

# Determinación del tamaño de muestra necesario para la evaluación de los diferentes niveles de cobertura en la Encuesta de Protección Social

Ramón Álvarez      Juan José Goyeneche      Guillermo Zoppolo  
Instituto de Estadística

E-mail: {ramon, jjgoye, gzoppolo}@iesta.edu.uy

## Resumen

En el marco del convenio realizado entre el BPS, INE y la Facultad de CCEEyA para la evaluación de la EPS (Encuesta de Protección Social), se construye una tipología de hogares para medir diferentes tipos de cobertura en seguridad social, para ser luego evaluada a través de una muestra probabilística.

Se propone un diseño muestral estratificado en 2 fases usando como marco muestral la ENHA 2006 y los 9 estratos creados que se combinan con 5 estratos según quintiles de ingreso.

Para poder determinar un tamaño de muestra adecuado se plantean diferentes escenarios donde se analiza la distribución en el muestro del error de medición asociado a cada una de los tipos de cobertura. En esos escenarios se generan mediante simulación montecarlo, determinada cantidad de muestras (replicaciones) con un tamaño de muestra fija.

**Palabras clave:** Protección Social, Nivel de Cobertura, Precisión, Distribución en el muestreo, Simulación.

# Índice

<b>1. Presentación de Encuesta de Protección Social</b>	<b>3</b>
1.1. Estrategia de relevamiento y formulario . . . . .	4
1.2. Algunas reflexiones sobre el formulario en la primera onda de la EPS . . . . .	5
1.3. Estructura del Formulario . . . . .	6
<b>2. Análisis del Diseño de Muestro</b>	<b>7</b>
2.1. Diseño Muestral . . . . .	7
2.2. Tamaño de la Muestra . . . . .	10
2.3. Escenarios Simulados . . . . .	10
2.4. Simulación Escenario propuesto por el INE . . . . .	12
<b>3. Anexos Estadísticos</b>	<b>14</b>

*Este documento consta de 2 partes: En la primera se presenta una introducción sobre la encuesta, el alcance que se desea que esta tenga, teniendo en cuenta que esta se intentaba implementar a en la primera parte del período pasado de gobierno donde se ponía a su vez en marcha el PANES a cargo del MIDES. En la sección 1 se resume parte del documento que elaboró IECON, con participación del Prof. Ec. Rodrigo Arim, para llevar adelante un diagnóstico de la propuesta de formulario a aplicar y la elaboración de las diferentes categorías de niveles de cobertura en seguridad social. En la segunda parte se presenta el diseño de muestreo propuesto por el IESTA y que surge de la propuesta de formulario y el tipo de análisis de cobertura planteado en 1. A su vez se presenta la metodología llevada adelante por el equipo del IESTA para la determinación del tamaño de muestra.*

## 1. Presentación de Encuesta de Protección Social

El primer país de América del Sur en llevar adelante estudios sobre los diferentes niveles de cobertura a nivel poblacional en seguridad Social con metodología de encuestas es Chile a través de la *Encuesta de Protección Social, Chile, 2002* [http://www.proteccionsocial.cl/\[1\],\[2\]](http://www.proteccionsocial.cl/[1],[2]).

La Encuesta de Protección Social (EPS) constituye una experiencia inédita en Uruguay desde dos puntos de vista, si el BPS pudiese implementarla .

En primer lugar, no existen antecedentes de relevamientos periódicos cuyo objetivo central es el análisis y caracterización de la cobertura del sistema protección social en el país. El principal instrumento de relevamiento socio-económico, la Encuesta continua de Hogares (ECH) realizada por el INE, si bien releva información valiosa sobre las principales variables de interés en esta área, no permite distinguir con precisión la cobertura y potenciales efectos de los diversos programas.<sup>1</sup>

Otro antecedente importante es la *Encuesta de Caracterización Social* financiada por el Banco Mundial. Este constituye un relevamiento ad-hoc, que, si bien recoge información más específica sobre los efectos de los programas que

---

<sup>1</sup>La Encuesta Nacional Ampliada de Hogares 2006 mejoró sustancialmente esta situación, en tanto incorpora en su grilla preguntas tendientes a identificar mejor a los beneficiarios de las políticas sociales.

integran la red de protección social, su finalidad central fue estudiar el papel de la red de protección social durante la crisis económica que sufrió el país en 2002. La EPS tiene su foco en el estudio de los programas sociales, en particular los administrados por el BPS, y pretende realizarse en forma periódica a partir del presente año, brindando información continua sobre la evolución de la cobertura, su vinculación con las principales características sociales de los beneficiarios y los vacíos de cobertura que potencialmente puedan persistir.

