



INSTITUTO NACIONAL DE ESTADÍSTICA



# Curso Selectivo 2016

## Marco de muestreo en las Encuestas de Hogares

*Departamento de Control de la Producción y  
Muestreo*

*8 de junio de 2016*



# Tipos de muestreo

## Muestreo probabilístico o científico

Basado en la Matemática Estadística (**Cálculo de Probabilidades**)  
Todas las unidades del marco tienen probabilidad conocida y mayor que cero de pertenecer a la muestra

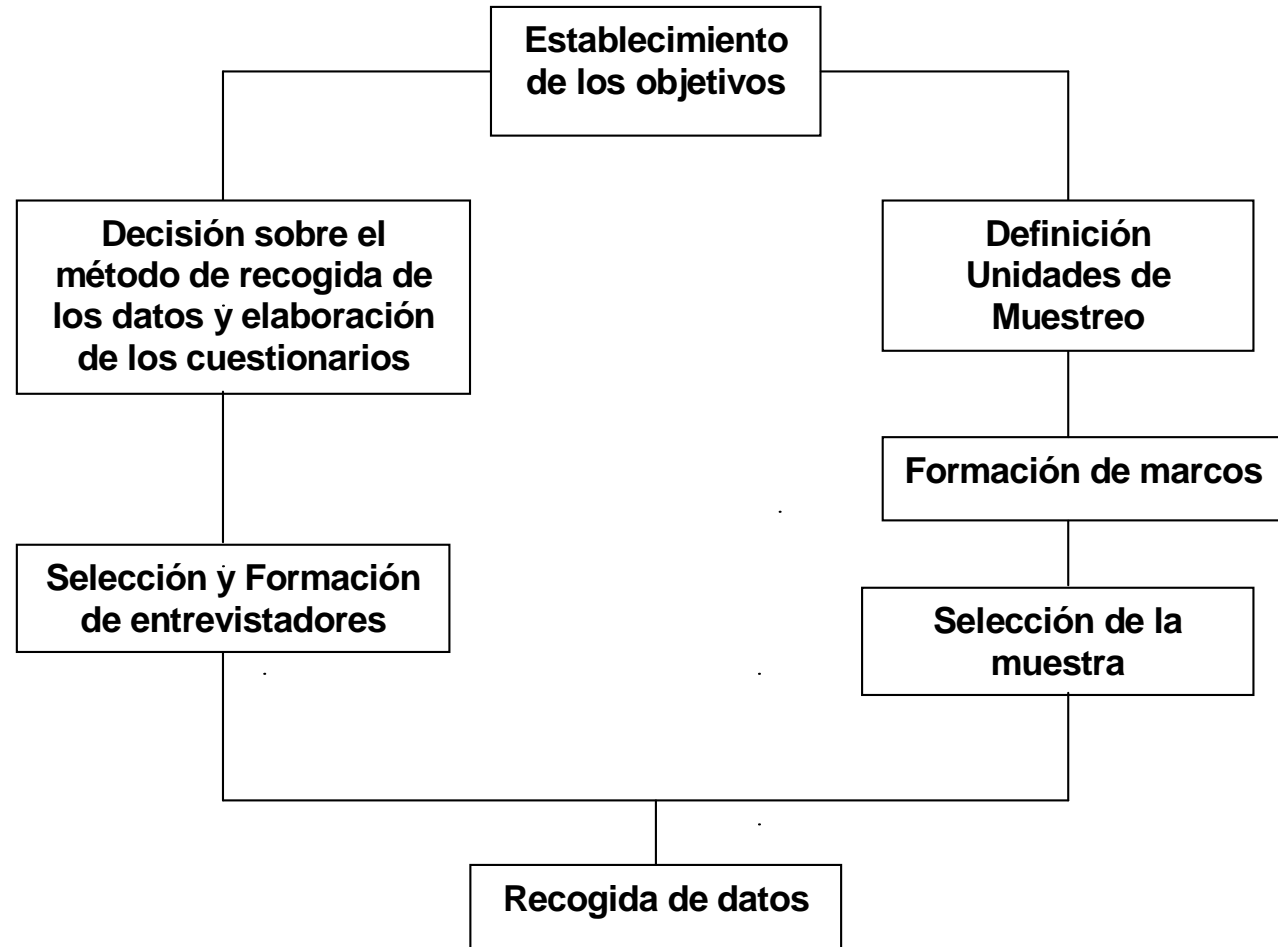
## Muestreo no probabilístico

La selección de la muestra no está sometida a criterios probabilísticos.

