



**Contratación de Estudio sobre el Grado de Conocimiento y  
Percepción de los Productos y Servicios de Seguros  
-GCPPSS-**

**Adjudicación Directa Selectiva N° 002/2012-SBS**

**Diseño de la Metodología Muestral**

DATUM Internacional S.A.

Departamento de Estadística y Procesamiento

Octubre / 2012.

## 1. ANTECEDENTES

El crecimiento de las primas de seguros en el Perú se ha mantenido en los últimos años, habiendo alcanzado 6,552 millones de nuevos soles a diciembre 2010, registrando una tasa de crecimiento nominal de 26.2% respecto al año anterior, que registró 5,190.4 millones de nuevos soles. Sin embargo, el índice de penetración de seguros aún se mantiene bajo al compararlo con los índices de otros países de la región. Este índice aumentó de 1.23% del PBI en diciembre de 2005 a 1.51% en diciembre de 2010<sup>1</sup>, mientras que en México fue de 1.93%, en Argentina 2.16%, en Colombia 2.8% y en Chile alcanzó el 4.07%<sup>2</sup>.

Las barreras de acceso al seguro en el mercado peruano lo constituyen, principalmente, la poca cultura de seguros que existe en la población y su baja conciencia hacia los riesgos, pero también la falta de una adecuada difusión de los seguros y la insuficiente oferta de productos para ciertos sectores de la población<sup>3</sup>. Actualmente, los mercados de seguros a nivel internacional ven en la baja penetración una oportunidad de negocio para desarrollar productos innovadores y nuevas estrategias de comercialización y venta de seguros, como es el caso de los microseguros, productos diseñados especialmente para la población de bajos ingresos. Esta es la población con mayor necesidad de un seguro, pero que tiene el menor acceso a este servicio, lo que se podría solucionar en parte con productos diseñados adecuadamente y desarrollando nuevas formas de comercialización<sup>4</sup>.

La importancia del seguro radica en permitir que las personas transfieran parte de sus riesgos a las compañías de seguros, a cambio de un pago periódico definido previamente, protegiéndose así de gastos inesperados y en algunos casos excesivos que, de otra forma, pondrían en peligro su subsistencia, llevándolas incluso a niveles de pobreza o pobreza extrema, como es el caso de inundaciones, terremotos y otros desastres naturales que afectan sobre todo a las familias pobres.

Los seguros contribuyen al crecimiento económico, asignando eficientemente los recursos, administrando los riesgos y movilizandolos a largo plazo. Además, reducen costos de transacción, generando liquidez y permitiendo economías de escala en las inversiones. Por ello, resulta necesario impulsar el crecimiento del mercado asegurador a través de nuevas alternativas de desarrollo, que pueden ir desde la innovación de sus productos y servicios, la acertada segmentación del mercado, las campañas de fomento del seguro hasta el uso de nuevos canales de comercialización. Si bien es cierto, los promotores de este desarrollo deben ser principalmente las propias compañías de seguros, corresponde también a la Superintendencia de Banca, Seguros y Administradoras Privadas de Fondos de Pensiones (SBS), en el marco de su misión institucional, fomentar la inclusión financiera y el desarrollo del mercado asegurador.

En la actualidad, las principales fuentes de información sobre el mercado de seguros analizan sobre todo el lado de la oferta; mientras que prácticamente no existe información sobre la situación del mercado objetivo, sus principales riesgos, el conocimiento y percepción del seguro en los potenciales clientes, y la posible existencia de una demanda insatisfecha en ciertos sectores de la población.

En ese contexto, el Banco Interamericano de Desarrollo (BID) ha aprobado el proyecto regional denominado “Aplicación de las normas reguladoras y supervisoras de los mercados de microseguros en América Latina”, financiado por el Fondo Multilateral de Inversiones (FOMIN) para fomentar la inclusión financiera en el ámbito de los seguros en los países de América Latina, entre otros, el Perú.

---

<sup>1</sup> Perú: Indicadores de Inclusión Financiera de los Sistemas Financiero, de Seguros y de Pensiones, Superintendencia de Banca, Seguros y AFP, Junio 2011.

<sup>2</sup> Boletín Estadístico de ASSAL, Diciembre 2010.

<sup>3</sup> Microseguros: Análisis de Experiencias en Latinoamérica y el Caribe, Fides, Fomin y Fundación Mapfre, 2011.

<sup>4</sup> Protegiendo a los pobres, Un compendio sobre microseguros, Munich Re Foundation, CGAP WorkingGroup on Microinsurance, ILO. Editado por Craig Churchill, España, 2009.

Por encargo del FOMIN, la Agencia de Cooperación Alemana al Desarrollo (GIZ) y la Iniciativa de Acceso al Seguro (A2II) deben implementar durante el presente año el proyecto antes señalado, en estrecha colaboración con los supervisores de seguros.

Uno de los componentes de este proyecto, es el análisis de la demanda de seguros en el país, con especial énfasis en la población de bajos ingresos, por lo que, al identificar la falta de estudios en nuestro país que muestren la realidad del mercado objetivo y constituyen una fuente de información adecuada, la SBS considera pertinente hacer un estudio en la población a nivel nacional para determinar el grado de conocimiento y percepción de los productos y servicios de seguros.