En segundo lugar, la EPS es la primera encuesta a personas y hogares que adquiriría una estructura longitudinal. Hasta el momento, salvo algunas encuestas realizadas por programas sociales específicos o por instituciones académicas – por ejemplo, Infamilia – cuyo universo de estudio es acotado, Uruguay no tiene experiencias en relevamientos realizados bajo un esquema de Encuestas de Panel. Esto es de por sí un aporte relevante de la EPS y un desafío en cuanto a su diseño. El carácter longitudinal permitirá estudiar la interacción dinámica entre los diversos programas que constituyen la red de protección social, los procesos de 'entradas' y 'salidas' del sistema y de sus diversos programas componentes, diseñar estrategias para identificar los cambios comportamentales inducidos por las políticas, su papel como mecanismos de contención frente a contingencias negativas que enfrentan los individuos, debilidades y fortalezas en su diseño actual, etc. En suma, la EPS constituye una innovación importante en las estadísticas del área socio-económica, habilitando un análisis más pormenorizado y completo de la realidad del sistema de protección social.

El formulario para el que se trabajó en el diseño muestral se toma a partir del usado en Chile (2002,2004) [1], cuya reformulación para Uruguay es el producto de un largo proceso, llevado adelante por cuadros técnicos del BPS encargados de diseñar y rediseñar su contenido, en forma conjunta con investigadores del IECON, IESTA y del INE.

### **1.1. Estrategia de relevamiento y formulario**

Luego de discutidas diversas alternativas, se decide que la estrategia de muestreo seleccionada sea la que considere como marco muestral el que se encuentra definido por los hogares que están incorporados en la Encuesta Nacional de Hogares Ampliada 2006. Obviamente esta opción, la que es evaluada como la más eficiente del punto de vista del muestreo, tiene implicaciones importantes en lo referente al rediseño del formulario.

## 1.2. Algunas reflexiones sobre el formulario en la primera onda de la EPS

Las características del sistema de información sobre seguridad social en Uruguay determinan que la EPS cuente con una impronta que la diferencian de otros instrumentos similares utilizados en otros países.

En primer lugar, el conjunto de políticas sobre las que se pretende relevar información que permita aplicar procesos de inferencia estadística es muy amplio, y abarca poblaciones diversas. El BPS es la institución de seguridad social encargado de abonar la mayor parte de las transferencias contributivas<sup>2</sup> y no contributivas.<sup>3</sup> Esto obliga a una estrategia de diseño muestral relativamente compleja, pero también a una longitud comparativamente mayor del formulario. Es posible que en pos de alcanzar una correcta representatividad de todos los programas y asegurar la calidad de las respuestas se necesite renunciar a una mayor precisión que en el caso de que el objetivo de la EPS fuera más acotado.

En segundo lugar, el BPS administra los registros administrativos de Historia Laboral, por lo que uno de los objetivos que usualmente se persiguen con este tipo de relevamiento – estudiar la densidad de cotizaciones en el tiempo – se encuentra ya cubierto por estos registros. El desafío es asegurar el apareo de la información que proporcionará la EPS con la información disponible en los registros administrativos. Esta situación permite reducir la participación de preguntas tendientes a captar las entradas y salidas a la cotización del régimen de seguridad social.

Uno de los objetivos que debería tener esta primera onda de la EPS es construir una 'línea de base' lo más precisa posible. Esto implica caracterizar a las personas/hogares que integrarán el panel en lo referente a su participación el sistema de protección social y en su trayectoria laboral, de forma de que la información dinámica que aporten las sucesivas olas cuente con un 'punto

---

<sup>2</sup>La excepción son las cajas paraestatales.

