{\beta} = \hat{\beta}_c\left(\hat{\sigma}^2_e, \hat{\sigma}^2_v\right)$ ha sido evaluada con las estimaciones de los componentes de varianza y $X_{d(p_r)}$ es el total de empleados en la comarca d para todos los establecimientos no muestreados.

2.2.2.1.- Estimadores de la media y del total por Territorio Histórico y para la C.A. de Euskadi

Lo mismo en la versión predictiva como en la proyectiva, las medias y totales por Territorio Histórico y para la C.A. de Euskadi se calculan como sigue.

El estimador de la media por Territorio Histórico viene dado por:

$$\hat{\bar{y}}_h = \frac{1}{N_h}\sum_{d\in h} N_d \hat{\bar{y}}_d = \frac{1}{N_h}\sum_{d\in h} \hat{t}_d , \quad (10)$$

donde $d \in h$ indica que la suma se efectúa en todas las comarcas del estrato h (en este caso $h = 1,2,3$ son los Territorios Históricos) y $N_h = \sum_{d\in h} N_d$ es el total poblacional del Territorio Histórico h .

El estimador del total por Territorio Histórico viene dado por:

$$\hat{t}_h = \sum_{d \in h} \hat{t}_d \ . \tag{11}$$

El estimador de la media para la C.A. de Euskadi viene dado por:

$$\hat{\bar{y}} = \frac{1}{N} \sum_{h=1}^{3} N_h \hat{\bar{y}}_h \ , \tag{12}$$

donde $N = \sum_{h=1}^{3} N_h$.

El estimador del total para la C.A. de Euskadi viene dado por:

$$\hat{t} = \sum_{h=1}^{3} N_h \hat{\bar{y}}_h = \sum_{h=1}^{3} \hat{t}_h \ ,. \tag{13}$$

2.2.3.- Estimación ponderada de los componentes de la varianza

Aunque existen varios métodos para estimar los componentes de varianza, en este documento se presenta la estimación por el método de los momentos ya que no depende de la hipótesis de normalidad que requieren otros métodos como el de máxima verosimilitud (MV) o máxima verosimilitud restringida (REML).

2.2.3.1.- Método de los momentos

El método de los momentos (Searle et al. 1992) puede aplicarse para estimar σ_v^2 y σ_e^2 igualando las esperanzas de las sumas de cuadrados a sus estimadores, de modo que:

$$\hat{\sigma}_e^2 = \frac{\varepsilon_s^{|} C_\varepsilon \varepsilon_s}{n - rank(X_s, Z_s)} = \frac{\varepsilon_s^{|} C_\varepsilon \varepsilon_s}{n - t_s - 2} \ , \tag{14}$$

donde $n = \sum_{d=1}^{t} n_d$ y $rank(X_s, Z_s)$ es el rango de la matriz ampliada (X_s, Z_s). Por consiguiente la estimación proporcionada puede obtenerse también calculando la varianza residual de la regresión ponderada entre los datos muestrales (variable dependiente) y los datos auxiliares (empleo) y los efectos aleatorios de la comarca como predictores (variables independientes). La varianza de los efectos aleatorios se calcula mediante la expresión:

$$\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0) = \max\left(\frac{1}{n_{*c}}\left\{\sum_{d=1}^{t_s}\sum_{j=1}^{n_d} c_{dj} s_{dj}^2 - (n-k-1)\hat{\sigma}_e^2\right\}, 0\right), \qquad (15)$$

donde:

$$n_{*c} = tr(M_c Z_s Z_s^{|}),$$

$$M_c = C_s - C_s X_s (X_x^{|} C_s X_s)^{-1} X_s^{|} C_s,$$

y, $s_{dj} = y_{dj} - x_{dj}\hat{\beta}_0 = y_{dj} - x_{dj}(X_s^{|} C_s X_s)^{-1} X_s^{|} C_s Y_s$, son los residuales de la regresión ponderada de Y_s sobre X_s con pesos $C_s = diag(c_{dj} = 1/x_{dj})$, $d = 1,\ldots,t_s$ y $j = 1,\ldots,n_d$. El estimador truncado $\hat{\sigma}_v^2 = \max(\tilde{\sigma}_v^2, 0)$ es sesgado pero consistente cuando t_s crece.

2.2.4.- Error cuadrático medio por comarcas

Kackar and Harville (1984) demostraron que bajo hipótesis de normalidad, el error cuadrático medio (MSE) del predictor genérico BLUP $t(\hat{\theta}, Y)$ viene dado por:

$$MSE\left[t(\hat{\theta}, Y)\right] = MSE\left[t(\theta, Y)\right] + E\left[t(\hat{\theta}, Y) - t(\theta, Y)\right]^2, \qquad (16)$$

donde $\theta = \sigma^2 = \left(\sigma_v^2, \sigma_e^2\right)$ y suponiendo que $\hat{\theta} = \hat{\sigma}^2 = \left(\hat{\sigma}_v^2, \hat{\sigma}_e^2\right)$ es invariante ante traslaciones. Para el predictor de la media $t(\hat{\theta}, Y) = \hat{\bar{y}}_d$, $\tilde{\bar{y}}_d$ es el predictor de $\bar{y}$ suponiendo conocidos los componentes de varianza y, por tanto, $t(\theta, Y) = \tilde{\bar{y}}_d$. Entonces el MSE del predictor de la media viene dado por:

$$MSE\left[\hat{\bar{y}}_d\right] = E\left[\bar{y}_d - \tilde{\bar{y}}_d\right]^2 + E\left[\hat{\bar{y}}_d - \tilde{\bar{y}}_d\right]^2. \qquad (17)$$