Durante el año 2012 se realizó de conformidad con lo señalado en el Decreto Legislativo N° 1017, Ley de Contrataciones del Estado, y el Decreto Supremo N° 184-2008-EF, Reglamento de la Ley de Contrataciones del Estado, en la fecha señalada en el cronograma la adjudicación directa selectiva N° 002/2012-SBS, para la contratación de un estudio sobre el grado de conocimiento y percepción de los productos y servicios de seguros, la cual a partir de este momento conoceremos como “GCPPSS 2012”. El otorgamiento de la Buena Pro se dio a Datum – Internacional S.A. el 27/03/2012, a través del SEACE.

En este sentido, el presente informe contiene una descripción metodológica sobre los requerimientos necesarios para la implementación de una muestra de viviendas, indicando para tales efectos los procedimientos aplicados. Asimismo, en este diseño muestral, se ha utilizado como información base el marco muestral de viviendas proveniente de los Censos Nacionales 2007: XI de Población y VI de Vivienda (INEI) y tiene como característica principal utilizar un método de selección probabilístico multietápico con estratificación.

## **2. OBJETIVOS**

El “Estudio sobre el Grado de Conocimiento y Percepción de los Productos y Servicios de Seguros” tiene por objetivo general, disponer de datos que permitan estimar los indicadores resultantes a nivel global (ámbito de estudio) y por departamento; obteniendo información sobre el grado de conocimiento y percepción que tiene la población urbana respecto a los productos y servicios de seguros que se ofrecen en el país.

Los resultados del estudio deberían permitir identificar las posibles barreras de acceso al seguro, así como las barreras de uso, que llevan a preferir no usar el servicio a pesar de tener técnicamente acceso a ello. Asimismo, al investigar los riesgos de la población de bajos ingresos, el uso actual y percepción de los seguros, se podría determinar la necesidad de adecuar los productos y servicios de seguros para esta población.

Para esto se ha determinado que el presente diseño muestral tenga como objetivos:

- a) Plantear los métodos probabilísticos adecuados para una correcta aplicación de la encuesta GCPPSS 2012.
- b) Proponer y aplicar los niveles de estratificación necesaria para una completa comprensión de la población objetivo.
- c) Determinar la población objetivo.
- d) Determinar un correcto marco muestral.
- e) Determinar los dominios, niveles de inferencia y ámbitos de estudios.
- f) Determinar un tamaño de muestra adecuado.
- g) Mostrar los pasos necesarios para obtener los resultados finales (ponderaciones y otros cálculos adicionales necesarios para una correcta presentación de los resultados).

### 3. POBLACIÓN OBJETIVO

La población de estudio son los hogares residentes en Lima Metropolitana, la Provincia Constitucional del Callao y los departamentos de Arequipa, Junín, La Libertad y Piura. Se excluye del estudio a las personas que residen en viviendas colectivas (hospitales, hoteles, asilos, claustros religiosos, cuarteles y cárceles).

El informante será el jefe de hogar. En caso no se encuentre el jefe del hogar se reprogramarán dos visitas posteriores.

- **Jefe del hogar:** Persona que aporta más ingresos en el hogar.
- **Hogar:** Es la persona o conjunto de personas, sean parientes o no, que ocupan en su totalidad o en parte una vivienda; comparten al menos las comidas principales y atienden en común otras necesidades básicas, con cargo a un presupuesto común.

### 4. ÁMBITO GEOGRÁFICO

La encuesta será aplicada en:

- Lima Metropolitana (como una sola unidad geográfica)<sup>5</sup>.

Y en 4 departamentos del interior del país:

- Departamento de Arequipa.
- Departamento de Junín.
- Departamento de La Libertad.
- Departamento de Piura.

Para el caso de Lima Metropolitana la distribución del número de viviendas es la siguiente:

	Número de viviendas – área urbana <sup>5</sup>
Lima Metropolitana	2066996

Y para el caso los departamentos del interior del país, el número de viviendas es el siguiente:

	Número de viviendas – área urbana <sup>7</sup>
Arequipa	296213
Junín	220877
La Libertad	299050
Piura	297728

---

<sup>5</sup> Lima Metropolitana se refiere a la urbe conformada por los distritos de Lima Metropolitana y de la Provincia Constitucional del Callao.

<sup>6</sup>Fuente: Censos Nacionales 2007.

<sup>7</sup>Fuente: Censos Nacionales 2007.

## 5. MARCO MUESTRAL

Este marco muestral es del tipo áreas y de lista. El marco de áreas está conformado por el material cartográfico e información estadística del Censo de Población y Vivienda de 2007<sup>8</sup>. El marco de lista será elaborado previo al levantamiento de información, en el cual se identificarán los conglomerados existentes en cada una de las cuatro áreas de estudio a partir de la información del Censo de Población y Vivienda de 2007. Se considera que este marco es óptimo porque contiene las variables necesarias para la identificación, ubicación y previa estratificación de las viviendas.

En el siguiente cuadro se puede observar la distribución de viviendas de cada uno de los dominios de estudio

AREAS	DISTRITOS	PROVINCIAS	URBANO
AREQUIPA	109	8	296213
JUNIN	124	9	220877
LA LIBERTAD	83	12	299050
PIURA	64	8	297728
LIMA METROPOLITANA	49	2	2066996

Tomando en cuenta que en el área urbana<sup>9</sup>, los distritos están divididos en zonas censales, y estos a su vez agrupan a secciones censales que cuentan en promedio con 120 viviendas, estas secciones censales, denominados “*conglomerados urbanos*” son las unidades de muestreo a ser seleccionados en la segunda etapa de muestreo, en el marco muestral, se registrará el número de viviendas en cada conglomerado urbano identificados por: un código de manzana<sup>10</sup>, estrato, distrito, provincia y departamento al que pertenece. Asimismo, antes de llevarse a cabo la selección de las unidades muestrales se ordenarán los registros por tamaño del conglomerado (número de viviendas).

