

MANUAL DE BUENAS PRÁCTICAS DE DISEÑO Y EXTRACCIÓN DE MUESTRAS



**EUSKAL ESTATISTIKA ERAKUNDEA
INSTITUTO VASCO DE ESTADÍSTICA**

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

RESUMEN

Este documento recoge los principios básicos del diseño y extracción de muestras y presenta los métodos más habitualmente usados en Eustat. Este manual de buenas prácticas pretende ser una herramienta para la gestión de la calidad que estaba previsto desarrollar dentro del Plan de Calidad de Eustat.

Se presentan en este documento los esquemas de muestreo (marcos utilizados, tipos de diseño,...) utilizados en las encuestas sociodemográficas y económicas, respectivamente. También se hace referencia al software utilizado en la extracción de las muestras y se especifican los procedimientos usados.

Por último, se detalla el diseño muestral de dos operaciones clave en el Instituto, una de tipo sociodemográfico y otra de tipo económico, la Encuesta de Población en Relación con la Actividad (PRA) y la Estadística Industrial, respectivamente.

Vitoria-Gasteiz, marzo de 2008

Indice

MANUAL DE BUENAS PRÁCTICAS DE DISEÑO Y EXTRACCIÓN DE MUESTRAS.....	1
RESUMEN.....	2
INDICE	3
1. INTRODUCCIÓN.....	4
2. TIPOS DE DISEÑO UTILIZADOS.....	6
3. MUESTRAS SOCIODEMOGRÁFICAS.....	8
4. MUESTRAS ECONÓMICAS.....	10
5. SOFTWARE PARA EXTRACCIÓN DE MUESTRAS	12
5.1 APLICACIÓN DE LOS PROCEDIMIENTOS DE SAS A LA OBTENCIÓN DE MUESTRAS DE EUSTAT	13
6. EJEMPLOS DE DISEÑOS MUESTRALES DE OPERACIONES REALIZADAS POR EUSTAT.....	15
6.1 DISEÑO MUESTRAL DE LA ENCUESTA DE POBLACIÓN EN RELACIÓN CON LA ACTIVIDAD (PRA).....	15
6.2 DISEÑO MUESTRAL DE LA ESTADÍSTICA INDUSTRIAL.....	17
7. REFERENCIAS	19

Introducción

Las encuestas basadas en muestras, en contraposición a los censos o a la utilización de fuentes administrativas en que se recopila información de todos los elementos de la población objeto de estudio, se fundamentan en la selección de una parte de la población a la que llamamos muestra y sobre la que se realiza el trabajo de recogida. A partir de la información recopilada de las unidades de la muestra se realizan inferencias relativas a toda la población.

Las muestras pueden ser o no probabilísticas, aunque habitualmente las oficinas de estadística trabajan con las primeras.

Si se siguen técnicas probabilísticas de selección por las que todas las unidades de la población tienen una probabilidad de selección positiva y conocida, se tiene la muestra representativa de la población de la que se extrae.

El método de muestreo utilizado y el tamaño de la muestra establecido dependen de múltiples factores como: el nivel de precisión exigido a las estimaciones por grupos temáticos o espacialmente, los marcos de población disponibles, la disponibilidad de información auxiliar para la estratificación de la población y selección de la muestra, el método de estimación a utilizar, los procedimientos posibles de acceso a la muestra, el tiempo de recogida, y el presupuesto disponible, entre otros.

La información auxiliar disponible puede ser utilizada también en la fase de estimación, con el fin de mejorar la precisión de las estimaciones.

El hecho de que las variables de estudio se midan únicamente sobre la muestra, y no sobre la población, produce en las estimaciones un error debido al muestreo. Si se repitiera varias veces una misma muestra, las estimaciones de muestra a muestra variarían en función de ese error de muestreo, el cual es debido al diseño muestral escogido y al procedimiento de estimación utilizado.

El error de muestreo varía de modo inverso al tamaño de la muestra, cuanto mayor es la muestra menor es el error de muestreo con el consiguiente mayor coste de recogida de información y con la posibilidad de que sean mayores otros errores ajenos al muestreo.

En todo este proceso es importante la utilización de toda la información auxiliar disponible para obtener un diseño eficiente, con la mayor precisión y el menor coste posible. En el concepto de coste también hay que incluir la carga que soportan los informantes por responder a la encuesta.

