



ARGENTINA

Dirección Nacional de Estadísticas Sociodemográficas
Serie: Documentos Técnicos
Documento Nro.:15

"SEGUNDA REUNION DEL COMITE DE EXPERTOS EN
ESTADISTICAS SOCIODEMOGRAFICAS"

Buenos Aires, 22 al 25 de agosto 1989

LA REFORMULACION MUESTRAL EN LA EPH:
LA DETERMINACION DE LOS DOMINIOS

Con la colaboración del Programa de Naciones Unidas
para Desarrollo y del Fondo de Población de
Naciones Unidas y de la Oficina Panamericana de la
Salud.

P R E S E N T A C I O N

Este documento contiene los avances realizados en este ámbito desde noviembre de 1988. Incluye aportes recibidos en oportunidad de la presentación en el Primer taller para rediseño de las estadísticas sociodemográficas realizado en Buenos Aires entre el 27/03 y el 07/04 de 1989.

Los responsables de la elaboración de este documento son: Marta R. de Messere, Clyde Charre de Trabuchi, Susana Martínez y Augusto Hozzorwski.

C O N T E N I D O

I.- INTRODUCCION	05
II.- AMBITO GEOGRAFICO	07
III.- DOMINIOS DE ESTUDIO	07
IV.- DETERMINACION DEL TAMAMO DE LA MUESTRA	09
V.- PLAN DE TABLAS MINIMAS A OBTENER PARA AGLOMERADOS PROVINCIALES	15
VI.- PLAZOS PARA LA APLICACION DE ESTE PROYECTO	17

ANEXOS:

-ANEXO 1: Promedio de económicamente activos por hogar	19
-ANEXO 2: Efectos Diseño para la variable desempleo en diferentes aglomerados	21
-ANEXO 3: Afijación de la muestra	23
-ANEXO 4: Distribución de la población según Cen 80	25

BIBLIOGRAFIA	27
--------------------	----

I.- INTRODUCCION

La EPH, se realiza en dos oportunidades en el año, mayo y octubre, en aproximadamente 35.000 hogares cada vez.

La República Argentina es un país federal, con 2.780.000 Km² y sólo 32.000.000 de habitantes según las últimas estimaciones. Está dividido en 22 provincias, un Territorio Nacional y un distrito federal. La población se distribuye en forma desigual (Ver Anexo 4) muy concentrada en el Gran Buenos Aires (30% del total) y algunos centros urbanos menores (Córdoba, Rosario, etc.). En contrapartida existen provincias muy extensas, con grandes zonas desérticas y escasa población.

La EPH comenzó a relevarse en 1972 en el Gran Buenos Aires y luego por diversas razones, se extendió en forma inorgánica a las capitales de provincia y algunos aglomerados importantes desde el punto de vista provincial. Posteriormente se realizó una experiencia en la provincia de Tucumán, abarcando su totalidad, es decir zona urbana y rural, durante cuatro años.

Desde 1977 se releva el Alto Valle del Río Negro, una pequeña región urbano-rural.

Como se ve, el alcance de la encuesta, no puede definirse de manera simple como una categoría de aglomerados urbanos, por ejemplo los aglomerados con 50.000 habitantes y más en el último censo, o el conjunto de capitales provinciales, ya que incluye capitales con mucha menor cantidad de habitantes que 50.000 y aglomerados que no son capitales. Cada una de las ciudades incluidas constituyen un dominio

de estudio, para el cual se deben obtener estimaciones separadas, por lo tanto el tamaño de muestra depende poco del aglomerado al cual representa.

Esta prioridad absoluta que se surgió a la información a nivel local originó los siguientes problemas:

- Las muestras, para cada localidad, son pequeñas para profundizar en el estudio de las estructuras y están sujetas a variaciones importantes debidas al muestreo. No existen en el país proyecciones demográficas que permitan ajustar estas estimaciones.
- A nivel nacional o regional no se puede definir ningún dominio de estudio preciso ni presentar resultados agregados.

Paradójicamente se ha llegado a la situación de tener implementada una enorme muestra de 70.000 hogares al año y no poder obtener resultados confiables a nivel global. Se concluye entonces que el diseño actual es costoso e ineficiente.