## 6. DOMINIOS GEOGRÁFICOS DE ESTUDIO.

Los dominios del estudio para la encuesta son:

- Lima Metropolitana.
- Departamento de Arequipa.
- Departamento de Junín.
- Departamento de La Libertad.
- Departamento de Piura.

## 7. NIVELES DE INFERENCIA

De los términos de referencia, el objetivo del presente estudio es obtener resultados para los siguientes niveles de inferencia:

- a) A nivel total del ámbito geográfico: Lima Metropolitana y áreas urbanas de los departamentos de Arequipa, Junín, La Libertad y Piura.

Adicionalmente, se pueden obtener los siguientes niveles de inferencia:

- a) A nivel del área de Lima Metropolitana.
- b) A nivel de cada uno de los departamentos seleccionados al interior del país.

---

<sup>8</sup> Según los términos de referencia “La muestra para la aplicación de la encuesta deberá ser seleccionada a partir del Censo de Población y Vivienda del año 2007”

<sup>9</sup> Definiciones censales del Instituto Nacional de Estadística e Informática.

<sup>10</sup> Una manzana agrupa en promedio a 50 viviendas.

## 8. TIPO DE SELECCIÓN DE LA MUESTRA

- La selección de la muestra se caracteriza por ser probabilística<sup>11</sup>, de áreas, estratificada<sup>12</sup>, multietápica, independiente en cada dominio de estudio.
- La muestra es probabilística porque las unidades de muestreo son seleccionadas mediante métodos aleatorios, con probabilidades conocidas, lo cual permite efectuar inferencias a la población sobre la base de la teoría de probabilidades.
- La muestra es de áreas, porque cada vivienda está asociada a un área geográfica (conglomerado) dentro de cada dominio de estudio.
- La muestra es estratificada, porque previamente a la selección, la población se ha dividido en estratos, con el objeto de mejorar su representatividad.
- La muestra es trietápica, porque se realizará una selección en tres etapas: en la primera y segunda etapa se utiliza la selección sistemática con probabilidad proporcional al tamaño (PPT) en número de viviendas. En la última etapa la selección es sistemática simple con arranque aleatorio.

## 9. DISPERSIÓN DE LA MUESTRA:

Este capítulo es de suma importancia dado que se tratará de la forma más adecuada la formación de las unidades de muestreo necesarias para la aplicación de la encuesta

Los términos de referencia marcan como zona de cobertura a Lima Metropolitana y 4 departamentos del interior del país, es necesario que se presenten métodos adecuados para poder dispersar óptimamente la muestra, para lo cual se ha tomado en cuenta que en Lima Metropolitana la muestra tendrá una metodología de dispersión diferente de la que se utilizará en los departamentos del interior del país, esto se realizará teniendo en cuenta las divisiones geopolíticas y realidades territoriales, las cuales son muy diferentes, lo que conlleva a tener poblaciones con diferentes características.

A continuación se presentan ambas realidades:

- **Lima Metropolitana:** conformada por 2 provincias y 49 distritos, que forman un bloque poblacional predominantemente urbano con la descripción de bloques o segmentos territoriales próximos o cercanos, es así que Lima Metropolitana se divide en seis grandes ciudades en formación; la mitad de la clase media vive fuera de Lima Moderna. Lima Norte es el cono más próspero y con gran potencial económico. Así, podemos decir que Lima Metropolitana (la urbe formada por Lima y Callao) está formada por seis zonas. Por su población, cada una de ellas es más grande que cualquier ciudad del interior del país (por esto mismo se hace la diferencia entre Lima Metropolitana y los departamentos seleccionados). Aquellos sectores que antes llamábamos “conos” son, ahora, auténticas ciudades en formación y cada vez más independientes entre sí.

---

<sup>11</sup> En las muestras probabilísticas se requieren probabilidades conocidas de selección, es decir se debe conocer la probabilidad de cada elemento incluido en el marco muestral, es decir cada unidad tiene una probabilidad de selección conocida y superior a cero. Este tipo de muestra permite calcular los errores muestrales y así poder generalizar los resultados de la muestra a la población de estudio, especificando la precisión de estas inferencias.

<sup>12</sup> El muestreo estratificado asegura una mayor confiabilidad de la muestra disminuyendo la varianza de las estimaciones. Una estratificación es más eficiente cuanto más homogéneas sean las unidades que pertenecen a un mismo estrato y más heterogéneos los estratos entre sí. Aun sin cumplir estrictamente con estas características, cualquier estratificación mejora la calidad de las estimaciones, con la única condición de que toda unidad de muestreo pertenezca a un estrato y sólo a uno y la unión de todos los estratos conformen la población total de estudio.

- **Departamentos del interior del país:** En promedio, cada departamento está conformado por 9 provincias y 11 distritos dentro de los cuales se ubican localidades con zonas urbanas y rurales -estas últimas fuera del alcance del estudio-. En este caso, se hace necesario considerar un diseño que permita asegurar dispersión y representatividad de la muestra con relación a la población urbana que habita en las provincias de los cuatro departamentos que son parte del ámbito de estudio y con características distintas a la población de Lima metropolitana.

Estas dos realidades conllevan finalmente a determinar estratos geográficos para los diferentes dominios, los cuales procederemos a presentar.

