

APÉNDICE A

DISEÑO Y COBERTURA DE LA MUESTRA

APÉNDICE A

DISEÑO Y COBERTURA DE LA MUESTRA

A.1 Introducción

La muestra utilizada para la encuesta **ENDES-96** es una muestra probabilística que fue seleccionada en tres etapas: la selección de centros poblados en la primera etapa, la selección de conglomerados en la segunda etapa, y la selección de viviendas en la etapa final. La muestra estuvo constituida por 1496 conglomerados, distribuidos proporcionalmente en el área urbana y rural de cada departamento, a excepción de los departamentos de Cajamarca, Cusco, Lima y Piura. La muestra total permite para las mujeres entre 15-49 años de edad sólidas estimaciones de un número de variables importantes a nivel de cada uno de los 24 departamentos y las tres regiones naturales del país (Costa, Sierra y Selva). También 1/12 de la muestra total permite estimaciones para la anemia en mujeres y para un número de variables de los hombres elegibles a nivel urbano, rural y total. De otro lado, aunque la muestra total dentro de cada departamento puede ser auto-ponderada, dicha muestra total no necesariamente es autoponderada a niveles de agregaciones mayores tal como regiones, zonas naturales, total urbano, total rural o de todo el país.

De la muestra total, fueron seleccionadas 33,498 direcciones de hogares para obtener 28,951 entrevistas completas de mujeres de 15 a 49 años de edad. Para la muestra total, el número de hogares visitados fue de 28,805, de los cuales sólo 28,122 fueron entrevistados. En los hogares entrevistados se encontró un total de 31,241 mujeres elegibles y de las cuales se obtuvo un total de 28,951 entrevistas completas. En la submuestra de 1/12 del total, el número de hogares visitados fue de 2,516 y de los cuales 2,473 fueron entrevistados. En los hogares entrevistados de la submuestra, se localizó a 2,818 mujeres elegibles y 2,942 hombres elegibles, de esas cifras se obtuvo 2,274 entrevistas completas para el estudio de la anemia y 2,487 entrevistas completas de hombres elegibles.

El universo de la muestra estudiada representa aproximadamente el 100 por ciento de la población total del país. Ninguna área fue deliberadamente omitida del marco muestral.

A.2 Diseño y Selección

Unidades de Muestreo

Las Unidades Primarias de Muestreo (**UPM**) fueron los centros poblados empadronados en el último censo de población del año de 1993. Cada **UPM** fue dividida en conglomerados (**USM**), y todo domicilio localizado en dicha área se considera como la unidad última de muestreo (**UUM**).

Estratificación

Además de la estratificación por departamento, en cada uno de ellos los centros poblados fueron clasificados en tres grupos: sedes urbanas, (20 mil o más habitantes) resto urbano (2 mil a menos de 20 mil) y rural (localidades con menos de 2 mil habitantes). Asimismo, las sedes urbanas con 100 mil o más habitantes en dos categorías: áreas periféricas y resto de la ciudad.

Marco Muestral

El Instituto Nacional de Estadística e Informática (INEI) realizó el último censo de población y de vivienda en 1993 y mantiene un listado computarizado de los centros poblados con información censal, lo cual permite el cálculo de los parámetros para el diseño de cualquier muestra utilizando dicho marco de unidades.

El archivo computarizado de centros poblados dentro de cada departamento ha sido organizado de tal manera que refleje la agrupación de ellos en sedes urbanas, en el resto urbano y en el área rural.

Como consecuencia se decidió usar este marco muestral para la encuesta **ENDES-96**. La razón principal es que es factible una subdivisión en conglomerados adecuada para los objetivos de la encuesta **ENDES-96**. La segunda razón, es la de hacer posible la implementación de los listados actualizados de viviendas en cada conglomerado seleccionado.