<sup>3</sup>Todas transferencias no contributivas monetarias son abonadas por el BPS. Las transferencias no contributivas en especie en general son provistas por otras instituciones, aunque el BPS controla un conjunto no despreciable de esta modalidad de transferencias.

de partida' informacional lo más sólido posible. Nuevamente, si se acepta este criterio, esto debería conducir a priorizar algunos bloques del formulario sobre otros.

### 1.3. Estructura del Formulario

Se parte de la estructura del formulario que se usó en la primera y segunda onda de la EPS, Chile en En total, la que constaba de 335 preguntas con una estructura modular como se detalla a continuación el formulario planteado para la encuesta piloto, el que luego necesariamente debería ser modificado en función de los resultados preliminares.

1. CARACTERÍSTICAS DE LA VIVIENDA Y EL HOGAR .Se incluye un conjunto de preguntas que apuntan a ver si el hogar se encuentra inscripto en el Plan Nacional de Emergencia Social. La modalidad a partir de la cual se genera el beneficio del PANES (relevamiento de información socio-económica en el hogar) puede inducir a pensar a los entrevistados beneficiarios del PANES que esta entrevista es un control realizado por el MIDES.
2. CARACTERÍSTICAS GENERALES DEL ENTREVISTADO (Se incorporan preguntas de educación y salud)
3. HISTORIA INDIVIDUAL Este es un módulo central de la EPS. Permite reconstruir la historia laboral del individuo y constituye una base imprescindible para su seguimiento en el tiempo.
4. CONDICIÓN DE ACTIVIDAD DEL ENTREVISTADO (Se incorporan módulos de *búsqueda de otro empleo* y *expectativas de reinserción laboral* como capítulos específicos dentro de este módulo)
5. SISTEMA PREVISIONAL (Se encuentra compuesto de varios sub-módulos divididos por prestación, módulo de atención materno infantil, módulo de asignaciones familiares, incluyendo preguntas sobre el conocimiento del mismo por parte del entrevistado).
6. MODULO DE PRESTACIONES IVS
7. INGRESOS
8. DATOS PARA UBICACIÓN DEL ENTREVISTADO EN EL FUTURO

## 2. Análisis del Diseño de Muestro

En esta sección se presenta en forma resumida la metodología de trabajo que involucra la construcción de la tipología de hogares que se usan luego en el diseño muestral y algunos resultados preliminares con respecto a la precisión y el tamaño de muestra considerado. Hay que recordar que la tipología de hogares responde a la necesidad de poder tener claramente identificadas las unidades de muestreo donde se pretende medir diferentes tipos de cobertura.

### 2.1. Diseño Muestral

El diseño muestral considerado es estratificado en 2 fases[3]. La primera fase viene dada por la muestra de la Encuesta de Hogares Ampliada (ENHA)[4]. Para la segunda fase, se construyen *estratos* de acuerdo a una *tipología de hogares* y a los *quintiles del ingreso per cápita del hogar*, lo que lleva a tener en este caso 45 estratos (9 que corresponden a la tipología y 5 del nivel de ingreso). La asignación del tamaño de muestra por estrato es proporcional al tamaño del estrato. En cada estrato el diseño es aleatorio simple de hogares y se relevan a todas las personas de los hogares seleccionados.

La tipología de hogares construída busca contemplar los diferentes niveles de cobertura la que queda en principio como

Estrato	Tipo de Hogar
1	unipersonal
2	pareja sola
3	biparental
4	monoparental nuclear
5	extendido biparental
6	extendido monoparental
7	compuesto

Tabla 1: Tipología de hogares

A su vez para estos tipos de hogares se toma en cuenta la edad del jefe del hogar haciendo una nueva clasificación

Se construyen luego las *variables objetivos* para las que se puede calcular los *totales* en el marco muestral. Esos totales pueden ser estimados a partir de

Tipo Hogar Original	Edad	Nuevo tipo Hogar
1=unipersonal	edadhog < 65	1
1=unipersonal	edadhog ≥ 65	2
2=pareja sola	edadhog < 65	3
2=pareja sola	edadhog ≥ 65	4
3=biparental		5
4=monoparental nuclear		6
5=extendido biparental		7
6=extendido monoparental		8
7=compuesto;		9

Tabla 2: Definición de Estratos

muestras extraídas con algún diseño muestral dado. Interesa por lo tanto ver como es la distribución en el muestreo de los estimadores propuestos, para los cuales se podrá analizar los errores asociados, teniendo en cuenta que para los resultados que se presentan no se considera la distribución espacial de los hogares muestreados en todo el territorio.

