

Diseño muestral de la Encuesta de Métodos de Producción Agrícola 2009

El diseño muestral de la Encuesta de Métodos de Producción Agrícola 2009 corresponde a un tipo de muestreo aleatorio estratificado. En cada Comunidad Autónoma, los estimadores están calibrados a las características más relevantes del Censo Agrario 2009 de dicha comunidad.

Los pasos seguidos en el diseño muestral han sido: En primer lugar, se determina las explotaciones exhaustivas y el resto de las explotaciones se distribuye entre los estratos. En segundo lugar, se calcula el tamaño muestral aplicando el criterio de afijación óptima, cumpliendo los requisitos de precisión establecidos en el Anexo IV del Reglamento CEE No 1166/2008. Posteriormente, se selecciona la muestra de manera aleatoria en cada estrato. Finalmente, se calculan los estimadores calibrados de los totales de las variables y los errores de muestreo. A continuación se detalla cada uno de estos pasos.

1. Determinación de las explotaciones exhaustivas

Para reducir la asimetría que presentan las principales características de las explotaciones agrícolas: SAU (Superficie Agrícola Útil), TL (Tierra Labrada) y UG (Unidades Ganaderas) se elige, en primer lugar, como explotaciones exhaustivas el 0,5% de las más grandes de cada Comunidad Autónoma, respecto a cada una de estas características.

Por otra parte, con el objetivo de reducir la dispersión de la población respecto a las características agrícolas y ganaderas más relevantes de cada Comunidad Autónoma, se usa un segundo criterio para seleccionar explotaciones exhaustivas: la regla de la desviación sigma (Julien y Mandala, 1990). Esta regla se aplica en cada Comunidad Autónoma y OTE2 a cada una de las características relevantes de la Comunidad Autónoma. Consiste en ordenar las explotaciones de menor a mayor según una cierta variable y tomar como exhaustivas las que siguen a la primera que cumple que la diferencia con la anterior es mayor que la desviación típica de dicha variable.

Las características agrícolas y ganaderas más relevantes de cada Comunidad Autónoma vienen determinadas por las condiciones establecidas en el Anexo IV del Reglamento CEE No 1166/2008.

2. Estratificación

El estrato se define por el cruce de la Comunidad Autónoma, OTE2 y 5 grupos de tamaño. Las variables de estratificación que definen el tamaño son la SAU y las TL en las OTEs de predominancia agrícola y la SAU y UG en las OTEs de predominancia ganadera.

Se lleva a cabo una pre-estratificación del marco inicial y una vez recogida la información censal, se actualiza la estratificación. En ambos casos, las variables usadas son las mismas.

Los grupos de tamaño se forman siguiendo la metodología de las encuestas sobre estructura de las explotaciones agrícolas. A cada una de las variables de estratificación se le aplica la regla acumulativa de la raíz cuadrada de la distribución de frecuencias (Cochran, 1977). Con ella se obtienen unas variables categóricas, asociadas a cada una de las variables de estratificación, que toman valores del 1 al 5 indicando el grupo de tamaño, en el que 1 representa a las explotaciones más pequeñas y 5 a las más grandes. Finalmente, el grupo de tamaño se define como el máximo valor que toman estas variables.

3. Tamaño de la muestra, afijación y selección muestral

Para calcular el tamaño de la muestra, se utiliza el criterio de afijación óptima siguiendo los requisitos de precisión para la determinación del tamaño muestral que vienen establecidos en el Anexo IV del Reglamento CEE No 1166/2008.

Estos requisitos consisten en fijar un coeficiente de variación, C' , del 10% para una serie de características agrícolas y ganaderas que cumplen las siguientes condiciones:

-Para las características agrícolas: aquellas que representan el 5% o más de la SAU de la Comunidad Autónoma

-Para las características ganaderas: las que representan el 10% o más de las UG de la Comunidad Autónoma y más del 5% del total nacional de dicha característica

Estas características se eligen entre las siguientes:

Características agrícolas:

- 1) Tierras para pastos
- 2) Cereales
- 3) Leguminosas, patatas y cultivos industriales
- 4) Cultivos forrajeros
- 5) Hortalizas y flores
- 6) Frutales, olivar, viñedo, viveros e invernadero

Características ganaderas:

- 1) Bovinos
- 2) Ovinos y caprinos
- 3) Porcinos
- 4) Aves

Para cada Comunidad Autónoma y OTE2, se calcula el tamaño muestral n_{cot} resolviendo el siguiente problema de optimización:

$$\begin{aligned} & \text{Min} \sum_{t:1}^5 n_{cot} \\ & \text{sujeto a} \\ & \frac{V(\hat{X}_{cov})}{X_{cov}^2} \leq C_{cov}^2 \end{aligned}$$

donde el subíndice 'c' indica la Comunidad Autónoma, 'o' la OTE2, 't' los grupos de tamaño, 'v' cada variable (características agrícolas y ganaderas determinadas por el Reglamento) y n_{cot} representa el tamaño muestral en el estrato indicado por el subíndice.