Composición de la muestra

Con el objetivo inicial de obtener un total aproximado de 28,000 entrevistas completas de mujeres elegibles se utilizó información de la encuesta **ENDES II, 1991-92**, como un instrumento de ayuda para estimar ciertos parámetros del diseño. Se estimó que se haría necesario un total aproximado de 33,000 viviendas a ser seleccionadas para obtener ese número de entrevistas completas. Ese total fue distribuido por departamentos pero no proporcionalmente a su tamaño poblacional sino para poder cumplir con los objetivos requeridos de la encuesta, tal como se puede observar en el Cuadro A.1.

Selección de la muestra

La muestra resultante para la **ENDES III** se obtuvo en tres etapas: la primera etapa es una selección sistemática de centros poblados con probabilidad proporcional a su tamaño. Después de la primera etapa de selección se implementó una actualización del levantamiento cartográfico existente y de sus viviendas en las áreas de nueva construcción que se habían generado después del último censo de población. Inmediatamente después de esta actualización cartográfica se dividió el centro poblado en conglomerados (de un tamaño promedio apropiado, más o menos alrededor de 100 viviendas cada uno). En los conglomerados seleccionados se realizó un levantamiento exhaustivo de viviendas. Finalmente fue seleccionado un número de viviendas en cada conglomerado, de tal manera de mantener una fracción muestral uniforme para cada vivienda (e implícitamente para cada hogar) dentro de cada dominio. Esta fracción no es la misma para cada dominio desde que la asignación del tamaño de muestra no fue proporcional a través de esos dominios de estudio.

Para el proceso de selección de los centros poblados se siguieron los siguientes pasos:

- El número de conglomerados a ser seleccionados en cada departamento fue calculado como la división del tamaño de muestra esperado en el departamento y del tamaño promedio de muestra esperado dentro del conglomerado.
- El total de muestra requerido en cada departamento fue distribuido entre sus subestratos (sede, resto urbano y rural) de acuerdo a su población censal, y también dentro de cada subestrato, sus centros poblados fueron ordenados geográficamente en una forma geográfica serpentina.
- Para cada subestrato, se elaboró un listado de sus centros poblados con sus correspondientes totales de población (viviendas) y también de sus correspondientes acumulados parciales de población.

Cuadro A.1. Número de conglomerados urbanos y rurales por departamentos

Número de conglomerados urbanos y rurales por departamento, Perú 1996

Departamento	Número esperado de entrevistas	Número de conglomerados				
		Sede	Resto urbano	Rural	Total	Periferia
Grau						
Tumbes	1000	26	18	6	50	0
Piura	1400	6	20	44	70	3
Grupo					50	
Resto					20	
Nor Oriental del Marañón						
Cajamarca	1400	3	9	58	70	2
Grupo					50	
Resto					20	
Amazonas	1000	6	10	34	50	0
Lambayeque	1000	24	16	10	50	12
San Martín						
San Martín	1000	18	10	22	50	0
La Libertad						
La Libertad	1040	24	12	16	52	7
Chavín						
Ancash	1040	23	8	21	52	8
Lima						
Lima	3820	228	12	8	248	80
Grupo					50	
Resto					198	
Los Libertadores						
Ica	1000	16	24	10	50	3
Ayacucho	1200	16	8	36	60	4
Huancavelica	1000	7	3	40	50	0
Huánuco	1000	12	7	31	50	4
Arequipa						
Arequipa	1120	40	8	8	56	14
José Carlos Mariategui						
Puno	1200	16	8	36	60	2
Moquegua	1000	16	24	10	50	0
Tacna	1000	38	6	6	50	10
Inka						
Cusco	1400	10	10	50	70	4
Grupo					50	
Resto					20	
Madre de Dios	1000	27	3	20	50	0
Apurímac	1000	8	8	34	50	0
Andrés A. Cáceres						
Junín	1000	14	16	20	50	2
Pasco	1000	16	12	22	50	0
Ucayali						
Ucayali	1000	26	8	16	50	0
Loreto						
Loreto	1160	26	10	22	58	0
Total	28780	646	270	580	1496	155

- . En cada subestrato, se realizó una selección (sistemática y con probabilidad proporcional a su tamaño poblacional) de un número de centros poblados requeridos para esta encuesta, digamos **a**, en el total de ellos en dicho subestrato.