- Muestreo opinático (selección de unidades tipo)
- Muestreo por cuotas



# ESQUEMA DE LAS FASES DE UN PROYECTO ESTADÍSTICO



## Marcos de muestreo

Relación de unidades de donde se selecciona la muestra junto con toda la información complementaria que se puede utilizar en el diseño de la muestra.

- Marco de unidades elementales
- Marco de áreas o listas (Muestreo de conglomerados)

Proporciona información fundamental:

- **Probabilidad de selección:** Factores de diseño
- **Información auxiliar:** Tratamiento de la falta de respuesta, **estratificación**, selección controlada, **calibrado**, simulaciones,...
- **Trabajos de campo:** Permite el acceso a la población objetivo. Dirección postal de la unidad seleccionada.



## Características del marco

- Buena cobertura: **Población Objetivo=Población Investigada**
  - Todos los elementos de la población objetivo deben de estar en el marco
  - Cada elemento debe estar presente una sola vez (Duplicados).
  - No debe tener elementos que no pertenecen a la población objetivo
  - Necesidad de depuración y actualización
- Las unidades del marco deben estar bien identificadas y ser localizadas si son seleccionadas para la muestra.
- Debe estar organizado de forma sencilla para facilitar la selección de la muestra.



## Importancia del marco

**Sin un buen marco no se puede diseñar una buena muestra.**

Para la realización de la encuesta es necesario disponer:

- Un marco adecuado a la población objetivo de la encuesta.
- Un marco actualizado ( y actualizable)

Fuentes: Censos, Padrón, Otros Registros administrativos, etc



## Marcos utilizados en las encuestas del INE.

- **Encuestas de Población:**

EPA y ECH: Censo de Población

Resto de encuestas: Padrón Continuo (EPF, TIC-H, ENS)

Futuro: SIGPT

- **Encuestas Económicas:** DIRCE

- **Otras marcos** (Personas sin hogar, ETEFIL, Censo Agrario,...)



## Marco en las encuestas de hogares

**Tipo de Muestreo:** Muestreo bietápico estratificado.

Se utilizan dos marcos:

**Marco de áreas geográficas:** Secciones censales.

- Comunidades Autónomas
- Provincias
- Municipios - Actualmente 8.200 aprox.
- Distritos municipales
- Secciones censales. Aproximadamente 36.000