Henderson (1975) dio una expresión para $MSE\left[\tilde{\bar{y}}_d\right] = g_{1d}(\sigma^2) + g_{2d}(\sigma^2)$, pero el segundo término de (17), llamado $g_{3d}(\sigma^2)$, no es fácil de calcular salvo en casos especiales. Kackar y Harville (1984) obtuvieron una expresión basada en el desarrollo en serie de Taylor:

$$E\left[\hat{\bar{y}}_d - \tilde{\bar{y}}_d\right]^2 \approx E\left[h_d(\theta)(\hat{\theta} - \theta)\right]^2,$$

con $h_d(\theta)=\dfrac{\partial t_d(\theta)}{\partial \theta}$. Prasad y Rao (1990) propusieron una aproximación posterior dada por:

$$tr\left\{A_d(\hat{\theta})E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^{|}\right]\right\}\approx tr\left\{(\vee b_d^{|})V_d(\vee b_d^{|})^{|}E\left[(\hat{\theta}-\theta)(\hat{\theta}-\theta)^{|}\right]\right\},$$

donde $\vee b_d^{|}=col_{1\le d\le p}\dfrac{\partial b_d^{|}}{\partial \theta_j}$ y p es el número de componentes de varianza. Los estimadores de $g_{2d}(\sigma^2)$ y $g_{3d}(\sigma^2)$ vienen dados por $g_{2d}(\hat{\sigma}^2)$ y $g_{3d}(\hat{\sigma}^2)$. Estos estimadores son correctos hasta el orden $O_p(t^{-1})$ (aquí t es el número de comarcas y no el predictor), ya que $\hat{\sigma}^2$ es un estimador consistente de σ^2. Sin embargo, $g_{1d}(\hat{\sigma}^2)$ no es un estimador correcto de $g_{1d}(\sigma^2)$, ya que su sesgo es de orden $O(t^{-1})$, y se obtiene usando un desarrollo en serie de Taylor de $g_{1d}(\sigma^2)$ en torno a σ^2 y tomando su esperanza. Después de realizar algunas operaciones se obtiene:

$$E\left[g_{1d}(\hat{\sigma}^2)\right]-g_{1d}(\sigma^2)=-g_{3d}(\hat{\sigma}^2)+O(t^{-1}).$$

2.2.4.1.- Versión proyectiva

Prasad y Rao (1990) proporcionan un estimador de (17) 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:

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

donde:

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

$$g_{2d}\left(\hat{\sigma}^2\right)=\left(\overline{X}_{d(p)}-\hat{\gamma}_{dc}\overline{x}_{dc}\right)'\hat{\Phi}_c\left(\overline{X}_{d(p)}-\hat{\gamma}_{dc}\overline{x}_{dc}\right),$$

$$g_{3d}\left(\hat{\sigma}^2\right)=\hat{\gamma}_{dc}\left(1-\hat{\gamma}_{dc}\right)^2\hat{\sigma}_e^{-4}\hat{\sigma}_v^{-2}h\left(\hat{\sigma}^2\right),$$

$$h\left(\hat{\sigma}^2\right)=\hat{\sigma}_e^4\,\hat{\text{var}}\left(\hat{\sigma}_v^2\right)+\hat{\sigma}_v^4\,\hat{\text{var}}\left(\hat{\sigma}_e^2\right)-2\hat{\sigma}_v^2\hat{\sigma}_e^2\,\hat{\text{cov}}\left(\hat{\sigma}_e^2\hat{\sigma}_v^2\right),$$

$$\hat{\text{var}}\left(\hat{\sigma}_e^2\right)=2\left(n-t_s-k\right)^{-1}\hat{\sigma}_e^4=2d_e^{-1}\hat{\sigma}_e^4,$$

$$\hat{\text{var}}\left(\hat{\sigma}_v^2\right)=2n_{*c}^{-2}\left[\left(n-t_s-k\right)^{-1}\left(t_s-1\right)\left(n-k-1\right)\hat{\sigma}_e^4+2n_*\hat{\sigma}_e^2\hat{\sigma}_v^2+n_{**c}\hat{\sigma}_v^4\right],$$

$$\hat{\text{cov}}\left(\hat{\sigma}_e^2,\hat{\sigma}_v^2\right)=-\left(t_s-1\right)n_{*c}^{-1}\,\hat{\text{var}}\left(\hat{\sigma}_e^2\right),$$

$$n_{*c} = tr\left(M_c Z_s Z_s'\right),$$

$$n_{**c} = tr\left(M_c Z_s Z_s'\right)^2,$$

$$M_c = C_s\left(I - X_s\left(X_s' C_s X_s\right)^{-1} X_s' C_s\right),$$

y k es el número de variables auxiliares, en este caso $k = 1$.

Frecuentemente se utiliza como medida de precisión del estimador $\hat{\bar{y}}_d$ la raíz cuadrada del error cuadrático medio (RMSE) dada por:

$$RMSE\left[\hat{\bar{y}}_d\right] = \sqrt{MSE\left[\hat{\bar{y}}_d\right]}.$$

El MSE del predictor del total para cada 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,

$$\hat{MSE}\left[\hat{t}_d\right] = N_d^2\left[g_{1d}\left(\hat{\sigma}^2\right) + g_{2d}\left(\hat{\sigma}^2\right) + 2g_{3d}\left(\hat{\sigma}^2\right)\right],$$

y la raíz cuadrada del error cuadrático medio del total se estima mediante la expresión:

$$R\hat{M}SE\left[\hat{t}_d\right] = \sqrt{MSE\left[\hat{t}_d\right]}.$$

El coeficiente de variación se define como:

$$CV\left[\hat{t}_d\right] = \frac{R\hat{M}SE\left[\hat{t}_d\right]}{\hat{t}_d}.$$

2.2.4.2.- Versión predictiva

En la versión predictiva, el estimador del error cuadrático medio del predictor (8), válido cuando los estimadores de los componentes de varianza se obtienen por el método REML o por el método de los momentos, viene dado por:

$$\hat{MSE}\left[\hat{\bar{y}}_d\right] = g_{1d}\left(\hat{\sigma}^2\right) + g_{2d}\left(\hat{\sigma}^2\right) + 2g_{3d}\left(\hat{\sigma}^2\right) + g_{4d}\left(\hat{\sigma}^2\right),$$

donde:

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

$$g_{2d}\left(\hat{\sigma}^2\right) = (1 - f_d)^2 \left[\left(\overline{X}_{d(p_r)} - \hat{\gamma}_{dc}\overline{x}_{dc}\right)' \hat{\Phi}_s \left(\overline{X}_{d(p_r)} - \hat{\gamma}_{dc}\overline{x}_{dc}\right)\right],$$

$$g_{3d}\left(\hat{\sigma}^2\right) = (1 - f_d)^2 c_d^{-1}\left(\hat{\sigma}_v^2 + \hat{\sigma}_e^2 / c_{d.}\right)^{-3}\left[\hat{\sigma}_e^4 \,\hat{\text{var}}\left(\hat{\sigma}_v^2\right) + \hat{\sigma}_v^4 \,\hat{\text{var}}\left(\hat{\sigma}_e^2\right) - 2\hat{\sigma}_e^2\hat{\sigma}_v^2 \,\hat{\text{cov}}\left(\hat{\sigma}_e^2, \hat{\sigma}_v^2\right)\right]$$

$$g_{4d}\left(\hat{\sigma}^2\right) = \sigma_e^2 N_d^{-2} \sum_{j \in p_r} c_{dj}^{-1} ,$$

son las contribuciones al ECM de la estimación de los efectos aleatorios, los efectos fijos, los componentes de varianza y los pesos del modelo. Además, p_r representa el dominio d de establecimientos censales no pertenecientes a la muestra y $c_{dj} = 1/x_{dj}$, donde x_{dj} es el empleo censal de los establecimientos no muestreados. Si se hacen agregaciones para conseguir un tamaño mínimo antes de proceder a las estimaciones, se considera como población no muestreada la de la subclase en la que se hace la proyección, y no la de la agregación que suele ser superior.

El MSE del predictor del total para cada 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,

$$M\hat{S}E\left[\hat{t}_d\right] = N_d^2\left[g_{1d}\left(\hat{\sigma}^2\right) + g_{2d}\left(\hat{\sigma}^2\right) + 2g_{3d}\left(\hat{\sigma}^2\right)\right], \tag{18}$$

y la raíz cuadrada del error cuadrático medio del total se estima mediante la expresión:

$$RM\hat{S}E\left[\hat{t}_d\right] = \sqrt{MSE\left[\hat{t}_d\right]}. \tag{19}$$

El coeficiente de variación, se define como:

$$CV\left[\hat{t}_d\right] = \frac{RM\hat{S}E\left[\hat{t}_d\right]}{\hat{t}_d}.$$

2.3.- Modelo lineal de efectos fijos

El modelo superpoblacional de efectos fijos viene dado por:

$$Y = X\beta + \varepsilon , \qquad \varepsilon \subset N\left(0, \sigma_e^2 C^{-1}\right), \tag{20}$$

donde $C = diag(c_{dj})$, es la matriz de pesos del modelo, d representa la comarca ($d = 1,\ldots,t$) y j es el establecimiento ($j = 1,\ldots,N_d$). El vector $Y = \left(Y_1^{\prime},\ldots,Y_t^{\prime}\right)$ es el vector $(N \times 1)$ cuyas componentes $Y_d^{\prime}$ son los valores observados de la variable de interés para cada comarca d, β es el único coeficiente fijo del modelo, X es el vector columna ($N \times 1$) de la variable auxiliar, es decir del empleo, y $\varepsilon^{\prime} = \left(\varepsilon_1^{\prime},\ldots,\varepsilon_t^{\prime}\right)$ donde $\varepsilon_d^{\prime} = \left(\varepsilon_{d_1},\ldots,\varepsilon_{d_{N_d}}\right)$ es el vector de errores aleatorios.

De forma similar a la descomposición realizada en el modelo mixto, se puede separar la parte muestreada y no muestreada del siguiente modo:

$$\begin{pmatrix} Y_s \\ Y_r \end{pmatrix} = \begin{pmatrix} X_s \\ X_r \end{pmatrix} \beta + \begin{pmatrix} \varepsilon_s \\ \varepsilon_r \end{pmatrix}$$

donde los subíndices s y r denotan los establecimientos muestreados y no muestreados respectivamente. Entonces el modelo muestral de efectos fijos puede escribir como

$$Y_s = X_s \beta + \varepsilon_s \qquad \varepsilon_s \subset N(0_s, \sigma_e^2 C_s^{-1}), \tag{21}$$

donde $C_s = diag\left(c_{dj}\right)$, $d = 1,\ldots,t_s$, $j = 1,\ldots,n_d$ y t_s es el número total de comarcas donde se ha muestreado. De forma extendida, el modelo (21) se expresa como:

$$y_{dj} = \beta x_{dj} + e_{dj} \qquad d = 1,\ldots,t,\ j = 1,\ldots,n_d \tag{22}$$

donde para el establecimiento j de la comarca d, y_{dj} es el valor que toma la variable de interés y x_{dj} es el número de empleados del establecimiento. El número total de establecimientos muestreados en la comarca d es n_d, β es el único efecto fijo del modelo y $e_{dj} \subset N\left(0, \sigma_e^2 c_{dj}^{-1}\right)$ son los errores aleatorios. Además, $e_{dj} \subset N\left(0, \sigma_e^2 c_{dj}^{-1}\right)$. Para corregir la heterocedasticidad presente en los datos se utilizan los pesos $c_{dj} = 1/x_{dj}$.

Si N_d es el número total de unidades del área d, la media poblacional del área d viene dada por

$$\bar{Y}_d = \frac{1}{N_d} \sum_{d=1}^{N_d} y_{dj} = f_d \bar{y}_{ds} + \left(1 - f_d\right) \bar{y}_{dr}, \tag{23}$$

donde $f_d = n_d / N_d$, $\bar{y}_{ds}$ es la media muestral de las unidades muestreadas y $\bar{y}_{dr}$ es la media muestreal de las no muestreadas. Dado que el segundo término de (23) no se ha observado, se sustituye por su valor estimado. Un estimador de (23) obtenido de manera similar al dado en (7) viene dado por:

$$\hat{\bar{y}}_d^F = f_d \bar{y}_{ds} + (1 - f_d)\bar{X}_{d(p_r)}\hat{\beta}\,, \tag{24}$$

donde $\bar{X}_{d(p_r)} = \sum_{j \in d_r} x_{dj} / (N_d - n_d)$ es la media poblacional del número de empleados no muestreados en el área d.