De cualquier manera vale la pena destacar que desde el punto de vista organizativo, este esquema impulsó una tarea de formación y entrenamiento en las Direcciones Provinciales de Estadística, que no sólo ha ido mejorando la captación de la encuesta, sino que los equipos así formados, han sido aprovechados en los operativos censales y otras actividades estadísticas.

Teniendo en cuenta lo dicho anteriormente resulta fácil comprender que en el proyecto general de revisión y reformulación de la EPH, uno de sus capítulos sustantivos lo constituye el nuevo diseño de muestra a implementar a partir de 1992, cuando se cuente con la información que brindará el Censo de 1990.

II.- AMBITO GEOGRAFICO

La encuesta abarcará todo el territorio nacional.

III.- DOMINIOS DE ESTUDIO

Las áreas para las que se harán tabulaciones separadas son:

- 1- Todo el país.
- 2- Cada una de las provincias y el territorio nacional.
- 3- Cada uno de los aglomerados con 500.000 y más habitantes en el Censo de 1990.
 - Según información del Censo de 1980:
 - Capital Federal
 - Partidos del conurbano
 - Córdoba
 - La Plata
 - Mendoza
 - Rosario
 - Tucumán
- 4- Total de aglomerados con más de 50.000 y menos de 500.000 habitantes en el Censo de 1990.
- 5- Resto del país.

IV.- DETERMINACION DEL TAMANO DE LA MUESTRA

Para determinar el tamaño de las muestras se fijó como condición que, a un nivel de confianza del 95%, una diferencia en menos del d% la estimación del cambio en la "desocupación" entre dos ondas de la encuesta pudiera ser admitida como debida al muestreo. Es decir:

$$P \left\{ \left| p_i - p_{i+1} \right| - \left| P_i - P_{i+1} \right| \leq k \cdot \sigma(p_i - p_{i+1}) \right\} = 0,95$$

Donde:

p_i es el desempleo en el período i y p_{i+1} es el desempleo en el período $(i+1)$.

El desvío standard de una diferencia es:

$$\sigma(p_i - p_{i+1}) = \sqrt{\sigma^2(p_i) + \sigma^2(p_{i+1}) - 2\rho \sigma_{p_i} \sigma(p_{i+1})}$$

ρ = correlación en el desempleo entre dos periodos de la encuesta que tienen α unidades de muestreo comunes.

Si se considera que el desempleo entre el período i e $(i+1)$ no varía:

$$\sigma(p_i - p_{i+1}) = \sqrt{2\sigma_{p_i}^2 - 2\rho \sigma_{p_i}^2}$$

$$\sigma(p_i - p_{i+1}) = \sqrt{2\sigma_{p_i}^2(1 - \rho)} \therefore$$

$$\sigma(p_i - p_{i+1}) = \sigma_{p_i} \sqrt{2(1 - \rho)}$$

en consecuencia: $\sigma_{p_i} = \frac{\sigma(p_i - p_{i+1})}{\sqrt{2(1 - \rho)}} \quad (1)$

Como se trata de una variable proporción y suponiendo que hay efecto diseño (e.d.) por conglomeración

$$\sqrt{V_{pi}} = e.d. \frac{p_i q_i}{n_o} \quad (2)$$

donde n es el tamaño muestral medido en personas económicamente activas establecemos que: $k \sqrt{(p_i - p_{i+1})} = d$

$$\sqrt{(p_i - p_{i+1})} = \frac{d}{k} \quad \text{luego de (1) y (2), reemplazando}$$

$$\left(\frac{d}{k \sqrt{2(1-\rho)}} \right)^2 = \frac{e.d. p_i q_i}{n_o}$$

despejando n_o

$$n_o = \frac{2 e.d. p_i q_i k^2 (1-\rho)}{d^2}$$

si \bar{E} = promedio de económicamente activos por hogar

$$n = \frac{n_o}{\bar{E}} \quad \text{tamaño muestral medido en hogares}$$

Se trabajó suponiendo:

$e.d. = 1,5$ (ver anexo 2 con estimaciones de EPH para la variable desempleo).