## 10. ESTRATOS.

Los estratos, al ser una partición de los dominios definidos para lograr mayor homogeneidad entre unidades permiten hacer más eficiente el diseño. Para este caso en particular, se propone lo siguiente:

### Para Lima Metropolitana:

- Se utiliza la estratificación sugerida en “Perfiles Zonales de la Gran Lima 2009”<sup>13</sup>. Los distritos de Lima Metropolitana y Callao, se clasifican en 6 zonas a partir de criterios relacionados a la situación socioeconómica de los hogares (nivel socioeconómico), entre las más importantes: Infraestructura y servicios básicos de la vivienda, tenencia de bienes y nivel de ingresos del hogar.

Si bien es cierto que la mejor estratificación sería aquella que considere como criterio a la variable de interés “*conocimiento de productos y servicios de seguros*”, cuyo comportamiento es desconocido y motivo del presente estudio, el uso de la agrupación de distritos en estos 6 estratos se podría justificar si se toma en cuenta que el conocimiento y tenencia de este tipo de productos podrían estar vinculados con el nivel socioeconómico del hogar.

En el siguiente cuadro se muestra la composición de estos 6 estratos:

CENSOS 2007: TOTAL DE VIVIENDAS, SEGUN POR ESTRATOS		
UBIGEO	DISTRITO	VIVIENDAS CENSADAS
<b>LIMA NORTE</b>		
150106	CARABAYLLO	51,284
150125	PUENTE PIEDRA	55,977
150110	COMAS	97,772
150117	LOS OLIVOS	76,593
150135	SAN MARTÍN DE PORRES	130,222
150112	INDEPENDENCIA	42,923
150102	ANCON	12,195
150139	SANTA ROSA	4,262
<b>LIMA ANTIGUA</b>		
150128	RIMAC	43,939
150101	LIMA	83,636
150105	BREÑA	24,101
150115	LA VICTORIA	52,966
150134	SAN LUIS	14,170

### CALLAO

<sup>13</sup> publicada por Ipsos Apoyo Opinión y Mercado.

70106	VENTANILLA	80,434
70101	CALLAO	90,865
70103	CÁRMEN DE LA LEGUA Y R.	8,710
70102	BELLAVISTA	17,000
70104	LA PERLA	14,438
70105	LA PUNTA	1,409
<b>LIMA ESTE</b>		
150132	SAN JUAN DE LURIG.	202,727
150111	EL AGUSTINO	38,219
150137	SANTA ANITA	41,456
150103	ATE	119,112
150107	CHACLACAYO	10,012
150109	CIENEGUILLA	5,541
150118	LURIGANCHO	44,779
<b>LIMA MODERNA</b>		
150136	SAN MIGUEL	35,067
150120	MAGDALENA DEL MAR	14,943
150113	JESÚS MARÍA	19,909
150121	MAGDALENA VIEJA	20,982
150116	LINCE	17,354
150131	SAN ISIDRO	20,681
150122	MIRAFLORES	33,092
150141	SURQUILLO	25,366
150130	SAN BORJA	30,812
150104	BARRANCO	10,439
150140	SANTIAGO DE SURCO	81,806
150114	LA MOLINA	35,108
<b>LIMA SUR</b>		
150143	VILLA MARIA DEL T.	88,571
150133	SAN JUAN DE MIRAF.	72,305
150142	VILLA EL SALVADOR	80,658
150108	CHORRILLOS	66,097
150119	LURÍN	16,211
150123	PACHACAMAC	19,606
150124	PUCUSANA	4,368
150126	PUNTA HERMOSA	3,092
150127	PUNTA NEGRA	2,587
150129	SAN BARTOLO	2,373
150138	SANTA MARÍA DEL MAR	827
<b>TOTAL</b>		<b>2,066,996</b>

#### Departamentos del interior del país:

- En cada uno de los departamentos restantes, las provincias se están clasificando en tres estratos, tomando en cuenta el total de su población urbana. Estos estratos han sido contruidos con base en los niveles de clasificación del Instituto Nacional de Estadística en la Encuesta Nacional de Hogares (ENAH0-1998)<sup>14</sup>.

<b>Estrato I:</b> áreas fuertemente metropolitanas, conformadas por aquellas provincias <sup>15</sup> con 100,000 y más habitantes,
<b>Estrato II:</b> centros urbanos de mediano tamaño, que incluye a provincias con 20,000 a menos de 100,000 habitantes,
<b>Estrato III:</b> áreas urbanas menores, conformadas por aquellos provincias con 2 000 a menos de 20,000 habitantes

<sup>14</sup> <http://www.inei.gob.pe/biblioineipub/bancopub/Est/Lib0330/ANEX-03.HTM>. En este documento, se hace referencia a cinco niveles, dos de los cuales corresponden al área rural.

<sup>15</sup> El INEI, hace la clasificación por localidad.