El estimador de β viene dado por

$$\hat{\beta} = \left(X_s^{\prime} C_s X_s\right)^{-1}\left(X_s^{\prime} C_s Y_s\right) = \sum_{d=1}^{t_s}\sum_{j=1}^{n_d} y_{dj} \Big/ \sum_{d=1}^{t_s}\sum_{j=1}^{n_d} x_{dj}$$

que es el estimador por mínimos cuadrados generalizados de β y

$$Var(\hat{\beta}) = \sigma^2\left(X_s^{\prime} C_s X_s\right)^{-1} = \sigma^2 \Big/ \sum_{d=1}^{t_s}\sum_{j=1}^{n_d} x_{dj}$$

es su matriz de varianzas y covarianzas.

El estimador del total para la comarca d se obtiene como:

$$\hat{t}_d^F = \sum_{j=1}^{n_d} y_{dj} + X_{d(p_r)}\hat{\beta}\,, \tag{25}$$

y el estimador de la media por Territorio Histórico viene dado por:

$$\hat{\bar{y}}_h^F = \frac{1}{N_h}\sum_{d \in h} N_d \hat{\bar{y}}_d^F\,, \tag{26}$$

donde $d \in h$ indica que la suma se efectúa en todas las áreas del estrato h (en este caso $h = 1,2,3$ son los Territorios Históricos) y $N_h = \sum_{d \in h} N_d$ es el total poblacional del Territorio Histórico h.

El estimador del total por Territorio Histórico viene dado por:

$$\hat{t}_h^F = \sum_{d \in h} N_d \hat{\bar{y}}_d^F = \sum_{d \in h} \hat{t}_d^F \tag{27}$$

donde $d \in h$ indica que la suma se efectúa en todas las áreas del estrato h (en este caso $h = 1,2,3$ son los Territorios Históricos) y $N_h = \sum_{d \in h} N_d$ es el total poblacional del Territorio Histórico h.

El estimador de la media y del total para la C.A. de Euskadi vienen dados, respectivamente por:

$$\hat{y}^F = \frac{1}{N}\sum_{h=1}^{3} N_h \hat{y}_h^F \, , \qquad \hat{t}^F = \sum_{h=1}^{3} \hat{t}_h^F \, . \qquad (28)\ (29)$$

Los errores cuadráticos medios de los estimadores de la media por comarcas (24), en su versión predictiva, vienen dados por:

$$M\hat{S}E\left[\hat{\bar{y}}_d^F\right] = E\left[\left(\hat{\bar{Y}}_d^F - \bar{Y}_d\right)^2\right] = (1-f_d)^2\left[\bar{X}'_{d(p_r)}\, \hat{\mathrm{var}}\left(\hat{\beta}\right)\bar{X}_{d(p_r)}\right] + \frac{\sigma^2}{N_d^2}\sum_{j\in p_r} x_{dj}$$

(30)

donde p_r es el empleo censal no muestreado. Si se hacen agregaciones para conseguir un tamaño mínimo antes de proceder a las estimaciones, se considera como población no muestreada la de la subclase en la que se hace la proyección, y no la de la agregación que suele ser superior.

El MSE para el total de la comarca se estima mediante:

$$M\hat{S}E\left[\hat{t}_d^F\right] = N_d^2(1-f_d)^2\left[\bar{X}_{d(p_r)}\, \hat{\mathrm{var}}\left(\hat{\beta}\right)\bar{X}_{d(p_r)}\right] + \sigma^2\sum_{j\in p_r} x_{dj} \qquad (31)$$

2.4.- Totales por sector

Los modelos presentados hasta ahora permiten obtener las predicciones de los totales por comarcas, Territorios Históricos y C.A. de Euskadi para cada subclase, así como sus errores estándar (RMSE) y coeficientes de variación. Para el cálculo de los totales por sector, por ejemplo para la clasificación A84, simplemente se realiza una agregación de totales obtenidos por comarcas para cada una de las subclases que lo conforman, agregación que puede hacerse por comarcas, Territorios Históricos y C.A. de Euskadi. Los errores estándar del sector se obtienen como raíz cuadrada de la suma de los MSE de cada subclase. Dentro del mismo sector puede ocurrir que unas subclases se hayan estimado por modelos mixtos y otras por fijos. Para obtener los resultados por un segundo nivel de agregación, por ejemplo para la clasificación A31 se agregan los totales de los sectores A84 que correspondan. El cálculo de errores estándar se hace también suponiendo que los sectores son independientes y por tanto para calcular el RMSE se calcula la raíz cuadrada de la suma de los MSE de los sectores correspondientes.

2.5.- Proceso de calibración

La calibración permite obtener exactamente los mismos totales que los proporcionados por el estimador de la Encuesta Industrial a nivel de Territorio Histórico y C.A. de Euskadi por A84 u otro nivel de agregación. Si llamamos t_d a la estimación del total obtenida bajo los modelos en la comarca d por un sector concreto, t_h a la estimación por modelos obtenida por Territorio Histórico y C_h a la estimación del total por Territorio Histórico obtenida por la Encuesta Industrial, entonces la nueva estimación calibrada por comarcas para cada sector vendrá dada por:

$$\tilde{t}_d = t_d \frac{C_h}{t_h}.$$

De este modo el total calibrado por modelos $\tilde{t}_h$ para cada Territorio Histórico en cada sector coincide con el total estimado C_h para cada Territorio Histórico con el estimador de la Encuesta Industrial para ese mismo sector, ya que:

$$\tilde{t}_h = \sum_d \tilde{t}_d = \sum_d t_d \frac{C_h}{t_h} = \frac{C_h}{t_h} \sum_d t_d = C_h.$$

2.6.- Plan de estimación en la Encuesta Industrial

Eustat ha programado una aplicación informática *ad hoc* en SAS para la introducción del cálculo de estimaciones de áreas pequeñas en la producción estadística. Se trata de un programa específico para la Encuesta Industrial pero que fácilmente puede adaptarse a otro tipo de encuestas económicas. Son diversas las decisiones que se han tomado y que se han programado:

- Tanto el modelo lineal mixto como el modelo lineal de efectos fijos se calculan en la versión predictiva. Ello es debido a que las poblaciones de algunas subclases son muy pequeñas, con lo que la fracción de muestreo no es despreciable.

