

PN-ABA-826

Best available copy -- pages 69 and 70 missing

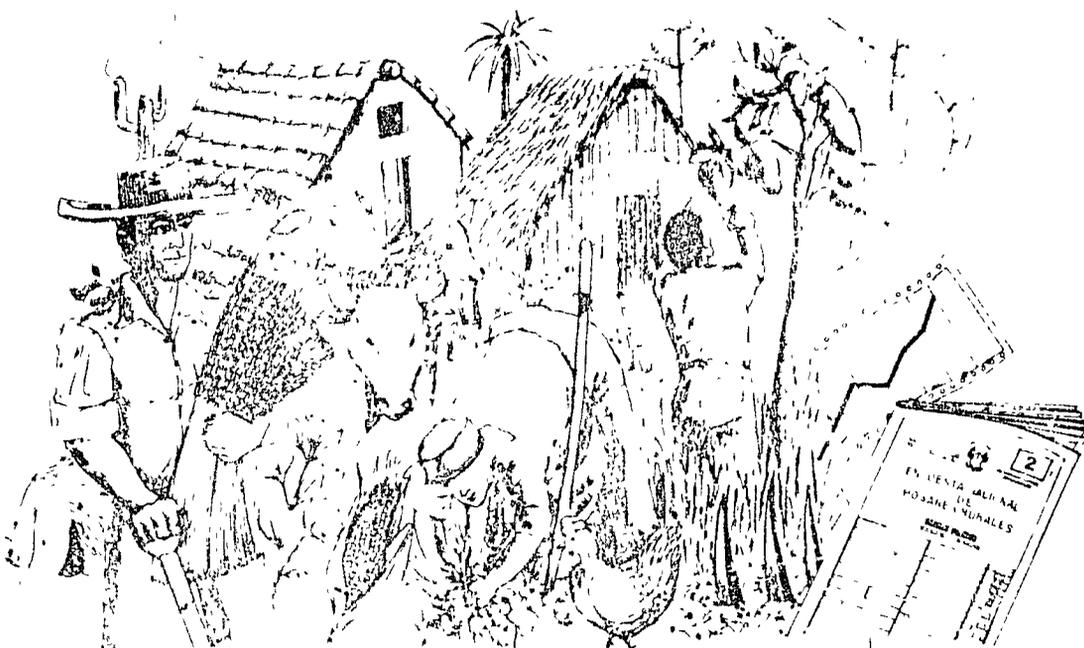


"Año del Cuatricentenario del Nacimiento de Santa Rosa de Lima"

INSTITUTO NACIONAL DE ESTADISTICA
DIRECCION GENERAL DE CENSOS Y ENCUESTAS

MINISTERIO DE AGRICULTURA
OFICINA SECTORIAL DE ESTADISTICA

HOGARES RURALES



Informe Metodológico

PLANEACION AGRICOLA Y DESARROLLO INSTITUCIONAL (PAI)
Proyecto AID-PRU No. 627-0238

DOCUMENTO N°2

Lima, Marzo de 1986

Doctor
ALAN GARCIA PEREZ
Presidente Constitucional de la República

**PRESIDENCIA DEL
CONSEJO DE MINISTROS**

Doctor
Luis Alva Castro
Presidente

**INSTITUTO NACIONAL
DE ESTADISTICA**

Doctora
Graciela Fernández Baca de Valdez
Jefa

Estadístico
Víctor Lozano Díaz
Director Técnico

Investigador Operativo
Atilio Pizarro Huamán
Director General de
Censos y Encuestas

LEY DEL SISTEMA ESTADISTICO NACIONAL
DECRETO LEY 21372 *

Artículo 1º- Establécese el Sistema Estadístico Nacional -SEN-, con la finalidad de asegurar que las actividades estadísticas oficiales se desarrollen en forma integrada, coordinada, racionalizada y bajo una normatividad común.

Artículo 4º- El SEN está conformado por: a.- El Instituto Nacional de Estadística-INE; b.- Las Oficinas Sectoriales de Estadística y demás Oficinas de Estadística de los Ministerios y las Oficinas de Estadística de los Organismos cuyos jefes tienen rango de Ministro, excepto las de la Fuerza Armada; c.- Las Oficinas de Estadística de los Organismos Públicos descentralizados; y d.- Las Oficinas de Estadística de los Gobiernos Locales.

Artículo 5º- En el SEN la actividad técnico-normativa será efectuada por el INE y la actividad operativa se efectuará por las Oficinas de Estadística integrantes del Sistema.

Artículo 9º- Inc. q.- Centralizar la elaboración de los sistemas de cuentas y ejecutar las actividades estadísticas que no pueden ser realizadas por las Oficinas de Estadísticas integrantes del SEN.

* El INE es el Organismo Público Descentralizado encargado de planear, dirigir, coordinar y supervisar las actividades estadísticas del Sistema (Artículo 13º del Decreto Ley 21372); dependiente de la Oficina del Presidente del Consejo de Ministros (Art. 174º del Decreto Legislativo 316).

INSTITUTO NACIONAL DE ESTADISTICA
Dirección General de Censos y Encuestas

MINISTERIO DE AGRICULTURA
Oficina Sectorial de Estadística

ENCUESTA NACIONAL DE HOGARES RURALES

Año Agrícola

Agosto 83 - Julio 84

INFORME METODOLOGICO

Documento Nº 2
Lima, Marzo de 1986

PREPARADO POR : Dirección General de Censos y Encuestas

Composición e Impresión en los Talleres de la Oficina de Comunicación,
Información e Impresiones del Instituto Nacional de Estadística (INE).
Edición 150 ejemplares

Domicilio, Redacción y Talleres: Av. 28 de Julio Nº 1056 - Lima

Abril 1986

555-86-OGCII-INE

PREFACIO

Este documento presenta los procedimientos metodológicos y la secuencia cronológica de las actividades de la Encuesta Nacional de Hogares Rurales (ENahr) que se llevó a cabo en el país. Se espera que las experiencias obtenidas durante su desarrollo sean de utilidad a personas e instituciones en la ejecución de futuras investigaciones, ya que este documento ha sido redactado para documentar la metodología, la organización y los procedimientos usados en la Encuesta Nacional de Hogares Rurales a fin de que los usuarios puedan apreciar tanto sus ventajas como sus limitaciones.

Teniendo en cuenta que este documento será utilizado por personas que no están familiarizadas con el proyecto en general, y con la Encuesta en particular, se ha incluido una descripción de los objetivos del proyecto, y de la ubicación de la Encuesta dentro de él.

Para el beneficio de los usuarios se han preparado cinco publicaciones, cuyo objetivo es el de presentar los procedimientos (expuestos en este documento), conceptos y definiciones, especificaciones para el procesamiento de los datos y los resultados de la ENahr. Las publicaciones son las siguientes:

1. Conceptos y Definiciones
2. Informe Metodológico
3. Especificaciones para el Procesamiento de Datos
4. Guía del Usuario
5. Resultados Definitivos

El presente documento fue redactado por especialistas del Instituto Nacional de Estadística (INE) y los asesores de la Oficina del Censo de los Estados Unidos de América.

Siglas utilizadas en el documento

- AID - Agencia para el Desarrollo Internacional (EE.UU.)
- BUCEN - Bureau of the Census (Oficina del Censo de EE.UU.)
- ENHR - Encuesta Nacional de Hogares Rurales
- GAPA - Grupo de Análisis de Política Agraria
- INE - Instituto Nacional de Estadística
- INIPA - Instituto Nacional de Investigación y Promoción Agropecuaria
- OSE - Oficina Sectorial de Estadística (Ministerio de Agricultura)
- OSPA - Oficina Sectorial de Planificación Agraria
(Ministerio de Agricultura)
- PADI - Planificación Agrícola y Desarrollo Institucional

Informe Metodológico de la Encuesta

	<u>Página</u>
Prefacio	1
Siglas Utilizadas en el Documento	11
CAPITULO I - INTRODUCCION	1
El Proyecto: "Planificación Agrícola y Desarrollo Institucional"	1
La Encuesta: "Encuesta Nacional de Hogares Rurales"	1
CAPITULO II - CONSIDERACIONES BASICAS DE LA ENCUESTA	4
Participación Institucional	4
Objetivos y Finalidad	4
Oportunidad de los Datos	5
Universo y Unidad de Análisis	5
Período de Referencia	6
Método de Recoger la Información	6
Explotación Agropecuaria	6
Ingreso Anual del Hogar	7
Inflación	7
CAPITULO III - DISEÑO DE LA MUESTRA	8
Población y Unidad de Análisis	8
Dominios de Estudio	8
Marco Muestral y Unidades de Muestreo	8
Definición de la Muestra	10
Estratificación	10
Determinación del Tamaño de la Muestra	10
Consideraciones para el Diseño Optimo	11
Selección de la Muestra	12
Resultados de la Selección de la Muestra	24
CAPITULO IV - METODOLOGIA DE ESTIMACION	25
Factor Básico de Expansión	25
Factores de Ajuste por No-entrevistas	25
Factor Final de Expansión	27
Estimación de las Variancias	27
CAPITULO V - DISEÑO DE CEDULAS Y OTROS FORMULARIOS	30
Introducción	30
Consideraciones del Diseño	30
Organización de la Cédula	33
Tipo de Preguntas y Método para Registrar las Respuestas	34
Ayuda Visual	35
Estructura Física	36
Diseño del Registro y otros Formularios	36

	<u>Página</u>
CAPITULO VI - DOCUMENTACION OPERACIONAL	38
Manual del Registrador	38
Manual del Jefe de Brigada (Registro)	39
Manual del Encuestador	39
Manual del Jefe de Brigada (Operación de Campo)	41
Manual de Codificación y Conversión	41
Manual de Crítica por Computadora	42
Manual de Entrada de Datos	44
CAPITULO VII - PRUEBA DE LA OPERACION DE REGISTRO Y DE LA CEDULA	45
Capacitación de los Registradores y Encuestadores	45
Primera Prueba de Campo	46
Evaluación de la Primera Prueba	47
Cambios Debido a la Primera Prueba	49
Segunda Prueba de Campo	50
Resultado y Evaluación	50
CAPITULO VIII - CAPACITACION Y TRABAJO DEL PERSONAL DE CAMPO	52
Capacitación de Registradores	52
Registro de Viviendas	53
Capacitación de los Instructores	54
Capacitación de Encuestadores, Jefes de Brigadas y Departamentales	55
Organización del Trabajo de Campo	56
Deligenciamiento de las Cédulas	57
Control de Calidad en el Campo	58
CAPITULO IX - CAPACITACION Y TRABAJO DEL PERSONAL DE OFICINA	60
Selección y Capacitación de los Crítico-Codificadores	60
Revisión de la Información	61
Codificación	61
Conversión	62
Análisis de Cobertura	62
Chequeo y Proceso de Conversión	63
Chequeo de Rangos	63
Análisis de Consistencia	64
Análisis de Tabulaciones	65
CAPITULO X - PROCESAMIENTO DE DATOS	67
Entrada de los Datos	67
Chequeo de Cobertura y Generación del Archivo de Datos	68
Conversión a Unidades Estándar	69
Crítica de Rango	69
Crítica de Consistencia	70
Ajuste por Inflación	70
Cálculo de Promedios e Imputación	70
Cálculo de Ingresos, Agrupaciones y Adiciones al Archivo.....	70
Tabulaciones	71
Cálculo de Variancias	71

	<u>Página</u>
CAPITULO XI - EVALUACION Y EXACTITUD DE LOS DATOS	72
Tipos de Errores en una Encuesta por Muestreo	72
Origen de los Errores de Muestreo	72
Estimaciones del Error Estándar y Cómo Deben ser Usados	73
Fuentes de Error Ajeno al Muestreo	73
Evaluación de los Errores Ajenos al Muestreo	76
Análisis de Errores	76
CAPITULO XII - RECOMENDACIONES PARA ENCUESTAS SIMILARES EN EL FUTURO	83
Dirección y Administración	83
Diseño Muestral	83
Diseño de Cédulas	84
Operación de Campo	84
Procesamiento de Datos	84
APENDICES -	
1. Departamentos que Conforman cada Dominio de Estudio..	86
2. Número de Segmentos y Viviendas por Dominio	88
3. Formulas para Calcular la Variancia	89
4. Rango de Valores Optimo	90
5. Estimación del C.V. para el Ingreso Promedio por Vivienda	91
6. Coeficientes de la Función Costo y Tasas de Costos para cada Grupo de Estratos	92
7. Descripción de la Función Costo	93
8. Especificaciones de Segmentos	96
9. Documento HR-06, Registro de Viviendas	101
10. Documento HR-07, Listado de las Viviendas Seleccionadas	102
11. Centro Poblados Seleccionados, Número de Encuestadores y Cédulas Diligenciadas	103
12. Distribución de los Segmentos Seleccionados y del Personal de Campo por Departamento para el Registro de Viviendas	104
13. Distribución del Personal de Campo por Región Agraria	105
14. Tasas de Viviendas Sin Encuestar	106
15. Flujogramas del Sistema de Procesamiento de Datos	107
16. Calendario de Principales Actividades	122
FUNCIONARIOS Y PERSONAL TECNICO QUE PARTICIPO EN LA ENCUESTA NACIONAL DE HOGARES RURALES	129

CAPITULO I
INTRODUCCION

El Proyecto: "Planificación Agrícola y Desarrollo Institucional".

El objetivo del proyecto Planificación Agrícola y Desarrollo Institucional (PADI) es el de fortalecer la capacidad del Gobierno del Perú para formular políticas sólidas orientadas al sector agropecuario y de dirigir la ejecución de éstas en forma efectiva. El Proyecto alcanzará su objetivo mediante:

- La satisfacción de la demanda actual y el estímulo de una mayor demanda de información y análisis para el establecimiento de políticas agropecuarias;
- El mejoramiento de la capacidad del gobierno para concebir y evaluar las ventajas y desventajas de las alternativas de políticas agropecuarias;
- El mejoramiento de la información de base para la formulación de ampliación de las políticas agropecuarias;
- El fortalecimiento de la capacidad en las instituciones públicas vinculadas con el sector agropecuario para cumplir eficientemente con sus funciones;
- La ampliación de la base de recursos humanos del Perú en cuanto a su capacidad técnica y administrativa;
- La creación de mecanismos para la participación sistemática y efectiva del sector privado en el proceso de formulación de políticas.