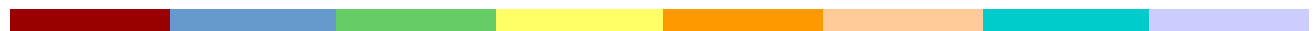


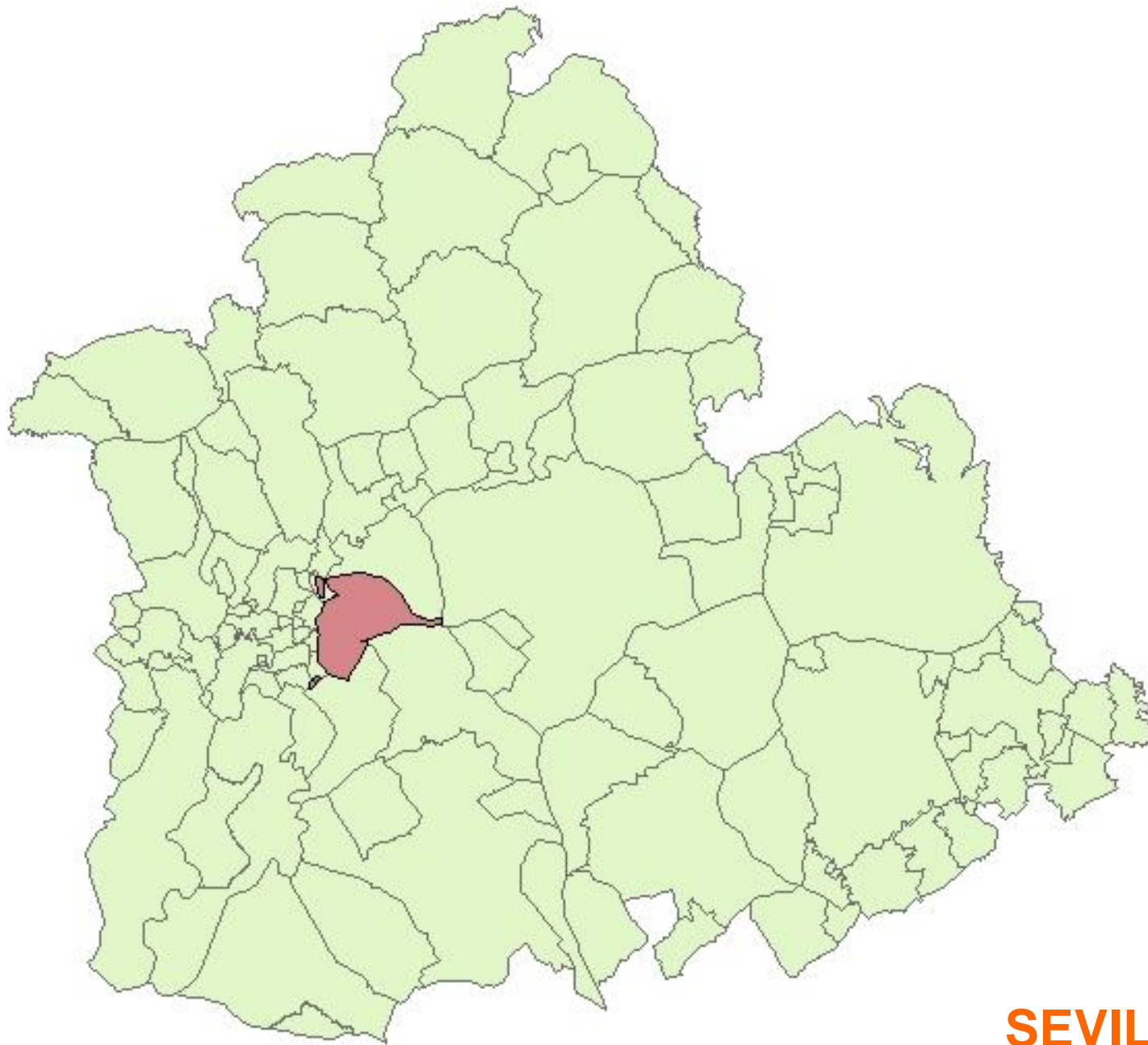
## Sección censal

**Área geográfica en que se divide el territorio nacional, utilizada con fines estadísticos y electorales**

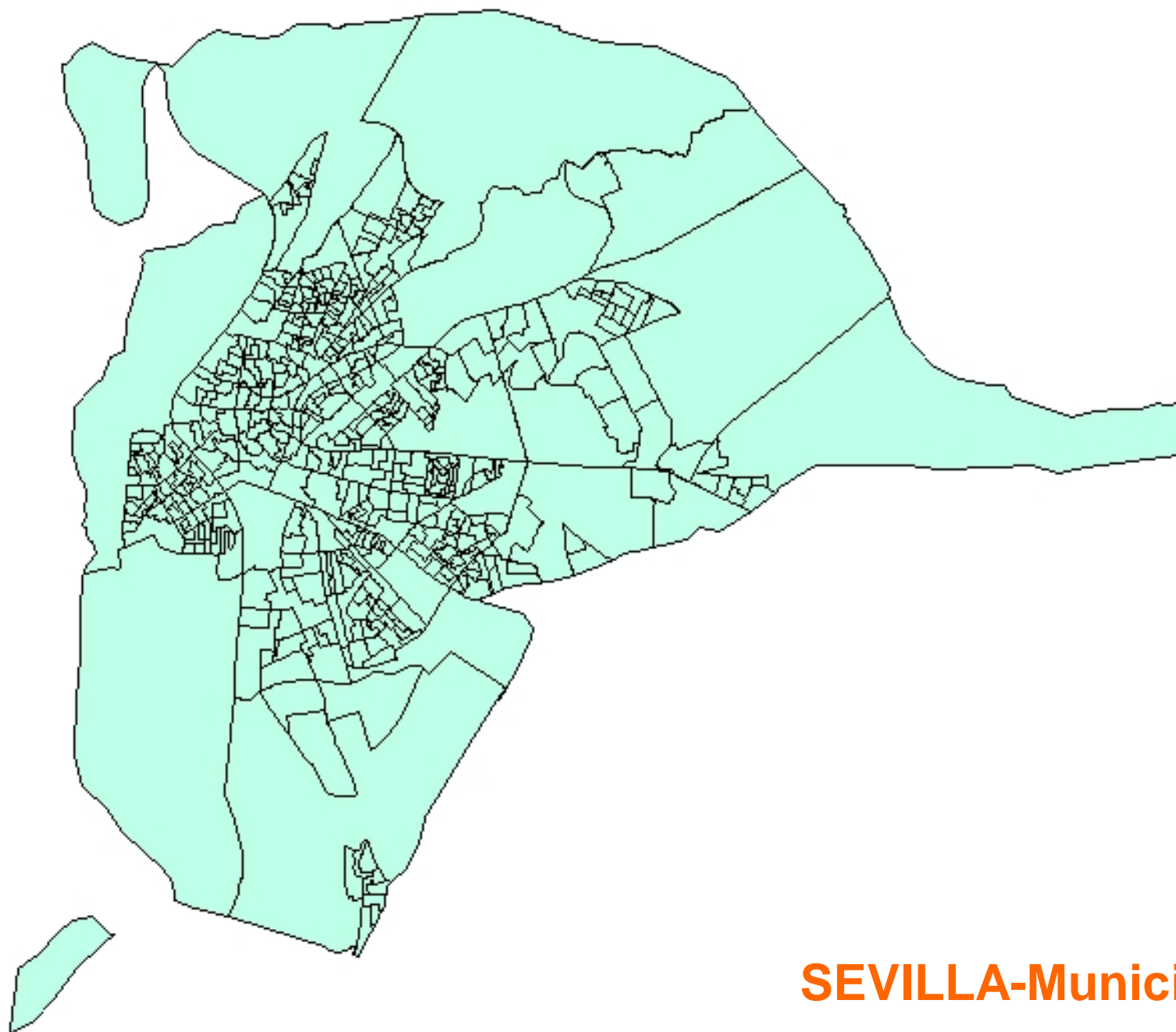
### **Características:**

- Perfectamente definida con límites fácilmente identificables
- El tamaño varía, de acuerdo con la Ley General Electoral, entre 500 y 2.000 electores
- Cualquier parte del territorio nacional debe pertenecer a una sola sección censal