A continuación, se presenta la estratificación de provincias en los cuatro departamentos de interés:

Departamento Arequipa

Provincia	Nro viviendas	Nro. Habitantes	Estrato
Arequipa	228146	842880	1
Caylloma	16710	49062	2
Islay	15682	47402	2
Camana	14181	44741	2
Caraveli	7515	23080	2
Castilla	7554	21913	2
Condesuyos	3152	7764	3
La Unión	3273	7550	3

Departamento Junín

Provincia	Nro viviendas	Nro. Habitantes	Estrato
Huancayo	102480	412051	1
Chanchamayo	25296	96191	2
Tarma	19821	72569	2
Jauja	21244	63941	2
Satipo	12418	53531	2
Yauli	14684	43818	2
Chupaca	8356	31328	2
Concepción	8900	30619	2
Junín	7678	21215	2

Departamento La Libertad

Provincia	Nro viviendas	Nro. Habitantes	Estrato
Trujillo	185857	792355	1
Ascope	26832	102230	1
Pacasmayo	24324	87718	2
Chepen	17084	62153	2
Viru	13836	59090	2
Sanchez Carrión	8367	36585	2
Otuzco	7303	21179	2
Pataz	4993	20462	2
Santiago de Chuco	5289	19966	3
Gran Chimú	2005	6544	3
Bolívar	1505	5754	3
Julcan	1655	4886	3

Departamento Piura

Provincia	Nro viviendas	Nro. Habitantes	Estrato
Piura	133113	573139	1
Sullana	59963	258723	1
Talara	30835	126866	1
Paita	25537	103615	1
Morropón	24265	91798	2
Sechura	15516	58497	2
Ayabaca	4113	15845	3
Huancabamba	4386	15358	3

## 11. LAS UNIDADES DE MUESTREO

### Formación de las Unidades de Muestreo.

Tomando en cuenta los puntos desarrollados en las secciones anteriores, procederemos a definir las unidades de muestreo en cada uno de los dominios de estudio, para las tres etapas consideradas en el diseño muestral:

**Unidad Primaria de Muestreo (UPM).**-Cada UPM es definida de manera diferente para Lima Metropolitana e interior del país.

Así, en Lima Metropolitana se considera una UPM a cada uno de los 49 distritos que lo conforman, ya clasificados en los 6 estratos propuestos: Lima antigua, Lima este, Lima moderna, Lima norte, Lima sur y Callao.

En los departamentos de Arequipa, Ayacucho, La Libertad y Piura, las UPM son las provincias que forman parte de estos ámbitos geográficos, clasificados en los tres estratos ya presentados con anterioridad:

- Provincias pertenecientes a la subdivisión: “Estrato I”
- Provincias pertenecientes a la subdivisión: “Estrato II”
- Provincias pertenecientes a la subdivisión: “Estrato III”

**Unidad Secundaria de Muestreo (USM).**- Las unidades muestrales de la segunda etapa (USM), son los conglomerados urbanos incluidos dentro de cada UPM (que como ya se mencionó en la sección 5, agrupa un promedio de 120 viviendas).

**Unidades Terciarias de Muestreo (UTM).**- Las unidades de la tercera etapa son las viviendas particulares habitadas permanentemente o aptas para habitarse que se encuentren en el momento de la elaboración de listado.

## 12. TAMAÑO MUESTRAL.

El tamaño de muestra es uno de los elementos más importantes que intervienen para que la información generada por una encuesta probabilística sea de calidad. Para cumplir con este propósito el tamaño de muestra de la GCPPSS debe ser suficiente para estimar todos los indicadores de interés, teniendo en cuenta los siguientes factores:

- Precisión de las estimaciones.
- Calidad de los datos recopilados por la encuesta.
- Tiempo de elaboración y divulgación de los datos.

Como tenemos el problema de carecer de antecedentes de las variables o indicadores de interés para la GCPPSS, nos prestamos entonces a evaluar los principales parámetros generados a partir del supuesto de máxima varianza para una distribución binomial.

La precisión de las estimaciones de las encuestas se expresa muchas veces en términos del coeficiente de variación de la estimación de interés. El coeficiente estimado de variación de una estimación de proporción, designado por  $CV(p)$ , viene dado por:

$$CV(p) = \sqrt{efd \frac{(1-p)}{np}}$$

Donde:

- CV(p): es el coeficiente de variación del estimador proporción.
- edf: es el efecto del diseño, se asume un valor igual a 2.
- P: es el indicador con valor de p=0.5 (valor con el que se obtiene la varianza máxima para la distribución binomial).
- n: es el número de viviendas de la muestra

De la fórmula anterior se puede obtener:

$$n = \left( \frac{1}{CV(p)} \right)^2 edf$$

Nota: Dado que en encuestas a hogares, siempre se tiene una pérdida por hogares que no acceden a responder, se hace necesario considerar una sobremuestra para controlar las precisiones esperadas. Para esto, se ha considerado el incluir el estimado de la tasa de no respuesta, o TNR, el cual es aproximado a 12.7% en Lima Metropolitana y 10.3% en los departamentos seleccionados del interior del país, es así que incluimos esta tasa de no respuesta o TNR a la fórmula anterior de esta manera:

$$n = \left( \frac{1}{CV(p)} \right)^2 edf \left( \frac{1}{1 - TNR} \right)$$

Cabe señalar, que los valores para la TNR ya mencionados, han sido elegidos con base en lo observado en estudios anteriores llevados a cabo por nuestra agencia. Además, guarda concordancia con los resultados presentados en la ENAHO 2011: Una TNR para Lima Metropolitana de 13,4% y de 11.2% en en área urbana.

Aplicando la fórmula anterior para valores fijados de p=0.5, edf=2 y TNR=12.7% para Lima y en 10.3% para los otros departamentos, se obtienen los tamaños de muestra de viviendas en cada dominio de estudio. Cabe señalar, que la precisión obtenida para estos tamaños de muestra es “buena”, observando los valores que toman los coeficientes de variación, de acuerdo a la calificación propuesta en el libro “Encuestas de hogares en los países en desarrollo y en transición – Naciones Unidas 2007”.