El Proyecto está constituido por cinco componentes:

1. Apoyo al Análisis de Políticas Agropecuarias.
2. Apoyo en Información.
3. Apoyo Administrativo.
4. Apoyo al Desarrollo de Recursos Humanos.
5. Apoyo al Sector Privado.

La Encuesta: "Encuesta Nacional de Hogares Rurales".

Actualmente la economía del Perú en general, y del sector agropecuario en particular, se encuentra en un proceso de transición. Orientar esta transición representa un desafío enorme para los funcionarios encargados de formular la política más adecuada que debe seguir el país. Sin embargo, antes de alcanzar una mejora significativa en el sector agropecuario, hay que solucionar primero los impedimentos que existen en él.

Uno de los principales obstáculos es la falta de datos estadísticos del sector agropecuario necesarios para la planificación a corto, mediano y largo plazo. Durante la década pasada, el gobierno peruano tomó tres iniciativas para mejorar las estadísticas agropecuarias y las relacionadas con los hogares rurales. En 1971 y 1972 se realizó la Encuesta Nacional de Consumo de Alimentos (ENCA) que incluyó aproximadamente 8,000 hogares de los cuales la mitad se encontraban en el área rural del país. ENCA proporcionó una fuente comprensiva de información sobre consumo de alimentos y nutrición. Aunque hubo propuestas por parte de analistas para hacer de la ENCA una encuesta quinquenal, hasta el presente no se ha vuelto a repetir.

En 1972, el Instituto Nacional de Estadística (INE) levantó el II Censo Nacional Agropecuario a un costo aproximado de 3.5 millones de dólares. El censo comprendió información sobre uso de la tierra, producción agrícola, número de animales y otros datos básicos obtenidos de 1'400,000 unidades agropecuarias aproximadamente. Sin embargo, debido a la crisis económica que el país ha venido soportando en los últimos años y el alto costo de realizar un censo agropecuario, no se ha ejecutado otro.

A mediados de la década del setenta, la Oficina Sectorial de Estadística (OSE) del Ministerio de Agricultura creó el Sistema Nacional de Estadística Alimentaria (SINEA). El sistema estaba compuesto por siete subsistemas que trataba con: (i) producción agrícola; (ii) costos de producción; (iii) producción pecuaria; (iv) procesamiento y distribución de productos agrícolas; (v) molinos de granos, camales, y unidades de almacenamiento; (vi) productores de insumos agrícolas; y (vii) consumo urbano. En 1977, el sistema contaba con los recursos financieros y de personal para el desarrollo de sus actividades, pero en los años siguientes los recursos financieros comenzaron a escasear. En 1979, el SINEA sólo pudo cumplir con una parte de sus metas. Posteriormente, el apoyo financiero siguió deteriorándose hasta que se desmanteló la infraestructura establecida.

Tomando en consideración la necesidad de contar con datos de referencia actualizados para formular políticas sectoriales vinculadas a la generación permanente del empleo rural, aumento de la productividad, producción agropecuaria, mejoramiento de las líneas del crédito agropecuario así como incentivar el uso de tecnología adecuada y la asistencia técnica en las actividades agropecuarias, el Gobierno del Perú decidió en 1983, realizar la Encuesta Nacional de Hogares Rurales (ENHR).

La ENHR, que es una de las actividades financiadas bajo el componente: Apoyo en Información del Proyecto "Planificación Agrícola y Desarrollo Institucional" (PADI), comenzó a diseñarse en agosto de 1983. El trabajo de campo, es decir la obtención de la información a través de las entrevistas, se realizó entre octubre y diciembre de 1984. El procesamiento de la información duró catorce meses, de diciembre de 1984 hasta febrero de 1986. Entre setiembre y diciembre de 1983, debido a la prohibición de contratar personal, no se pudo contar con los recursos humanos necesarios para desarrollar las actividades programadas, lo que afectó en forma significativa el desarrollo de la encuesta. La ENHR también se vio afectada por media docena de paros y huelgas en el sector público. Varias de éstas duraron hasta seis semanas, paralizando las actividades emprendidas.

El costo de la ENAHR fue de \$1'466,000 dólares, de los cuales \$1'258,000 dólares fue aportado por la Agencia para el Desarrollo Internacional (AID) de EE. UU., como préstamo y una parte como donación. El Gobierno del Perú contribuyó con \$208,000 dólares para la realización de este esfuerzo de interés nacional.

CAPITULO II

CONSIDERACIONES BASICAS DE LA ENCUESTA

Participación Institucional

En primer término fue necesario hacer una evaluación de las dos principales instituciones productoras de estadísticas a fin de determinar cuál de ellas tenía la capacidad de dirigir este esfuerzo. El INE contaba con personal de planta calificado en técnicas de encuestas, pero no contaba con el número suficiente de personal de campo con experiencia en encuestas agropecuarias. Por otra parte, la OSE disponía del personal de sus Oficinas Regionales los cuales están encargados de recoger datos agropecuarios. Después de analizar las ventajas y desventajas, se optó por encargar a ambas instituciones la realización de la ENAHR en cumplimiento a lo dispuesto en el D.L. 21372 que establece en el Perú el Sistema Estadístico Nacional (SEN).

El INE tuvo la responsabilidad de los aspectos metodológicos de la encuesta tales como el muestreo, diseño de la cédula, capacitación del personal de oficina, preparación de manuales y formularios de control, procesamiento de datos y publicación de los resultados. La OSE tuvo la responsabilidad de la ejecución del trabajo de campo, proporcionando los vehículos y el personal de campo para las entrevistas. La responsabilidad de capacitar el personal de campo y revisar los resultados definitivos de la encuesta fue compartida por ambas instituciones. Esta forma de ejecución permitió que dos instituciones productoras de estadísticas dentro del Sistema Estadístico Nacional trabajaran coordinadamente hacia una meta común, forjando a su vez lazos de cooperación entre ellas.

Además del INE y la OSE, el Grupo de Análisis de Política Agraria (GAPA) que se creó dentro del Ministerio de Agricultura con el objeto de utilizar la información proporcionada por la ENAHR y toda aquella obtenida bajo el componente de Apoyo en Información del proyecto PADI, y la Oficina Sectorial de Planificación Agrícola (OSPA) debían haber desempeñado, según lo acordado en el proyecto, un papel significativo durante el diseño analítico de la encuesta. Específicamente, el GAPA y la OSPA, conjuntamente con el INE y la OSE, debían haber encarado los problemas conceptuales relacionados con las unidades de análisis a estudiarse, variables específicas a incluirse, definición de términos, y otros asuntos que tenían que ver con el diseño analítico de la encuesta. Desafortunadamente el papel que desempeñaron el GAPA y la OSPA fue mínimo. El primero porque no fue constituido oportunamente y el segundo porque estaba desarrollando otro proyecto dentro del PADI.

Por último, la Oficina del Censo de EE. UU. prestó apoyo técnico a lo largo de toda la encuesta.

Objetivos y Finalidad

Para asegurar que los resultados obtenidos a través de la ENAHR cumplieran con los requerimientos de información de los usuarios, se efectuaron una serie de reuniones antes de diseñar la muestra y elaborar los documentos básicos de la encuesta como el Programa (Definiciones y Conceptos), Plan de Tabulación y la

cédula respectiva. El propósito de estas reuniones, donde participaron funcionarios de varias instituciones públicas nacionales y extranjeras, fue el de delinear los objetivos de la encuesta y de definir las variables a investigarse. Como objetivo fundamental se acordó que la ENAHR proporcionara información estadística de las principales características sociales y económicas de la población rural así como de la población urbana cuya actividad estuviese relacionada con la agricultura o ganadería. Los objetivos específicos fueron: (1) proporcionar información sobre empleo e ingresos de los miembros del hogar, características del hogar y sus miembros, producción agropecuaria, crédito agropecuario, comercialización, tecnología y asistencia técnica; y (2) proporcionar información estadística complementaria a la obtenida por las estadísticas agropecuarias en relación con los hogares.

Dentro del marco de estas reuniones se propuso una cuantiosa lista de variables a investigarse. Esta lista de variables y la relación entre ellas sirvió como base para definir los conceptos y diseñar los cuadros preliminares. El desarrollo de los cuadros preliminares fue una de las primeras actividades abordada en la ENAHR, ya que era esencial para diseñar la cédula y asegurar que los datos a obtenerse permitieran alcanzar los objetivos de la encuesta.

La finalidad principal de la encuesta era la de satisfacer los requerimientos de información estadística prioritaria para la formulación y análisis de políticas y planes de desarrollo rural.

Oportunidad de los Datos

Dentro del componente de Apoyo en Información del proyecto PADI, la ENAHR fue la única que se diseñó con el propósito de obtener información sobre las características socio-económicas de los hogares rurales. Teniendo presente la necesidad de encarar la falta de información y proporcionar los resultados de la encuesta lo antes posible, se tomaron varias medidas que acortaron el tiempo que usualmente toma realizar una encuesta de la índole de la ENAHR. Las medidas que se tomaron, que serán explicadas más adelante, fueron en las actividades de diseño de cédula y procesamiento de datos por computadora.

Universo y Unidad de Análisis

Durante las reuniones iniciales entre los funcionarios de las diferentes instituciones nacionales y extranjeras que integraban el grupo de diseño de la encuesta, se acordó que el universo de la ENAHR estuviese constituido por todos los hogares ubicados en el área rural del país, y en el área urbana por los hogares donde por lo menos uno de sus miembros fuese productor agropecuario. La razón principal que se definiera el universo de esta manera era que abarcaba las tres unidades de análisis de mayor interés para los usuarios. Estas son: (1) hogares rurales con productor agropecuario, (2) hogares rurales sin productor agropecuario, y (3) hogares urbanos con productor agropecuario.

Como el propósito de la ENAHR era obtener información de los hogares y no de medir la producción agropecuaria del país, las unidades productoras que no eran hogares tales como las Cooperativas Agrícolas de Producción y otras Empresas Asociativas fueron excluidas del universo.

Período de Referencia

Uno de los principales objetivos de la encuesta fue el de medir el ingreso de los hogares rurales. Tomando en consideración este objetivo y también que la mayoría de los hogares rurales dependen de la agricultura para su sostén, se decidió utilizar el Año Agrícola como período de referencia en lugar del año calendario.

El Año Agrícola como período de referencia tiene principalmente dos ventajas sobre el año calendario en las encuestas de hogares rurales. La primera es que facilita al productor (informante) a ubicar correctamente los hechos ocurridos dentro del tiempo señalado. El productor, se puede decir, está virtualmente sujeto al Año Agrícola cuando llega el momento de tomar decisiones sobre la siembra y cosecha de sus cultivos. Por lo tanto, toda información retrospectiva sobre los cultivos puede ser más precisa cuando el período de referencia coincide con el Año Agrícola. De haberse optado por el año calendario, el productor hubiese tenido que separar mentalmente las actividades que ocurrían en un año y las que ocurrían en el otro debido a que el período seleccionado no correspondía con el del cultivo. La otra ventaja está basada en el análisis y la publicación de los datos. Un análisis de producción y gastos de cultivos, por ejemplo, se puede realizar si los datos se refieren al mismo período de referencia. Para publicar los datos, lo indicado y lógico es utilizar el mismo período de referencia; cosa que no hubiera sido posible si los gastos de producción hubieran sido efectuados en diferentes años calendario.

Método de Recoger la Información

Basado en los fondos disponibles y la necesidad de los usuarios de obtener los resultados lo más pronto posible, se diseñó la ENAHR como una encuesta retrospectiva de visita única. Esto significa que se visitó a los hogares seleccionados una sola vez a fin de que proporcionaran la información solicitada.

Explotación Agropecuaria

Para efectos de la ENAHR se consideró como explotación agropecuaria "...todo terreno o conjunto de terrenos cuya superficie es de un mínimo de 1000 metros cuadrados, aprovechado total o parcialmente para la producción agropecuaria y que es explotado por una persona o grupo de personas, sin consideración de la ubicación ni el régimen de tenencia". Esta definición cumplía con los objetivos de la encuesta debido a que el ingreso del hogar proviene de la producción total de todos los terrenos no importando si están o no dentro del mismo distrito (última división político-administrativa) o si están explotados bajo cualquier modalidad de tenencia de las tierras.

Además del concepto mencionado de explotación agropecuaria, se acordó establecer otros criterios que permitían incluir a aquellas que no tenían tierras o tenían menos de 1000 metros cuadrados. Los límites que se establecieron fueron los siguientes: 3 vacunos, 6 5 porcinos, 6 10 ovinos, 6 10 caprinos, 6 10 auquénidos (alpacas y llamas), 6 100 aves de corral, 6 100

cuyes 6 100 conejos'. La razón por la cual se fijaron estos límites fue que los usuarios estaban particularmente interesados en analizar datos estadísticos de pequeñas explotaciones agropecuarias a fin de formular políticas que permitan un mejoramiento substancial en cuanto a la calidad y nivel de vida de esta población.

Llegar a un consenso general de lo que era una explotación agropecuaria, y los criterios mínimos que tenía que reunir para ser calificada como tal, no fue fácil pero se logró mediante el esfuerzo y la cooperación de los técnicos de las instituciones participantes.

Ingreso Anual del Hogar

A solicitud de los principales usuarios se estableció que el ingreso a calcularse para los hogares sería el ingreso neto de los hogares analizado, por lo que en la elaboración de la cédula se consideró el nivel de desagregación que permitiera deducir los gastos de operación y otros pagos similares. Para el cálculo del ingreso neto anual se consideró todos los recursos monetarios o en especie que percibieron los miembros del hogar cualquiera haya sido su origen. Fueron incluidos los ingresos que se recibieron regular y periódicamente por concepto de trabajo, rentas de la propiedad, transferencias regulares, etc.

Para los hogares con productor agropecuario se consideraron los ingresos percibidos por la venta de la producción agrícola, de la pecuaria y la de los productos elaborados.

Para aquella parte de la producción agropecuaria que fue consumida en el hogar y/o por los animales, la que fue perdida por cualquier razón, la que se dió en trueque, y/o la que se utilizó para semilla (en el caso de la agrícola), se decidió imputarle un valor. Este valor fue derivado de la multiplicación de la cantidad y el precio de venta promedio recibido dentro del mismo dominio por ese producto o animal.