Así, el nivel de precisión final de las estimaciones calculadas depende de ese error de muestreo y de otros errores ajenos al muestreo.

Entre los errores ajenos al muestreo están los producidos por la utilización de un marco no suficientemente actualizado. Esto es, que la población objetivo difiera de forma considerable del marco de población utilizado. Pueden ocurrir errores de sobrecobertura y de infracobertura del marco.

Los errores de sobrecobertura se producen cuando el marco contiene elementos que no pertenecen ya a la población objetivo.

Los errores de infracobertura ocurren cuando el marco no contiene todos los elementos de la población objetivo, lo que suele ser más problemático ya que es más difícil de detectar.

Cuando el tamaño de muestra disponible no es suficiente para la estimación de ciertos dominios, se ha de recurrir a técnicas especiales de estimación en áreas pequeñas.

En una encuesta por muestreo normalmente se obtiene información de muchas variables al mismo tiempo. El tamaño y diseño de la muestra dependen de la distribución de las variables de estudio en la población. Por ello, a la hora de definir el diseño habrá que fijarse en algunas pocas variables consideradas principales.

En el proceso de estimación hay que tener en cuenta que la muestra recogida puede no corresponder con la muestra diseñada debido a: problemas de actualización del marco, no-respuesta, errores de clasificación de las unidades en los estratos,..., por lo que todo ello habrá que considerarlo e intentar corregirlo en la estimación.

En encuestas continuas ó de panel habrá aspectos especiales a tener en cuenta a la hora de diseñar la muestra. En función del fenómeno a investigar y la demanda de información se establecerán los períodos de reencuestación.

En el ámbito estadístico oficial se suelen utilizar con mayor frecuencia los diseños muestrales basados en paneles rotantes, donde una parte del panel es fija de un período al siguiente y una parte se renueva en cada período. El fenómeno de la llamada 'fatiga de encuesta', que cada vez parece pesar más entre las causas de rechazo, apoya también el desarrollo de este tipo de diseños.

De forma secuencial se establece un orden de entrada y de salida a la muestra, una vez fijado el período de permanencia o su inverso, la tasa de renovación.

Para determinar la tasa de renovación de período a período hay que llegar a una solución de compromiso entre las ventajas e inconvenientes producidos:

- ❖ Por un lado, una tasa de renovación más pequeña (permanencia más prolongada en la encuestación) produce una mayor estabilidad en la estimación de los cambios.
- ❖ Por otro lado, una tasa de renovación más alta (permanencia más reducida en la encuestación) conduce a una mejor estimación de los niveles, ya que permite una mejor corrección de los errores de marco y es favorecida además, por una menor fatiga de encuesta sobre los informantes.

Tipos de diseño utilizados

El muestreo más sencillo utilizado es el **muestreo aleatorio simple** en el que todas las unidades tienen la misma probabilidad de ser seleccionadas.

En muchos casos se dispone de variables auxiliares que permiten la estratificación de las unidades de la población. Estas variables pueden estar relacionadas con las variables de estudio ó pueden determinar los dominios de estudio (ó de difusión). Se utiliza entonces el **muestreo estratificado** en el que la selección se hace en cada estrato de forma independiente.

El tamaño global de la muestra se reparte por estrato mediante distintos medios de asignación: igual, proporcional al tamaño del estrato, afijación de varianza mínima (que requiere tener mejor conocimiento de las características poblacionales de cada estrato, ya que requiere saber no sólo su tamaño sino también conocer ó estimar su varianza).

En distribuciones asimétricas como las de las variables económicas, una proporción pequeña de unidades grandes de la población proporciona gran parte de la estimación final y, en ese caso es conveniente incluir un estrato censal de unidades grandes que entran en la muestra con probabilidad 1. A estas unidades se les denomina **unidades autorrepresentadas**. Es uso de muestreo habitual en las encuestas económicas determinar como estratos censales los estratos correspondientes a tamaño de empleo mayor que una cierta cantidad.

Otro tipo de muestreo utilizado es el **muestreo por conglomerados** en el que se seleccionan conglomerados ó unidades agregadas de las unidades últimas en lugar de seleccionarse directamente las unidades últimas.

Esto en la práctica puede estar debido a que no se dispone de una lista de todas las unidades últimas de la población y/o a que se quiere obtener una muestra concentrada en algunas de las unidades agregadas para reducir el coste de recogida.