Las variables objetivos tiene que ver con el tipo de cobertura que cada persona puede recibir y que consiste en

- ASIGNACION FAMILIAR (AS)
- APORTES (AP)
- JUBILADO (JU)
- PENSION a la VEJEZ (PV)
- PENSION de SOBREVIVENCIA (PS)

Teniendo en cuenta esos tipos de cobertura a nivel de las personas, las edades y el sexo de las personas la clasificación queda como



Variable	Cobertura	Sexo	Tramo de Edad
apf1829	APORTES	f	[18 a 29]
apf3045	APORTES	f	[30 a 45]
apf4664	APORTES	f	[46 a 64]
apm1829	APORTES	m	[18 a 29]
apm3045	APORTES	m	[30 a 45]
apm4664	APORTES	m	[46 a 64]
asf05	ASIGNACION FAMILIAR	f	[0 a 5]
asf611	ASIGNACION FAMILIAR	f	[6 a 11]
asf1217	ASIGNACION FAMILIAR	f	[12 a 17]
asm05	ASIGNACION FAMILIAR	m	[0 a 5]
asm611	ASIGNACION FAMILIAR	m	[6 a 11]
asm1217	ASIGNACION FAMILIAR	m	[12 a 17]
juf	JUBILACION	f	
jum	JUBILACION	m	
psf	PENSION SUPERVIVENCIA	f	
psm	PENSION SUPERVIVENCIA	m	
pvf	PENSION VEJEZ	f	
pvm	PENSION VEJEZ	m	

Tabla 3: Clasificación del tipo de cobertura por persona

## 2.2. Tamaño de la Muestra

Para poder determinar un tamaño de muestra adecuado se determinan diferentes escenarios donde se puede ver cual es la distribución en el muestreo del error de medición asociado a cada una de los tipos de cobertura. En esos escenarios se generan mediante simulación montecarlo determinada cantidad de muestras (replicaciones) con un tamaño de muestra fija ( $Tm$ ), donde se calculan totales para cada 'unidad tipo'. Por lo tanto es importante ver cual es la sensibilidad de la precisión en función de esos 2 parámetros.

## 2.3. Escenarios Simulados

Los resultados que se presentan surgen de los modelos de simulación que se implementaron con el software [5], en los que se aplican para las replicaciones la subrutina *PROC SURVEYSELECT*. En la tabla 4 se ven las diferentes combinaciones entre el número de replicaciones y el tamaño de muestra (la cantidad de hogares)

<b>Escenario1</b>	500 muestras de tamaño 5000 hogares con 5 Estratos por quintil de ingresos
<b>Escenario2</b>	500 muestras de tamaño 5000 hogares con 5 Estratos por quintil de ingresos y 8 tipos de hogar
<b>Escenario3</b>	500 muestras de tamaño 5000 hogares con 5 Estratos por quintil de ingresos y 9 tipos de hogar
<b>Escenario4</b>	500 muestras de tamaño 8000 hogares con 5 Estratos por quintil de ingresos y 9 tipos de hogar
<b>Escenario5</b>	500 muestras de tamaño 10000 hogares con 5 Estratos por quintil de ingresos y 9 tipos de hogar

Tabla 4: Definición de Escenarios

Se ve en la tabla 5 que se logra bajar el error cambiando en forma importante el tamaño de muestra de cada replicación. Los cálculos están hechos en base a los ES que se reportan en las tablas de los anexos estadísticos 3, donde aparecen los extremos de los intervalos de confianza calculados para un *95 % de confianza*.