Otros ingresos considerados fueron aquellos que se percibieron por haber dado en alquiler maquinaria, equipo o implementos como tractores, sembradoras, desgranadoras, mochilas, o animales para tracción

Inflación

El Perú ha venido padeciendo una inflación superior al 100% durante los últimos tres años. Si la información relacionada al ingreso que se recogió a través de la encuesta no se hubiera ajustado por inflación, los resultados estarían totalmente sesgados. Este es el caso en particular del sector agrario, donde el lapso de tiempo entre la siembra, la cosecha y, eventualmente, la venta de un cultivo puede tomar 6 o más meses. La diferencia entre el gasto ocurrido y el monto recibido podría ser, por lo tanto, mayor de un 50% causado principalmente por la inflación.

Cuando se diseñó la ENAHR se tomó en cuenta el efecto que podría causar la inflación en los resultados de la encuesta. Para obviar esta situación, se solicitó el mes en que ocurrió cualquier transacción para más tarde utilizarla en el ajuste de los datos.

CAPITULO III

DISEÑO DE LA MUESTRA

Población y Unidad de Análisis

La población bajo estudio estuvo conformada por todas las viviendas particulares y sus ocupantes residentes en el área rural, la misma que fue definida como la población ubicada en las áreas de empadronamiento rural (AER) utilizadas en el Censo de Población y Vivienda de 1981 y en los centros poblados urbanos que tenían menos de 2,000 habitantes. Incluyó asimismo las viviendas particulares con productor agropecuario ubicadas en los centros poblados urbanos que tenían 2,000 ó más habitantes y por lo menos el 20% de viviendas con productor agropecuario.

La unidad de análisis es el hogar con productor agropecuario que corresponde a una explotación agropecuaria en las áreas urbanas y rurales, como también los hogares sin productor agropecuario en las áreas rurales.

Dominios de Estudio

Frecuentemente los analistas requieren información por separado para cada parte o sector de la población bajo estudio, para los cuales es necesario obtener estimaciones. Cada parte se le conoce como dominio. Comúnmente, dichos dominios se refieren a regiones y/o zonas geográficas, como fue en el caso de la ENAHR donde se consideraron los siguientes dominios:¹

1. Costa Urbana (con productor agropecuario)
2. Sierra Urbana (con productor agropecuario)
3. Selva Alta Urbana (con productor agropecuario)
4. Selva Baja Urbana (con productor agropecuario)
- 5-6 Costa Norte Rural (con y sin productor agropecuario)
- 7-8 Costa Centro Rural (con y sin productor agropecuario)
- 9-10 Costa Sur Rural (con y sin productor agropecuario)
- 11-12 Sierra Norte Rural (con y sin productor agropecuario)
- 13-14 Sierra Centro Rural (con y sin productor agropecuario)
- 15-16 Sierra Sur Rural (con y sin productor agropecuario)
- 17-18 Selva Alta Norte Rural (con y sin productor agropecuario)
- 19-20 Selva Alta Centro Rural (con y sin productor agropecuario)
- 21-22 Selva Alta Sur Rural (con y sin productor agropecuario)
- 23-24 Selva Baja Rural (con y sin productor agropecuario)

Marco Muestral y Unidades de Muestreo

El marco muestral se basó en los datos del Censo de Población y Vivienda de 1981, como también en los materiales cartográficos y documentos auxiliares utilizados en el empadronamiento. Se utilizó un marco de "áreas," destinado a cubrir el área rural completa del país y también el área de los centros

^{1/} Los departamentos o la parte de ellos que integran cada dominio son presentados en el Apéndice 1.

poblados que tenían una población de 2,000 ó más habitantes y por lo menos 20% de viviendas con productor agropecuario. Dado que el diseño de la muestra fue multietápico, el marco muestral en el área urbana y rural se dividió en unidades administrativas o conglomerados, correspondientes a cada etapa de selección.

En el área urbana, las unidades de muestreo para cada etapa se definieron de la siguiente manera:

- Primera etapa: Se consideró como unidad primaria de muestreo (UPM) urbana a todo centro poblado urbano que tenía, según los resultados del Censo de 1981, 2,000 ó más habitantes y por lo menos el 20% de las viviendas con productor agropecuario.
- Segunda etapa: Se consideró como unidad secundaria de muestreo (USM) urbana al conglomerado formado por una o más manzanas, con un promedio de 100 viviendas particulares.
- Tercera etapa: Se consideró como unidad terciaria de muestreo (UTM) urbana a la vivienda particular con productor agropecuario.

En la parte del marco que corresponde al área rural y en lo referente a los centros poblados urbanos con menos de 2,000 habitantes, se definieron las unidades de muestreo de la siguiente manera:

- Primera etapa: Se consideró como UPM a cada centro poblado urbano con menos de 2,000 habitantes.
- Segunda etapa: Se consideró como USM a cada conglomerado formado por una o más manzanas con promedio de 100 viviendas particulares.
- Tercera etapa: Se consideró como UTM a cada vivienda particular con o sin productor agropecuario.

En la parte del marco que corresponde al área rural y en lo referente a los AER, se definieron las unidades de muestreo de la siguiente manera:

- Primera etapa: Se consideró como UPM a cada AER del Censo de 1981 que tenía en promedio 100 viviendas particulares distribuidas en uno o más centros poblados rurales. En el caso de que el AER sea muy grande o difícil de registrar fue necesario subdividirlo para obtener tamaños de conglomerados adecuados para el registro de viviendas.
- Segunda etapa: Las USM en el área rural estaban constituidas por las viviendas particulares con o sin productor agropecuario ubicadas dentro de las USM rurales seleccionadas.

Definición de la Muestra

La muestra fue probabilística, multietápica y estratificada. Se consideró probabilística porque las unidades fueron seleccionadas mediante métodos aleatorios, lo cual permite efectuar inferencias a la población basadas en la teoría de las probabilidades. La muestra fue independiente en cada estrato. En el área urbana, se utilizó una muestra trietápica 2/, y en el área rural se utilizó una muestra bietápica. Se decidió usar una muestra bietápica con la finalidad de dispersar más los segmentos dentro de cada dominio, considerando la factibilidad del trabajo de campo.

Cabe señalar que las unidades en cada etapa fueron seleccionadas con probabilidades proporcionales al tamaño, excepto en la última etapa, con el objeto de mantener una muestra aproximadamente autoponderada dentro de cada estrato. En la última etapa se seleccionó un número fijo de viviendas para cada subestrato (vivienda con productor o sin productor) dentro del segmento muestreado.

Estratificación

Para propósitos de estratificación, se dividió la población bajo estudio en 14 estratos según las regiones geográficas y áreas urbana y rural. Dichos estratos corresponden a los dominios geográficos definidos para la encuesta (ver Apéndice 1). Dado la homogeneidad de la población dentro de cada dominio, esta estratificación contribuyó significativamente a la eficiencia de la muestra. Dentro de cada estrato se subestratificaron los segmentos por departamentos y, al interior de ellos, se ordenaron las UPM según el porcentaje de viviendas con productor agropecuario, a fin de obtener una subestratificación implícita como resultado de la aplicación de la selección sistemática, mejorando así la representatividad de la muestra.

Determinación del Tamaño de la Muestra

El límite superior para el tamaño de la muestra inicialmente fue fijada en 5,000 viviendas. Posteriormente el tamaño fue incrementado a 7,000 viviendas, cuando se decidió incluir la Selva Baja y Selva Alta, dividida en Norte, Centro y Sur. En base a experiencias previas, se consideró que un promedio de 300 viviendas seleccionadas por dominio sería adecuada para satisfacer los principales objetivos de análisis de los usuarios. Para dicho análisis de los datos, se estableció tener un coeficiente de variación (CV) máxima de 20% para las variables importantes, tal como el ingreso a nivel de dominio. El tamaño de muestra necesario en cada dominio para conseguir un CV de 20% basada en estimaciones de la variancia y correlación intraclase para ingreso promedio por hogar se determinó usando los datos del Censo de Población y Vivienda de 1981. Con la restricción de que el tamaño total de la muestra no fuera mayor que 7,200, se fijó este número en función de los recursos humanos y financieros disponibles para la encuesta, así mismo, este tamaño permitía cubrir una tasa de no respuesta razonable.

2/ Es pertinente señalar que los centros poblados urbanos con más del 20% de viviendas con productor se utilizaron únicamente como "unidades de conteo" a fin de muestrearlos en una etapa preliminar en forma sistemática y con probabilidades proporcionales al tamaño; este procedimiento equivale a una selección bietápica.

Consideraciones para el Diseño Óptimo

Preliminarmente se concibió un diseño que consistió en seleccionar 30 conglomerados de muestreo en cada estrato con un promedio de 10 viviendas seleccionadas por conglomerado. En el caso de los estratos rurales, los mismos conglomerados seleccionados en la muestra fueron utilizados para seleccionar viviendas con productor y sin productor. El número de conglomerados y viviendas a seleccionar dentro de cada estrato en el diseño preliminar de la muestra se presenta en el Apéndice 2.

A base de la información del Censo de Población y Vivienda de 1981, se hicieron estimaciones de la variancia y del coeficiente de correlación intraclase para el ingreso promedio por vivienda en cada estrato, con el objeto de determinar un diseño aproximadamente óptimo que permitiera ajustar el diseño preliminar en el supuesto caso que variaría significativamente del óptimo. Los datos de ingreso fueron obtenidos de la muestra del 25% de viviendas que le correspondió la Cédula 2 (muestral) en el Censo de Población y Vivienda de 1981. Considerando las dificultades que presenta la medición de la variable ingreso, no se esperaba que los datos del Censo de Población y Vivienda de 1981, proporcionaran cifras precisas sobre la variancia y el coeficiente de correlación intraclase para el ingreso por vivienda; sino, simplemente tener una idea general de su magnitud a la que se pueda tomar como referencia.

Las fórmulas usadas para calcular la variancia y los coeficientes de correlación intraclase se encuentran especificadas en el Apéndice 3, las cuales fueron derivadas de fórmulas estándares del libro "Sample Survey Methods and Theory", Volumen 1, de Hansen, Hurwitz y Madow (páginas 259 y 260). La fórmula para el coeficiente de correlación intraclase asume que los conglomerados dentro de cada estrato son del mismo tamaño; sin embargo, en la realidad se encontró que los tamaños (números de viviendas) de los conglomerados son variables por su naturaleza. La limitación de esta fórmula es que el componente del coeficiente de correlación intraclase que corresponde a la variancia entre conglomerados no toma en cuenta las ponderaciones diferenciales para tamaños variables de dichos conglomerados. Dadas estas limitaciones de los datos y los procedimientos de estimación, el diseño óptimo de la muestra basado en tabulaciones del censo fue tomada en cuenta sola como referencia.

El coeficiente de correlación intraclase para cada dominio (δ_h) estimado para la variable ingreso, se presenta en el Apéndice 4 en donde se puede ver que en la mayoría de los casos las estimaciones de los coeficientes son razonables, pero existen algunos casos particulares. Por ejemplo, existen 2 coeficientes con signo negativo y otro que es ligeramente mayor que 1. Estos casos probablemente corresponden a dominios con gran variabilidad en el número de viviendas por conglomerado y/o a una alta tasa de no respuesta para la variable ingreso. Es muy improbable que el verdadero coeficiente de correlación intraclase para cualquier dominio sea negativa o mayor que 0.50. Por lo tanto, para los propósitos del diseño óptimo del estudio, se decidió fijar un

límite inferior de 0.04 y un límite superior de 0.50 para todos los coeficientes. Los valores extremos fuera de este rango fueron fijados en los valores máximos o mínimos según como correspondía. En los apéndices 4 y 5 se presentan los valores ajustados del coeficiente de correlación intraclase entre parénesis.

A fin de determinar el mínimo de viviendas a seleccionar por conglomerado que sería aproximadamente óptimo para cada dominio, se calcularon tasas de costos para 8 grupos de estratos, basado en estimaciones aproximadas de los coeficientes de la función costo para cada grupo. Dichos coeficientes y las tasas correspondientes son presentados en el Apéndice 6. El Apéndice 7 contiene una descripción de la función costo y la estimación de los coeficientes.

Las tasas de costo (C_0^2/CC_2 y C_1/C_2) fueron usadas conjuntamente con las estimaciones del coeficiente de correlación intraclase para determinar el número óptimo de viviendas a seleccionar por conglomerado para cada estrato (\bar{n}_h), como también el rango de valores de \bar{n}_h que no aumentarían la variancia mínima en más del 10%. Usando la tabla 7 de la página 96 del libro "Sample Survey Methods and Theory," se usaron interpolaciones aproximadas para determinar los rangos. El rango de valores óptimo para \bar{n}_h para cada dominio se presenta en el Apéndice 4. Se puede ver que para la mayoría de los dominios, el rango de valores óptimos de \bar{n}_h incluye 10, lo cual ha demostrado que el diseño preliminar de la muestra basado en \bar{n}_h igual a 10 viviendas por conglomerado para cada dominio fue razonable. Los rangos de \bar{n}_h que no incluyen 10 viviendas por conglomerado corresponden a los dominios con un coeficiente de correlación intraclase estimado en 0.50 o más. En el Apéndice 4 se puede ver que si se debe usar un n_h óptimo, éste sería menor para dichos dominios. Dados los problemas relacionados con la estimación de los coeficientes de correlación intraclase, estos coeficientes estimados (0.50 ó más) no deben ser muy confiables. Sin embargo, es interesante señalar que en el caso de la Costa Norte Rural, Costa Centro Rural y Selva Baja Rural, los dominios con productor tienen un coeficiente de correlación intraclase bastante alto (más de 0.50), mientras que los dominios sin productor tienen un coeficiente de correlación intraclase bajo (0.04 - 0.06). Los datos del ingreso a base del Censo pueden ser menos confiables para las viviendas con productor que para las viviendas sin productor, puesto que es lógico pensar que la mayoría de los agricultores no tienen un sistema adecuado de contabilidad que le permita estimar su ingreso. Esto dió por resultado estimaciones no confiables del coeficiente de correlación intraclase para las viviendas con productor. Por lo tanto, no se modificó el diseño preliminar de la muestra basado únicamente en estas estimaciones.