Dominio	Efecto del Diseño (edf)	TNR	(p)	Tamaño de Muestra de Viviendas	CV(p) %	PRECISION OBTENIDA
Lima Metropolitana	2	12.7%	0.5	400	7.6%	Buena
Departamento de Arequipa.	2	10.3%	0.5	310	8.5%	Buena
Departamento de Junín.	2	10.3%	0.5	310	8.5%	Buena
Departamento de La Libertad.	2	10.3%	0.5	310	8.5%	Buena
Departamento de Piura.	2	10.3%	0.5	310	8.5%	Buena

Es importante señalar, que los niveles de precisión son los recomendados para las estimaciones requeridas en cada dominio de estudio, siguiendo la recomendación de los autores del libro “Diseño de muestras para encuestas de hogares Directrices prácticas – Naciones Unidas 2009”

*(...)El tamaño de la muestra necesario para alcanzar el criterio del 10% de margen de error relativo es, por tanto, la cuarta parte de lo que sería en un caso en el que el margen de error relativo se hubiera fijado en el 5%. Un margen de error relativo del 20% suele considerarse el máximo permisible para indicadores importantes (aunque no se recomienda). Esto se debe a que el intervalo de confianza en torno a las estimaciones con mayor tolerancia de errores es demasiado amplio para alcanzar unos resultados que satisfagan la mayoría de las necesidades analíticas o de política. En general, si el presupuesto lo permite, **recomendamos aplicar un 5%-10% de errores relativos** para los principales indicadores.*

*Capítulo: Estrategias de muestreo – página 37.*

Si bien es cierto, que siempre es posible obtener muestras con mejores niveles de precisión, lo es también el incremento del presupuesto asignado, costo de oportunidad que es necesario tomar en cuenta en toda estudio por encuestas.

Finalmente, los tamaños de muestra en cada dominio de estudio son los siguientes:

<b>Dominio</b>	<b>Muestra</b>
Lima Metropolitana y Callao	400
Arequipa	310
Junín	310
La Libertad	310
Piura	310
<b>Total</b>	<b>1640</b>

### 13. SELECCIÓN DE LA MUESTRA

En esta sección se presenta el método de selección a utilizar para la elección de las unidades de muestreo en cada una de las tres etapas para los cinco dominios de estudio.

#### **Selección para Lima Metropolitana:**

Como la muestra es trietápica, de selección sistemática proporcional al tamaño de viviendas en cada estrato en la primera y segunda etapa y selección simple con arranque aleatorio en la tercera etapa

- **Selección de las UPM (distritos)**

En cada uno de los seis estratos ya definidos, se seleccionarán cuatro UPM (distritos).

- **Selección de las USM (conglomerado urbano)**

La USM es el conglomerado urbano, los cuales serán ordenados por el número de viviendas. Luego se procede a seleccionar las USM mediante el muestreo sistemático con probabilidad proporcional al número de viviendas.

Dentro de cada dominio de estudio se seleccionará el siguiente número de conglomerados:

<b>Dominio</b>	<b>Conglomerados</b>
Lima Metropolitana y Callao	80

- **Selección de las UTM (viviendas)**

La unidad de selección en la tercera etapa es la vivienda particular. En cada USM seleccionada, se procede a la selección de 5 viviendas utilizando un arranque aleatorio y en forma sistemática simple al azar, luego de actualizar el total de viviendas.

- **Procedimiento de selección de la vivienda (UTM):**

Para la elección final de la vivienda se va a tener en cuenta que cada USM seleccionada tiene un mapa, en el cual se señala un punto de partida geográfico. Este punto es la esquina de la cuadra que apunta más hacia el nor-oeste, a partir de este punto el encuestador empezará con el conteo de viviendas.

Luego, reportará telefónicamente el número de viviendas hábiles en la USM. Con esta información, el coordinador de campo procederá a calcular el salto sistemático “K” y hará entrega al encuestador de los números aleatorios que corresponden a las viviendas seleccionadas.

El punto de inicio para la aplicación de encuestas, es el mismo que se utilizó para el conteo de viviendas. A partir de este punto se empieza a marcar un recorrido en sentido horario tanto dentro de manzana como entre manzanas, hasta ubicar las viviendas seleccionadas.

#### **Selección para los departamentos seleccionados del Interior del país:**

En cada dominio de estudio la muestra es trietápica, de selección sistemática proporcional al tamaño de viviendas en la primera y segunda etapa, y sistemática simple con arranque aleatorio en la tercera etapa

- **Selección de UPM (provincias)**

Se procede a seleccionar las UPM mediante el muestreo sistemático con probabilidad proporcional al tamaño en número de viviendas. Previo a la selección de las UPM, se utilizará la estratificación presentada en la sección 10, la cual toma en cuenta los niveles de estratificación por tamaño (Estrato I, Estrato II Y Estrato III).

<b>Dominios</b>	<b>Total provincias</b>	<b>Estrato I</b>	<b>Estrato II</b>	<b>Estrato III</b>
Arequipa	8	1	5	2
Junin	9	1	8	0
La libertad	12	2	6	4
Piura	8	4	2	2

Así, se ha fijado seleccionar aleatoriamente  $\frac{1}{4}$  del total de provincias **de cada estrato** para de esta manera asegurar que los tres estratos estén representados en la muestra. El número de UPM a seleccionar en cada departamento es:

Dominios	Total muestra (UPM)	Estrato I	Estrato II	Estrato III
Arequipa	3	1	1	1
Junin	3	1	2	-.-
La libertad	4	1	2	1
Piura	3	1	1	1

- Selección de las USM (Conglomerado urbano)**

Se procede a seleccionar las USM en las UPM escogidas en la primera etapa, mediante el muestreo sistemático con probabilidad proporcional al tamaño en número de viviendas. De este modo, el número de conglomerados a sortear en cada departamento es:

Dominio	Conglomerados
Arequipa	62
Junín	62
La Libertad	62
Piura	62
<b>Total</b>	<b>248</b>

- Selección de las UTM (Viviendas)**

La unidad de selección en la tercera etapa es la vivienda particular. En cada USM seleccionada, se procede a la selección de 5 viviendas con arranque aleatorio y en forma sistemática simple al azar.