- La probabilidad de selección del i-ésimo centro poblado para la **ENDES III** es:

$$P_{1i} = (a * m_i) / (\sum m_i)$$

donde

- a : es el número de centros poblados a ser seleccionados en el subestrato.
- m_i : es la población total de i-ésimo centro poblado de acuerdo al censo de 1993,
- $\sum m_i$: es la población total del subestrato según el censo de 1993.

Previo a la siguiente etapa de selección, cada centro poblado seleccionado (previa actualización cartográfica) es dividido en áreas compactas denominadas conglomerados.

- En cada centro poblado, una selección de conglomerados (en general uno por cada centro poblado) se realizó en una forma similar a la selección de los centros poblados .
- La probabilidad final del j-ésimo conglomerado para la encuesta ENDES III está dada como:

$$P_{1i} = (a * m_i) / (\sum m_i) (k_j / K_i)$$

donde

- k_j : es el número de conglomerados seleccionado en el i-ésimo centro poblado,
- K_i : es el número total de conglomerados en el i-ésimo centro poblado.

Después de obtener las listas actualizadas de viviendas para cada conglomerado seleccionado, se determinó el número final de viviendas de acuerdo a las siguientes fórmulas:

$$P_{2ij} = (n_{ij} / L_{ij})$$

$$f = P_{1i} * P_{2ij}$$

y de las cuales se obtiene

$$n_{ij} = (f * L_{ij}) / (P_{1i})$$

donde

- n_{ij} : es el número de viviendas seleccionado en el j-ésimo conglomerado del i-ésimo centro poblado,
- L_{ij} : es el número total de viviendas listadas en el j-ésimo conglomerado del i-ésimo centro poblado durante 1996,
- f : es la fracción de muestreo dentro del departamento.
- P_{1i} : es la probabilidad de selección del conglomerado para la encuesta ENDES III, y cuyo valor es calculado en base de los parámetros del diseño de la muestra.

Una vez establecido el número total de viviendas a ser seleccionado, el procedimiento de selección de ellos depende si el conglomerado es urbano o rural.

Para cada conglomerado urbano, la selección debe ser en forma sistemática con un intervalo de selección de $I_{ij} = L_{ij} / n_{ij}$. La primera selección es hecha al azar dentro del rango del intervalo de selección.

Para cada conglomerado rural, la selección se hace en grupos compactos de ocho hogares contiguos sobre el listado de hogares de acuerdo al siguiente procedimiento:

- se enumeran las viviendas en forma consecutiva en el listado del conglomerado rural.
- después de cada múltiplo de ocho (es decir 8, 16, 24, etc) se traza una línea horizontal (con un marcador de color). Cada grupo está formado por las viviendas entre cada dos líneas horizontales consecutivas y el primer grupo está constituido por la primera vivienda en la lista hasta el anterior inmediato a la primera línea horizontal, y en el caso de existir más de 4 viviendas después de la última línea horizontal, el último grupo estará constituido por esas viviendas después de la última línea horizontal. Todos estos grupos así formados se denominan grupos compactos; sea G_{ij} ese total de ellos.
- si hubieran viviendas (menos de cuatro) después de la última línea horizontal, ellas serán distribuidas uniformemente entre los grupos compactos anteriores en forma sistemática y con un inicio al azar,
- se calcula el número de grupo compactos a seleccionar como

$$g_{ij} = \text{redondeo al entero más cercano de } (n_{ij}/8)$$

- se selecciona entonces g_{ij} grupos compactos del total G_{ij} en una forma sistemática y con un comienzo al azar.
- todas las viviendas dentro de los grupos compactos en el conglomerado rural será considerado como la muestra de viviendas de dicho conglomerado.

Un caso muy particular, en el diseño para la **ENDES III** fue la inclusión de los conglomerados rurales usados en la **ENDES II** en el departamento de Ancash.

