

Encuesta de Presupuestos de Tiempo 2003. Cálculo de Errores de Muestreo.

INDICE

1. Introducción.....	3
2. Métodos de Replicación.....	3
3. Cálculo de errores EPT 2003.....	4
3.1 Diseño Muestral.....	4
3.2 Cálculo de las replicadas.....	5
3.3 Cálculo de tablas y estadísticos anuales. EPT- 2003.....	5
3.4 Resultados e Interpretación.....	8
Bibliografía.....	9

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 esa parte de la población (diseño muestral), del número de unidades que se extraen (tamaño de la muestra), 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 más buenos 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. Una de estas soluciones consiste en aplicar técnicas de replicación que proporcionan de una forma rápida y sencilla, estimaciones de la varianza para cualquier tipo de estadístico (medias, totales, proporciones,...).

En lo que sigue introduciremos estos métodos y su aplicación concreta en el caso de la Encuesta de Presupuestos de Tiempo (en adelante EPT) para los datos referentes a 2003.

2. Métodos de Replicación

El cálculo de errores de muestreo mediante replicación [2] es un método útil para aproximar la varianza de estadísticos no-lineales. Además, produce resultados asintóticamente equivalentes a la linealización.

El método consiste en elegir submuestras (replicadas) de la muestra total y calcular el estadístico de interés en cada una de las submuestras. La varianza del estadístico de la muestra total se estima usando la variabilidad de las estimaciones calculadas para cada submuestra.

Existen distintos métodos de replicación : Jackknife (JK1, JK2, JK_n), Replicación de Equilibrio Repetido (BRR), métodos Bootstrap. La elección del método más adecuado depende del diseño muestral.

Todos ellos basan el cálculo del error muestral en la siguiente expresión:

$$\text{Var}(\hat{\theta}) = c \sum_{g=1}^G h_g f_g (\hat{\theta}_{(g)} - \hat{\theta})^2 \quad (2)$$

con :

$\hat{\theta}_{(g)}$ = Estimación de θ para la replicada g

G = Número total de replicadas formadas

c = Constante que depende del método de replicación

h_g = Factores de ajuste (JKn)

f_g = Factor de corrección de poblaciones finitas

El método que más se adapta a los diseños muestrales más comunes en EUSTAT es el Jackknife (n) que se caracteriza por:

- Ajustarse a diseños muestrales estratificados con dos o más unidades en cada estrato.
- Cada replicada se forma excluyendo una unidad de la muestra original cada vez.
- $G = n$ replicadas (n =número de unidades primarias en la muestra original).
- $c = 1$
- $h_g = (n_h - 1)/n_h$ siendo h el estrato asociado a la replicada g .

El software WesVar de la empresa americana Westat, Inc. [4] implementa, entre otros, este método de replicación y será la herramienta que se utilizará para el cálculo de las replicadas y sus correspondientes estimadas.

3. Cálculo de errores. EPT 2003

3.1 Diseño Muestral [3]

La EPT es una operación quinquenal que tiene por objeto informar sobre la distribución y uso del tiempo en las distintas actividades de ámbito económico, social y cultural de la población de la C.A.E.

El tamaño de la muestra es de 5016 familias, para el cuestionario familiar, y luego una persona de 16 o más años por familia, para el diario. Para distribuir la muestra se procede de la siguiente manera:

- Se reparte la muestra por Territorio Histórico de forma proporcional a la raíz cuadrada del número de viviendas ocupadas, con el fin de equilibrar una distribución que de otra forma acusaría una excesiva concentración de la muestra en Bizkaia.
- En cada territorio se distribuye por zonas de forma proporcional al número de viviendas ocupadas. Una vez fijado el tamaño de cada zona se divide en dos mitades, una para cada toma (otoño y primavera).
- Para elegir la muestra en cada zona se utiliza un muestreo probabilístico en dos etapas: en la primera se eligen las secciones, por muestreo proporcional a su tamaño, y en la segunda se seleccionan las viviendas por muestreo sistemático. Previamente, con el fin de multiplicar las ventajas del muestreo probabilístico, se agrupan las secciones en tipologías o grupos homogéneos.