**Procedimiento de selección de la vivienda (UTM - urbana):**

Para la selección de viviendas se utilizará el mismo procedimiento explicado para el dominio "Lima Metropolitana".

#### 14. PONDERACIÓN Y FACTOR DE EXPANSIÓN

Para que las estimaciones de la encuesta sean representativas de la población investigada, es necesario multiplicar los datos de cada vivienda en la base de datos por el peso muestral o factor de expansión. Como todos los datos de la encuesta serán procesados por computadora, es fácil agregar un peso al registro de cada vivienda muestral en los archivos; con esto, los programas de tabulación pueden ponderar los datos automáticamente.

El factor de expansión básico para cada vivienda muestral es determinado por el diseño de la muestra. Equivale al inverso de su probabilidad final de selección, el mismo que es producto de las probabilidades de cada etapa seleccionada (selección de conglomerados y selección de viviendas). Para ilustrar el cálculo de la probabilidad de selección y los factores de expansión correspondientes, puede considerarse el caso de las viviendas seleccionadas en unidades primarias de muestreo de un departamento. La probabilidad de selección de cada vivienda será calculada a través de la fórmula siguiente:

$$Phij = \left[ \frac{(nh)}{(Mhi)} \right] / Mh \times \left[ \frac{(ghi)}{(Mhij)} \right] / Mhi \times \left[ \frac{mhij}{Mhij} \right]$$

Donde:

- Phi = Probabilidad final de selección.

**Primera etapa de selección:**

- nh = Número de UPM a seleccionarse en h-estrato.
- Mh = Número total de viviendas en el marco en el h-estrato.
- Mhi = Total de viviendas según marco muestral para la i-UPM seleccionada en el h-estrato.

**Segunda etapa de selección:**

- ghi = Número de USM a seleccionarse en la i-UPM.
- Mhij = Total de viviendas en el marco para la j-USM seleccionada dentro de la i-UPM seleccionada del departamento.

**Tercera etapa de selección:**

- mhij = Número de viviendas seleccionadas en la j –ésima USM seleccionada dentro de la i-UPM seleccionada en el departamento.
- Mhij = Total de viviendas listadas en la j-USM seleccionada dentro de la i-UPM seleccionada en el departamento.

El factor de expansión básico es calculado como el inverso de la probabilidad final de selección:

$$Whij = 1 / Phij$$

Es importante ajustar los factores de expansión teniendo en cuenta la magnitud de la no respuesta. Dado que los factores de expansión son calculados a nivel de cada USM seleccionada, es ventajoso ajustar los factores de expansión a este nivel. En este caso, los factores de expansión final (Whij) para las viviendas seleccionadas en la j-USM seleccionada dentro de la i-UPM seleccionada en el departamento se puede expresar como:

$$Whij = ( WhijB B) (mhij / mhij)$$

- mhij = Total de viviendas seleccionadas en la j-USM seleccionada dentro de lai-UPM seleccionada en el departamento (es decir, el número de entrevistas mas el número de no entrevistas).
- mhij = Total de viviendas entrevistados en la j-USM seleccionada dentro de lai-UPM seleccionada en departamento.

## 15. VARIANZA Y PRECISIÓN DE LOS RESULTADOS: ERRORES DE MUESTREO

Para el cálculo de la varianza se ha previsto tener en cuenta utilizar:

$$V_c(p) = \left(1 - \frac{n}{N}\right) (CV(p))^2 p^2$$

Donde a posteriori también se calcularán las estimaciones de las principales variables investigadas. Para cada parámetro especificado y dominio de estimación, se produce un cuadro de salida con los indicadores siguientes:

- El valor estimado del parámetro (estimación puntual)
- El error estándar
- El error relativo o desviación relativa muestral (coeficiente de variación - cv)

El intervalo con 95 por ciento de confianza

- El número de observaciones sobre el cual se basa la estimación, para estas estimaciones se trabajara con el modulo de muestras complejas del spss y el cenvar.

La precisión de una estimación se mide a través del error muestral o desviación relativa muestral, el cual mide la precisión de una estimación muestral (de un promedio, total, tasa o proporción) respecto a la que podría obtenerse si se eligieran todos los elementos de la población. Es decir, el error muestral puede atribuirse al hecho que no se eligió a la totalidad de la población (universo).