Rendimiento de la muestra

En los cuadros siguientes se presentan las tasas de respuesta por departamento y por zona de residencia para la encuesta de hogares y para la encuesta de mujeres. De un total de 1,496 conglomerados seleccionados, se logró implementar la encuesta en 1,492 de ellos.

De los 33,498 hogares que fueron seleccionados, el 9.7 por ciento de las viviendas estaban desocupadas, destruidas o en otra condición. En los restantes (hogares residenciales) se obtuvieron 28,122 entrevistas completas, con una tasa de respuesta del 98 por ciento entre los hogares residenciales.

De un total de 31,241 mujeres elegibles se logró entrevistar 28,951, obteniéndose una tasa de respuesta individual del 93 por ciento. La tasa de respuesta combinada de hogares y mujeres fue del 91 por ciento, la cual es bastante satisfactoria para este tipo de encuesta. La mejor tasa de cobertura se logró en el departamento de San Martín con casi un 97 por ciento y la más baja en el departamento de Arequipa con casi un 80 por ciento.

Si bien no se detalla en el Cuadro A.2, de los 2,818 hogares que fueron seleccionados para la encuesta de hombres, el 7.5 por ciento de las viviendas estaban desocupadas, destruidas o en otra condición. En los restantes (hogares residenciales) se obtuvieron 2,474 entrevistas completas, con una tasa de respuesta del 98 por ciento entre los hogares residenciales. De un total de 2,942 hombres elegibles se logró entrevistar 2,487 obteniéndose una tasa de respuesta individual del 85 por ciento. La tasa de respuesta combinada de hogares y hombres en la submuestra fue del 83 por ciento, la cual es también bastante satisfactoria.

Cuadro A.2 Implementación de la muestra

Distribución porcentual de hogares y de mujeres elegibles en la muestra por resultado de la entrevista y tasas totales de respuesta, por departamento y área de residencia, Perú 1996