El coeficiente C_{cov} se establece como sigue:

$$C_{cov} = \frac{C'}{(X_{cov}/X_{cv})H^{1/2}}$$

donde H es el número de OTE2s en la Comunidad Autónoma.

De esta forma se garantiza que el coeficiente de variación para el total de la variable 'v' en la Comunidad Autónoma 'c' sea menor de C', dado que:

$$V(\hat{X}_{cv}) \leq \sum_{o=1}^H C_{cov}^2 X_{cov}^2 = \sum_{o=1}^H \frac{C'^2 X_{cov}^2}{(X_{cov}/X_{cv})^2 H} = C'^2 X_{cv}^2$$

Para resolver este problema de optimización se emplea el algoritmo de Bethel (Répartition de l'échantillon dans les enquêtes à plusieurs variables, Techniques d'enquêtes, 1989, vol.15 n° 1, pp. 49-60). El tamaño muestral obtenido se incrementa en previsión de la falta de respuesta.

La muestra se selecciona dentro de cada estrato de forma aleatoria.

En el anexo I se presenta la distribución de la población y muestra efectiva para el total nacional por OTE2 y grupos de tamaño. El grupo de tamaño 6 corresponde a las explotaciones exhaustivas.

4. Estimadores

4.1. Estimadores de los totales de las variables

En cada Comunidad Autónoma, se construye estimadores calibrados usando como información auxiliar las características más relevantes del Censo Agrario 2009.

Sea Y la variable de interés de la cual se quiere estimar su total poblacional:

$$Y = \sum_{h:1}^L Y_h \quad Y_h = \sum_{k:1}^{N_h} y_{hk}$$

donde L denota el número de estratos, N_h el tamaño de la población en el estrato h e y_{hk} es el valor de la variable de interés en la explotación k .

El estimador calibrado viene dado por

$$\hat{Y}_w = \sum_{h:1}^L \hat{Y}_{hw} \quad \hat{Y}_{hw} = \sum_{k:1}^{n_h} w_{hk} y_{hk}$$

donde n_h es el tamaño muestral en el estrato h y los pesos w_{hk} se calculan resolviendo el siguiente problema:

$$\begin{aligned} & \text{Min}_{w_{hk}} \sum_{k \in s_c} d_{hk} G(w_{hk} / d_{hk}) \\ & \text{sujeto a} \\ & \sum_{k \in s_c} w_{hk} x_{jkh} = X_{cj} \quad \forall j = 1 \dots J \end{aligned}$$

donde las variables $X_{c,j:1...J}$ representan las características censales utilizadas como variables auxiliares en el calibrado en la Comunidad Autónoma 'c', s_c es la muestra de la Comunidad Autónoma 'c', d_{hk} denota el peso de diseño usado en el estimador de Horvitz-Thompson del total de Y_h , que se obtiene como la inversa de la probabilidad de selección y $G(w_{hk}/d_{hk})$ es la función distancia logarítmica truncada dada por la siguiente expresión

$$G(w_{hk} / d_{hk}) = G(r) = \left[(r - I) \text{Log} \frac{r - I}{1 - I} + (U - r) \text{Log} \frac{U - r}{U - 1} \right] \frac{1}{A} \quad \text{si } I < r < U$$

donde $A = \frac{U - I}{(1 - I)(U - 1)}$

Las constantes I y U representan los límites inferior y superior respectivamente de los cocientes w_{hk}/d_{hk}

Los límites establecidos varían de 1/3 ó 1/6 para I y de 3 ó 6 para U . Las características censales utilizadas en el calibrado han sido las más relevantes de cada Comunidad Autónoma y correlacionadas con las variables objeto de estudio, observadas en la muestra.

Para resolver este problema se aplica la macro CALMAR realizada en el software SAS por el INSEE francés.

4.2. Estimadores de los errores de muestreo

El estimador de la varianza del estimador calibrado se calcula aplicando la técnica de los residuos ponderados (Särndal et al, Biometrika 1989, 76,3, pp 527-37).

En el caso de muestreo aleatorio estratificado este estimador en el estrato h viene dado por la siguiente expresión:

$$\hat{V}(\hat{Y}_{hw}) = N_h^2 \frac{1 - f_h}{n_h} \frac{\sum_{k \in s_h} (w_{hk} f_h e_{hk})^2}{n_h - 1}$$

donde $f_h = \frac{n_h}{N_h}$ $e_{hk} = y_{hk} - x'_{hk} \hat{B}_{hw}$

$$\hat{B}_{hw} = \frac{\sum_{s_h} w_{hk} x_{hk} y_{hk}}{\sum_{s_h} w_{hk} x_{hk} x'_{hk}}$$

El estimador de la varianza del estimador calibrado del total para una determinada suma de estratos viene dado por la suma de los estimadores de las varianzas de los estratos.