Para la ENAHR se determinó el ingreso para los hogares con productor indirectamente, es decir, estimando el valor y costo de producción. Quizás un estudio del coeficiente de correlación intraclase se puede hacer posteriormente utilizando los datos de la encuesta.

El Apéndice 5 presenta los coeficientes de variación³ para la estimación del

$$\text{3/ } CV(\bar{x}_h) = \sqrt{\frac{\sigma_h^2}{m_h \bar{n}_h \bar{x}_h^2} [1 - \delta_h (\bar{n}_h - 1)]}$$

ingreso promedio por vivienda para cada dominio que resultaría del diseño preliminar de la muestra ($m_h = 30$, $\bar{n}_h = 10$) basado en las estimaciones de δ_h y σ_h^2 , de los datos del Censo. Se puede ver que los coeficientes de variación estimados para la mayoría de los estratos son bajos, y el más alto se encuentra por debajo del límite superior del 20% especificado por los analistas. Dicho CV (0.19) corresponde al dominio de la Sierra Sur Rural con productor, que tenía un estimado para \hat{c}_h (1.01) de baja confiabilidad.

El Apéndice 5 también indica el número de conglomerados a seleccionar en cada dominio para obtener un CV de 20% para el ingreso promedio por vivienda, asumiendo un $\bar{n}_h = 10$. En razón de que las estimaciones de los CV que resultan de 30 conglomerados por dominio son bajos, consecuentemente las estimaciones de m_h serán igualmente bajas. El valor más alto para m_h , que es de 26, es también lógicamente para el dominio de Sierra Sur Rural con productor. Aunque algunos dominios tienen un $m_h = 1$, es pertinente señalar que no existe diseño muestral que no incluya por lo menos dos conglomerados por dominio.

De acuerdo con los resultados del Apéndice 5, el diseño preliminar de la muestra debe satisfacer los requerimientos para la precisión de las estimaciones más importantes de la encuesta, como el ingreso promedio por hogar y, posiblemente, se hubiese podido reducir un poco la cifra preliminar de 7,200 viviendas. Sin embargo, fue recomendable mantener este tamaño máximo de la muestra, porque algunas estimaciones de la encuesta van a ser requeridos para características poco comunes para las cuales los resultados del estudio no podrían garantizar que el CV no sea mayor del 20%. En términos generales, este tamaño máximo de la muestra va a mejorar la precisión de las estimaciones de las características medidas por la encuesta. Otra razón de utilizar el tamaño máximo de la muestra es que las estimaciones de los CV y m_h pueden estar subestimados, considerando los problemas relacionados con la calidad de los datos del Censo; específicamente, sobre la variable ingreso, característica ésta muy difícil de medir con precisión. Aparte del problema de falta de respuesta para la variable ingreso, también se supuso que había problemas de subestimaciones de los ingresos, dado que por lo general las personas tienen reserva de proporcionar dicho tipo de información a los que ellos perciben como representantes, agentes o servidores del gobierno.

La razón principal de los bajos CV en el Apéndice 5 es la baja variancia relativa (σ_x^2/x^2) para la variable ingreso. La variancia relativa promedio por dominio era de 0.29; sin embargo, en el año 1977, esta variancia relativa promedio en la Encuesta de Hogares del Valle Sur en Bolivia fue de 2.73, lo cual estaría indicando mucho menos variabilidad del ingreso en el Perú. La subestimación de los ingresos altos introduciría un sesgo negativo en las estimaciones de la σ_x^2 y el CV.

La conclusión general a que se llega de los apéndices 4 y 5 es que el diseño preliminar de la muestra que consiste de 30 conglomerados seleccionados por estrato con un promedio de 10 viviendas seleccionadas por conglomerado, es muy razonable y consistente respecto al diseño óptimo. También, es importante tener un diseño más o menos estandarizado para todos los estratos, con la finalidad de simplificar el procesamiento y las estimaciones de las variancias.

Selección de la Muestra

Anteriormente se definió que las unidades primarias de muestreo (UPM) eran de dos tipos: Centros Poblados Urbanos (CCPP) y Areas de Empadronamiento Rural (AER). En los dominios urbanos las UPM son los CCPP con 2,000 ó más habitantes y con un mínimo de 20% de viviendas con productor agropecuario. En los dominios rurales las UPM son de dos tipos: por un lado están las AER que son las áreas típicamente rurales; y por otro, los CCPP de 500 a menos de 2,000 habitantes pero con más de 100 viviendas (límite establecido por los Censos de 1981 para considerar a un centro poblado urbano). Sin embargo, se debe señalar que dentro de la definición de centro poblado urbano, el Censo incluyó también a los centros poblados capitales de distritos no importando si tenían menos de 100 viviendas. Es pertinente indicar que esta última consideración traía como consecuencia algunos riesgos. Por ejemplo, si un centro poblado tenía muy pocas viviendas, era probable que dicha UPM no pudiera satisfacer el mínimo requerido de 10 viviendas con productor y 10 viviendas sin productor.

Las unidades secundarias de muestreo (USM) para los centros poblados urbanos estuvieron conformadas por una o más manzanas que en conjunto daban un total aproximado de 100 viviendas y, en las AER, las USM vienen a ser las últimas unidades de muestreo, o sea, las viviendas entrevistados.

1. Primera Etapa para la Muestra Urbana

En el área urbana la selección de la muestra se realizó en tres etapas, siendo la primera una etapa previa en la cual los CCPP han sido usados como "unidades de conteo." Para la selección de las UPM se efectuó el siguiente procedimiento:

- (1) se elaboró un listado de CCPP urbanos (UPM) con 2,000 y más habitantes y con por lo menos 20% de viviendas con productor agropecuario y clasificadas por dominio;
- (2) se dividió el total de viviendas con productor agropecuario del dominio entre el número de UPM a seleccionarse, para determinar el primer intervalo;
- (3) las UPM con un número de viviendas con productor mayor que el primer intervalo, pasaron a ser UPM autorepresentadas (probabilidad = 1) y el resto de las UPM las denominamos no autorepresentadas, es decir, entraron al proceso de selección propiamente dicho;
- (4) para determinar el número de USM a seleccionarse en las UPM autorepresentadas y no autorepresentadas, se efectuaron los siguientes pasos:

Paso No. 1: Se elaboró un listado de las UPM autorepresentadas y se sumó el número total de viviendas con productor agropecuario de dichas UPM (N_{AR}).

Paso No. 2: Se confeccionó un listado de las UPM no autorepresentadas y se acumuló el número de viviendas con productor agropecuario para cada UPM (N_{NAR}).

Paso No. 3: Se distribuyó el número de USM a seleccionarse proporcionalmente al número de viviendas con productor entre las UPM autorepresentadas y no autorepresentadas, utilizando las siguientes fórmulas:

- Número de USM a seleccionar en las UPM autorepresentadas (m_{AR}).

$$m_{AR} = m \times \frac{N_{AR}}{N_{AR} + N_{NAR}}$$

- Número de USM a seleccionar en las UPM no autorepresentadas (m_{NAR})

$$m_{NAR} = m \times \frac{N_{NAR}}{N_{AR} + N_{NAR}}$$

donde:

m = número de USM a seleccionar para cada estrato.

Paso No. 4: Se distribuyeron las USM a seleccionarse (m_{AR}) entre las UPM autorepresentadas en forma proporcional al número de viviendas con productor agropecuario, utilizando la siguiente fórmula:

$$m_i = \frac{N_i}{N_{AR}} \times m_{AR}$$

donde:

m_i = número de USM a seleccionarse para la UPM autorepresentada "i"

N_i = número de viviendas con productor agropecuario en la UPM autorepresentada "i"

- (5) Para la selección de las UPM urbanas no autorepresentadas dentro de su respectivo dominio, se cumplieron los siguientes pasos:

Paso No. 1: Se elaboró un listado de las UPM no autorepresentadas y se acumuló el número de viviendas con productor para cada UPM (N_{AR}).

Paso No. 2: Se dividió N_{NAR} entre m_{NAR} para obtener el segundo intervalo (I_{NAR}) para la selección de las UPM no autorepresentadas.

Paso No. 3: Se escogió un número aleatorio (A_{NAR}) entre 0 y I_{NAR} , para el arranque de la selección sistemática.

Paso No. 4: Se efectuó la selección sistemática de las UPM utilizando la siguiente fórmula:

$$S_i = A_{NAR} + (i - 1) I_{NAR} \quad , \quad \text{redondeando al próximo entero.}$$

$$i = 1, 2, \dots, m_{NAR}$$

2. Segunda Etapa para la Muestra Urbana

Para la selección de las USM se ejecutaron los siguientes pasos:

Paso No. 1: Una vez seleccionadas las UPM autorepresentadas y no autorepresentadas, se obtuvo un listado de las USM que las conformaban, ordenado de acuerdo al porcentaje de viviendas con productor para cada UPM, y acumulando el número de viviendas con productor para cada USM.

Paso No. 2: Se dividió el acumulado del número de viviendas con productor para la UPM entre el número de USM a seleccionarse previamente establecidos, para obtener el intervalo de selección (I_g).

Paso No. 3: Se escogió un número aleatorio (A_g) entre 0 e I_g para el arranque de la selección sistemática.

Paso No. 4: Se efectuó la selección sistemática de las USM utilizando la siguiente fórmula:

$$S_j = A_g + (j - 1) I_g, \quad \text{redondeando al próximo entero.}$$

$$j = 1, 2, \dots, m_i$$

La USM correspondiente al número seleccionado S_j es la que tiene el acumulado más próximo a S_j sin excederlo.

3. Tercera Etapa para la Muestra Urbana

Para la selección de las unidades últimas de muestreo (viviendas), se ejecutaron los siguientes pasos:

Paso No. 1: Las USM seleccionadas fueron objeto de un registro detallado de las viviendas que se localizaban dentro de los límites del segmento seleccionado. El objetivo fue obtener información adecuada para muestrear solamente las viviendas con productor agropecuario.

Paso No. 2: A base de la información obtenida en dicho registro, se procedió a identificar las viviendas con productor, para lo cual previamente se tuvo que convertir las unidades de medida a una unidad estándar (hectáreas).

Paso No 3: Se procedió a seleccionar 10 viviendas con productor para cada segmento. Sin embargo, en los estratos en donde el número de segmentos con menos de 10 viviendas con productor era significativo, se procedía a incrementar el número de viviendas a muestrear en los otros segmentos. Tal fue el caso del Estrato 2 (Sierra Urbana), antes descrito, y del Estrato 4 (Selva Baja Urbana), en el cual se incrementó el número de viviendas por segmento a 11.

Paso No. 4: Aunque originalmente se consideró seleccionar una muestra autoponderada, después de una evaluación de la relación costo/beneficio se decidió que era mas importante fijar dentro de cada estrato el número de viviendas a seleccionar en cada segmento. Esto resultó en que la probabilidad de selección de una vivienda fue:

$$P_{hij} = P_1 \cdot P_2 \cdot P_3 = \frac{N_{hi}}{N_h/30} \cdot \frac{N_{hij}}{N_{hi}} \cdot \frac{10}{N'_{hij}} = \frac{300}{N_h} \cdot \frac{N_{hij}}{N'_{hij}},$$

donde:

- N_h = número total de viviendas con productor en el estrato urbano "h"
- N_{hi} = número de viviendas con productor en el centro poblado urbano "i" (UPM_i) del estrato urbano "h"
- N_{hij} = número de viviendas con productor en el conglomerado "j" (USM_j) del centro poblado urbano "i" del estrato urbano "h"
- N'_{hij} = número de viviendas con productor encontrados en el registro para la encuesta en el conglomerado "j" del centro poblado urbano "i" del estrato urbano "h"

Determinada la probabilidad de selección por vivienda se procedió a la selección de las mismas siguiendo un procedimiento similar al que se describe para el área rural.

4. Primera Etapa para la Muestra Rural

Tal como se definió anteriormente, en los dominios rurales las UPM son de dos tipos: centros poblados de 500 a menos de 2,000 habitantes que fueron considerados urbanos en el Censo de 1981, y las áreas de empadronamiento rural (AER). Para la selección de las UPM se cumplió el siguiente procedimiento:

- (1) Las UPM fueron agrupadas a nivel de cada dominio en dos subpoblaciones: centros poblados urbanos con menos de 2,000 habitantes y AER.
- (2) Las UPM que corresponden a ambas subpoblaciones fueron clasificadas por departamentos y, al interior de éstos, se ordenaron según el porcentaje de viviendas con productor agropecuario.
- (3) Se listaron las UPM así clasificadas y ordenadas, acumulando el número de viviendas para cada subpoblación.

- (4) Luego se distribuyó el número de UPM a seleccionarse en forma proporcional al número total de viviendas de la subpoblación CCPP urbanos y de la subpoblación AER, utilizando las siguientes fórmulas:

- Número de UPM a seleccionarse en la subpoblación de CCPP urbanos:

$$m_u = m \times \frac{N_u}{N_u + N_r}$$

- Número de UPM a seleccionarse en la subpoblación de AER.

$$m_r = m \times \frac{N_r}{N_u + N_r}$$

donde:

m = Número de UPM a seleccionarse para el estrato

N_u = Número total de viviendas en la subpoblación CCPP urbanos

N_r = Número total de viviendas en la subpoblación AER

- (5) Para la selección de las UPM en la subpoblación de CCPP urbanos, se cumplieron los siguientes pasos:

Paso No. 1: Se calculó el intervalo de selección ($I_u = N_u/m_u$) y se revisó el tamaño de los CCPP a fin de verificar si existía alguno que sobrepasara dicho intervalo con el objeto de considerarlo en la categoría de autorepresentado. Por consiguiente, se llegó a comprobar que no existían en el marco de estratos rurales, CCPP autorepresentados, de tal manera que todos los centros poblados en cada uno de sus estratos entraron en el proceso de selección sistemática.

Paso No. 2: Habiendo calculado el intervalo de selección de CCPP (I_u) se procedió a escoger un número aleatorio (A_u) entre 0 e I_u .