Resultado	Ama- zonas	Apu- canch	Are- quipa	Aya- cucho	Caja- marca	Huanca- cusco	Huá- velica	Huá- nuco	Ica	Junín	La Li- bertad	Lamba- yeque	Lima	Loreto	Madre		Piura	Puno	San		Uca- yali	Área Urbana	Área Rural	Total			
															Dios	Mo- quegua			Pasco	Marín					Tacna	Tumbes	
Hogares																											
Entrevista completa	84.9	82.8	84.6	83.8	77.4	79.9	86.7	81.0	83.2	84.7	91.2	84.5	92.0	85.8	88.5	73.1	77.6	85.8	84.9	76.0	90.9	81.4	87.4	81.8	86.7	80.3	84.0
Sin persona calificada	1.4	0.5	3.7	1.6	1.3	0.5	2.0	0.9	0.3	0.1	0.3	1.0	0.5	0.7	1.4	1.9	0.3	0.3	1.1	0.2	0.1	0.3	0.2	0.6	0.6	1.2	0.9
Pospuesta	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	
Rechazo	0.0	0.2	0.1	0.7	0.3	0.7	0.0	0.3	0.0	0.2	0.2	0.1	0.4	1.8	0.3	0.6	0.9	0.2	0.2	0.1	0.0	0.2	0.1	0.1	0.7	0.1	0.5
Vivienda no encontrada	0.4	0.3	0.5	1.0	1.7	1.3	0.3	0.3	0.4	0.5	0.0	0.2	0.0	0.8	0.1	2.3	0.1	1.9	0.7	0.4	0.2	0.4	0.2	0.8	0.5	0.9	0.6
Ausente	5.3	3.4	3.8	6.8	9.6	5.0	3.9	3.3	3.4	1.5	3.3	5.6	2.8	2.1	3.4	13.4	3.3	2.9	3.7	7.7	2.6	1.7	1.8	7.1	2.7	6.2	4.2
Vivienda desocupada	7.4	12.1	6.4	5.6	7.2	11.1	6.3	13.1	11.2	12.1	4.4	7.7	4.1	8.1	5.8	7.5	17.1	7.9	8.4	14.9	6.0	15.9	9.7	8.2	8.2	9.9	8.9
Vivienda destruida	0.3	0.2	0.2	0.0	0.5	0.3	0.4	0.8	0.3	0.4	0.2	0.4	0.2	0.4	0.3	0.3	0.0	0.4	0.1	0.3	0.2	0.0	0.2	0.8	0.2	0.5	0.3
Otra respuesta	0.3	0.6	0.7	0.2	1.9	1.2	0.5	0.3	1.1	0.5	0.3	0.5	0.0	0.3	0.2	0.8	0.7	0.7	0.9	0.4	0.0	0.2	0.3	0.5	0.2	0.9	0.5
Total Porcentaje	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Número	1,008	1,058	1,223	1,349	1,273	1,495	1,723	1,181	1,169	1,298	1,148	1,235	958	4,679	1,493	1,156	1,071	1,119	1,824	1,381	1,144	1,191	1,202	1,120	18,920	14,578	33,498
Tasa de respuesta	97.9	98.9	95.1	95.9	95.8	96.9	97.5	98.3	99.1	99.1	99.4	98.5	99.0	96.2	98.0	93.7	98.2	97.3	97.7	99.1	99.7	99.0	99.4	98.2	97.9	97.3	97.6
Mujeres																											
Entrevista completa	91.9	95.3	94.3	83.5	89.6	90.5	92.9	96.0	92.5	94.7	93.5	92.4	90.4	91.2	94.6	87.6	94.8	96.4	90.3	91.9	97.6	96.9	98.0	95.1	92.7	92.6	92.7
Mujer ausente	6.1	2.7	5.0	11.9	7.6	6.2	5.2	2.3	5.5	3.2	4.3	6.4	6.3	5.2	3.0	9.4	3.0	2.2	7.2	5.1	2.0	1.8	1.2	3.6	4.7	5.5	4.9
Entrevista pospuesta	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.1	0.0	0.2	0.0	0.0	0.1	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
Rechazo	0.2	0.6	0.2	1.3	0.4	1.2	0.7	1.1	1.1	0.8	1.0	0.2	1.5	2.2	1.1	1.4	1.6	0.6	0.6	1.0	0.1	0.8	0.2	0.5	1.4	0.4	1.0
Entrevista incompleta	0.4	0.0	0.0	1.8	0.9	0.1	0.4	0.1	0.1	0.2	0.2	0.0	0.2	0.3	0.1	0.4	0.0	0.2	0.4	1.1	0.1	0.2	0.1	0.0	0.3	0.3	0.3
Mujer incapacitada	1.4	1.3	0.4	1.0	1.2	1.5	0.2	0.5	0.6	1.0	0.6	0.9	1.4	0.8	0.8	1.3	0.6	0.5	1.4	0.4	0.3	0.3	0.5	0.7	0.8	1.0	0.8
Otra respuesta	0.0	0.1	0.0	0.4	0.3	0.4	0.4	0.0	0.2	0.1	0.2	0.0	0.2	0.2	0.2	0.0	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.1	0.1	0.2	0.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Número de mujeres	802	975	843	1,336	926	1,362	1,383	857	1,009	1,298	1,177	1,215	1,213	5,241	1,658	853	833	940	1,777	938	1,105	1,132	1,272	1,096	20,637	10,604	31,241
Tasa de respuesta	91.9	95.3	94.3	83.5	89.6	90.5	92.9	96	92.5	94.7	93.5	92.4	90.4	91.2	94.6	87.6	94.8	96.4	90.3	91.9	97.6	96.9	98.0	95.1	92.7	92.6	92.7
Tasa de respuesta global	90.0	94.2	89.7	80.1	85.9	87.7	90.6	94.4	91.6	93.8	93.0	91.0	89.4	87.7	92.7	82.0	93.2	93.7	88.3	91.0	97.3	95.9	97.5	93.3	90.7	90.1	90.5

APÉNDICE B

LA PRECISIÓN DE LOS RESULTADOS: LOS ERRORES DE MUESTREO