$\bar{E} = 1,5$ (ver anexo 1 con estimaciones de EPH)

$\rho = 0,20$ (valor para la correlación del desempleo entre dos ondas de la encuesta cuando tienen 75% de la muestra en común, ver [6])

$k = 2$ ya que se quiere trabajar con 95% de confianza.

Surge así el cuadro 1 que presenta los tamaños de muestra necesarios para diferentes valores del desempleo (p) y de (d), diferencia admisible en la estimación del cambio entre dos periodos de la encuesta.

CUADRO 1: TAMAÑOS MUESTRALES, DADO p_i y d

d	0,005	0,01	0,015	0,02	0,025	0,03	0,035
p_i							
0,02	5.018	1.254	558	314	201	139	102
0,03	7.450	1.862	928	466	298	207	152
0,04	9.830	2.458	1.092	614	393	273	201
0,05	12.160	3.040	1.351	760	486	338	248
0,06	14.438	3.610	1.604	902	578	401	295
0,07	16.666	4.166	1.852	1.042	667	463	340
0,08	18.842	4.710	2.094	1.178	754	523	385
0,09	20.966	5.242	2.330	1.310	839	582	428
0,10	23.040	5.760	2.560	1.440	922	640	470
0,12	27.034	6.758	3.004	1.690	1.081	751	552
0,15	32.640	8.160	3.627	2.040	1.306	907	666
0,20	40.960	10.240	4.551	2.560	1.638	1.138	836
0,25	48.000	12.000	5.333	3.000	1.920	1.333	980
0,30	53.760	13.440	5.973	3.360	2.150	1.493	1.097
0,35	58.240	14.560	6.471	3.640	2.330	1.618	1.189
0,40	61.440	15.360	6.827	3.840	2.458	1.707	1.254
0,50	64.000	16.000	7.111	4.000	2.560	1.778	1.306

p_i : PROPORCION ESTIMADA.
 d : DIFERENCIA A DETECTAR.

Al analizar los valores del cuadro 1, se observa que $n = 900$ que es un tamaño razonable desde el punto de vista operativo, brinda estimaciones aceptables del cambio para los niveles de desempleo de la EPH (6%). Luego se decide asignar $n = 1.000$ a cada uno de los dominios provinciales y aglomerados autorrepresentados (dominios 2 y 3). Se incrementa el tamaño de muestra para compensar el efecto de la no respuesta.

Asimismo se estudiaron los tamaños de muestra necesarios para estimaciones puntuales del desempleo para diferentes valores de p_i y coeficientes de variación (C.V.). Si:

$$\sigma_{p_i} = \sqrt{\frac{p_i \cdot q_i}{n_0} (e.d.)}$$

$$C.V. p_i = \frac{\sqrt{\frac{p_i \cdot q_i}{n_0} (e.d.)}}{p_i}$$

$$C.V.^2 = \frac{p_i q_i (e.d.)}{n_0 p_i^2}$$

despejando:

$$n_0 = \frac{q_i \cdot (e.d.)}{p_i \cdot (C.V.)^2}$$

$$n = \frac{n_0}{\bar{E}}$$

$$n = \frac{q_i \cdot (e.d.)}{p_i (C.V.)^2 \bar{E}}$$

Considerando $e.d.$ y \bar{E} con los mismos valores supuestos para el cuadro 1, se construyó el CUADRO 2.

CUADRO 2: TAMAÑOS MUESTRALES, DADO p_i y C.V.