3.2 Cálculo de las replicadas

La información de la EPT se encuentra almacenada en un fichero de actividades y tiempos acumulados para cada actividad que requiere un procesamiento previo para convertirlo en un fichero de individuos que pueda ser leído por el software de cálculo de errores Wesvar. También será necesario generar las variables que delimitan el estrato y las unidades primarias de muestreo (PSU's) necesarias para el cálculo de los errores.

El cálculo de las replicadas debe respetar el diseño muestral original. Es decir, cada submuestra debe ser una "réplica" en lo que a diseño se refiere, de la muestra original. En el caso de la EPT las replicadas deberían mantener la estratificación por zona y tipología de secciones, por lo que se creará la variable estrato cruzando estas dos variables.

Las unidades primarias de muestreo son las secciones de población, identificando a cada una dentro del estrato mediante un número correlativo. En aquellos estratos donde hay una única sección, no es posible aplicar el método de replicación Jk(n) por lo que habrá que definir a las unidades pertenecientes a dicha sección como unidad primaria de muestreo. Se asignará un número correlativo a cada individuo dentro de la sección correspondiente.

El cálculo de las replicadas se realiza automáticamente con el software WesVar, previa especificación de la variable que conforma el estrato, en nuestro caso tipología por zona, y la que define las unidades primarias de muestreo.

3.3 Cálculo de tablas y estadísticos anuales. EPT- 2003.

Los estadísticos para los cuales se va a estimar el error, serán función de los tiempos dedicados a cada actividad concreta. Existen 8 grandes grupos de actividades: Necesidades Fisiológicas, Trabajo y Formación, Trabajos Domésticos, Cuidados a Personas del Hogar, Vida Social, Ocio Activo y Deportes, Ocio Pasivo y Trayectos. A su vez, estos grandes grupos contienen sub-actividades para las cuales también se calcularán los estadísticos y errores correspondientes.

Existen 3 tipos de estadísticos de interés:

1) X_i : Tiempo Medio Social: Tiempo medio diario que la sociedad en su conjunto dedica a una actividad.

$$X_i = \frac{\sum t_i}{P_{\geq 16}}$$

2) Y_i : Tiempo Medio Participante: Tiempo medio diario que dedican a una actividad aquellos que la realizan.

$$Y_i = \frac{\sum t_i}{PP_{\geq 16}}$$

3) T_i : Tasa Participación: Tanto por ciento de personas que realizan diariamente una actividad.

$$T_i = \frac{X_i}{Y_i} \times 100 = \frac{PP_{\geq 16}}{P_{\geq 16}}$$

La **primera medida** era simplemente una media ponderada de los tiempos de cada actividad, y se utilizó para ello la función MEAN() de WesVar.

$$X_i = MEAN(act) = \frac{\sum w_i t_i}{\sum w_i}$$

donde w_i son los pesos muestrales, y t_i los tiempos dedicados a cierta actividad act .

En la **segunda medida** sólo intervienen los individuos que realizan una determinada actividad por lo que es necesario crear previamente unos indicadores de actividad que tomarán valor 1 si el individuo participa de la actividad y 0 en otro caso. Por ejemplo, para una actividad genérica (act) se calcula:

```
If tiempo_act=0 then i_act=0;
else i_act=1;
```

Entonces, la segunda medida se calcula en WesVar de la siguiente manera:

$$Y_i = MEAN(act) \times \frac{elev_tot}{elev_act} = \frac{\sum w_i t_i}{\sum w_i} \times \frac{\sum w_i}{\sum i_act \cdot w_i} = \frac{\sum w_i t_i}{\sum i_act \cdot w_i}$$

donde $elev_tot$ es el elevador de la población total, $elev_act$ es el elevador de la actividad (act), e i_act es el indicador de realización de esa actividad. Así, la medida queda como una media ponderada, pero ahora solo entran en juego los individuos que realizan la actividad.

La **tercera medida**, la tasa, se calcula simplemente dividiendo la primera medida por la segunda y multiplicando por 100.