En ese caso sólo se requiere la lista de las unidades agregadas ó unidades primarias de muestreo (junto con alguna información auxiliar de las mismas según el tipo de extracción, por ejemplo, un indicador del tamaño de las unidades para una selección proporcional al tamaño) y las listas de las unidades últimas correspondientes a las unidades agregadas seleccionadas en la muestra.

Este tipo de muestreo se utiliza cuando se seleccionan secciones censales para encuestar las viviendas familiares en esas secciones y cuando se seleccionan viviendas para encuestar las personas que viven en esas viviendas.

El hecho de que las variables utilizadas para la estratificación estén relacionadas con las variables de estudio conduce a que las unidades dentro de cada estrato sean homogéneas. En general, los estratos se caracterizan por ser homogéneos dentro de

los mismos y heterogéneos entre ellos. Por el contrario, los conglomerados son heterogéneos dentro de ellos y homogéneos entre ellos.

La estratificación conduce a una reducción de la varianza de la estimación global ó un aumento de la precisión, y la utilización de conglomerados por el contrario, tiene un efecto de aumentar esa varianza (efecto conglomerado).

El diseño de la muestra se puede obtener en **una sola etapa** ó ser **multietápico**, en el que en una primera fase se seleccionan los conglomerados y en una segunda etapa se seleccionan algunos elementos de esos conglomerados. Y así sucesivamente en el caso de más etapas. Tendremos entonces unidades de muestra de primera etapa, de segunda etapa y sucesivas.

La selección de las unidades de la muestra puede no ser con igual probabilidad para todas las unidades sino con probabilidad proporcional al tamaño de las unidades a extraer.

Éste puede ser el caso de conglomerados como las secciones censales, cuando éstas se extraen con probabilidad proporcional a su población. También puede ser el caso de establecimientos económicos, cuando se seleccionan con probabilidad dependiente de un indicador de su tamaño.

A la hora de determinar el reparto de la muestra en un muestreo estratificado hay que considerar el problema de las unidades que cambian de estrato. Al hacer la afijación se pueden utilizar tasas esperadas de unidades mal clasificadas en cada estrato. En caso de no disponer esas tasas habrá que intentar corregir ese problema en la etapa de estimación.

Muestras sociodemográficas

Las muestras sociodemográficas suelen ser muestras de viviendas familiares obtenidas en 1, 2 ó 3 etapas. El marco poblacional en el caso de Eustat lo constituye el Directorio de Viviendas. En la mayoría de casos el marco son las viviendas familiares ocupadas y para ello se tiene conectado el Directorio de Viviendas con el Registro Estadístico de Población (REP) a través del identificador de la vivienda (IDEV).

El tamaño global de la muestra suele estar determinado por las recomendaciones de órganos estadísticos internacionales, por el nivel de desagregación en que se quieran tener representados con precisión colectivos sensibles de población o territoriales, por las referencias disponibles de encuestas similares propias o ajenas, por la disponibilidad de un diseño muestral estándar, por la información obtenida de los pretest que se hayan podido realizar y casi siempre en mayor o menor medida por la capacidad técnica disponible y el presupuesto.

Las muestras obtenidas en una etapa son normalmente estratificadas, al menos por territorio histórico y en algún caso por alguna variable más correlacionada con las variables de interés y disponible en el Directorio de Viviendas

La asignación ó reparto del tamaño global de la muestra por estratos suele ser normalmente proporcional a la población en cada estrato ó a la raíz cuadrada de la población del estrato.

El reparto por territorio es normalmente proporcional a la raíz cuadrada de la población para favorecer a los territorios con menor población y obtener igual nivel de representatividad en todos ellos.

En cada estrato se obtiene una muestra independiente. La selección puede ser aleatoria simple en el estrato ó puede obtenerse de modo sistemático.

La muestra sistemática se obtiene habiendo ordenado previamente el marco por dirección postal de modo que implícitamente se obtiene una muestra estratificada geográficamente.

Si la información recogida en la vivienda es la de los individuos que residen en ella, se trata de una muestra de conglomerados, ya que se seleccionan viviendas para encuestar a personas.

Las muestras de viviendas en dos etapas constan de una primera etapa de selección de secciones censales y una segunda de selección de viviendas. En las viviendas se encuesta al total de los residentes.