p_i	C.V.	0,05	0,08	0,10	0,12	0,15	0,17	0,20	0,25	0,30
0,02		19.600	7.656	4.900	3.403	2.178	1.696	1.225	784	544
0,03		12.933	5.052	3.233	2.245	1.437	1.119	808	517	359
0,04		9.600	3.750	2.400	1.667	1.067	830	600	384	267
0,05		7.600	2.969	1.900	1.319	844	657	475	304	211
0,06		6.267	2.448	1.567	1.088	696	542	392	251	174
0,07		5.314	2.076	1.329	923	590	460	332	213	148
0,08		4.600	1.797	1.150	799	511	398	288	184	128
0,09		4.044	1.580	1.011	702	449	350	253	162	112
0,10		3.600	1.406	900	625	400	311	225	144	100
0,12		2.933	1.146	733	509	326	254	183	117	81
0,15		2.267	885	567	394	252	196	142	91	63
0,20		1.600	625	400	278	178	138	100	64	44
0,25		1.200	469	300	208	133	104	75	48	33
0,30		933	365	233	162	104	81	58	37	26
0,35		743	290	186	129	83	64	46	30	21
0,40		600	234	150	104	67	52	38	24	17
0,50		400	156	100	69	44	35	25	16	11

p_i : PROPORCION ESTIMADA.

C.V.: COEFICIENTE DE VARIACION.

Se observa que con una muestra de $n=900$, se mide un desempleo de $p=0.06$ con un coeficiente de variación de 13% que es aceptable.

Luego se concluye que $n=1000$ es asimismo un tamaño razonable para brindar estimaciones puntuales para los dominios provinciales, y para los aglomerados autorrepresentados.

Dentro de cada provincia a fin de atender las demandas de información y no interrumpir las series de variables que se vienen relevando, se decidió asignar a la ciudad que tiene implementada EPH, $n=600$ que observando los cuadros 1 y 2 puede brindar estimaciones puntuales y del cambio, razonables. El resto de la muestra asignada a la provincia, se adjudica en forma proporcional a la población restante de la misma. Esta afijación está detallada en el anexo 3. Es así como se llega, por agregación, a un tamaño total de muestra $n=30.600$.

En cuanto al dominio 4, "total de aglomerados con más de 50.000 y menos de 500.000 habitantes en el censo de 1990", el n afijado es de aproximadamente 10.500 hogares que, para un valor de desempleo del 6%, mide cambios iguales o mayores a 0,3%.

V.- PLAN DE TABLAS MINIMAS A OBTENER PARA AGLOMERADOS PROVINCIALES QUE YA INTEGRAN EL PROGRAMA EPH

En estos aglomerados con $n=600$, se podrá obtener, en principio, información básica sobre:

- población por sexo y grupos de edad
- condición de actividad por sexo
- condición de actividad por grupo de edad
- nivel de instrucción por sexo
- nivel de instrucción por grupo de edad
- nivel de instrucción por condición de actividad
- población ocupada por grandes ramas de actividad
- distribución decilica de ingreso

VI.- PLAZOS PARA LA APLICACION DE ESTE PROYECTO

Se prevee iniciar el cambio muestral en mayo de 1992 completando el dominio 3, es decir cada aglomerado con más de 500.000 habitantes según censo '90.

Lograr completar el dominio 4 en octubre de 1992, de esta forma se podrá tener información básica para cada aglomerado actualmente en la muestra y, a su vez, obtener una razonable información a nivel nacional, la correspondiente a aglomerados con más de 50.000 habitantes.

Los niveles provinciales (dominios 2) se irán alcanzando entre mayo de 1993 y octubre de 1995 según orden de prioridad a ir fijando oportunamente.

Cuando se concluya la incorporación de las provincias, se logrará simultáneamente completar los dominios 5, resto del país, y 1, total nacional.