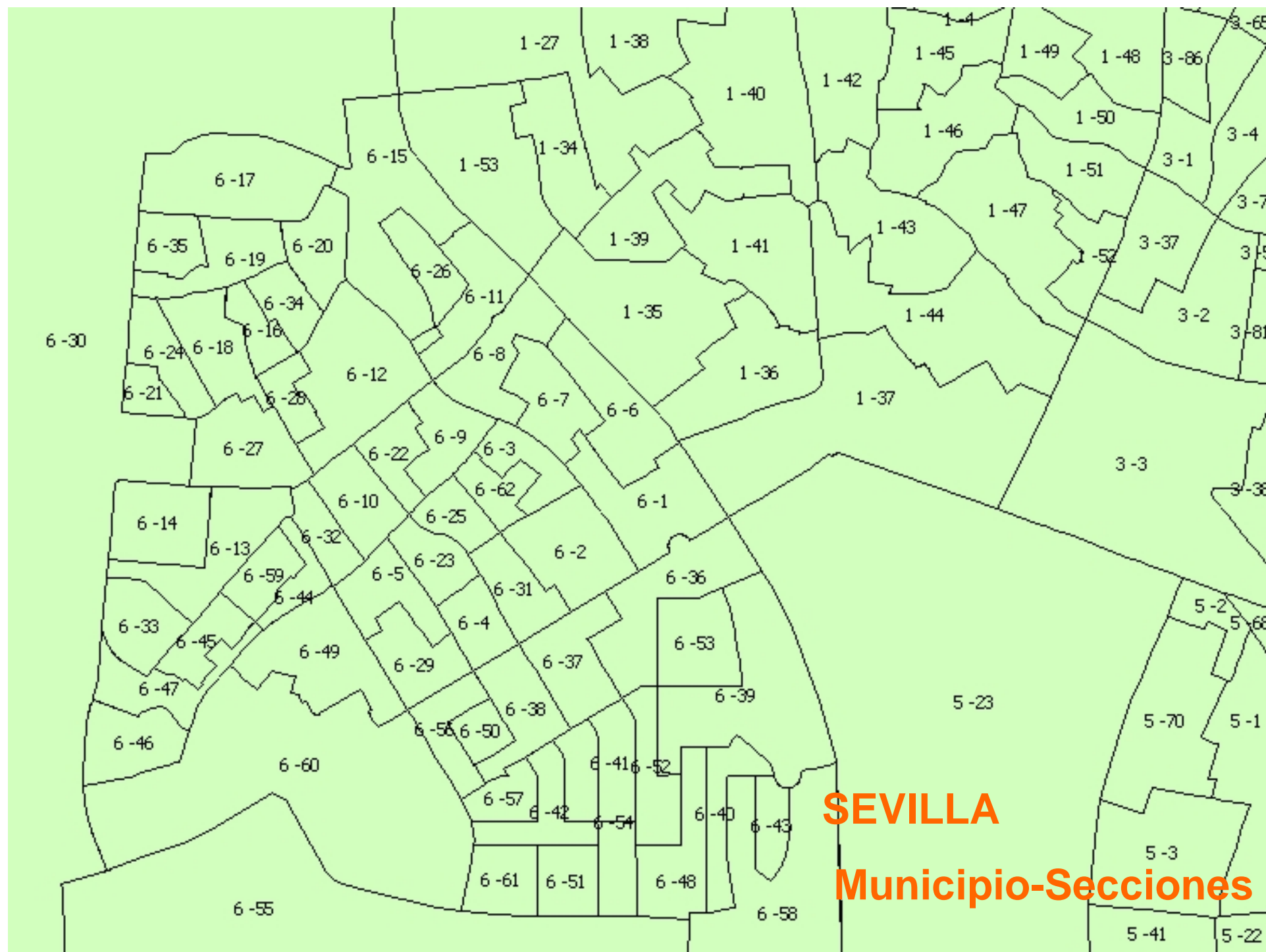




**SEVILLA**



**SEVILLA-Municipio**



## Marco de unidades de segunda etapa

- **Relación de todas unidades seleccionables** con sus direcciones postales, en cada una de las secciones censales **seleccionadas para la encuesta**
  - En el caso de la EPA es la relación **de viviendas familiares**, en la ECH es la relación de **Huecos**.
  - Se obtiene en cada **censo** a partir de los recorridos censales y se elabora el marco una vez finalizada la fase de recogida de los cuestionarios censales. **Se actualiza periódicamente**
  - En las encuestas esporádicas , son las viviendas familiares y se obtiene de una explotación del **Padrón Continuo**



## Utilización del Censo en la formación del marco de las Encuestas de hogares

Fuente de información desagregada a **nivel de unidades primarias** de muestreo:

- Variables de estratificación: Tamaño del municipio
- Variables de subestratificación: Categoría socioeconómica de la población activa de la sección.

Instrumento para la **formación del marco de unidades de segunda etapa.**



**INE**

**INSTITUTO NACIONAL DE ESTADISTICA**

CPRO	CMUN	DIST	NSECC	Población	% de jóvenes (0-19)	% de jóvenes (15-24)	% de Mayores	% de parados en la sección	% de inactivos	% de ocupados	% de extranjeros
41	091	01	022	1.146,0	9,34	21,29	20,24	10,38	53,66	35,95	3,14
41	091	01	023	1.487,0	9,75	21,52	16,75	11,97	49,83	37,26	2,69
41	091	01	024	1.261,0	10,55	17,76	20,38	10,47	54,48	34,66	2,38
41	091	01	025	2.036,0	11,25	19,40	17,58	9,48	49,85	37,28	2,65
41	091	01	027	1.391,0	9,99	22,00	21,21	5,97	54,57	39,47	1,01
41	091	01	028	773,0	12,55	20,83	17,21	11,25	52,65	34,67	2,85
41	091	01	029	1.915,0	9,92	23,86	13,68	11,96	47,42	35,67	1,04
41	091	01	030	762,0	8,27	23,23	22,18	6,96	53,67	37,53	0,79
41	091	01	031	758,0	8,84	17,81	26,65	10,16	56,20	33,64	1,72