La primera etapa consiste normalmente en una selección en los estratos determinados por el territorio histórico y el tipo de la sección.

Este tipo de sección viene de un análisis previo de clasificación de las secciones en función de sus características censales. Dado que por ahora sólo se dispone de fuentes censales quinquenales y que los seccionados varían de forma continua, resulta necesario asignar tipos a las secciones nuevas.

La afijación ó reparto de número de secciones por estrato se hace previamente. En general, el reparto suele ser por territorio proporcional a la raíz de la población en cada territorio y dentro de cada uno proporcional a la población de cada tipo (esto es, la población de todas las secciones que tienen asignado ese tipo). La selección de la muestra de secciones en cada estrato es independiente.

La probabilidad de selección de las secciones es proporcional a la población de las mismas.

La segunda etapa es la selección de las viviendas, normalmente un número fijo por sección y con la misma probabilidad para todas las viviendas de la sección.

La probabilidad final de selección de una vivienda, producto de las probabilidades de primera y segunda etapa, es la misma para todas las viviendas dentro de cada estrato. A esta muestra le llamamos autoponderada porque todas las unidades de la misma tendrán el mismo peso.

Estas muestras de dos etapas corresponden por tanto a un muestreo de conglomerados (secciones censales) y estratificado (las secciones censales se distribuyen en estratos).

El marco de selección de la primera etapa es la lista de las secciones censales de la C.A. de Euskadi, con un indicador del tamaño de las mismas que suele ser la población de la sección censal (ó el número de viviendas en la misma).

El marco de selección de la segunda etapa es la lista de viviendas según el Directorio en las secciones extraídas en la primera etapa.

En algunas encuestas hay una etapa posterior a la selección de viviendas, que es la selección de personas dentro de la vivienda. Se puede hacer previamente, una vez asignadas las personas contenidas en el REP a su vivienda (con la posibilidad de desajustes temporales del marco) o en el propio momento de la encuestación; si la recogida se sustenta en procedimientos CATI, CAPI o CAWI, la selección aleatoria del individuo o individuos a encuestar se puede programar y si se trata de PAPI se suele solventar con la utilización de tablas aleatorias (Tabla de Kish, etc.).

Si se selecciona sólo una persona, la probabilidad de selección de esa persona dentro de la vivienda, en general, es la misma para todas las personas de la vivienda.

La probabilidad final de selección de cada persona es producto de las probabilidades de selección de cada una de las etapas hasta llegar a ellas.

Muestras económicas

En las encuestas económicas de Eustat el marco es el Directorio de Actividades Económicas (DIRAE) que incluye a todos los establecimientos de la C.A. de Euskadi.

El tamaño global de la muestra suele estar determinado por el nivel de precisión que se requiere para las estimaciones y por el presupuesto disponible.

En general el diseño es estratificado, siendo los estratos considerados estratos de empleo y estratos de actividades económicas.

Los estratos de empleo son agrupaciones de la variable empleo, variable disponible en el Directorio, que clasifican las unidades en función de su tamaño.

Los estratos de actividades económicas son agrupaciones de actividades a 5 dígitos correspondientes a la Clasificación Nacional de Actividades Económicas (CNAE).

La estratificación considerada en cada caso dependerá de la encuesta concreta.

El ámbito de aplicación de la encuesta determinará también el marco a aplicar, normalmente determinadas actividades económicas.

Los estratos de empleo mayores suelen ser censales, esto es, entran en la muestra todos los establecimientos cuyo empleo está en uno de esos estratos. Ello es debido a que gran parte de las variables económicas que se quieren estimar provienen de una parte pequeña de la población y esa corresponde a los estratos de mayor empleo.

Los estratos de empleo pequeños, a los cuales pertenece la mayor parte de establecimientos del Directorio, habrán de ser muestrales.

La afijación utilizada en el diseño estratificado, ó el reparto de las unidades de la muestra en los estratos, sigue una serie de criterios generales.

El reparto suele ser inicialmente por territorio histórico de modo proporcional a la raíz cuadrada del número de establecimientos ó del empleo en cada territorio. Ello es debido a que se quiere obtener el mismo grado de precisión en todos los territorios y la asignación en función de la raíz cuadrada favorece a los de menor tamaño.

Dentro de cada territorio, el reparto por estrato de empleo y por sector de actividad depende de cada encuesta concreta.

En algunos casos, el reparto es proporcional al número de establecimientos ó al empleo en cada estrato.