Una vez calculadas las replicadas o submuestras, es necesario estimar el valor del estadístico de interés en cada una de ellas para los cruces y subgrupos que se estimen necesarios para, posteriormente, calcular su variabilidad.

En caso de la EPT los estadísticos y cruces para los cuales interesa calcular el error de muestreo son:

Estadístico\Variable de cruce	Tipo de actividad	Territorio Histórico
Tiempo Medio Social (hh:mm)	X	X
Tiempo Medio por Participante (hh:mm)	X	X
Tasa de Participación (%)	X	X

Nota: En todos los casos este cruce es simultáneo.

El tipo de actividad se corresponde con la siguiente clasificación:

1 NECESIDADES FISIOLÓGICAS

- Sueño.
- Cuidados personales
- Cuidados médicos
- Comidas a domicilio
- Comidas fuera
- Recepciones

2 TRABAJO PROFESIONAL Y TIEMPO DE FORMACIÓN

- Trabajo profesional principal
 - Trabajo profesional secundario
 - Pausas en el trabajo
 - Formación no reglada
 - Formación reglada
 - Otras actividades educativas de formación
- 3 TRABAJOS DOMÉSTICOS
- Preparar comidas
 - Limpieza
 - Ropa de casa
 - Otros arreglos
 - Compras de bienes y servicios
 - Gestiones administrativas
 - Semi-ocios
- 4 CUIDADOS A LAS PERSONAS DEL HOGAR
- Cuidados materiales y médicos dedicados a los niños
 - Juegos e instrucción
 - Cuidados materiales y médicos para los adultos
- 5 VIDA SOCIAL
- Recepciones y salidas
 - Conversaciones
 - Participación religiosa
 - Participación civil, desinteresada y ayudas
 - Trabajo desinteresado al servicio de una organización. Trabajo para la organización (no para terceros, ni para personas)
 - Trabajo a través de una organización (trabajo con la gente)
 - Ayudas informales a otros hogares (Trabajo al servicio de terceras personas)
- 6 OCIO ACTIVO: DEPORTES, ACTIVIDADES AL AIRE LIBRE Y LIGADAS AL PC E INTERNET
- Práctica de un deporte
 - Excursiones, paseos
 - Uso de ordenador personal , videoconsolas y otros equipamientos informáticos
 - Uso de Internet
- 7 OCIO PASIVO Y AFICIONES
- Medios de difusión e información
 - Sin actividad
 - Asistencia a espectáculos
 - Juegos
 - Aficiones artísticas y otras
- 8 TRAYECTOS
- Trayectos al trabajo
 - Trayectos al lugar de estudios
 - Trayectos derivados del cuidado del hogar
 - Trayectos para llevar o acompañar a alguien.
 - Trayectos derivados del servicio a organizaciones
 - Trayectos derivados de la ayuda a otros hogares
 - Trayectos derivados de la vida social
 - Trayectos derivados del ocio, cultura y deportes
 - Otros trayectos
 - Esperas y colas en los trayectos

También se calculan errores a un nivel más desagregado (actividades a tres dígitos) y por día de la semana, con el fin de controlar la calidad de las estimaciones utilizadas para explotaciones específicas y peticiones externas.

3.4 Resultados e Interpretación.

Aparte de la estimación del error de muestreo (2), WesVar proporciona otras medidas del error que son de utilidad y ayudan a la interpretación del mismo. Estas medidas se detallan en el documento explicativo del cálculo de errores de la PRA, siendo la interpretación de las mismas igualmente válidas para el caso de la EPT. Ver detalles en [1].

Bibliografía

[1] "Cálculo de errores muestrales. PRA 2004". EUSTAT.

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

[2] "El método de replicación para la estimación de errores de muestreo". D. Morganstein, "Seminario Internacional de Estadística, 37".EUSTAT. 1998. http://www.eustat.es/prodserv/vol37_c.html

[3] "Encuesta de Presupuestos de Tiempo . Ficha metodológica." EUSTAT.

http://www.eustat.es/document/presupuestos_c.html

[4] "Manual del usuario WesVar 4.2" Copyright 2002. WESTAT, Inc.