

Estimación de los coeficientes de variación en la Encuesta Industrial

Índice

1. Breve descripción de la Encuesta Industrial	
1.1. Objetivos de la operación	3
1.2. Ámbito de la operación.....	3
1.3. Diseño muestral	4
1.4. Postestratificación.....	5
2. Sistema de estimación en la Encuesta Industrial	
2.1. Introducción.....	5
2.2. Estimador indirecto de razón y su varianza	5

1. Breve descripción de la Encuesta Industrial

1.1 Objetivos de la operación

Los objetivos principales de la Encuesta Industrial de la C.A. de Euskadi se pueden resumir en:

- conocer las características estructurales de la industria vasca y su evolución;
- estimar las principales macromagnitudes económicas del sector industrial, base esencial para la elaboración de las Tablas Input-Output y las Cuentas Económicas;
- desarrollar una batería de indicadores que determinen los aspectos relevantes del sector y que sirvan como base para la obtención de indicadores coyunturales elaborados por EUSTAT como el Índice de Producción Industrial y el Índice de Precios Industriales;
- facilitar un instrumento útil y válido a los agentes sociales para el análisis económico, tan vital en esta época de ciclos fluctuantes;
- utilización de los resultados de la operación para la elaboración del Avance de Cuentas Económicas.

1.2 Ámbito de la operación

Ámbito poblacional: 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

Ámbito geográfico: Las unidades estadísticas que están ubicadas en el ámbito geográfico de la Comunidad Autónoma de Euskadi, aun cuando su sede social o gerencia se encuentre fuera de la CAE.

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

1.3 Diseño Muestral

Marco de la encuesta. El Directorio de Actividades Económicas de Eustat ha sido la base para establecer el marco de la encuesta. Su utilización permite la elaboración de un muestreo probabilístico que acote los errores muestrales.

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 (se incluyen también entre las autoponderadas las altas del Directorio del año t); en la segunda fase se realiza un muestreo aleatorio estratificado, también en dos subfases:

▪ **Subfase 1.** Las variables de estratificación utilizadas son:

- Territorio Histórico: Araba, Bizkaia y Gipuzkoa.
- Actividad: clasificación normalizada de EUSTAT A-84. Se calcula la afijación óptima de cada estrato con la restricción presupuestaria correspondiente a cada año, que en este caso es que el tamaño muestral global no exceda de 3.500 unidades estadísticas. La afijación óptima se calcula sobrerrepresentando Araba con una afijación proporcional a la raíz cuadrada del número de establecimientos de cada Territorio Histórico.

La distribución sectorial dentro de cada territorio histórico se realiza con la siguiente formulación:

$$n_{s,h} = n_h \frac{T_{s,h}}{\sum T_{s,h}}, \quad (1)$$

Con $T_{s,h} = N_{s,h} \sqrt{E_{s,h}}$, donde $N_{s,h}$ es el número de establecimientos en el sector s en el Territorio Histórico h , $E_{s,h}$ es el empleo medio en el sector s en el Territorio Histórico h y n_h es el tamaño muestral en el Territorio Histórico h , siendo $n = \sum n_h$ el tamaño muestral total.

- **Subfase 2.** La distribución de la afijación óptima en cada sector de actividad se hace, siguiendo un criterio de afijación óptima, proporcionalmente al peso de cada subclase en el total del sector. El Servicio de Análisis Sectorial del Departamento de Agricultura y Pesca en las actividades 15.1, 15.5, 15.321, 15.333, 15.411, 15.413, 15.7, 15.93, 15.94, 15.95 y 20.201 sobrerrepresenta la muestra. Aquellos establecimientos industriales que realizan actividades de I+D y pertenezcan al estrato de menores de 20 empleados se incluyen automáticamente en la muestra. En esta segunda subfase, y con el objetivo de

disminuir la tasa de no respuesta, la selección aleatoria que se realizó el año 1993 se ha ido sustituyendo por un panel de establecimientos que contestaron el año anterior a la encuesta más una selección aleatoria hasta completar la afijación óptima.

1.4 Postestratificación

Las unidades estadísticas, una vez recogida la información, se postestratifican y se recalculan las poblaciones, ya que el diseño muestral se ha realizado con el Directorio del año $t-1$.

Las poblaciones se calculan:

1. Para los estratos de empleo 20-49, 50-99, 100-499 y ≥ 500 tabulando el Directorio de Actividades Económicas del año t .
2. Para el estrato < 20 empleados, se realiza una proyección de la población t para cada estrato, multiplicando a la población del año $t - 1$ la variación del empleo de los establecimientos muestrales que permanecen en el estrato y actualizados entre el año t y el $t-1$.

2. Sistema de estimación en la Encuesta Industrial

2.1 Introducción

Como se ha dicho anteriormente, todas las unidades estadísticas con más de 19 empleados son autoponderadas, por lo que principalmente el interés radica en las estimaciones dentro del estrato de empleo de 1-19 empleados.

Las encuestas económicas de EUSTAT utilizan distintos tipos de estimadores a la hora de extrapolar la información muestral a la población. Por un lado, los estimadores directos basados en el diseño muestral (estimador de Horvitz-Thompson, estimador de la Razón, ...) y por otro, estimadores asistidos por modelos que utilizan información auxiliar de otros dominios para estimar en dominios donde la muestra es escasa. Estos últimos, tienen la ventaja de disminuir el error muestral al estimar en áreas pequeñas, pero a la vez pueden introducir un sesgo importante si la información auxiliar en los diferentes dominios (o estratos) no es homogénea. Por lo tanto, una solución óptima es la utilización de estimadores que compensen por un lado, la inestabilidad de los estimadores directos y por otro, el sesgo de los indirectos. Ver [1] y [2].

2.2 Estimadores compuestos y su varianza

El tipo de estimadores referidos en la última parte del apartado anterior es el utilizado por la Encuesta Industrial. Se denominan estimadores compuestos y tienen la siguiente expresión genérica:

$$\hat{\theta}_{COMPUESTO} = \phi \hat{\theta}_{DIRECTO} + (1 - \phi) \hat{\theta}_{INDIRECTO} \quad \text{con } 0 \leq \phi \leq 1 \quad (2)$$

La expresión del Error Cuadrático Medio para este tipo de estimadores no es sencilla y se propone una aproximación de ésta que tiene la siguiente forma:

$$ECM(\hat{\theta}_{COMPUESTO}) = \phi^2 ECM(\hat{\theta}_{DIRECTO}) + (1 - \phi)^2 ECM(\hat{\theta}_{INDIRECTO}) - 2\phi(1 - \phi)[ECM(\hat{\theta}_{DIRECTO}) - \hat{\theta}_{INDIRECTO} * \text{Sesgo}] \quad (3)$$

Tanto la expresión del estimador como la de su error cuadrático medio están implementadas en una macro de SAS programada al efecto. Más detalles sobre el origen y cálculo de las expresiones anteriores se pueden consultar en la referencia de la bibliografía.

Bibliografía

[1] EUSTAT (2005). “*Cálculo de coeficientes de variación para diferentes estimadores directos e indirectos utilizados en las encuestas económicas de Eustat.*”

http://www.eustat.es/document/datos/Errores_c.pdf

[2] EUSTAT (2005). “*Estimación de Áreas Pequeñas en la Encuesta Industrial de la C.A. de Euskadi.*”

http://www.eustat.es/document/datos/ct_14_.pdf