En el anexo II se presentan los errores relativos de muestreo o coeficientes de variación, definidos como el cociente entre la raíz cuadrada del estimador de la varianza de la estimación y la estimación, de las variables más relevantes de la encuesta para el total nacional y por Comunidades Autónomas

Anexo I

Población y muestra efectiva de la MP09 total nacional

	TAMAÑOS:1		2		3		4		5		6	TOTAL	
OTE2	POBLACI	MUESTR	POBLACI	MUESTR	POBLACI	MUESTR	POBLACI	MUESTR	POBLACI	MUESTR	POBLACI	POBLACION	MUESTRA
15	57385	1717	32413	1594	19409	1494	12008	1602	5980	1408	1600	128795	9415
16	35781	1513	18525	888	10186	707	5770	664	3543	704	1579	75384	6055
21	12577	395	3056	171	2012	123	1183	101	387	51	308	19523	1149
22	6337	456	3209	290	1534	210	881	150	539	108	271	12771	1485
23	1261	52	872	49	471	39	322	30	204	30	137	3267	337
35	34324	1046	16376	738	9441	666	4992	509	1900	319	294	67327	3572
36	94599	1199	34781	1106	20157	913	8205	789	2636	486	569	160947	5062
37	127784	1127	51725	714	21714	581	8523	516	2990	451	382	213118	3771
38	19643	275	11490	267	5912	233	3011	218	1322	216	174	41552	1383
45	4822	332	5489	352	5260	428	4135	544	2236	438	1182	23124	3276
46	15365	1572	15157	1845	11577	1682	7434	1636	4437	1721	1446	55416	9902
47	780	64	828	77	750	84	520	81	450	92	97	3425	495
48	16833	2241	12643	1777	9622	1484	6447	1151	3937	1046	1277	50759	8976
51	3348	263	2978	387	2803	517	2706	742	2161	851	2197	16193	4957
52	1235	119	1132	192	1156	287	1033	380	800	222	883	6239	2083
53	833	68	1192	107	623	78	480	65	312	72	95	3535	485
61	20473	784	11723	576	6885	465	3736	404	1951	456	520	45288	3205
73	2539	239	2468	279	1735	246	1179	253	725	216	242	8888	1475
74	1258	60	807	84	550	82	431	96	426	126	220	3692	668
83	2725	102	2452	146	1856	131	1323	187	1042	234	446	9844	1246
84	10408	540	5963	576	3276	456	2051	362	1310	356	317	23325	2607
90	10190	444	3677	131	1952	155	1001	93	541	70	23	17384	916
Total	480500	14608	238956	12346	138881	11061	77371	10573	39829	9673	14259	989796	72520

Anexo II

COEFICIENTES DE VARIACIÓN (EN TANTOS POR CIENTO) DE LAS VARIABLES MÁS RELEVANTES

	Variables									
	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10
Total nacional	0,73	1,22	1,05	0,86	0,05	1,34	1,82	0,07	0,07	2,29
Andalucía	0,89	4,53	3,79	3,68	0,05	3,23	6,77	0,16	0,16	5,06
Aragón	4,22	2,69	2,99	1,78	0,14	5,49	7,77	0,28	0,28	7,82
Asturias	1,65	1,93	5,53	1,85	0,37	4,37	4,73	1,12	1,12	21,12
Baleares	1,01	3,62	5,92	1,48	0,13	6,24	4,78	0,67	0,67	11,11
Canarias	1,08	5,18	3,4	1,8	0,06	3,02	8,83	1,16	1,16	7,11
Cantabria	1,96	2,19	1,65	1	0,08	7,38	4,14	0,53	0,53	12,79
Castilla-La Mancha	2,86	4,41	3,78	1,32	0,27	4,09	6,75	0,14	0,14	5,19
Castilla-León	3,68	6,1	1,61	0,33	0,18	3,4	10,75	0,05	0,05	5,15
Cataluña	3,32	4,09	1,61	5,06	0,12	3,54	4,86	0,59	0,59	6,13
Com. Valenciana	1,2	4,34	4,72	0,26	0,05	3,4	5,58	0,38	0,38	5,93
Extremadura	0,91	5,08	3,66	1,39	0,13	4,79	9,08	0,22	0,22	7,64
Galicia	1,12	1,07	2,42	3,43	0,11	3,39	1,99	0,11	0,11	4,61
Madrid	1,73	6,05	0,18	0,02	0,2	6,47	14,91	0,33	0,33	11,33
Murcia	6,45	4,2	1,81	2,75	0,05	3,58	7,71	0,23	0,23	9,76
Navarra	6,54	6,57	2,33	0,41	0,23	7,88	10,17	0,13	0,13	11,23
País Vasco*	0	0	0	0	0	0	0	0	0	0
La Rioja	11,81	3,69	0,68	0,54	0,21	8,95	19,86	0,13	0,13	11,48

Variables:

Y1: Superficie total pastada
 Y2: Número de plazas de bovinos
 Y3: Número de plazas de porcinos
 Y4: Número de plazas de gallinas ponedoras
 Y5: SAU media regada en los últimos tres años
 Y6: Superficie de la explotación abonada con estiércol sólido
 Y7: Superficie de la explotación abonada con estiércol semilíquido o lisier
 Y8: Laboreo del terreno para cultivos herbáceos
 Y9: Cubierta invernal del suelo en cultivos herbáceos
 Y10: Cultivos herbáceos fuera de la rotación prevista

* CENSO