- La versión predictiva separa la predicción en dos partes. La obtenida por el modelo para los establecimientos no muestreados y la observada para los establecimientos muestreados. Esta forma es especialmente útil en la Encuesta Industrial debido a que en algunas subclases hay presencia de establecimientos atípicos, es decir, de establecimientos cuyo comportamiento es muy diferente del resto, y que pueden distorsionar considerablemente las estimaciones. Estos establecimientos se clasifican como muestreados no válidos y al utilizar la versión predictiva no se tienen en cuenta en los procedimientos de estimación, pero sí en la predicción final. Los establecimientos clasificados como no válidos se suman al igual que el resto de la muestra, a la predicción de los no observados para obtener la predicción del total.

- Se ha considerado necesario establecer un número mínimo de establecimientos para proceder al cálculo de los modelos mixtos o fijos. Si no se dispone de este número mínimo de establecimientos, se procede a hacer agregaciones de CNAEs con un dígito menos. Primero se estima el modelo mixto a ese nivel de agregación y si $\sigma_v^2 = 0$ ó $\sigma_e^2 = 0$ entonces se estima el modelo de efectos fijos. Este número mínimo se ha fijado en la actualidad (puede ser variado) en 5 establecimientos.

- Se ha tomado la decisión de utilizar el modelo de efectos fijos cuando el modelo mixto no es válido debido a que se considera prioritaria la decisión de no agrupar subclases. Es por ello, que se prefiere un modelo de efectos fijos con 5 dígitos por ejemplo, a un modelo mixto de 3 dígitos.

- Cuando se realiza una agregación, ésta permite estimar los coeficientes del modelo pero las predicciones se hacen de forma particularizada a la subclase considerada.

- El uso de la variable auxiliar "número de empleados" introduce heterocedasticidad en los modelos, ya que habitualmente la variable respuesta y tiene mayor variabilidad a medida que aumenta el número de empleados. Por ello, todos los modelos de efectos fijos y mixtos consideran que la varianza del error es proporcional al número de empleados.

- En cada subclase los totales por Territorio Histórico y por C.A. de Euskadi se obtienen de manera agregada a partir de las estimaciones por comarcas. Los totales por sector A84 se obtienen agregando las predicciones obtenidas a nivel de subclase. Lo mismo sucede para los totales por Territorios Históricos y C.A. de Euskadi. Se procede de igual modo para otros tipos de agregaciones.

- En cada subclase, para calcular las raíces cuadradas de los errores cuadráticos medios de las predicciones a nivel de Territorio Histórico, se aplican fórmulas específicas, ya que no se obtiene como raíz cuadrada de la suma de los errores cuadráticos medios de las predicciones por comarcas. Ello es debido a que en cada subclase las estimaciones por comarcas no son independientes en ninguno de los modelos.

- Sin embargo, una vez hechas las estimaciones de los RMSE por subclases para cada Territorio Histórico, el cálculo de las estimaciones por C.A. de Euskadi es directo, a que ahora se cumple la hipótesis de independencia. A partir de los cálculos por subclases los RMSE de la variable A84 o de cualquier otra agrupación se obtienen de forma directa, es decir, calculando la raíz cuadrada de la suma de los MSE de las subclases que forman cada sector.

Capítulo

3

Aplicación de técnicas de estimación en áreas pequeñas a la Encuesta Industrial de la C.A. de Euskadi. 2002 y 2003.

3.1.- Estimaciones comarcales en la Encuesta Industrial de la C.A. de Euskadi. 2002 y 2003

A continuación se presentan las estimaciones obtenidas utilizando el sistema de estimación antes expuesto en la Encuesta Industrial de la C.A. de Euskadi correspondientes a los años 2002 y 2003. Las macromagnitudes escogidas para su publicación han sido:

- Valor añadido a coste de factores
- Ventas netas
- Costes de personal
- Excedente bruto de explotación
- Inversión
- Resultado antes de impuestos

Se ofrecen también las estimaciones del Empleo de las comarcas dado que ésta ha sido la variable exógena utilizada en los modelos de áreas pequeñas utilizados en la Encuesta Industrial de la C.A. de Euskadi.

Personal ocupado de la industria por territorio histórico y comarca. 2002-2003

	2002	2003	Δ 03/02
C.A. de Euskadi	**246.063**	**248.922**	**1,2**
Alava	**47.804**	**49.019**	**2,5**
Valles Alaveses	1.889	1.953	3,4
Llanada Alavesa	30.978	31.912	3,0
Montaña Alavesa	354	294	-16,9
Rioja Alavesa	3.778	4.042	7,0
Estribaciones del Gorbea	3.942	3.949	0,2
Cantábrica Alavesa	6.863	6.869	0,1
Bizkaia	**103.749**	**104.619**	**0,8**
Arratia-Nervión	5.173	5.021	-2,9
Gran Bilbao	60.622	61.312	1,1
Duranguesado	21.614	21.736	0,6
Encartaciones	2.128	2.139	0,5
Gernika-Bermeo	4.418	4.612	4,4
Markina-Ondarroa	3.892	4.056	4,2
Plentzia-Mungia	5.902	5.743	-2,7
Gipuzkoa	**94.510**	**95.284**	**0,8**
Bajo Bidasoa	5.490	5.602	2,0
Bajo Deba	10.452	10.643	1,8
Alto Deba	20.275	20.091	-0,9
Donostia-San Sebastián	25.464	25.508	0,2
Goierri	12.831	13.233	3,1
Tolosa	8.001	8.185	2,3
Urola Costa	11.997	12.022	0,2

