

Encuesta económica del sector hostelero. Cálculo de Errores de Muestreo.

INDICE

1. Introducción.....	3
2. Breve descripción de la encuesta.....	3
2.1 Definición.....	3
2.2 Diseño Muestral.....	4
3. Sistema de estimación y cálculo de errores.....	5
3.1 Introducción.....	5
3.2 Estimador compuesto y su varianza.....	5
3.3 Tablas de estimaciones y coeficientes de variación.....	5
Bibliografía.....	7

1. Introducción

Podemos definir error de muestreo como la imprecisión que se comete al estimar una característica de la población de estudio (parámetro) mediante el valor obtenido a partir de una parte o muestra de esa población (estadístico).

Este error depende de muchos factores, entre ellos, del procedimiento de extracción de la muestra, del número de unidades que se extraen, del método de estimación, de la naturaleza de la característica a estimar, etc. Una expresión generalizada del error de muestreo sería la siguiente:

$$\text{Error de muestreo} = \sqrt{\text{Var}(\hat{\theta})} \quad (1)$$

Siendo $\hat{\theta}$ el estadístico de interés (media, total, proporción,...). Este estadístico tomará valores distintos dependiendo de la muestra extraída. La variabilidad del estadístico en el muestreo determinará el error muestral.

La expresión de este error cambiará dependiendo de la técnica de muestreo utilizada, haciéndose más complejo su cálculo conforme más complicado sea el diseño muestral. La mayoría de las encuestas de EUSTAT tienen un diseño muestral complejo que incluye estratificación, probabilidades de selección desiguales, etc. Estos diseños se aplican con el fin de producir estimadores puntuales lo mejores posibles, pero en la práctica complican sobremanera la estimación de los errores de muestreo.

La literatura ha sugerido algunas alternativas a los métodos convencionales de cálculo de errores muestrales. De entre éstas, las técnicas de replicación [1] y linealización [5], [6] proporcionan de una forma rápida y sencilla, estimaciones de la varianza para cualquier tipo de estadístico (medias, totales, proporciones,...).

No obstante, y para determinados supuestos, será necesario calcular el Error Cuadrático Medio (ECM) que tiene en cuenta no sólo la varianza muestral del estadístico sino posibles sesgos en las estimaciones debidos a factores ajenos a la muestra (p.ej: uso de información auxiliar). Este es el caso de algunas encuestas económicas de EUSTAT, que utilizan la siguiente expresión para estimar el error total cometido al inferir los datos poblacionales [2]:

$$\text{ECM}(\hat{\theta}) = \text{Var}(\hat{\theta}) + \text{Sesgo}^2 \quad (2)$$

La Encuesta Económica del Sector Hostelero utiliza esta última expresión para estimar el error de muestreo. En lo que sigue, introduciremos el sistema de estimación y cálculo de errores para el caso concreto de estas encuestas.

2. Breve descripción de la encuesta de servicios

2.1 Definición

La Encuesta de Hostelería y otros Servicios es la agrupación de tres encuestas diferenciadas:

- Encuesta Económica del Sector Hostelero
- Encuesta Económica de otros Servicios
- Encuesta de Servicios Profesionales y Empresariales

En concreto, y para el cálculo de los errores muestrales, nos vamos a centrar en la primera de ellas, que se describe a continuación:

Encuesta Económica del Sector Hostelero

Ámbito Poblacional: La población objeto de estudio esta constituida por los establecimientos cuya actividad exclusiva o principal, según CNAE93, sea la división 55 (Hostelería).

Ámbito Geográfico: La encuesta se extiende a aquellas unidades estadísticas ubicadas en el ámbito geográfico de la Comunidad Autónoma de Euskadi, aún cuando su sede social o gerencia se encuentre fuera de la Comunidad.

Ámbito Temporal: El periodo de referencia de la encuesta es el año anterior al de la recogida de información.

2.2 Diseño Muestral

En el año 2004, el diseño muestral se hizo de forma conjunta para la Encuesta Económica del Sector Hostelero, la Encuesta Económica de Otros Servicios y la Encuesta de Servicios Profesionales y Empresariales.