ANEXO 1: PROMEDIO DE ECONOMICAMENTE ACTIVOS POR HOGAR

AGLOMERADO	ABRIL	OCTUBRE	ABRIL	OCTUBRE	ABRIL
	1986	1986	1987	1987	1988
-GRAN BUENOS AIRES	1,39	1,39	1,42	1,40	1,40
-GRAN LA PLATA	1,25	1,26	1,30	1,26	1,21
-BAHIA BLANCA	1,30	1,32	1,27	1,30	1,28
-PARANA	1,37	1,36	1,36	1,36	1,33
-SANTA FE y SANTO TOME	1,39	1,36	1,42	1,38	1,38
-GRAN CORDOBA	1,39	1,39	1,37	1,38	1,38
-GRAN ROSARIO	1,36	1,33	1,31	1,34	1,30
-GRAN MENDOZA	1,41	1,41	1,38	1,42	1,43
-GRAN SAN JUAN	1,47	1,46	1,45	1,47	1,44
-SAN LUIS y EL CHORRILLO	1,46	1,47	1,43	1,51	1,44
-GRAN CATAMARCA	1,56	1,49	1,44	1,46	1,44
-SAN SALVADOR DE JUJUY	1,52	1,53	1,45	1,48	1,49
-LA RIOJA	1,59	1,65	1,60	1,64	1,61
-SALTA	1,62	1,60	1,56	1,57	1,55
-STGO. DEL ESTERO y LA BANDA	1,51	1,50	1,55	1,54	1,51
-GRAN TUCUMAN y TAFI VIEJO	1,57	1,60	1,73	1,70	1,64
-CORRIENTES	1,40	1,40	1,42	1,31	1,34
-GRAN RESISTENCIA	1,50	1,49	1,61	1,62	1,59
-FORMOSA	1,55	1,51	1,49	1,42	1,43
-POSADAS	1,44	1,38	1,47	1,49	1,51
-NEUQUEN	1,53	1,55	1,52	1,52	1,52
-COMODORO RIVADAVIA	1,42	1,42	1,42	1,39	1,48
-RIO GALLEGOS	1,50	1,47	1,51	1,54	1,54

fuente: EPH

ANEXO 2: EFECTOS DISEÑO PARA LA VARIABLE DESEMPLEO EN DIFERENTES AGLOMERADOS DE EPH

AGLOMERADO	ESTIMACION	CV (%)	EFFECTO DISEÑO
- GRAN BUENOS AIRES	0,050	6,2	1,33
- LA PLATA	0,075	16,0	1,53
- CORDOBA	0,059	12,4	1,51
- ROSARIO	0,073	10,8	1,31
- SAN JUAN	0,091	10,1	1,49

Fuente: EPH

NOTA: para cálculo del tamaño de muestra se tomó como hipótesis de máxima e.d. = 1,5 para la variable desempleo.

ANEXO 3: AFIJACION DE LA MUESTRA

REGION	PROVINCIA	AGLOMERADOS CON POBLACION (EN MILES)						TOTAL	TOTAL
		> 500	500-100	99-50	49-20	19-2	< 2	PROVINCIA	REGION
	-BUENOS AIRES	4.000	191	180	140	249	240	5.000	5.000
LITORAL	-ENTRE RIOS	----	600 (1)	77	38	131	154	1.000	3.000
	-SANTA FE	1.000	600 (1)	18	72	163	147	2.000	
CENTRO	-CORDOBA	1.000	78	88	117	454	263	2.000	2.000
CUYO	-MENDOZA	1.000	----	117	85	155	643	2.000	4.000
	-SAN JUAN	----	600 (1)	----	----	106	294	1.000	
	-SAN LUIS	----	----	600 (1)	----	120	280	1.000	
N. O. A.	-CATAMARCA	----	----	600 (1)	----	114	286	1.000	7.000
	-JUJUY	----	600 (1)	----	134	118	148	1.000	
	-LA RIOJA	----	----	600 (1)	----	140	260	1.000	
	-SALTA	----	600 (1)	----	64	152	184	1.000	
	-STGO. DEL ESTERO	----	600 (1)	----	79	64	257	1.000	
	-TUCUMAN (*)	1.000	----	----	160	145	695	2.000	
N. E. A.	-CORRIENTES	----	600 (1)	----	97	106	197	1.000	4.000
	-CHACO	----	600 (1)	----	62	107	231	1.000	
	-FORMOSA	----	----	600 (1)	42	99	259	1.000	
	-MISIONES	----	600 (1)	----	45	96	259	1.000	
COMAHUE	-NEUQUEN	----	----	600 (1)	67	181	152	1.000	2.000
	-LA PAMPA	----	----	600 (1)	79	130	191	1.000	
PATAGONIA	-SANTA CRUZ	----	----	----	600 (1)	279	121	1.000	3.600
	-CHUBUT	----	----	600+125(1)	505	107	118	1.000	
	-RIO NEGRO	----	----	----	600 (1)	200	200	1.000	
	-T. DEL FUEGO	----	----	----	----	600 (1)	----	600	
TOTAL		8.000	5.669	4.805	2.531	4.016	5.579	30.600	30.600

(*) PROVINCIAS QUE CONTIENEN AGLOMERADOS EN EL DOMINIO 3.