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Valor añadido a coste de factores de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**13.008.214**	**0,01**	**13.371.649**	**0,01**	**2,8**
Alava	**2.676.614**	**0,00**	**2.747.477**	**0,00**	**2,6**
Valles Alaveses	89.743	0,03	106.061	0,02	18,2
Llanada Alavesa	1.708.733	0,01	1.763.392	0,01	3,2
Montaña Alavesa	17.455	0,19	14.211	0,16	-18,6
Rioja Alavesa	243.155	0,02	270.594	0,04	11,3
Estribaciones del Gorbea	216.196	0,02	210.776	0,01	-2,5
Cantábrica Alavesa	401.332	0,01	382.443	0,01	-4,7
Bizkaia	**5.371.448**	**0,01**	**5.561.854**	**0,01**	**3,5**
Arratia-Nervión	252.658	0,02	259.622	0,01	2,8
Gran Bilbao	3.259.555	0,01	3.395.279	0,01	4,2
Duranguesado	1.109.460	0,01	1.113.048	0,01	0,3
Encartaciones	154.460	0,05	155.544	0,04	0,7
Gernika-Bermeo	180.951	0,03	199.627	0,02	10,3
Markina-Ondarroa	149.595	0,03	167.664	0,02	12,1
Plentzia-Mungia	264.769	0,01	271.070	0,01	2,4
Gipuzkoa	**4.960.152**	**0,01**	**5.062.318**	**0,01**	**2,1**
Bajo Bidasoa	218.659	0,02	231.525	0,02	5,9
Bajo Deba	482.592	0,02	482.471	0,02	0,0
Alto Deba	996.600	0,01	1.013.516	0,00	1,7
Donostia-San Sebastián	1.585.031	0,01	1.591.077	0,01	0,4
Goierri	655.356	0,01	682.139	0,01	4,1
Tolosa	422.004	0,02	442.768	0,01	4,9
Urola Costa	599.910	0,01	618.822	0,01	3,2

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Ventas netas de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**42.393.031**	**0,01**	**43.768.410**	**0,01**	**3,2**
Alava	**8.336.340**	**0,01**	**8.614.962**	**0,01**	**3,3**
Valles Alaveses	321.016	0,02	390.287	0,02	21,6
Llanada Alavesa	5.521.407	0,01	5.564.479	0,01	0,8
Montaña Alavesa	46.122	0,17	40.494	0,17	-12,2
Rioja Alavesa	603.353	0,02	702.555	0,03	16,4
Estribaciones del Gorbea	705.855	0,02	763.485	0,01	8,2
Cantábrica Alavesa	1.138.586	0,01	1.153.661	0,01	1,3
Bizkaia	**19.341.307**	**0,01**	**20.078.450**	**0,01**	**3,8**
Arratia-Nervión	743.317	0,03	822.265	0,02	10,6
Gran Bilbao	12.733.700	0,01	13.268.830	0,01	4,2
Duranguesado	3.518.902	0,02	3.562.602	0,01	1,2
Encartaciones	422.018	0,05	417.302	0,03	-1,1
Gernika-Bermeo	673.681	0,03	727.858	0,02	8,0
Markina-Ondarroa	495.976	0,03	524.603	0,03	5,8
Plentzia-Mungia	753.713	0,03	754.989	0,02	0,2
Gipuzkoa	**14.715.384**	**0,01**	**15.074.998**	**0,01**	**2,4**
Bajo Bidasoa	581.630	0,03	630.369	0,02	8,4
Bajo Deba	1.273.758	0,03	1.238.176	0,04	-2,8
Alto Deba	3.134.142	0,01	3.202.536	0,01	2,2
Donostia-San Sebastián	4.600.028	0,01	4.557.629	0,01	-0,9
Goierri	2.039.115	0,01	2.224.916	0,01	9,1
Tolosa	1.239.249	0,04	1.315.686	0,02	6,2
Urola Costa	1.847.462	0,02	1.905.687	0,03	3,2

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Costes de personal de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**7.398.012**	**0,00**	**7.808.500**	**0,00**	**5,5**
Alava	**1.477.956**	**0,00**	**1.576.926**	**0,00**	**6,7**
Valles Alaveses	57.176	0,01	61.286	0,01	7,2
Llanada Alavesa	969.544	0,00	1.024.924	0,00	5,7
Montaña Alavesa	7.968	0,06	7.258	0,04	-8,9
Rioja Alavesa	88.179	0,03	106.669	0,03	21,0
Estribaciones del Gorbea	119.351	0,01	126.500	0,01	6,0
Cantábrica Alavesa	235.738	0,00	250.291	0,00	6,2
Bizkaia	**3.108.386**	**0,00**	**3.273.633**	**0,00**	**5,3**
Arratia-Nervión	154.341	0,01	155.158	0,01	0,5
Gran Bilbao	1.879.256	0,01	1.976.019	0,01	5,1
Duranguesado	638.435	0,01	675.569	0,01	5,8
Encartaciones	57.201	0,02	62.725	0,02	9,7
Gernika-Bermeo	114.381	0,01	125.955	0,01	10,1
Markina-Ondarroa	93.588	0,01	102.151	0,01	9,1
Plentzia-Mungia	171.183	0,02	176.056	0,01	2,8
Gipuzkoa	**2.811.670**	**0,00**	**2.957.941**	**0,00**	**5,2**
Bajo Bidasoa	138.692	0,01	150.936	0,01	8,8
Bajo Deba	303.162	0,01	319.017	0,01	5,2
Alto Deba	625.672	0,00	652.605	0,00	4,3
Donostia-San Sebastián	749.887	0,01	775.827	0,01	3,5
Goierri	403.843	0,00	436.067	0,01	8,0
Tolosa	240.291	0,01	259.569	0,01	8,0
Urola Costa	350.123	0,01	363.920	0,01	3,9