Los establecimientos con más de 10 empleados son censales. El resto de la muestra se extrae de los estratos de empleo pequeños (0-2 y 3-9). El reparto de este tamaño muestral (en los estratos de empleo pequeños) se realiza de la siguiente forma:

- Reparto del total de la Comunidad Autónoma de Euskadi (CAE) por Territorio Histórico (TH): Afijación proporcional a la raíz del número de establecimientos por TH.
- Reparto de los totales de cada TH por estratos de empleo: por afijación proporcional a la raíz del número de establecimientos.
- Reparto de los totales de TH y estrato de empleo por la clasificación de actividad A84 [4]: Afijación óptima de Neyman: Se modifican los resultados para obtener muestra para todas las agrupaciones A84. Se reparte la muestra dentro de cada agrupación, de forma que de la agrupación/es con una muestra asignada mayor se le/s quita un establecimiento a muestrear que se le asigna a la/s agrupación/es que no le/s ha correspondido muestra.
- Reparto del total de TH, estrato de empleo y A84 por CNAE a 5 dígitos: Afijación proporcional al empleo.

3. Sistema de estimación y cálculo de errores

3.1 Introducción

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 [2] y [3].

3.2 Estimadores compuestos y su varianza

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

$$\hat{\theta}_{\text{COMPUESTO}} = \phi \hat{\theta}_{\text{DIRECTO}} + (1 - \phi) \hat{\theta}_{\text{INDIRECTO}} \quad \text{con} \quad 0 \leq \phi \leq 1 \quad (3)$$

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:

$$ECM(\hat{\theta}_{\text{COMPUESTO}}) = \phi^2 ECM(\hat{\theta}_{\text{DIRECTO}}) + (1 - \phi)^2 ECM(\hat{\theta}_{\text{INDIRECTO}}) - 2\phi(1 - \phi)[ECM(\hat{\theta}_{\text{DIRECTO}}) - \hat{\theta}_{\text{INDIRECTO}} * \text{Sesgo}] \quad (4)$$

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 [2] de la bibliografía.

3.3 Tablas de estimaciones y coeficientes de variación.

La información más relevante proporcionada por la Encuesta Económica del Sector Hostelero, hace referencia a las principales macromagnitudes económicas de los sectores de actividad que abarca y a la cuenta de Pérdidas y Ganancias de dicho sector. Por lo tanto, las tablas de estimaciones y errores a publicar serán las siguientes:

Hostelería

- Coeficientes de Variación para macromagnitudes del sector de hostelero, por Territorio Histórico.
- Coeficientes de Variación para la cuenta de pérdidas y ganancias del sector hostelero, por Territorio Histórico.

El **Coefficiente de Variación** es una medida relativa del error que permite comparar precisiones entre distintos grupos o poblaciones. Se trata de una magnitud adimensional cuya expresión es:

$$CV(\hat{\theta}) = \frac{\sqrt{ECM(\hat{\theta})}}{\hat{\theta}} \quad (5)$$

Otra forma de interpretar esta información consiste en calcular el **error relativo al 95% de confianza**, que se obtiene al multiplicar el percentil 1,96¹ por el Coeficiente de Variación. Este error relativo nos permite hablar en términos de puntos porcentuales del valor de la estimación.

Es decir, a un nivel de confianza del 95% se puede afirmar que el verdadero valor de la magnitud económica en la población se encuentra en el intervalo:

$$(\hat{\theta} \pm \text{error relativo} * \hat{\theta}) = (\hat{\theta} \pm 1,96 * \hat{\theta})$$

¹ Se trata del percentil de la distribución Normal(0,1) que corresponde a un 95% de probabilidad.

Es importante señalar aquellas estimaciones que sobrepasen un determinado porcentaje del error relativo al 95%, para que el usuario tome las debidas precauciones a la hora de interpretar la información dada. Un umbral razonable estaría en aquellas estimaciones que sobrepasen el 20% de error relativo (C.V. > 10% aprox.), señalando de forma especial aquellas casillas donde este error sea mayor que el 30% (C.V. > 15% aprox.).

Bibliografía

- [1] EUSTAT (1998). "El método de replicación para la estimación de errores de muestreo". D. Morganstein, "Seminario Internacional de Estadística, 37". http://www.eustat.es/prodserv/vol37_c.html
- [2] 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
- [3] EUSTAT (2005). "Estimación de Áreas Pequeñas en la Encuesta Industrial de la C.A. de Euskadi." http://www.eustat.es/document/datos/ct_14_c.pdf
- [4] EUSTAT (2007). Clasificaciones Sectoriales. http://www.eustat.es/document/datos/codigos/clasificacion_sectorial.xls
- [5] Fuller, W. A. (1975), "Regression Analysis for Sample Survey," Sankhy \bar{A} , 37, Series C, Pt. 3, 117 - 132.
- [6] Woodruff, R. S. (1971), "A Simple Method for Approximating the Variance of a Complicated Estimate" Journal of the American Statistical Association, 66, 411 -414.