Paso No. 3: Se procedió a la selección sistemática de los CCPP utilizando la siguiente fórmula:

$$S_i = A_u + (i - 1) I_u \quad \text{redondeando al próximo entero}$$

$$i = 1, 2, \dots, m_u$$

El centro poblado correspondiente al número de selección S_i , es el que tiene el acumulado más próximo a S_i , sin excederlo.

(6) Para la selección de las UPM en la subpoblación de las AER, se cumplieron los siguientes pasos:

Paso No. 1: Se dividió el acumulado del número de viviendas para la subpoblación (N_r) entre m_r , para determinar el intervalo de selección de las AER (I_r).

Paso No. 2: Se obtuvo un arranque aleatorio (A_r) entre 0 e I_r .

Paso No. 3: Se procedió a la selección sistemática de los AER utilizando la siguiente fórmula:

$$S_i = A_r + (i - 1) I_r \quad \text{redondeando al próximo entero}$$

$$i = 1, 2, \dots, m_r$$

El AER correspondiente al número seleccionado S_i , es el que tiene el acumulado más próximo a S_i , sin excederlo.

5. Segunda Etapa para la Muestra Rural

Para la selección de las USM se siguieron dos procedimientos diferentes dependiendo de la UPM de la cual se seleccionó. Tal como se indicó anteriormente, las UPM en los estratos rurales fueron de dos tipos: centros poblados urbanos que tenían menos de 2,000 habitantes y áreas de empadronamiento rural. En el caso de los centros poblados urbanos, las USM estuvieron conformadas por una o más manzanas que en total sumaban aproximadamente 100 viviendas y para la selección de ellas se siguieron los siguientes pasos:

Paso No. 1: Para cada UPM seleccionada se obtuvo un listado de las USM que la conformaban ordenado de acuerdo al porcentaje de viviendas con productor y acumulando el número de viviendas de cada USM.

Paso No. 2: Se escogió un número aleatorio (A_g) entre 0 y el número de viviendas de la UPM para la selección de la USM.

Paso No. 3: Las USM seleccionadas fueron objeto de un detallado registro de viviendas de donde después de identificar las viviendas con productor y sin productor en forma separada se procedió a la selección de las mismas en forma sistemática y siguiendo el mismo procedimiento que se describe en la selección de la muestra para las AER.

En el caso de los estratos rurales, es muy difícil obtener una muestra autoponderada, ya que es necesario seleccionar las viviendas con productor y sin productor en los mismos segmentos. La probabilidad de selección de una vivienda con productor es:

$$P_{hi(cp)} = P_1 \cdot P_2 = \frac{N_{hi}}{N_h/30} \cdot \frac{10}{N_{hi(cp)}} = \frac{300}{N_h} \cdot \frac{N_{hi}}{N_{hi(cp)}}$$

donde:

N_h = Número total de viviendas en el estrato rural "h".

N_{hi} = Número total de viviendas en el UPM_i del estrato rural "h".

$N_{hi}(cp)$ = Número de viviendas con productor encontradas en el registro para la encuesta en la UPM_i del estrato rural "h".

De la misma forma, la probabilidad de selección de una vivienda sin productor es:

$$P_{hi}(sp) = P_1 \cdot P_2 = \frac{N_{hi}}{N_h/30} \cdot \frac{10}{N_{hi}(sp)} = \frac{300}{N_h} \cdot \frac{N_{hi}}{N_{hi}(sp)}$$

donde:

$N_{hi}(sp)$ = Número de viviendas sin productor encontradas en el registro para la encuesta en la UPM_i del estrato rural "h".

De la observación de las probabilidades se concluye que la muestra rural solamente podría haber sido autoponderada en cada dominio si teóricamente se hubiera considerado que todos los AER hubieran tenido 50% de las viviendas con productor y 50% sin productor, o si se hubiera fijado la probabilidad de la última etapa para conseguir una muestra autoponderada de 600 viviendas sin tomar en cuenta lo característico con y sin productor (alternativa no viable porque se necesitaba hacer estimaciones al nivel de dominio). Por lo tanto, se decidió por una muestra con probabilidades desiguales.

Para la selección de las USM (viviendas) se cumplieron con los siguientes pasos:

Paso No. 1: Las UPM seleccionadas fueron objeto de un registro detallado de las viviendas que se localizaban dentro de los límites del segmento seleccionado, con la finalidad de obtener información adecuada para muestrear separadamente viviendas con productor y sin productor. El registro tuvo como objetivo determinar la dirección de la vivienda, el nombre del jefe de familia y si alguno de los ocupantes de esa vivienda era productor agropecuario o no.

Paso No. 2: Concluida la labor del registro de viviendas, se procedió a su revisión encontrando los siguientes puntos saltantes:

(1) Diez segmentos no pudieron ser trabajados por problemas socio-políticos y de narcotráfico, de los cuales seis segmentos se ubicaron en el departamento de Ayacucho, provincia de La Mar (estrato 12); uno en el departamento de Apurímac, provincia de Aymaraes (estrato 10); uno en el departamento de San Martín, provincia de Mariscal Cáceres (estrato 11); y dos en el departamento del Cusco, provincia de la Convención (estratos 13 y 14). De acuerdo a las informaciones detalladas de los jefes de brigada la situación era tan riesgosa en algunos segmentos que corría peligro la vida misma del personal de

campo.⁴ Frente a esta situación se estudiaron varias alternativas. Una de ellas fue la de esperar un momento más adecuado que podría ser en la etapa de entrevista y efectuar, simultáneamente al registro de viviendas, la selección de las mismas y la entrevista. La otra alternativa fue la de coordinar con el Ministerio de Educación para ver si era posible contar con la colaboración de los maestros; y por último, si después de agotar todos los recursos no era posible efectuar el registro, se podría reconstruir un registro en base de la información del Censo, del cual se seleccionaría la muestra con todas las limitaciones del caso que ello significa. Se decidió por esta última alternativa, aunque sin éxito en el momento de la entrevista.

- (2) En otro segmento se había podido determinar que las viviendas que figuraban en los documentos auxiliares⁵ pertenecían a otro segmento, resultando que de las 123 viviendas que se suponía deberían existir, solo fueron registradas 18. Después de efectuar las averiguaciones del caso y de analizar la situación de la cartografía y su comparación con el documento CPV-03.51, se decidió que el registrador debería delimitar, en el mismo terreno, el ámbito del segmento y luego proceder a efectuar el registro.
- (3) Otra situación, aunque similar a la anterior, fue de un segmento que había salido seleccionado con 112 viviendas y que solamente se habían registrado 29, porque dentro del segmento existía un centro poblado en su mayoría con cantinas y gente de dudosa conducta. En este caso se dispuso, igualmente, que se retornara al campo a fin de completar el registro de dicho centro poblado, para lo cual se solicitó el apoyo de la municipalidad.
- (4) Hubo también un segmento donde el registrador había entendido que debería registrar únicamente las viviendas con productor; es decir, había dejado de lado a las viviendas sin productor. Después de analizar los segmentos vecinos se pudo ver que el número de viviendas sin productor en dicha área era insignificante y por los altos costos que acarrearía no se justificaba volver al terreno.
- (5) Existió un segmento en el cual se observó que la única vía para llegar era por avión. Se decidió sin embargo efectuar el registro. Existieron casos similares en donde para llegar se requería varios días a lomo de bestias, deslizador o a pie, y segmentos cuyo acceso era por un departamento diferente al que políticamente pertenecía.

⁴/ Existen informes de campo realmente dramáticos al respecto (Informe Técnico No. 05-84-Ayacucho-VLB.)

⁵/ Uno de los documentos auxiliares que se le entregó al personal de campo para el registro fue el documento CPV-03.51, en donde figuran los nombres de los jefes de cada hogar en el segmento seleccionado según el Censo de 1981.

(6) Otro problema que se presentó en la identificación de la vivienda, en cuanto a su calificación de vivienda con o sin productor agropecuario, estuvo relacionado en saber si las tierras eran trabajadas en sociedad o no. Se presentaron casos en que estas preguntas no tenían respuestas; en tal situación, se optó por tomarla como vivienda con productor toda vez que al momento de hacer la entrevista y según la pregunta 29 de la cédula tipo 2, se establecía la situación real.

Es necesario mencionar que cuando la información no permitía calificar a la vivienda de acuerdo a si era con o sin productor agropecuario, se optaba por el evento de mayor probabilidad dentro del segmento.

Paso No. 3: En base de la información obtenida en dicho registro se procedió a identificar las viviendas con productor y sin productor. Luego de identificadas las viviendas con y sin productor, se procedió a numerar correctamente cada grupo por separado, tanto las viviendas con productor $N_j(cp)$ como las viviendas sin productor $N_j(sp)$ para cada segmento.

Paso No. 4: De acuerdo al diseño de la muestra se había establecido seleccionar 10 viviendas por segmento para cada subestrato, dando un total estimado de 300 viviendas con productor y 300 viviendas sin productor en los estratos rurales. Sin embargo, en los estratos en donde el número de segmentos con menos de 10 viviendas con o sin productor era significativo, o donde habían problemas socio-políticos, o de narcotráfico, se procedió a incrementar el número de viviendas a muestrear en cada segmento y según el subestrato, resultando en los siguientes tamaños:

NUMERO DE VIVIENDAS SELECCIONADAS POR SEGMENTO
PARA CADA ESTRATO Y SUBESTRATO

<u>E s t r a t o</u>	<u>Número de Viviendas Seleccionadas</u>	
	<u>Con Productor</u>	<u>Sin Productor</u>
1. Costa Urbana	10	-
2. Sierra Urbana	11	-
3. Selva Alta Urbana	10	-
4. Selva Baja Urbana	11	-
5. Costa Norte Rural	10	11
6. Costa Centro Rural	12	10
7. Costa Sur Rural	12	10
8. Sierra Norte Rural	10	15
9. Sierra Centro Rural	10	15
10. Sierra Sur Rural	11	15
11. Selva Alta Norte Rural	10	13
12. Selva Alta Centro Rural	12	15
13. Selva Alta Sur Rural	11	15
14. Selva Baja Rural	11	15

Tal como se puede observar, los mayores problemas se presentaron en el subestrato de viviendas sin productor, lo cual es explicable si se tiene en cuenta que en el área rural la actividad predominante es la agricultura caracterizada por el minifundismo. Es pertinente señalar que en el caso específico de los segmentos con problemas socio-políticos y de narcotráfico, se decidió seleccionar 15 viviendas con productor y 15 viviendas sin productor previendo las dificultades que se presentarían en el campo.

Paso No. 5: Para la selección de viviendas con productor dentro de cada segmento muestreado, se cumplió con los siguientes pasos:

- (1) Si el número de viviendas con productor $N_{j(cp)}$ en el segmento era igual o menor al número de viviendas con productor fijado para ser seleccionada, $\bar{n}_{h(cp)}$, se incluían todas las viviendas con productor del segmento en la muestra (probabilidad "1" en la última etapa).
- (2) Si el número de viviendas con productor en el segmento era mayor al número fijado a seleccionarse, se seguía el procedimiento que a continuación se detalla:
 - a) Se determinaba el intervalo de selección, $I_{v(cp)}$ para lograr el número de viviendas fijado:

$$I_{vcp} = \frac{N_{j(cp)}}{\bar{n}_{h(cp)}}$$

donde:

$N_{j(cp)}$ = Número de viviendas con productor agropecuario dentro del segmento.

$\bar{n}_{h(cp)}$ = Número de viviendas con productor agropecuario a muestrear dentro del segmento.

- b) Luego se escogía un número aleatorio (A_k) entre 0 e $I_{v(cp)}$
- c) Se procedía a seleccionar sistemáticamente las viviendas utilizando la siguiente fórmula:

$$V_k = A_k + (k - 1) I_{v(cp)}, \text{ redondeando al próximo entero,}$$

donde:

$$k = 1, 2, \dots, \bar{n}_{h(cp)}$$

V_k = Número de la vivienda seleccionada

Paso No. 6: Para la selección de viviendas sin productor para cada segmento muestreado, se procedió en forma similar que para las viviendas con productor, tanto para el caso de segmentos con un número de viviendas igual o menor al fijado, como para el caso de segmentos con un número de viviendas mayor al establecido, siendo en este último caso el intervalo

$$I_v(sp) = \frac{N_j(sp)}{n_h(sp)}$$

donde:

$N_j(sp)$ = Número de viviendas sin productor agropecuario dentro del segmento.

$\bar{n}_h(sp)$ = Número de viviendas sin productor agropecuario a muestrear dentro del segmento.

Resultados de la Selección de la Muestra

De la aplicación de los procedimientos de la selección de viviendas antes descrita se lograron los siguientes resultados a nivel de cada estrato:

NUMERO DE VIVIENDAS SELECCIONADAS POR ESTRATO Y SUBESTRATO

Estrato	Número de Viviendas Seleccionadas		Total
	Con Productor	Sin Productor	
1. Costa Urbana	300	-	300
2. Sierra Urbana	319	-	319
3. Selva Alta Urbana	300	-	300
4. Selva Baja Urbana	325	-	325
5. Costa Norte Rural	300	310	610
6. Costa Centro Rural	307	300	607
7. Costa Sur Rural	322	299	621
8. Sierra Norte Rural	300	180	480
9. Sierra Centro Rural	300	266	566
10. Sierra Sur Rural	334	203	537
11. Selva Alta Norte Rural	305	303	608
12. Selva Alta Centro Rural	378	305	683
13. Selva Alta Sur Rural	334	180	514
14. Selva Baja Rural	<u>334</u>	<u>192</u>	<u>526</u>
Total	4458 $\bar{X}=318$	2538 $\bar{X}=254$	6996

Tal como se puede observar y como se había señalado anteriormente, el subestrato de viviendas con productor no presentó problemas en cuanto al tamaño de la muestra por segmento. Se seleccionó en promedio 318 viviendas con productor en cada estrato. Los pocos casos que se presentaron de segmentos con una cantidad de viviendas inferior a las fijadas, se debió principalmente a centros poblados típicamente pesqueros o centros poblados urbanos con por lo menos 20% de viviendas con productor pero como resultado de crecimiento urbano produjo un efecto de disminución o extinción de las viviendas con productor agropecuario.

CAPITULO IV

METODOLOGIA DE ESTIMACION

La metodología de estimación para procesar los datos de la encuesta involucra el uso de un peso o factor de expansión para cada registro correspondiente a un hogar seleccionado que sería multiplicado por todos los datos conformantes del registro.