El error muestral se calcula estadísticamente a partir de los datos de la muestra y se determina por el error estándar, que es igual a la raíz cuadrada de la varianza del estimador. El error estándar se usa para obtener el coeficiente de variación (cv), denominado error muestral o desviación relativa muestral. Permite al usuario evaluar la precisión del estimador en términos relativos y comparar niveles de precisión para estimaciones de diferentes parámetros en diferentes poblaciones:

Las estimaciones que se obtienen a partir de una muestra pueden tener la precisión siguiente:

**C.V. CALCULADA**

Hasta 5%  
5% a 10%  
10% a 15%  
Mas de 15%

**PRECISION OBTENIDA**

Muy Buena  
Buena  
Aceptable  
No confiable (referencial)



## **Anexo A**

### **Sobre el trabajo de campo**

#### **1. Selección y capacitación de proveedores de campo**

Para la convocatoria a proveedores de campo –encuestadores y supervisores- primero se seleccionará de nuestra base de datos, a las personas que cuenten con experiencia previa no menor a un año en la aplicación de encuestas, que de preferencia tengan estudios superiores y que no hayan sido sancionados en los últimos seis meses.

Luego, se llevará a cabo una capacitación de dos días en la cual se les hará entrega a todos los participantes del manual del cuestionario. Estas reuniones estarán dirigidas por el Investigador responsable del estudio y contará además con la participación del coordinador de campo, la responsable de la edición y codificación de encuestas y del coordinador de digitación.

Durante la capacitación, se explicará en detalle todas las preguntas del cuestionario así como la metodología a seguir para la selección de viviendas. Al finalizar, se pasará a evaluar a los participantes quienes deben aprobar de manera satisfactoria (nota mínima de 14 puntos) antes de salir a campo.

#### **2. Supervisión**

Un supervisor depende directamente del coordinador de campo, persona responsable de esta fase del estudio. Todo estudio tiene obligatoriamente como monto mínimo de verificación el 30% del total de encuestas realizadas, distribuidas de acuerdo a las zonas de trabajo. Se utilizarán las técnicas de verificación coincidental y por recordación.

**Verificación coincidental:** El verificador acompañará al proveedor durante el recorrido por lo menos mientras completa el 50% de su ruta y es parte activa del recojo de información, el resto de entrevistas de esa misma ruta pueden ser verificadas por ejemplo telefónicamente en oficina, para complementar el trabajo. Durante esta forma de verificación hará una revisita en las viviendas que no estuvo con el proveedor y confirmará la validez de la información.

**Verificación por recordación:** El verificador acude a la zona trabajada por el proveedor y además de revisar la metodología con la que trabajó, acude al 100% de las viviendas donde el proveedor haya hecho alguna entrevista. Repreguntando los aspectos claves del cuestionario, además de tiempo de duración, si se hizo entender claramente, entre otras preguntas que puedan ser de interés para el estudio.

#### **Pasos operativos**

- En las verificaciones, se utilizará lapicero verde para evitar que se confunda con la información recogida por el proveedor.
- Todos los proveedores de campo deberán ser supervisados.
- Se debe verificar el porcentaje consignado en la ficha técnica.

Considerar como parte de la verificación:

- Recorrido esté de acuerdo a la metodología propuesta.
- Se haya respetado el punto de partida.

- Cada una de las viviendas seleccionadas en la zona de trabajo, corresponda al NSE establecido en la hoja de ruta.
- El entrevistado cumpla con las características requeridas.
- El cuestionario se haya aplicado correctamente.
- Se deberá de verificar los datos de control, preguntas claves de acuerdo a los objetivos del estudio y las baterías de preguntas.
- Programar las verificaciones (coincidentales y por recordación) a la par del reporte de encuestas.
- Ningún proveedor encargado de las verificaciones de un estudio, puede participar como aplicador de encuestas. Salvo durante el primer día de campo para asegurar el manejo del cuestionario.

### **Reportes**

Elaborados por los verificadores y consolidados por el asistente de campo quien debe revisar que el proveedor:

- Utilice el formato de verificación.
- En el informe haya detallado los resultados de cada visita realizada.

En todas las encuestas verificadas, deben ir los datos personales del verificador y la fecha que se realizó la evaluación

### **Controles**

A ser llenados en formato Excel por el asistente de campo para el reporte a investigación. Es el paso previo a la entrega del material al grupo de edición. Asegurándose que la información cumpla con los requisitos solicitados en la ficha técnica.

### **3. Edición e imputación**

Luego del reporte de las encuestas realizadas en el día, estas pasarán al área de edición donde se revisará minuciosamente el 100% del material, como paso previo a la codificación de encuestas.

El paso siguiente será la codificación y consistencia manual de las encuestas. La responsable del área, deberá hacer la verificación de un número de encuestas codificadas, contrastando las respuestas con las alternativas de respuesta aperturadas antes de hacer entrega del material al área de digitación.

### **4. Captura de Datos y verificación de la digitación**

Previo a la creación del programa de ingreso de datos, se elaborará una tabla de consistencia considerando la relación lógica entre las preguntas del cuestionario, saltos, número de respuestas permitidas, entre otros.

Haciendo uso de este insumo, se construirá el programa con los controles necesarios que permitan agilizar el trabajo del digitador y minimizar los errores de digitación. Luego, se hace una prueba de consistencia con las primeras encuestas editadas y codificadas.

Para el control del ingreso de datos, se seleccionará una muestra aleatoria de las encuestas digitadas, asegurando incluir como mínimo el 10% de la cuota de trabajo de cada digitador.

Estas encuestas deben ser “re-digitadas” por un operador diferente para luego comparar los resultados con la primera base digitada. Si se encuentra un porcentaje alto de errores en el ingreso de un proveedor se vuelve a digitar el total de encuestas asignadas a esta persona.

Cabe señalar, que este control de calidad se aplica sistemáticamente mientras dura la tarea de digitación con la finalidad de volver a capacitar a cada proveedor de acuerdo a los errores cometidos y así, minimizar los índices de error.