**% de personas con nivel de estudios**

CPRO	CMUN	DIST	Nº SECC	inferiores	medios	superiores	Renta total por vivienda con percentores	Renta por desempleo entre renta total	Renta Capital mobiliario e inmobiliario sobre renta total	Renta agraria sobre renta total	Subestrato
41	091	01	022	39,70	37,87	22,43	19160,6	2,0	4,6	0,1	4
41	091	01	023	36,85	42,57	19,64	17464,7	2,2	3,1	0,0	4
41	091	01	024	44,96	33,23	21,41	19662,2	1,6	5,2	0,3	4
41	091	01	025	43,71	34,33	18,57	18711,8	1,8	3,7	0,3	4
41	091	01	027	26,82	37,60	35,59	44987,0	0,5	23,4	1,2	6
41	091	01	028	46,18	33,12	19,28	19579,7	1,5	4,6	0,4	4
41	091	01	029	37,08	39,95	18,02	19480,2	1,7	4,6	0,3	4
41	091	01	030	29,27	43,96	24,93	33633,7	1,2	6,6	0,0	4
41	091	01	031	41,29	39,58	19,13	17857,5	2,7	4,1	0,1	4







Instituto Nacional de Estadística

Encuesta de Población Activa

EGP-A2

Página /1 58

Ent. Colectiva 0

Ent. Singular 2

Núcleo/Dis 1

SEVILLA

Código de encuesta4102106 2

Provincia 41

Municipio 91

Sección 25

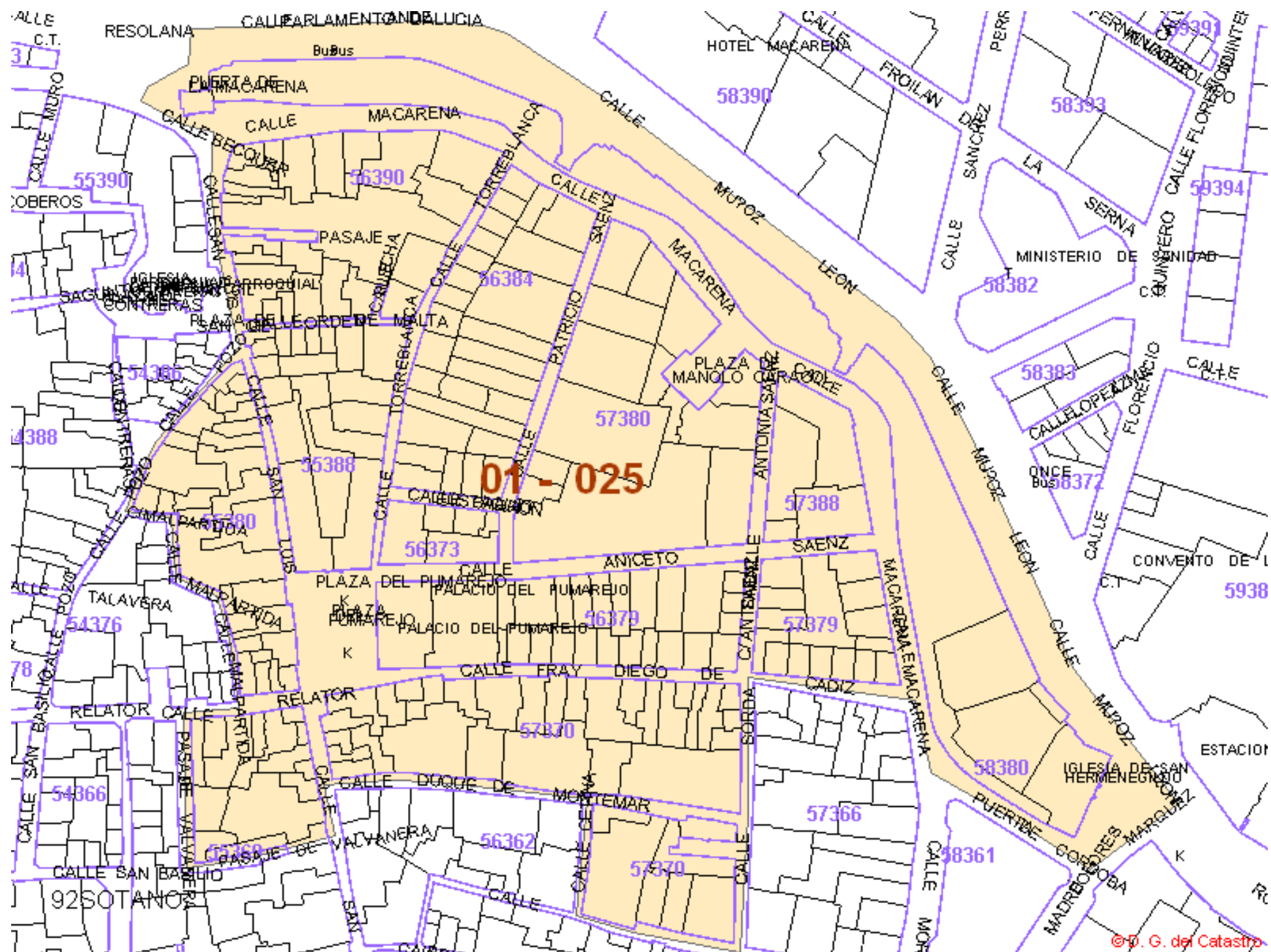
SEVILLA

SEVILLA

Num de manz	TVIA NVIA	Dirección Postal							Núm. de orden	Cod de Hueco	Viviendas	Colectivos	Locales	Ocupante Principal	Núm de Ins	Núm de Sel	Encuestas Anteriores
		TP	NUM	BQ	PR	ES	PLN	PUER									
	CALLE ANICETO SAENZ	N	9 Z			PBJ			006	0045		C		.....	1	0	0
	CALLE ANICETO SAENZ	N	21			1 P01	F		0013	00100	H			.....	2	1	0
	CALLE ANICETO SAENZ	N	21			1 P01	G		0013	00101	H			.....	3	2	0
	CALLE ANICETO SAENZ	N	21			1 P02	G		0013	00102	S			.....	4	0	0
	CALLE ANICETO SAENZ	N	21			1 P02	IZ		0013	00103	H			.....	5	3	0
	CALLE ANICETO SAENZ	N	21			1 P02	J1		0013	00104	H			.....	6	4	0