El factor final para cada hogar tiene dos componentes: el factor básico de muestreo y los factores de ajuste por no-entrevista.

Factor Básico de Expansión

El factor básico de expansión para cada hogar muestral es determinado por el diseño de la muestra; equivale al inverso de su probabilidad final de selección, el mismo que es el producto de las probabilidades de selección en cada etapa.

El diseño de la muestra, tal como se expuso en el Capítulo III, involucra hasta tres etapas de muestreo donde las unidades fueron seleccionadas con probabilidades proporcionales al tamaño (PPT) excepto la última etapa, con el objeto de mantener una muestra aproximadamente autoponderada dentro de cada estrato. En la última etapa se seleccionó un número fijo de viviendas para cada subestrato (viviendas con productor o sin productor) dentro del segmento.

Los factores básicos de muestreo varían por segmento y por subestrato. En el caso de los segmentos muestrales dentro de un estrato urbano, los pesos varían por los factores N'_{hi}/N_{hi} , donde: N'_{hi} es el número total de viviendas con productor, registrados para la encuesta en el segmento "i" del estrato urbano h y N_{hi} es la cifra correspondiente de los datos del Censo de 1981. En el caso de los estratos rurales los factores varían de acuerdo a la proporción de viviendas con productor en el segmento. Por esta razón se calculó un factor básico distinto para cada segmento y subestrato.

Factores de Ajuste por No-entrevistas

1. Ajuste de primera etapa para hogares no-entrevistados

Es importante ajustar los factores básicos de muestreo de los hogares entrevistados para tomar en consideración los hogares no entrevistados con la finalidad de reducir el sesgo correspondiente. Ese ajuste a los factores básicos de muestreo puede ser hecho a nivel de estrato o por segmento y subestrato.

Normalmente se calcula el factor de ajuste por no-entrevista por cada estrato o segmento de la siguiente manera:

$$\frac{\text{Número total de entrevistas o no-entrevistas en el estrato o segmento.}}{\text{Número total de entrevistas en el estrato o segmento}}$$

Las no-entrevistas aquí se refieren a unidades de muestreo válidas, es decir, que no pudieron ser entrevistadas a causa de rechazos, ausencias, etc. y no incluye unidades no válidas seleccionadas como viviendas abandonadas, destruidas, etc.

Luego con el objeto de obtener un mayor beneficio de la correlación entre hogares al interior del segmento, se decidió hacer el ajuste por no-entrevistas de la primera etapa a nivel de segmentos y subestrato. Este procedimiento también fue conveniente ya que los pesos básicos varían por segmento y subestrato, y el número de entrevistas y no-entrevistas fue procesada a ese nivel.

2. Ajuste de Segunda Etapa por Segmentos No-entrevistados

Un sesgo más serio es introducido cuando no se puede entrevistar íntegramente un segmento seleccionado, por ejemplo, los segmentos no trabajados por problemas sociopolíticos y de narcotráfico. La distribución de dichos segmentos no-entrevistados por estrato y los correspondientes tasas de no entrevista se muestran en el siguiente cuadro:

DISTRIBUCION DE SEGMENTOS NO-ENTREVISTADOS, POR ESTRATO Y FACTORES DE AJUSTE PARA SEGUNDA ETAPA POR SEGMENTOS NO-ENTREVISTADOS

Estrato	No. de Segmentos No-entrevistados	Porcentaje de segmentos No-entrevistados	Factor de Ajuste de 2da. Etapa por Segmentos No-entrevistados
10. Sierra Sur Rural	1	3.3	1.03
11. Selva Alta Norte Rural	1	3.3	1.03
12. Selva Alta Centro Rural	7	23.3	1.30
13. Selva Alta Sur Rural	1	3.3	1.03
14. Selva Baja Rural	1	3.3	1.03

Como se puede ver en dicho cuadro, el estrato "Selva Alta Centro Rural" tiene la tasa de no-entrevista más alta (23.3%). El sesgo resultante de las estimaciones para este estrato es potencialmente grande dado que la mayoría de los segmentos omitidos son de la misma provincia, la cual no estaría representada en los datos.

Un método de ajustar los factores para segmentos no-entrevistados que en algunos casos no puede usar, involucra el aparear los segmentos seleccionados no-entrevistados con segmentos similares que fueron entrevistados y ajustar los factores de expansión correspondientes. Sin embargo, considerando que los segmentos no-entrevistados son en la mayoría de los casos de la misma área geográfica, no fue factible utilizar este método. Por lo tanto, dado que no contamos con información para estos segmentos, solamente es posible hacer un ajuste simple para dichos segmentos no-entrevistados. Un ajuste sencillo a nivel de estrato se puede hacer usando el siguiente factor:

$$\frac{\text{Número total de segmentos seleccionados en el estrato}}{\text{Número total de segmentos seleccionados entrevistados en el estrato}}$$

El factor de ajuste de la segunda etapa para cada estrato con segmentos no-entrevistados se presenta en el cuadro anterior en la columna (3).

El ajuste de segunda etapa para no-entrevistas está basado en el supuesto que los segmentos no-entrevistados son similares a los entrevistados, pero el hecho de que dichos segmentos no pudieron ser entrevistados por razones sociopolíticas, nos estaría indicando que son segmentos distintos. Por lo tanto, las estimaciones para estratos con una alta tasa de no-entrevista (tal como el caso de la Selva Alta Centro Rural) puede estar afectado por un sesgo significativo pero no cuantificado. Por esta razón, los resultados de esos estratos deben ser usadas con mucha cautela.

Factor Final de Expansión

El factor final para cada hogar fue determinado multiplicando el factor b_g correspondiente por los factores por no-entrevista de primera y segunda etapa. Luego, los factores finales de expansión fueron añadidos a los registros correspondientes a cada hogar muestral por segmento y subestrato.

Estimación de los Variancias

Es muy importante que se calcule el error de muestreo para las principales estimaciones de la encuesta o fin de determinar el nivel de precisión correspondiente en el análisis de los datos de la encuesta (ver Capítulo XI, "Evaluación de la Exactitud de los Datos"). Dada la complejidad de desarrollar programas especiales para calcular variancias, así como el tiempo que ellos demandarían, se decidió que lo más recomendable sería usar un paquete adecuado y disponible para efectuar dichos cálculos. Es muy importante que el estimador de variancias tome en cuenta el diseño de la muestra. Un paquete con tales características es SUPER CARP (Cluster Analysis and Regression Program), desarrollado por la Universidad de Iowa State y diseñado para calcular variancias para muestros estratificadas polietápicas, tal como la ENAHR. Este paquete puede calcular variancias de totales, promedios, proporciones y otras razones, como también coeficientes de regresión. La mayoría de las estimaciones de la encuesta están en forma de razones, aunque también aparecen en los cuadros estimaciones de totales. El análisis de regresión también puede ser utilizado en el análisis de los datos. Un grupo representativo de estimaciones de la encuesta a diferentes niveles de desagregación fue seleccionado de los cuadros para calcular los correspondientes errores de muestreo usando SUPER CARP.

El paquete SUPER CARP utiliza un tipo de estimador de variancia llamado "conglomerados últimos" en el sentido que los cálculos son basados a nivel de segmentos. El estimador de variancias para razones es basado en una expansión de serie Taylor. Las fórmulas de variancia en SUPER CARP son presentadas en el manual en forma de matrices. Las siguientes fórmulas fueron usadas para calcular la variancia de totales y razones (sin un factor de corrección para población finita), presentadas de una manera simplificada:

(1) Variancia para una estimación de total (\hat{X})

$$\text{Var}(\hat{X}) = \sum_h \frac{m_h}{m_h - 1} \sum_{i=1}^{m_h} \left(\hat{x}_{hi} - \frac{\hat{X}_h}{m_h} \right)^2$$

donde:

m_h = número de segmentos seleccionados en estrato h.

$$\hat{x}_{hi} = \sum_{j=1}^{n_{hi}} w_{hij} x_{hij} = \text{Total ponderado para variable X para el segmento seleccionado i en el estrato h.}$$

n_{hi} = número de hogares seleccionados en el segmento seleccionado i en el estrato h.

w_{hij} = factor de expansión final para los hogares seleccionados j del segmento i del estrato h.

x_{hij} = valor de la variable X para el hogar seleccionado j del segmento i del estrato h.

$$\hat{X}_h = \sum_{i=1}^{m_h} \hat{x}_{hi} = \text{total ponderado de X para estrato h.}$$

(2) Variancia de una estimación de razón (\hat{Y}/\hat{X})

$$\text{Var} \left(\frac{\hat{Y}}{\hat{X}} \right) = \sum_h \frac{m_h}{m_h - 1} \left(\frac{1}{\hat{X}^2} \right) \sum_{i=1}^{m_h} \left[\left(\hat{y}_{hi} - \frac{\hat{Y}_h}{m_h} \right)^2 + \left(\frac{\hat{Y}}{\hat{X}} \right)^2 \left(\hat{x}_{hi} - \frac{\hat{X}_h}{m_h} \right)^2 - 2 \left(\frac{\hat{Y}}{\hat{X}} \right) \times \left(\hat{y}_{hi} - \frac{\hat{Y}_h}{m_h} \right) \left(\hat{x}_{hi} - \frac{\hat{X}_h}{m_h} \right) \right],$$

donde: Y, Y_h e Y_{hi} son definidas para la variable Y de la misma manera que las X, X_h , X_{hi} ; respectivamente.

Aunque SUPER CARP tiene la opción de un factor de corrección de población finita, éste está basado en el supuesto que las UPM son seleccionadas con probabilidades iguales; sin embargo, las UPM para

la encuesta fueron seleccionadas con probabilidades proporcionales al tamaño (PPT). Por esta razón, el factor de corrección de población finita no fue incluido en los cálculos de variancia. De cualquier manera, eso no debe afectar significativamente las variancias ya que la tasa de muestreo fue relativamente baja.

Fue necesario crear un nuevo archivo de datos para SUPER CARP que correspondía al formato requerido por el paquete, y consistía en un registro para cada hogar seleccionado, empezando con el número de estrato, número de segmento y peso, seguido por las variables de clasificación y análisis a usarse en los cálculos de variancias.

En este archivo para SUPER CARP los números de estratos fueron recodificados para mejorar la estimación de variancias. Aunque el marco de muestreo fue dividido en 14 estratos geográficos, algunos estratos urbanos tenían UPM autorepresentadas (AR), (incluidos en la muestra con probabilidad = 1), que son considerados estratos individuales. El número de segmentos a seleccionarse en cada UPM urbano AR fue determinado dividiendo su medida de tamaño (número de viviendas con productor agropecuario) por el intervalo de selección y redondeado al entero más próximo. En el caso de UPM-AR con dos o más segmentos seleccionados, la UPM fue recodificada como un estrato individual. Las UPM-AR con solamente un segmento seleccionado fueron agregadas en estratos colapsados de 2 ó 3 segmentos cada uno. Para reflejar las ganancias en precisión de la estratificación implícita resultante del muestreo sistemático, las UPM no autorepresentadas también fueron divididas en subestratos con 2 ó 3 segmentos cada uno (en lo posible dentro del mismo departamento). De esa manera, un total de 186 estratos fueron creados recodificando los números de estrato en el archivo de SUPER CARP.

Estos nuevos estratos están especificados en el Apéndice 8. Sin embargo, en el caso de las estimaciones relacionadas a cultivos individuales, las variancias fueron calculadas basadas en los 14 estratos geográficos originales, en razón del número relativamente bajo de observaciones de algunos cultivos en ciertas regiones.

CAPITULO V

DISEÑO DE CEDULAS Y OTROS FORMULARIOS

Introducción

La cédula es el instrumento más importante de una encuesta, por lo tanto, es imprescindible que su diseño se efectúe cuidadosamente. Durante el diseño de la cédula se debe tener presente los objetivos de la encuesta, las definiciones y conceptos utilizados, los temas a investigarse y la información prevista en el plan de tabulaciones. Igualmente, debe tomarse en cuenta la capacidad y actitud de los informantes para contestar las preguntas, la redacción, contenido y estructura de las mismas. En la ENAHR se siguió al pie de la letra estos lineamientos, dando como resultado un buen diseño de las cédulas lo que contribuyó al alcance de los objetivos de la encuesta.

Para la ENAHR se diseñaron dos cédulas, una para los hogares donde existía al menos un productor agropecuario, y la otra, para aquellos que estando ubicados en el área rural no tenían al menos una persona que condujera o trabajara una explotación agropecuaria por cuenta propia. El diseño de las cédulas fue un esfuerzo conjunto de funcionarios del INE, la OSE y del BUCEN. El primer borrador fue diseñado en las oficinas del INE entre los meses de octubre y noviembre de 1983. Este borrador fue evaluado durante una pequeña prueba que se realizó en diciembre de ese mismo año en la provincia de Chancay, departamento de Lima. Una vez que se hicieron los ajustes necesarios de acuerdo con los resultados obtenidos, ambas cédulas fueron, en febrero de 1984, nuevamente sometidas a prueba dentro de la actividad "Prueba de Campo", que se realizó en los departamentos de Lambayeque, Junín y Huánuco. Siendo esta última una prueba ambiciosa ejecutada en las tres regiones naturales del país, la Costa, la Sierra y la Selva, los resultados obtenidos fueron significativos lo que dió lugar a rediseñar las cédulas, especialmente, la que era para los hogares con productor agropecuario.

Teniendo como limitante el factor tiempo, ya que la capacitación del personal de campo estaba programada originalmente para el mes de julio de 1984, se decidió realizar el diseño final de las cédulas en las oficinas de BUCEN en Washington, D.C. Funcionarios del INE viajaron a Washington D.C. para participar en esta labor siendo valiosa su participación.

Debido a los cambios significativos que se introdujeron al rediseñar las cédulas, en particular la de los hogares con productor agropecuario, fue necesario realizar una prueba adicional antes de que la cédula a diligenciarse en los hogares con productor agropecuario fuera utilizada en la operación de campo. Esta última prueba se efectuó en las zonas de Tarma y Chanchamayo en el departamento de Junín durante los últimos días del mes de mayo de 1984. Fueron pocos los cambios que resultaron y rápidamente se incorporaron al diseño final.