V. Objetivo	Error Relativo				
	Escenario 1	Escenario 2	Escenario 3	Escenario 4	Escenario 5
apf1829	13	13	12	10	9
apf3045	8	8	8	6	5
apf4664	9	9	9	7	6
apm1829	11	11	11	8	7
apm3045	7	7	7	5	5
apm4664	8	8	7	5	5
asf05	12	12	12	9	8
asf1217	12	12	12	9	8
asf611	11	11	10	9	8
asm05	12	12	12	10	8
asm1217	11	11	11	9	8
asm611	11	10	10	8	7
juf	7	7	7	5	5
jum	7	6	6	5	4
psf	8	7	7	5	4
psm	37	38	36	31	27
pvf	19	19	20	16	13
pvm	21	21	21	17	14

Tabla 5: Error relativo (en porcentajes) por tipo de escenario

## 2.4. Simulación Escenario propuesto por el INE

A pedido del INE se elabora un escenario donde se hace una reformulación de la tipología de hogares

1. tipología de hogares
  - a) Unipersonal
  - b) Pareja Sola
  - c) Nuclear
  - d) Monoparental
  - e) Extendido compuesto
  
2. 5 tramos de ingresos (quintiles)

Con esta nueva clasificación de unidades a muestrear se hicieron 5000 repeticiones de muestras de tamaño 7000 (son 7 tipos de Hogar con 5 tramos de ingresos cada uno de ellos)

Para tomar una decisión acerca de la mejor estrategia queda pendiente ver como impacta en el costo la distribución espacial de las muestras, pudiéndose considerar en una fase previa estratos *geográficos*, cuyos costos debieran ser establecidos por el INE, de quien dependería el trabajo de campo. Se podrían ver de esta manera cuales son los *tipos de cobertura* que tienen estimaciones con precisiones mas pequeñas de lo deseado y llegar a una solución de compromiso que haga mas eficiente el diseño y controlable el trabajo de relevamiento y su posterior seguimiento entre ondas de la encuesta.

<b>V. Objetivo</b>	$t_0$	$SD_t$	$Inf$	$Sup$	$\%error$
apf1829	75971	4070.6	67830	84113	0.11
apf3045	152079	5234.2	141611	162547	0.07
apf4664	128996	4790.4	119415	138577	0.07
apm1829	107894	4736.4	98421	117367	0.09
apm3045	189946	5505.6	178934	200957	0.06
apm4664	169201	5160.5	158880	179522	0.06
asf05	84150	4270.0	75610	92690	0.10
asf1217	94844	4543.7	85757	103931	0.10
asf611	103135	4768.5	93598	112672	0.09
asm05	86626	4240.2	78146	95107	0.10
asm1217	93081	4554.4	83973	102190	0.10
asm611	103816	4604.4	94607	113025	0.09
juf	190743	5353.5	180036	201450	0.06
jum	199200	5285.2	188629	209770	0.05
psf	198847	5873.1	187101	210593	0.06
psm	7555	1226.0	5103	10007	0.32
pvf	28196	2259.3	23677	32714	0.16
pvm	22011	1916.5	18178	25844	0.17

Tabla 6: Escenario propuesto por el INE

### 3. Anexos Estadísticos

V. Objetivo	$t_0$	$SD_t$	$Inf$	$Sup$	%error
apf1829	75939	4826	66286	85592	0.13
apf3045	152600	6274	140053	165147	0.08
apf4664	129081	5854	117374	140789	0.09
apm1829	107899	5827	96245	119553	0.11
apm3045	190501	7059	176384	204618	0.07
apm4664	169401	6505	156392	182410	0.08
asf05	84778	4923	74931	94625	0.12
asf1217	94787	5638	83511	106064	0.12
asf611	102908	5498	91913	113904	0.11
asm05	86673	5268	76138	97208	0.12
asm1217	93140	5008	83123	103156	0.11
asm611	103779	5529	92721	114836	0.11
juf	190624	6673	177278	203970	0.07
jum	198979	7265	184450	213508	0.07
psf	198880	7461	183959	213802	0.08
psm	7684	1431	4823	10546	0.37
pvf	28101	2637	22826	33375	0.19
pvm	21936	2296	17344	26528	0.21

Tabla 7: Escenario 1

<b>V. Objetivo</b>	$t_0$	$SD_t$	$Inf$	$Sup$	$\%error$
apf1829	76116	4801	66513	85719	0.13
apf3045	151899	6209	139480	164318	0.08
apf4664	129106	5578	117951	140261	0.09
apm1829	107699	5732	96234	119164	0.11
apm3045	190164	6586	176991	203337	0.07
apm4664	168915	6345	156225	181605	0.08
asf05	84296	4903	74490	94102	0.12
asf1217	94592	5664	83265	105919	0.12
asf611	103309	5623	92063	114555	0.11
asm05	86472	5027	76418	96526	0.12
asm1217	92891	5291	82309	103472	0.11
asm611	103936	5048	93839	114033	0.10
juf	190542	6549	177445	203640	0.07
jum	199274	6444	186386	212161	0.06
psf	198815	6945	184925	212705	0.07
psm	7502	1429	4643	10360	0.38
pvf	28023	2716	22590	33456	0.19
pvm	21947	2285	17378	26516	0.21