	CALLE ANICETO SAENZ	N	21			1 PBJ	1		0013	0099			L	.....	7	0	0
1	CALLE ANICETO SAENZ	N	21			1 PBJ	A1		0013	0097	H				8	5	0
1	CALLE ANICETO SAENZ	N	21			1 PBJ	B1		0013	0098	V				9	0	0
1	CALLE ANICETO SAENZ	N	21			2 P01	DR		0013	109	H				10	6	0
1	CALLE ANICETO SAENZ	N	21			2 P02	D		0013	00110	H				11	7	0
1	CALLE ANICETO SAENZ	N	21			2 P02	H		0013	00111	H				12	8	0
1	CALLE ANICETO SAENZ	N	21			2 P02	IZ		0013	00112	H				13	9	0
1	CALLE ANICETO SAENZ	N	21			2 PBJ	A		0013	00106	H				14	10	0
1	CALLE ANICETO SAENZ	N	21			2 PBJ	B2		0013	00105	H				15	11	0







## Utilización del Padrón Continuo como marco de encuestas a hogares.

- El Padrón Municipal es un **Registro administrativo** donde figuran todos los habitantes del País. La formación, mantenimiento, revisión y custodia corresponde a los Ayuntamientos.
- En 1996 se aprueba la nueva normativa padronal y se establece un nuevo sistema de **gestión continua** de los padrones coordinada por el INE. Este nuevo sistema tiene como punto de arranque el Padrón Municipal con referencia a 1 de Mayo de 1996.
- Se obtienen cifras oficiales a primero de cada año



El Padrón Continuo proporciona una **relación actualizada de viviendas habitadas**. Se obtiene mediante un programa informático desarrollado en el INE.

### **Ventajas:**

- Disponible de forma inmediata.
- Nombres actualizados de las personas que habitan las viviendas.
- Uso como marco de personas.

### **Inconvenientes:**

- Es un fichero de personas.
- Las personas mal empadronadas no tiene probabilidad de ser seleccionadas.
- Escasa información.