La impresión de aproximadamente 10,000 cédulas se realizó en los talleres de imprenta del INE siendo de buena calidad el trabajo efectuado.

Consideraciones del Diseño

De acuerdo a los objetivos de la ENAHR, la cédula que iba a ser diligenciada en los hogares con productor agropecuario se diseñó de una manera que permitiera

recoger información sobre las características básicas del hogar, sus ingresos y los de la explotación agropecuaria, al igual que la información relacionada con el acceso al mercado, asistencia técnica recibida y uso del crédito agropecuario. Aunque la encuesta no estaba enfocada a obtener el rendimiento de los cultivos, la mayoría de las preguntas en la cédula estaban relacionadas a ellos, puesto que la actividad principal de la mayoría de los hogares en el área rural está estrechamente vinculada a la agricultura.

Para la mayoría de las preguntas, el período de referencia fue el Año Agrícola que comenzó el 1º de agosto de 1983 y concluyó el 31 de julio de 1984. Sin embargo, de acuerdo a la naturaleza de ciertas preguntas, fue necesario cambiar el período de referencia a fin de captar de manera más precisa la información requerida.

Las preguntas sobre empleo, ocupación e ingresos percibidos fuera de la explotación agropecuaria fueron consideradas con distintos períodos de referencia. Por ejemplo, para ocupación se utilizó la semana previa a la de la entrevista; para ingresos por concepto de pensión, jubilación o cesantía, remesas periódicas de otros hogares, por divorcio y alimentos por hijos, se optó el mes previo al de la entrevista. Para otros casos donde se solicitaba información sobre ingresos percibidos fuera de la explotación agropecuaria, se utilizaron los últimos doce meses previos al mes de la entrevista.

Con respecto a este último período de referencia, se optó los últimos doce meses en vez del Año Agrícola debido a que las entrevistas originalmente estaban programadas para que se iniciaran el 01 de agosto de 1984, los "doce últimos meses", por consiguiente, coincidían exactamente con el Año Agrícola. Otra consideración fue que los informantes que percibían ingresos no relacionados con la agricultura, el concepto de los "últimos doce meses" era más fácil de comprender que el Año Agrícola. Desafortunadamente, como el inicio del trabajo de campo se postergó dos meses debido a paralizaciones en el Sector Público, ocasionando además que el trabajo de campo se prolongara más de lo programado, el período de referencia de los últimos doce meses ya no coincidió con el del Año Agrícola. Sin embargo, el sesgo de la información relacionada con este desfase es poco significativo si se considera que no existe mucha diferencia entre los ingresos recibidos en los hogares entre el período que se quería captar y el que, en efecto, ocurrió.

Otras consideraciones que se tomaron en cuenta para obtener información más precisa, fue la de redactar preguntas que, dentro del marco del período de referencia del Año Agrícola, solicitaban información de un lapso más corto de tiempo. Entre este grupo de preguntas estaban aquellas relacionadas con la producción de leche donde se solicitaba la cantidad que en promedio se producía diariamente, las relacionadas con la de la producción de huevos, lana y productos derivados de la leche donde las cantidades producidas se referían a una semana; y otras donde las preguntas se formulaban de manera que el informante indicaba el período de referencia de su respuesta, por ejemplo, mensual, trimestral, semestral o anual.

En el diseño de la cédula para los hogares con productor agropecuario se introdujo el uso del "Dibujo" o bosquejo general de todas las parcelas que conforman la explotación agropecuaria durante el período de referencia. El propósito

principal del "Dibujo" fue el de mejorar la calidad de la información recopilada con respecto al uso de la tierra, los cultivos cosechados y perdidos, la superficie de los mismos y la forma como fueron cultivados. Se exigió que ambos, encuestador e informante, participaran en elaborar el dibujo de las parcelas de la explotación agropecuaria a fin de asegurar que la información requerida no fuese omitida. El dibujo resultó ser un método sumamente efectivo. Fue utilizado durante la entrevista para ayudar a recordar al productor sobre los cultivos cosechados y, sucesivamente, su destino ya fuese para la venta, consumo en el hogar u otro uso. Además de ser útil para la operación de campo, el dibujo también sirvió durante la crítica de los datos en la oficina para resolver problemas de inconsistencia entre las variables de la cédula.

Desde el punto de vista del procesamiento de datos una de las consideraciones más importantes del diseño de la cédula fue la facilidad con la cual ésta podía ser procesada. En la ENAHR los programadores de computadora contribuyeron en el diseño dando sugerencias sobre el formato y la identificación de la cédula. También recomendaron el uso de "códigos de fuente" para facilitar el procesamiento. Particularmente, cada código de fuente identifica el casillero donde es anotada la respuesta en la cédula; es decir, cada espacio para respuesta tiene un código numérico específico. Al momento de digitarse la respuesta, el digitador introduce el código de fuente y el valor anotado. De estar en blanco, pasa al siguiente casillero que contenga un valor. Como sólo los casilleros con información son digitados, se ahorra tiempo durante la digitación, se eliminan los errores introducidos por un espaciado incorrecto por parte de los digitadores al registrar las respuestas y se reduce el tamaño del registro de datos cuando solamente ciertas secciones o preguntas contienen información. Como se anticipaba que para la ENAHR la mayoría de los casilleros iban venir en blanco, el código de fuente como método de identificación para los casilleros fue el indicado. Al mismo tiempo, el desarrollo de las especificaciones para la crítica de consistencia por computadora se facilitó ya que cada casillero tenía una identificación única, permitiendo así señalar fácilmente la relación entre las variables.

Para las dos cédulas diseñadas se asignaron un total de 3,955 códigos de fuente, de los cuales los códigos de 1 al 3,111 fueron para la cédula de los hogares con productor agropecuario y del 3,200 al 3,955 para la cédula de los hogares sin productor agropecuario.

La secuencia de códigos, sin embargo, no fue continua. Entre sección y sección de cada cédula existen unos 20 códigos que se dejaron a propósito sin asignar. La asignación se efectuó de esa manera ya que permitía que más adelante, durante la etapa del procesamiento de datos, se utilizaran aquellos códigos no asignados para almacenar valores adicionales necesarios para la tabulación de los datos.

Otra de las consideraciones del diseño fue la de incluir un casillero para anotar la unidad de medida a continuación del casillero donde se registraba la superficie o cantidad declarada. Fue necesario tomar esta decisión debido a que a lo largo del país se utilizan algo más de 60 unidades de medida para la superficie. Igualmente, como la unidad de medida varía de acuerdo al cultivo (por ejemplo: camionada, saco, atado, caja, etc.), se incluyeron casilleros para la unidad de medida local en relación con la cantidad destinada a la venta, al consumo del hogar o los animales, a la utilizada para semilla

o al trueque. También fue necesario incluir casilleros referidos a la unidad de medida local para registrar la producción de leche y los productos derivados de ella, la producción de lana o fibra y la producción de huevos de gallina. A excepción de la producción de leche y huevos de gallina, se añadió otro casillero denominado "kilos por unidad de medida". Para la producción de leche y la de huevos, los casilleros adicionales fueron denominados "litros por unidad de medida" y "número de huevos por unidad de medida" respectivamente. El propósito de haber incluido este último casillero fue el de utilizarlo durante la conversión de diferentes unidades de medida a una estándar. Se incluyeron también casilleros para "Uso de Oficina". Además de la codificación de la ubicación geográfica, se utilizaron casilleros para la codificación de superficie, cultivos, ocupación actividad económica y razón principal por la que el productor agropecuario o algún miembro del hogar dejó de pertenecer a alguna asociación agrícola.

Organización de la Cédula

La cédula para los hogares con productor agropecuario, denominada tipo 2, contiene 220 preguntas ordenadas en 11 secciones. La cédula para los hogares sin productor agropecuario, denominada tipo 1, contiene 45 preguntas distribuidas en 3 secciones. Hubieron dos razones por la cual la primera cédula contenía un gran número de preguntas: (1) la necesidad de los usuarios de obtener información sobre el ingreso neto de aquellos hogares que dependían de la agricultura para su sostén, lo que implicó obtener no sólo información sobre la producción y venta de cultivos, sino también sobre los gastos efectuados para lograr esa producción; y (2) la necesidad de usuarios y analistas de contar con información confiable y actualizada de las variables complementarias.

La cédula (tipo 2) para los hogares con productor agropecuario está constituida de las siguientes secciones:

Carátula: Identificación y Ubicación Geográfica

Sección I - Características de la Explotación Agropecuaria y Uso de la Tierra

II - Producción Agrícola y su Destino

III - Gasto de Insumos para Cultivos

IV - Inventario de Animales, Producción Pecuaria y su Destino

V - Otros Gastos e Ingresos Agropecuarios

VI - Crédito Agropecuario

VII - Asistencia Técnica

VIII - Diversos

IX - Características del Hogar y sus Miembros

X - Empleo, Ocupación e Ingresos

XI - Capacitación e Idiomas

La cédula para los hogares sin productor agropecuario (tipo 1) está constituida de la siguiente manera:

Carátula: Identificación y Ubicación Geográfica

Sección I - Características del Hogar y sus Miembros

II - Empleo, Ocupación e Ingresos

III - Diversos

En ambas cédulas las secciones de "Características del Hogar y sus Miembros" y "Empleo, Ocupación e Ingresos" contienen las mismas preguntas. Sin embargo, la sección "Diversos" es distinta para cada cédula. En la cédula tipo 1, "Diversos" incluye compra y venta de tierras, capacitación recibida por el jefe del hogar e idioma principal del hogar; para la cédula tipo 2, las preguntas están relacionadas con la explotación agropecuaria.

El ordenamiento de las secciones se hizo de acuerdo a los resultados obtenidos en las anteriormente citadas pruebas de campo. También se tuvo cuidado en ubicar al inicio de ambas cédulas las secciones más fáciles de responder. La sección sobre ingreso personal de los miembros del hogar se dejó hacia el final, pero no fue la última pues hubiese dejado una sensación negativa hacia la encuesta. Las preguntas sobre ingresos fueron precedidas por preguntas relacionadas con el nombre de los miembros del hogar, edad, grado de instrucción y actividad que desempeñaron la semana previa a la entrevista. Estas preguntas, siendo fáciles de contestar, establecía un ambiente agradable hacia la entrevista y, por consiguiente, reducía la posibilidad de que el informante se negara a contestar las preguntas relacionadas con los ingresos.

Tipo de Preguntas y Método para Registrar las Respuestas

En el contenido de ambas cédulas, una de las cosas que resalta de inmediato es que las preguntas fueron redactadas en su totalidad, es decir, palabra por palabra con el objeto de que el encuestador las leyera textualmente al informante. Claro está que las preguntas redactadas en su totalidad ocupan mayor espacio en la cédula pero aseguran que los encuestadores formulen las mismas preguntas, logrando así uniformidad entre ellos y eliminando el sesgo que hubiese sido introducido si se permitía que los mismos encuestadores formularan las preguntas de acuerdo a sus propios criterios o experiencias. Asimismo, se puede afirmar que la mecánica operativa que se adoptó facilitó significativamente el flujo de la entrevista tanto para el informante como para el encuestador.

Como se puede apreciar en las cédulas de la ENAHR, las preguntas fueron diseñadas de manera que las respuestas pudieran clasificarse en tres grupos:

(1) respuestas literales que fácilmente se podrían anotar en un espacio determinado, (2) respuestas numéricas, y (3) respuestas que se podían marcar mediante un aspa o cruz en una de las opciones predeterminadas. El primer grupo incluía preguntas relacionadas a nombre de los miembros del hogar, ocupación y rama de actividad económica. También se incluyó en este grupo una pregunta que solicitaba la razón por la que algún miembro del hogar dejó de pertenecer a una asociación agrícola. Las respuestas numéricas constituyeron aproximadamente un 70% de los códigos de fuente. Aunque el porcentaje fue alto hay que tener presente que la mayoría de las preguntas estaban relacionadas con producción, ingresos y gastos que, en sí, implican mucha variabilidad y son difíciles de precodificar. Por último, se diseñaron preguntas donde las opciones de respuesta fueron incluidas en la cédula. Muchas de las respuestas, que dieron motivo a esta precodificación, se obtuvieron mediante las pruebas de las cédulas en el campo. Para aquellas preguntas donde fue imposible determinar toda la gama de respuestas posible, se incluyó una categoría denominada "Otro - Especifique". Si durante el procesamiento de datos se hubiese detectado que esa categoría tenía una frecuencia relativamente alta comparado a las otras opciones de respuesta, estuvo previsto investigarla a fin de asegurar que no correspondiera a una de las otras categorías. De no ser ese el caso, se hubiese considerado independientemente al presentar los resultados.

Para dirigir la entrevista y evitar que se formularan preguntas no aplicables, se incluyeron en la cédula instrucciones. Por ejemplo, en la pregunta "En los últimos 12 meses, ¿ha recibido ingresos monetarios o en especie por concepto de trabajo fuera de la explotación agropecuaria, alquiler de propiedad,....? La respuesta podía ser "sí" o "no". Si la respuesta era negativa, se consignó la instrucción: "Pase a la próxima persona", ya que las 19 siguientes preguntas no debían formularse; de lo contrario, el entrevistado hubiese pensado que el encuestador no le había creído o que el encuestador no estaba atento a lo que el informante había dicho, lo que en algunos casos podría haber motivado el rechazo a la entrevista.

Ayuda Visual

Como la labor que debía realizar el encuestador no era fácil, se trató al diseñar la cédula de darle toda clase de ayuda visual para lo cual el tipo de letra utilizada para las preguntas era de tamaño y molde diferente al de las instrucciones. En algunos casos se utilizaron flechas para indicarle al encuestador dónde tenía que pasar o escribir la respuesta. Cuando el encuestador tenía que sumar y anotar el total, como en el caso de uso de la tierra, la anotación debía hacerla en un casillero cuyos bordes eran más gruesos. Se utilizaron puntitos que iban desde la pregunta hasta el lugar que se anotaba la respuesta para guiar al encuestador. Los casilleros denominados "Uso de Oficina" estaban bien delineados, siendo fácil diferenciarlos de los demás. Las preguntas relacionadas a la superficie tenían una línea para anotar los números enteros y dos líneas para los decimales. Los casilleros para anotar valores monetarios estaban siempre indicados con la palabra "soles". La mayoría de las instrucciones estaban precedidas por una flecha de diferente color en cuyo interior se encerraba la palabra "Encuestador".