(1) TAMARO MINIMO PARA CAPITAL DE PROVINCIA O AGLOMERADO DE INTERES PROVINCIAL.

ANEXO 4: DISTRIBUCION DE LA POBLACION SEGUN CEN - 80

PROVINCIA	SUPERFICIE EN Km ²	TOTAL POBLACION SEGUN CEN '80	% POBLACION QUE VIVE EN AGLOMERADOS:		TOTAL AGLOMER. CON CANTIDAD DE HABITANTES:					
			CON 2000 HAB. y MAS	CON MENOS 2000 HAB.	2.000	10.000	50.000	100.000	500.000	TOTALES
					a'	a'	a'	a'	y M A S	
-BUENOS AIRES (*)	307.770	13.788.237	94,6	5,4		45	12	2	2	153
-CATAMARCA	100.967	207.717	57,5	42,5	6	--	1	--	--	7
-CORDOBA	168.766	2.407.754	80,7	19,3	87	18	2	1	1	109
-CORRIENTES	88.190	661.454	64,4	35,6	15	8	--	1	--	24
-CHACO	99.633	701.392	60,9	39,1	22	4	--	1	--	27
-CHUBUT	224.686	263.116	81,4	18,6	4	3	2	--	--	9
-ENTRE RIOS	78.781	908.313	68,8	31,2	24	11	2	1	--	38
-FORMOSA	72.066	295.887	55,7	44,3	12	1	1	--	--	14
-JUJUY	53.219	410.008	73,6	26,4	18	4	--	1	--	23
-LA PAMPA	143.440	208.260	64,9	35,1	16	1	1	--	--	18
-LA RIOJA	89.680	164.217	61,7	38,3	5	1	1	--	--	7
-MENDOZA	148.827	1.196.228	68,9	31,1	15	5	1	--	1	22
-MISIONES	29.801	588.977	50,4	49,6	19	5	--	1	--	25
-NEUQUEN	94.078	243.850	76,1	23,9	7	3	1	--	--	11
-RIO NEGRO	203.013	383.354	71,8	28,2	13	8	--	--	--	21
-SALTA	154.775	662.870	71,8	28,2	24	5	--	1	--	30
-SAN JUAN	89.651	465.976	72,0	28,0	9	1	--	1	--	11
-SAN LUIS	76.748	214.416	70,0	30,0	8	--	2	--	--	10
-SANTA CRUZ	243.943	114.941	86,8	13,2	8	2	--	--	--	10
-SANTA FE	133.007	2.465.546	82,0	18,0	82	21	1	1	1	106
-STGO. DEL ESTERO	135.254	594.920	51,9	48,1	14	3	1	--	--	18
-TUCUMAN	22.524	972.655	70,9	29,1	18	6	--	--	1	25
-T. DEL FUEGO	21.263	27.358	88,6	11,4	--	2	--	--	--	2

(*) INCLUIDO: AGLOMERADO GRAN BUENOS AIRES

B I B L I O G R A F I A

- [1] KISH, L. (1965) Survey Sampling (Wiley)
- [2] COCHRAN W. (1977) Sampling Techniques (Wiley)
- [3] HANSEN M., Hurwitz W., Madow W. (1935). Sample Survey Methods and Theory (Wiley).
- [4] Programa para desarrollar la capacidad nacional de efectuar Encuestas de Hogares. Marco de muestreo y diseños muestrales (1987). Naciones Unidas. Oficina de Estadísticas.
- [5] Popstan. Un estudio de caso para los censos de población y habitación de 1980. (Bureau of the Census).
- [6] GRAIS, Bernard (1987). Un plan de rotation de l'échantillon pour l'enquête communautaire sur les forces de travail: les enseignements de l'expérience française. (INSEE - Paris).