En otros casos también se utiliza una afijación de mínima varianza utilizando la raíz cuadrada del empleo medio del estrato como indicador de la variabilidad del mismo.

En general, el tipo de muestreo en esos estratos muestrales suele ser aleatorio simple, todos los establecimientos tienen la misma probabilidad de ser seleccionados.

Aunque en la práctica en la mayor parte de encuestas se fija de antemano las unidades ya en la muestra y encuestadas en el período anterior (en muchos casos en el año anterior), se asignan a los estratos que corresponden y se completa la muestra según el diseño. Ello responde a la necesidad de mantener un panel de establecimientos y así aumentar la estabilidad de las estimaciones.

Capítulo
5

Software para extracción de muestras

El software utilizado en Eustat para la extracción de muestras es el programa estadístico SAS¹.

Anteriormente se ha determinado el diseño de la muestra, fijado el tamaño y en su caso el reparto por estratos. El estudio de la información disponible para el diseño (marco disponible, información auxiliar,...) y la determinación del tamaño y su reparto se suele hacer en Excel ó en otro paquete estadístico (como el propio SAS, el programa R²,...).

La versión 8 de SAS incorporó varios procedimientos para extraer y analizar datos de muestras. Estos procedimientos están incluidos en el módulo SAS/STAT del que dispone Eustat y el de extracción de muestras es en concreto el procedimiento SURVEYSELECT.

Este procedimiento contempla todos los tipos de muestreo que mencionamos en este documento, como se puede comprobar en su documentación.

Los parámetros necesarios a indicar al procedimiento dependerán del tipo de muestreo que sea.

Si se trata de un muestreo estratificado, habrá que incluir una sentencia STRATA con las variables que determinan los estratos.

Si el muestreo es proporcional al tamaño de las unidades a extraer, se habrá de incluir la sentencia SIZE.

Es conveniente en cualquier caso incluir la sentencia ID que identifique a las unidades, en el fichero de entrada y en el de salida.

La sentencia PROC SURVEYSELECT contendrá los parámetros esenciales en forma de opciones:

- la opción DATA= con el nombre del data-set de entrada que es el marco del que se selecciona la muestra (deberá contener por tanto todas las variables que referenciemos en el resto de sentencias);
- la opción OUT= contendrá el nombre del data-set de salida donde se escribirá la muestra;
- la opción METHOD=, por defecto SRS (simple random sampling) para el muestreo aleatorio simple, SYS(systematic) para el muestreo

¹ Statistical Analysis Software

² Software estadístico de distribución libre

- sistemático, PPS (probability proportional to size) cuando se trate de un muestreo proporcional al tamaño (caso de secciones censales);
- y la opción SAMPSIZE= ó N= para indicar el tamaño ó tamaños (caso de estratos) de la muestra, bien directamente bien en un data-set creado previamente.

5.1 Aplicación de los procedimientos de SAS a la obtención de muestras de Eustat

Aplicando lo anterior al caso de muestras de establecimientos económicos que se extraigan de un directorio de acuerdo a una estratificación previa, la sintaxis del procedimiento se aproxima a:

PROC SURVEYSELECT data=direc sampsize=tamaño out=muestra;

STRATA var1;

ID ident;

run;

con:

'direc' un listado del directorio que únicamente ha de contener una ó varias variables de identificación ó variables que se incluirán en el fichero de la muestra y las variables de estratificación;

'tamaño' un listado de los tamaños de muestra de cada estrato;

'muestra' el nombre del fichero de la muestra obtenida;

'var1' las variables de estratificación (territorio, estrato de empleo,..);

'ident' indica la variable ó variables que se escriben en el fichero de salida de la muestra.

La obtención de una muestra de viviendas en los diseños de 2 etapas supondría dos extracciones, primero una de secciones censales y otra de viviendas.

La extracción de secciones censales llevaría la siguiente sintaxis:

PROC SURVEYSELECT data=direc method=PPS sampsize=tamaño out=muestra;

STRATA var1;

SIZE var2;

ID ident;

run;

con:

'direc' un listado de todas las secciones censales conteniendo
un identificador 'ident' de las mismas,
la variables 'var1' de estratificación (normalmente territorio y tipología) y
la variable 'var2' de tamaño de la sección ó población.