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Excedente bruto de explotación de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**5.610.202**	**0,01**	**5.563.149**	**0,01**	**-0,8**
Alava	**1.198.658**	**0,01**	**1.170.551**	**0,01**	**-2,3**
Valles Alaveses	32.931	0,07	44.762	0,05	35,9
Llanada Alavesa	739.335	0,01	737.566	0,02	-0,2
Montaña Alavesa	9.530	0,34	7.186	0,29	-24,6
Rioja Alavesa	156.177	0,03	164.114	0,05	5,1
Estribaciones del Gorbea	95.502	0,04	83.599	0,03	-12,5
Cantábrica Alavesa	165.183	0,01	133.325	0,03	-19,3
Bizkaia	**2.263.062**	**0,01**	**2.288.221**	**0,01**	**1,1**
Arratia-Nervión	98.254	0,04	104.809	0,03	6,7
Gran Bilbao	1.376.066	0,02	1.418.524	0,01	3,1
Duranguesado	474.025	0,02	438.507	0,02	-7,5
Encartaciones	97.238	0,07	92.094	0,05	-5,3
Gernika-Bermeo	66.908	0,07	73.657	0,05	10,1
Markina-Ondarroa	56.424	0,06	64.772	0,05	14,8
Plentzia-Mungia	94.146	0,03	95.858	0,04	1,8
Gipuzkoa	**2.148.482**	**0,01**	**2.104.377**	**0,01**	**-2,1**
Bajo Bidasoa	79.729	0,05	79.886	0,04	0,2
Bajo Deba	178.555	0,04	164.393	0,05	-7,9
Alto Deba	373.600	0,02	358.072	0,02	-4,2
Donostia-San Sebastián	831.687	0,02	814.317	0,02	-2,1
Goierri	254.308	0,02	247.468	0,02	-2,7
Tolosa	181.731	0,04	185.976	0,04	2,3
Urola Costa	248.871	0,03	254.264	0,03	2,2

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Inversión realizada de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**3.211.667**	**0,02**	**2.717.879**	**0,01**	**-15,4**
Alava	**660.740**	**0,02**	**510.428**	**0,01**	**-22,7**
Valles Alaveses	36.549	0,55	26.231	0,03	-28,2
Llanada Alavesa	457.031	0,02	332.071	0,02	-27,3
Montaña Alavesa	2.396	0,17	2.198	0,20	-8,3
Rioja Alavesa	58.364	0,15	60.380	0,05	3,5
Estribaciones del Gorbea	55.831	0,45	34.254	0,04	-38,6
Cantábrica Alavesa	50.569	0,04	55.295	0,02	9,3
Bizkaia	**1.459.249**	**0,02**	**1.364.705**	**0,01**	**-6,5**
Arratia-Nervión	41.686	0,35	41.403	0,05	-0,7
Gran Bilbao	1.030.310	0,02	1.079.300	0,01	4,8
Duranguesado	248.724	0,27	165.607	0,03	-33,4
Encartaciones	42.907	1,80	18.728	0,08	-56,4
Gernika-Bermeo	27.135	0,54	8.636	0,18	-68,2
Markina-Ondarroa	25.626	0,07	24.139	0,06	-5,8
Plentzia-Mungia	42.860	0,35	26.891	0,06	-37,3
Gipuzkoa	**1.091.678**	**0,02**	**842.746**	**0,02**	**-22,8**
Bajo Bidasoa	53.852	0,54	37.669	0,04	-30,1
Bajo Deba	93.377	0,23	79.199	0,05	-15,2
Alto Deba	193.433	0,02	213.272	0,01	10,3
Donostia-San Sebastián	370.970	0,04	216.867	0,03	-41,5
Goierri	142.977	0,30	115.465	0,03	-19,2
Tolosa	137.092	1,17	79.849	0,15	-41,8
Urola Costa	99.976	0,05	100.425	0,04	0,4

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

Resultado antes de impuestos de la industria y coeficiente de variación (cv) por territorio histórico y comarca (miles de €). 2002-2003

	2002	cv	2003	cv	Δ 03/02
C.A. de Euskadi	**3.213.405**	**0,02**	**2.904.928 0**	**0,02**	**-9,6**
Alava	**772.874**	**0,02**	**579.541**	**0,02**	**-25,0**
Valles Alaveses	13.060	0,15	18.734	0,10	43,5
Llanada Alavesa	468.551	0,02	316.075	0,04	-32,5
Montaña Alavesa	6.934	0,46	6.568	0,31	-5,3
Rioja Alavesa	145.513	0,02	103.511	0,07	-28,9
Estribaciones del Gorbea	40.156	0,07	61.471	0,03	53,1
Cantábrica Alavesa	98.660	0,02	73.182	0,05	-25,8
Bizkaia	**1.344.669**	**0,02**	**1.274.144**	**0,02**	**-5,2**
Arratia-Nervión	64.725	0,06	67.487	0,04	4,3
Gran Bilbao	838.007	0,02	774.138	0,02	-7,6
Duranguesado	263.632	0,03	243.464	0,03	-7,7
Encartaciones	63.818	0,10	62.697	0,08	-1,8
Gernika-Bermeo	37.564	0,11	31.403	0,11	-16,4
Markina-Ondarroa	23.769	0,13	34.147	0,08	43,7
Plentzia-Mungia	53.155	0,04	60.808	0,06	14,4
Gipuzkoa	**1.095.862**	**0,03**	**1.051.243**	**0,03**	**-4,1**
Bajo Bidasoa	37.231	0,10	35.167	0,09	-5,5
Bajo Deba	101.389	0,07	91.380	0,09	-9,9
Alto Deba	139.336	0,04	127.648	0,04	-8,4
Donostia-San Sebastián	429.737	0,03	418.766	0,03	-2,6
Goierri	127.438	0,04	108.258	0,05	-15,1
Tolosa	108.981	0,06	123.785	0,05	13,6
Urola Costa	151.750	0,05	146.238	0,05	-3,6

Fuente: EUSTAT. Cuentas Industriales y de la Construcción

3.2.- Valor añadido a coste de factores y Personal ocupado por comarcas en la Encuesta Industrial de la C.A. de Euskadi. 2002 y 2003. Conclusiones.

A continuación se extraen las principales conclusiones del análisis de los resultados presentados comarcalmente para las dos variables más relevantes de la Encuesta Industrial: Valor añadido a coste de factores y Personal ocupado.

De la estimación de los datos de la Industria vasca por comarcas para el año 2003, destaca que la comarca del Gran Bilbao representa una cuarta parte del valor añadido industrial (25,4%) de la C. A. de Euskadi. Le siguen la Llanada Alavesa con el 13,2% y la comarca de Donostia-San Sebastián, con el 11,9%. Estas tres comarcas aportaron algo más de la mitad del valor añadido industrial (50,5%) de la C. A. de Euskadi en 2003 y el 50,4% en el 2002. En términos de personal empleado industrial, estas tres comarcas ocupaban al 47,7% de los efectivos industriales de la C. A. de Euskadi.