Tabla 8: Escenario2

<b>V. Objetivo</b>	$t_0$	$SD_t$	$Inf$	$Sup$	$\%error$
apf1829	76041	4641	66758	85324	0.12
apf3045	151865	6382	139101	164629	0.08
apf4664	129061	5841	117380	140743	0.09
apm1829	107643	5698	96248	119038	0.11
apm3045	189909	6593	176724	203094	0.07
apm4664	169308	5980	157348	181268	0.07
asf05	84569	5134	74301	94837	0.12
asf1217	95150	5498	84153	106147	0.12
asf611	103273	5410	92454	114092	0.10
asm05	86737	5054	76629	96844	0.12
asm1217	92872	5181	82511	103234	0.11
asm611	103943	5396	93151	114736	0.10
juf	190718	6398	177922	203514	0.07
jum	198912	6323	186267	211558	0.06
psf	198257	6989	184279	212234	0.07
psm	7570	1360	4850	10291	0.36
pvf	28302	2792	22719	33886	0.20
pvm	21985	2320	17345	26624	0.21

Tabla 9: Escenario 3



<b>V. Objetivo</b>	$t_0$	$SD_t$	$Inf$	$Sup$	$\%error$
apf1829	76202	3741	68720	83684	0.10
apf3045	152262	4869	142524	162001	0.06
apf4664	129060	4430	120200	137920	0.07
apm1829	107820	4416	98988	116652	0.08
apm3045	190301	5077	180147	200455	0.05
apm4664	169071	4435	160201	177941	0.05
asf05	84323	3917	76488	92157	0.09
asf1217	94971	4230	86511	103431	0.09
asf611	102785	4664	93457	112113	0.09
asm05	86341	4163	78015	94666	0.10
asm1217	92872	3980	84912	100832	0.09
asm611	103774	4355	95063	112484	0.08
juf	191084	5200	180683	201484	0.05
jum	199201	5113	188975	209427	0.05
psf	198861	5435	187991	209731	0.05
psm	7557	1168	5222	9892	0.31
pvf	28191	2221	23749	32633	0.16
pvm	22074	1830	18414	25734	0.17

Tabla 10: Escenario 4

## Índice de tablas

1.	Tipología de hogares . . . . .	7
2.	Definición de Estratos . . . . .	8
3.	Clasificación del tipo de cobertura por persona . . . . .	9
4.	Definición de Escenarios . . . . .	10
5.	Error relativo (en porcentajes) por tipo de escenario . . . . .	11
6.	Escenario propuesto por el INE . . . . .	13
7.	Escenario 1 . . . . .	14
8.	Escenario2 . . . . .	15
9.	Escenario 3 . . . . .	16
10.	Escenario 4 . . . . .	17

## Referencias Bibliográficas

- [1] Cuestionarios para Encuesta de Protección Social [página web]. Chile: Encuesta de Protección Social; c2000-01. Available from: <http://www.proteccionsocial.cl/> [updated 2002 May 16; cited 2002 Jul 9].
- [2] Metodología y diseño muestral para Encuesta de Protección Social [página web]. Chile: Encuesta de Protección Social; c2000-01. Available from: <http://www.proteccionsocial.cl/> [updated 2010 Set 10; cited 2010 Set 9].
- [3] Särndal CE, Swensson B, Wretman J. *Model Assisted Survey Sampling*. Springer-Verlag; 1992.
- [4] Metodología de la Encuesta Nacional de Hogares Ampliada, 2006 [página web]. Uruguay; 2006. Available from: <http://www.ine.gub.uy/biblioteca/ech/ech2006/HOGARES%202006.pdf>.
- [5] SAS/Stat 9.2 User's Guide: Survey Data Analysis (Book Excerpt). Sas Publishing; 2009.