La extracción de viviendas podría ser de un listado de las viviendas de las secciones censales obtenidas en la muestra de secciones y considerando cada sección como un estrato y valdría la sintaxis del muestreo estratificado.

```
PROC SURVEYSELECT data=direc sampsize=tamaño out=muestra;
```

```
STRATA var1;
```

```
ID ident;
```

```
run;
```

La selección de viviendas en un diseño estratificado sería similar a la de los establecimientos en las muestras económicas que utilizan una estratificación previa.

Ejemplos de diseños muestrales de operaciones realizadas por Eustat

A continuación se describen, a modo de ejemplo, los diseños de muestras de dos operaciones del Instituto, una sociodemográfica y otra económica.

6.1 Diseño muestral de la Encuesta de Población en Relación con la Actividad (PRA)

El objetivo principal de la encuesta es producir información estadística continua sobre la participación de la población en las distintas actividades, laborales o no, con especial hincapié en las de carácter económico satisfaciendo las necesidades de los agentes socioeconómicos.

En 2005 tiene lugar un cambio de diseño y del tamaño de muestra, y pasa de ser una muestra de 3.750 a 5.088 viviendas y de ser una muestra bietápica (primero se seleccionan secciones censales y luego viviendas) a una muestra de una etapa (se seleccionan viviendas directamente).

Se define como un panel rotante de viviendas, pasándose de una tasa de renovación de 1/6 de la muestra anterior a 1/8 en la muestra actual.

En adelante se describe el diseño actual.

La muestra de viviendas familiares se estratifica por territorio histórico. Se obtiene de modo independiente una muestra sistemática dentro de cada territorio.

A través de un informante adecuado se recoge la información de todos los individuos de la vivienda. Es por ello una muestra de una sola etapa.

Se trata también de una muestra de conglomerados, ya que las unidades seleccionadas son viviendas y los resultados se refieren a los individuos.

Se ha determinado en base a experiencias previas y recomendaciones internacionales, además de la disponibilidad presupuestaria que el tamaño de la muestra ha de ser aproximadamente 5.000 viviendas cada trimestre, repartido de forma proporcional a la raíz cuadrada del número de viviendas en cada territorio.

En principio, el reparto de 5.000 viviendas era: 1.114 para Álava, 2.196 para Bizkaia y 1.690 para Gipuzkoa. Estas cantidades se tienen que ajustar para que se cumplan los requisitos de los turnos de rotación y la distribución en las 12 semanas del trimestre, por lo que, la muestra final es de 5.088 viviendas, repartidas como sigue: 1.152 para Álava, 2.208 para Bizkaia y 1.728 para Gipuzkoa. En cada turno de rotación hay 636 viviendas. Más abajo, se presenta cómo se ajusta la muestra por TH, semanas y rotación.

Opciones de distribución de la muestra de la PRA

Total	Muestra total				Muestra 1/8			
	Alava	Bizkaia	Gipuzkoa	C.A.E.	Alava	Bizkaia	Gipuzkoa	C.A.E.
1	96	184	144	424	12	23	18	53
2	96	184	144	424	12	23	18	53
3	96	184	144	424	12	23	18	53
4	96	184	144	424	12	23	18	53
5	96	184	144	424	12	23	18	53
6	96	184	144	424	12	23	18	53
7	96	184	144	424	12	23	18	53
8	96	184	144	424	12	23	18	53
9	96	184	144	424	12	23	18	53
10	96	184	144	424	12	23	18	53
11	96	184	144	424	12	23	18	53
12	96	184	144	424	12	23	18	53
Ajustado	1.152	2.208	1.728	5.088	144	276	216	636
Diferencia	38	12	38	88	4,75	1,5	4,75	11

La muestra se distribuye en un trimestre de forma continua, de modo que cada semana del trimestre constituye una muestra independiente y sistemática en cada territorio de tamaño 1/12 del total.

La muestra es un panel compuesto de 8 turnos de rotación. Hay una asignación aleatoria del turno de rotación a cada 1/8 de la muestra. En cada trimestre se renueva 1/8 de la muestra completa.

Cada nuevo turno de rotación es una nueva muestra sistemática de viviendas en cada territorio y cada semana.

La extracción de la muestra sistemática dentro de cada estrato será considerando la lista circular y con arranque aleatorio, ordenada la misma por toda la dirección postal, de modo que dentro de cada estrato la muestra resulta también en cierto modo estratificada geográficamente.