## Mejoras en el marco de las encuestas a hogares.

### Estudio Demográfico Longitudinal (EDL).

(Se inicia en 2001)

- Proporcionar información demográfica de tipo longitudinal integrando distintas fuentes sociodemográficas.
- Ofrecer un marco muestral óptimo para las encuestas dirigidas a los hogares o personas :
  1. Permanentemente actualizado (Padrón)
  2. Gran cantidad de información auxiliar (Otras fuentes)
- Crear la infraestructura necesaria para la elaboración del censo de población 2011.



## Fichero Precensal (FPC)

- Nueva denominación del EDL. Base de datos centralizada obtenida con el máximo aprovechamiento de los registros administrativos tomando al padrón como elemento básico de su estructura.
- Contiene a todas las personas de Padrón en las direcciones de Padrón.
- En una estructura de huecos, con base el Censo 2001. Faltan los creados desde 2001 y que nunca hayan sido vivienda principal.
- Se van añadiendo huecos desde 2007 en adelante a partir de las modificaciones de Padrón.
- Base para el recorrido censal de 2011



## SIGPT

- **Sistema de Información Geográfica de Población y Territorio**
- Aprovechamiento de Registros Administrativos.
- Contiene a las direcciones georreferenciadas
- Permanentemente actualizado.
- Contiene la relación de Huecos de la sección con identificación de su tipo.
- Contenido Normalizado, Apps, Huecos, códigos de vías, etc



## MDG

- **Marco de Direcciones georreferenciadas**
- Se trata de un Marco a fecha noviembre de 2011, aprovechando el recorrido Censal (ficheros de grabación y coordenadas recogidas en Campo) y la información procedente del Fichero Precensal.
- Contiene referencia geográfica para las aproximaciones postales y permite una integración directa con el sistema IDE\_INE.
- Representa un territorio completo a una fecha fija, noviembre de 2011, es decir, todos los elementos o jerarquías territoriales están referenciados a la misma fecha, independientemente de cuándo se recogiera su información.
- Contiene las personas existentes en Padrón a noviembre de 2011 asignadas a algún elemento de la estructura territorial anterior (vía-app-hueco).



**I. Actualización en el marco de viviendas**, restringida a las secciones que pertenecen a la muestra.

**II. Actualización en el marco de unidades primarias** (secciones censales) como consecuencia de modificaciones en las unidades primarias seleccionadas para la muestra.

**III. Actualización con carácter general**, relativa a todas las secciones y viviendas, se realiza cada tres años y se actualizan las probabilidades de selección.





## Actualización del marco de viviendas

Afecta a las encuestas continuas.

Si el marco **procede del censo (EPA)**:

Se realiza un recorrido completo de la sección para:

- Dar de alta a edificios y viviendas de nueva creación.
- Visitar las viviendas vacías, locales o establecimientos colectivos por si hubiera cambiado su situación para asignarle probabilidad de selección.
- Verificar los límites de la sección, nuevos caminos, etc.

Si el marco **procede del Padrón**:

Cada año se genera un marco actualizado de viviendas.



## Actualización en el marco de secciones

- Tipos de incidencias que se producen:
  - " Cambio de denominación
  - " Partición
  - " Fusión
  - " Variación de límites
- Forma de proceder:
  - " Cálculo de las probabilidades de selección
  - " Selección de la nueva sección
  - " Determinación del número de viviendas a entrevistar



## Actualización de las probabilidades de selección.

- Procedimiento probabilístico que permite la **permanencia** de secciones en la muestra con la probabilidad **actualizada**.
- Procedimiento alternativo: Selección de una nueva muestra con probabilidades actualizadas.
- Inconveniente: Ruptura total con la muestra anterior.
- Metodología de Keyfitz y Kish siguiendo técnicas de selección controlada.



## Probabilidad de selección de una vivienda de la muestra

**Secciones:** Probabilidad proporcional al tamaño medido por el número de viviendas familiares principales

**Viviendas:** Probabilidad igual(muestreo sistemático)

De esta forma en cada estrato, las viviendas familiares tienen la misma probabilidad de pertenecer a la muestra (**muestra autoponderada**)

$$P(V_{ijh}) = P(S_{jh}) \cdot P\left(\frac{V_{ijh}}{S_{jh}}\right) = K_h \cdot \frac{V_{jh}}{V_h} \cdot \frac{m}{V_{jh}} = \frac{K_h \cdot m}{V_h}$$