El Duranguesado, el Alto Deba y la Cantábrica Alavesa aportaban el 18,8% del valor añadido industrial de la C. A. de Euskadi y representaban el 19,6% del personal ocupado de dicho sector. Por lo tanto, el 69,3% de la industria se concentra en seis de las veinte comarcas de la C. A. de Euskadi. Sin embargo, este fenómeno de concentración-dispersión varía mucho de un territorio a otro.

En el Territorio Histórico de Álava, casi las dos terceras partes del valor añadido industrial (en concreto el 64,2%) se concentran en la Llanada Alavesa, mientras en la comarca de la Montaña Alavesa se ubica tan sólo el 0,5%. Las cuatro comarcas restantes de este territorio tienen un peso industrial que se sitúa entre el 13,9% de Cantábrica Alavesa y el 3,9% de los Valles Alaveses.

El efecto de la concentración comarcal de la industria es también fuerte en Bizkaia, donde el 81,0% de la misma se acumula en dos comarcas, el Gran Bilbao (61,0%) y el Duranguesado (20,0%), mientras que la comarca de las Encartaciones apenas cubre el 2,8% de la actividad industrial del territorio. La tercera comarca de mayor peso industrial es la de Arratia-Nervión, con el 4,7% del valor añadido industrial de este territorio.

En Gipuzkoa la concentración es menor y el equilibrio intercomarcal más ponderado que en los otros dos territorios. La mitad de la industria se concentra en dos comarcas: Donostia-San Sebastián (31,4%) y el Bajo Deba (20,0%), pero, al mismo tiempo, existen comarcas como la del Goierri y Urola Costa, con el 13,5% y el 12,2% de peso industrial relativo territorial, respectivamente. Por otro lado, la comarca de menor peso industrial es la del Bajo Bidasoa, en Gipuzkoa, pero llega a cubrir el 4,6% de la industria del territorio, frente al 0,5% de la Montaña Alavesa y al 2,8% de las Encartaciones.

La Industria de la C. A. de Euskadi tuvo un crecimiento nominal del 2,8% en el 2003 y esta media de crecimiento ha sido superada por la comarcas de Valles Alaveses (18,2%), Rioja Alavesa (11,3%) y Llanada Alavesa (3,2%), en Álava; por Marquina-Ondarroa (12,1%), Gernika-Bermeo (10,3%) y Gran Bilbao (4,2%) en Bizkaia; y por Bajo Bidasoa (5,9%), Tolosa (4,9%), Goierri (4,1%) y Urola Costa (3,2%) en Gipuzkoa.

En cuanto a la evolución del personal empleado en el 2003, todas las comarcas alavesas muestran un perfil positivo, a excepción de la Montaña Alavesa. Solamente dos comarcas de Bizkaia (Arratia-Nervión y Plentzia-Mungia) retroceden en el empleo industrial, mientras que todas las demás aumentan la ocupación industrial. En lo referente a Gipuzkoa, únicamente una comarca (Alto Deba) reduce los efectivos industriales, incrementando, en mayor o menor grado las seis restantes.

Capítulo

4

Conclusiones

Los modelos de áreas pequeñas descritos en este documento permiten obtener estimaciones de totales de las diferentes macromagnitudes de la Encuesta Industrial a nivel de comarcas. Los resultados obtenidos son altamente satisfactorios con carácter general, dado que los coeficientes de variación estimados son ciertamente moderados en aquellas comarcas en las que la representación muestral y el tamaño poblacional es relativamente elevado. En comarcas donde esto no ocurre, los coeficientes de variación estimados son mayores.

Eustat ofrecerá a partir del presente año estimaciones comarcales de la Encuesta Industrial a un nivel de desagregación mayor o igual que el presentado en este documento con carácter anual. Esto permitirá disponer, en un futuro cercano, de series temporales de estimaciones comarcales, que permitirán realizar análisis coyunturales más completos que los actuales, facilitando un mayor conocimiento de la evolución de las principales macromagnitudes económicas dentro de la C.A. de Euskadi. Toda esta información podrá ser de gran utilidad en la confección de las diferentes políticas territoriales.

Es preciso añadir, que la realización de este proyecto de estimación de áreas pequeñas ha conllevado una rigurosa revisión de los procesos de elevación de la Encuesta Industrial, que se ha traducido en mejoras metodológicas en el cálculo de estimadores y en el cálculo de los errores muestrales.

Eustat sigue investigando la metodología de los modelos de estimación en áreas pequeñas con objeto de dar más adelante un paso más y poder ofrecer estimaciones de calidad a un nivel de desagregación aún mayor. Se está también trabajando en la utilización de esta metodología en otras encuestas de Eustat, no sólo en encuestas relacionadas con el ámbito industrial, si no también en encuestas donde todas o parte de las variables a estimar son discretas.

Bibliografía

[1] Battese, G.E., Harter, R.M and Fuller, W.A. (1988). An Error-Components Model for Prediction of Country Crop Areas Using Survey and Satellite Data. *Journal of the American Statistical Association*, **83**, 28-36.

[1] Henderson, C.R. (1975). Best Linear Unbiased Estimation and Prediction Under a Selection Model. *Biometrics*, **31**, 423-447.

[2] Kackar, R.N. and Harville, D.A.. (1984). Approximations for standard errors of estimators of fixed and random effects in mixed linear models. Journal of the American Statistical Assocition.

[2] Pfefferman, D. (2002). Small Area Estimation – New Developments and Directions. International Statistical Review, 70, 125-143.

[3] Prasad, N.G.N and Rao, J.N.K (1990). The Estimation of Mean Squared Error of Small Area Estimators. Journal of the American Statistical Association, 85, 163-171.

[4] Rao, J.N.K. (2003). Small Area Estimation. Wiley Series in Survey Methodology.

[5] Särndal, C.E., Hidirogloy, M.A. (1989). Small Domain Estimation: A Conditional Analysis. Journal of the American Statistical Association, 84, 266-275.

[6] Searle, S.R.., Casella, G. and McCullogh, C.E. (1992). Variance Components. Wiley Series in Probability and Statistics.