La muestra la componen la información de la vivienda y la de los individuos que residen en ella, según el Registro Estadístico de Población (eliminando las bajas). Los datos individuales son: fecha de nacimiento, nombre, apellidos y DNI, junto con las claves necesarias (UPB).

Una vivienda permanece por tanto en la muestra durante 2 años completos.

La muestra efectiva de la PRA sirve para recoger también, el segundo y el cuarto trimestre del año, la Encuesta sobre la Sociedad de la Información a Familias –ESIF- La encuestación se realiza una semana después de la recogida de la PRA.

Las viviendas de renovación y las sustitutas

Sobre la misma base muestral (eliminando las “titulares”), hay que sortear otras 636 viviendas para la renovación.

El procedimiento es el mismo, solo que se extrae 1/8 en cada TH para cada semana.

Salen del panel todas las viviendas del turno 8 y las que entran de renovación tendrán turno 1. Esto es en cada muestra de cada semana.

Eliminando las dos anteriores (“titulares” la primera vez que se saca y las de renovación) de forma aleatoria se eligen 5.088 sustitutas, de modo estratificado (TH).

Igual que las anteriores, el turno de rotación y la semana que tomen estas viviendas será los de la que van a sustituir.

Parte común de la muestra trimestral

Ahora la parte común de trimestre a trimestre es de 7/8 (teóricamente), es decir, 4.452 viviendas.

Antes con el diseño anterior a 2005 teníamos 3.125 de 3.750.

En la muestra nueva, entre el primer trimestre de 2005 y el de 2006, por ejemplo, es común la mitad de la muestra trimestral (4/8), es decir, 2.544 viviendas.

En los 4 trimestres, la parte común de la muestra trimestral por lo tanto serían 10.176 viviendas.

Para el estudio de la relación con la actividad, por ejemplo, las viviendas serían más de 10.000, sin descontar las incidencias. Aunque hay que tener en cuenta que de ellas son distintas 4.452.

En el sistema antiguo, teníamos 5.000 viviendas comunes en dos años, de las cuales 3.125 eran distintas.

6.2 Diseño muestral de la Estadística Industrial

El Directorio de Actividades Económicas de Eustat es la base para establecer el marco de la encuesta.

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, es decir son autorrepresentadas (se incluyen también entre las autorrepresentadas las altas del Directorio del año t);

-una segunda fase donde se realiza un muestreo aleatorio estratificado, también en dos subfases:

1^a subfase) Las variables de estratificación utilizadas son:

- a) Territorio Histórico (3): Álava, Bizkaia y Gipuzkoa.
- b) Actividad: clasificación normalizada de EUSTAT A-84.

Se calcula la afijación óptima de cada estrato con la restricción presupuestaria correspondiente a cada año.

La afijación óptima se calcula sobrerepresentando Álava con una afijación proporcional a la raíz cuadrada del número de establecimientos de cada Territorio Histórico.

La distribución sectorial se realiza con la siguiente formulación:

$$n_j = n * \left[\frac{T_j}{\sum T_j} \right] \quad \text{siendo :} \quad T_j = N_j * \sqrt{E_j}$$

N_j = Número de establecimientos
 E_j = Empleo medio
 n = tamaño muestral

2^a subfase) La distribución de la afijación óptima en términos de empleo en cada sector de actividad se hace en proporción al peso de cada subclase en el total del sector, realizándose una selección aleatoria.

El Organismo Estadístico Específico del Departamento de Agricultura, Pesca y Alimentación 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 de la Clasificación Nacional de Actividades Económicas (CNAE) sobrerepresenta la muestra.

Con el objetivo de disminuir la tasa de no respuesta, la selección aleatoria que se realizó en el año 2000 se ha ido sustituyendo por un panel de establecimientos que contestaron el año anterior a la encuesta, facilitando así la estabilidad de las estimaciones, más una selección aleatoria hasta completar la afijación óptima.

Referencias

- Buenas prácticas en la elaboración de estadísticas oficiales. INE, mayo 2004. Biblioteca Eustat.
- Statistics Canada Quality Guidelines. Octubre 2003. <http://www.statcan.ca/english/freepub/12-539-XIE/12-539-XIE03001.pdf>
- Proyecto técnico de la PRA.. Eustat
- Proyecto técnico de la Estadística Industrial. Eustat