Posiblemente lo más notable del diseño fue la utilización del efecto de diferentes tonos para diferenciar el lugar destinado a anotar las respuestas.

Este lugar, que contenía la información que iba a ser digitada, era blanco. El espacio donde se imprimió las preguntas era azul celeste y las preguntas mismas estaban escritas con tono azul oscuro. Esta ayuda visual hizo la tarea del personal de campo al igual que el de oficina mucho más fácil.

Otras ayudas visuales que formaban parte del diseño de la cédula fueron el uso del sombreado para evitar que el encuestador hiciera alguna anotación donde no tenía que hacerla, y el número y título respectivo al inicio de cada sección.

Estructura Física

La cédula tipo 1 constaba de 20 páginas. También, contenía dos solapas que permitían registrar los nombres y apellidos de 11 miembros del hogar en cada uno, habiendo espacio para un total de 22 miembros como máximo. La cédula tipo 2 constaba a su vez de 72 páginas y tres solapas; dos de estas, como en la cédula tipo 1, eran para registrar los nombres y apellidos de los miembros del hogar; y la otra para anotar los cultivos cosechados o perdidos en la explotación agropecuaria. El propósito de la solapa fue el de tener a la vista los nombres de las personas o los cultivos anotados a que las preguntas siguientes se referían sin necesidad de voltear las hojas de la cédula para verlos o transcribirlos de hoja en hoja. Al mismo tiempo, el revés de cada solapa contenía definiciones y conceptos utilizados en las preguntas pertinentes de cada sección.

Diseño del Registro y otros Formularios

Además de las cédulas tipo 1 y 2, se diseñaron para la ENAHR otros formularios. Entre estos se encuentran el "Registro de Viviendas" y el "Listado de las Viviendas Seleccionadas". Al igual que las cédulas tipo 1 y 2, estos dos formularios fueron diseñados originalmente en el INE. Posteriormente, el "Registro de Viviendas" y el "Listado de las Viviendas Seleccionadas" fueron modificados en las oficinas de BUCEN en Washington, D.C. por funcionarios de esa entidad y del INE en base a los resultados de la prueba de campo.

El "Registro de Viviendas", como su nombre indica, se utilizó para el registro de todas las viviendas en los segmentos seleccionados. El registro consistía de 16 columnas en las cuales se anotaban la dirección o ubicación de cada una de las viviendas, el nombre y apellidos del jefe de familia en cada una, e incluía una serie de preguntas para determinar si alguien en la vivienda era productor agropecuario o no. La información captada a través de este registro permitió clasificar las viviendas de acuerdo a si tenían o no productores agropecuarios formando así la base para la selección de viviendas a encuestarse. El registro tenía espacio para anotar 22 viviendas y, en la esquina superior derecha de cada uno, se anotaba en forma correlativa el número de cada una de las hojas diligenciadas seguido del número total de hojas que se utilizó para registrar todas las viviendas dentro del segmento seleccionado. En algunas preguntas como ayuda visual se utilizó flechas para indicar los pases a otras preguntas, y se usó diferentes tipos de letras para diferenciar las preguntas de las instrucciones. Así mismo, se trazó con líneas anchas la columna reservada para uso de oficina (ver Apéndice 9).

El "Listado de las Viviendas Seleccionadas" consistió de 10 columnas que se utilizaron para indicar la dirección ó ubicación de cada vivienda seleccionada y el nombre y apellidos del jefe de familia asociado con la misma a fin de que el encuestador la ubicara y encuestara. El listado fue diseñado de una manera muy similar a la del registro de viviendas, es decir, se utilizaron ayudas visuales en el diseño de su formato para facilitar la tarea de todo el personal que tenía que trabajar con él (ver Apéndice 10).

Entre los otros formularios que se diseñaron para la ENAHR se encuentran:

1. Informe del Encuestador
2. Informe Final del Encuestador
3. Control de Entrega y Devolución de Documentos y Material
4. Informe del Registrador
5. Hoja de Control del Jefe de Brigada
6. Hoja de Modificaciones.

CAPITULO VI

DOCUMENTACION OPERACIONAL

Para la ENAHR se preparó una serie de manuales operacionales que sirvieron no sólo como documentos de referencia durante la fase del trabajo correspondiente a cada uno de ellos, sino también como instructivos durante los seminarios y talleres de capacitación impartidos, y como documentación histórica de las normas y procedimientos establecidos para dicha encuesta y los cuales podrán ser utilizados en un futuro en encuestas similares.

Todos los manuales utilizados en la ENAHR fueron elaborados por un equipo de técnicos del INE con la colaboración de asesores del BUCEN.

Manual del Registrador

Una de las operaciones de campo más importante de la ENAHR fue el registro de las viviendas en los segmentos seleccionados. Esta operación contó con la participación de 137 funcionarios de la OSE. Para asegurar que todo el personal comprendiera cabalmente los conceptos, definiciones y procedimientos para la ejecución del registro, se redactó el Manual del Registrador.

El manual definitivo contó de 9 capítulos y anexos. Específicamente, el manual les dió a conocer a los registradores el objetivo y finalidad de la encuesta y de la operación del registro de viviendas, los conceptos y definiciones empleados para la ejecución de su labor, y los documentos auxiliares necesarios para realizar el trabajo. También les ofreció instrucciones para la actualización y recorrido de los segmentos urbanos seleccionados y para el levantamiento y/o modificaciones del croquis y recorrido de los segmentos rurales seleccionados. Se incluyeron instrucciones para el diligenciamiento del documento titulado "Registro de Viviendas" que básicamente fue el instrumento principal de la operación. Por último, el manual cubrió otros aspectos importantes tales como la revisión y devolución al jefe de brigada (supervisor) de los documentos de trabajo.

Como era necesario que los registradores determinaran en metros cuadrados la cantidad de tierra trabajada por los miembros de cada vivienda en los segmentos seleccionados, se incluyó como anexo al manual una lista de las diferentes unidades de medida de superficie y sus equivalencias en el sistema métrico decimal. También se incluyeron ilustraciones para explicar algunos términos cartográficos tales como: línea divisoria de aguas, curva de nivel, cota y los signos convencionales utilizados en los planos. Otras ilustraciones y ejemplos utilizados en el manual sirvieron para demostrar a los registradores como tenían que hacer el recorrido en los segmentos urbanos y rurales, y como debían completar cada hoja del registro de viviendas.

Manual del Jefe de Brigada (Registro de Viviendas)

El principal objetivo del Manual del Jefe de Brigada para la operación del registro de viviendas fue el de presentar los procedimientos para la supervisión del trabajo, especialmente en: (1) la entrega y devolución de documentos y materiales utilizados por los registradores, (2) la observación del trabajo de campo, y (3) la revisión de los registros con el propósito de corregir errores.

En el referido manual se consideró lo siguiente:

- Objetivo y finalidad del Registro de Viviendas así como las responsabilidades, funciones y prohibiciones del jefe de brigada en la ejecución del trabajo.
- Los documentos y materiales a utilizarse en el registro de viviendas, los procedimientos para la distribución de los mismos a los registradores y la forma de ser recepcionados para ser posteriormente devueltos a los jefes departamentales de la encuesta.
- Descripción de la mecánica para supervisar en el terreno el trabajo de los registradores a fin de detectar posibles errores y verificar el correcto diligenciamiento del documento "Registro de Viviendas".
- El procedimiento a seguir para la revisión de los documentos diligenciados incidiendo especialmente en la información de preguntas interrelacionadas a fin de establecer la coherencia de los datos y controlar la consistencia de la información recopilada.
- Las acciones que debía tomar el jefe de brigada en casos de que el registrador hubiese cometido errores, disponiendo ya sea la recuperación de los datos, o repetición del registro.
- Las instrucciones para el diligenciamiento del documento "Hoja de Control del Jefe de Brigada". Este documento permitía al jefe de brigada ejercer un control del avance del trabajo de cada registrador a su cargo. En él se registraba el número del segmento concluido, la fecha de recepción, devolución y entrega, así como el número de viviendas registradas en cada segmento.

Manual del Encuestador

De todos los manuales que se elaboraron para la ENAHR, el Manual del Encuestador fue el más importante. El manual definitivo consistía de 7 capítulos y 4 anexos, y tenía un total de 154 páginas. Los capítulos desarrollados fueron los siguientes:

- (1) Objetivos, Finalidad y Metodología de la Encuesta;
- (2) Organización para la Encuesta;
- (3) Conceptos y Definiciones Básicas;
- (4) Documentos y Materiales a Utilizarse;

- (5) Normas y Técnicas de la Entrevista;
- (6) Instrucciones para el Diligenciamiento de las Cédulas;
- (7) Revisión y Devolución de los Documentos de Trabajo; y
- (8) Anexos.

Aunque todos los capítulos se consideraban importantes y tenían que ser estudiados y aplicados a plenitud por el personal encargado de efectuar las entrevistas a fin de garantizar que las tareas se ejecutaran bajo una normatividad común, el capítulo 6 se puede decir fue el principal del manual. Este capítulo encerraba las instrucciones para el diligenciamiento de las cédulas. Las instrucciones estaban a su vez divididas entre las que eran de carácter general y las que eran específicas. Dentro de las instrucciones generales, se encontraban aquellas que permitían al personal de campo familiarizarse con la cédula, o sea, con su formato y su manejo. Las instrucciones específicas, por otro lado, cubrían todas las preguntas de ambas cédulas. El propósito de cada pregunta fue aclarado, y el tipo de respuesta deseada y su relación con otras preguntas fue discutido. Esta parte del manual abarcó el 71% del texto. Numerosos ejemplos con ilustraciones pertinentes fueron incluidos en un esfuerzo por asegurar interpretaciones correctas y consistentes entre todos los encuestadores.

El capítulo 6 contó con las instrucciones para hacer el dibujo o bosquejo general de todas las parcelas de la explotación agropecuaria, y para anotar los cultivos que se habían cosechado o perdido en ellas durante el período de referencia. De hecho, el dibujo sirvió como guía al encuestador y al informante para el llenado de las preguntas sobre la producción agrícola y su destino, y de referencia y consulta en etapas posteriores a la operación de campo.

Otro capítulo de gran utilidad, del punto de vista de conducir la entrevista, fue el de "Normas y Técnicas de la Entrevista". Este capítulo le enseñaba a los encuestadores como prepararse para la entrevista, como formular las preguntas, como obtener respuestas satisfactorias, como anotarlas, como controlar y terminar la entrevista, y como tratar los casos especiales.

El Manual del Encuestador fue el único redactado para las pruebas de campo. Las pruebas de campo demostraron la necesidad de rediseñar las cédulas y, por consiguiente, de revisar extensivamente el manual. Aproximadamente 200 ejemplares fueron impresos en los talleres del INE y distribuidos al personal de la OSE quien estaba encargado de la ejecución de la ENHR en el campo.

Durante la capacitación de los encuestadores surgieron varios problemas que no habían sido ni previstos en el manual ni detectados en las pruebas de campo. De hecho se decidió redactar instrucciones adicionales y complementarias al Manual del Encuestador en forma de directiva y distribuirla a los encuestadores, jefes de brigada y jefes departamentales al final del curso de capacitación.

Manual del Jefe de Brigada (Operación de Campo)

El Manual del Jefe de Brigada se elaboró con el propósito de que las personas responsables del trabajo de los encuestadores tuvieran un conocimiento completo de las tareas y labores relacionadas a la operación de campo. Uno de los cuatro capítulos y tres anexos que contenía el manual encaraba el tema de la revisión de las cédulas. Las instrucciones desarrolladas para este capítulo cubrían cada pregunta de la cédula, sección por sección. Su objetivo era detectar cualquier error que existiera a fin de corregirlo antes de que las cédulas diligenciadas fueran enviadas a Lima para ser procesadas.

Otro capítulo detallaba como eliminar el problema de "la falta de respuesta" el cual afecta de manera significativa la calidad de la encuesta. Este problema surge cuando el encuestador no logra obtener información alguna sobre el hogar seleccionado o que logre obtener sólo información incompleta. También se incluyeron en el manual instrucciones para la distribución de los documentos y materiales a los encuestadores y su recepción posteriormente, y la elaboración del informe final del jefe de brigada. La programación de tareas y labores para la operación de campo, al igual que copias de los documentos sobre el informe final y hoja de control del jefe de brigada, se presentó en los anexos.

Manual de Codificación y Conversión

Aunque la mayoría de las preguntas en la cédula exigía un respuesta numérica, habían otras que tenían que ser anotadas literalmente. En este grupo se encontraban las preguntas relacionadas a la ocupación y actividad económica de cada miembro del hogar, la ubicación geográfica del hogar, la unidad de medida para superficie, y la de los nombres de los cultivos cosechados o perdidos dentro del período de referencia. Las respuestas para este grupo, por consiguiente, tuvieron que asignarse códigos o claves numéricas en la oficina para entrar los datos a la computadora ya que no permitía respuestas literales. Por otro lado, el sistema de procesamiento de datos para la ENAHR se diseñó de manera que no aceptaba decimales a excepción de las preguntas sobre superficie. El resultado fue que hubieron muchas respuestas que se anotaron utilizando decimales, especialmente aquellas pertinentes al peso, y que por necesidad tenían que convertirse a enteros antes de la digitación. También fue necesario revisar las cédulas diligenciadas para verificar que las anotaciones en ellas se efectuaran correctamente y de acuerdo a los procedimientos establecidos.

Dado que era imprescindible efectuar todas las tareas mencionadas antes de que las cédulas pasaran a digitación, se elaboró el Manual de Codificación y Conversión. El manual definía para los crítico-codificadores las normas, procedimientos e instrucciones a seguirse en la ejecución de la codificación, conversión y revisión de la información contenida en las cédulas de la ENAHR. El manual contenía 7 capítulos y 2 anexos de los cuales algunos de ellos trataba de instrucciones generales y otros de específicas. El capítulo sobre revisión de las cédulas diligenciadas explicaba los chequeos generales a efectuarse tales como leer las anotaciones para determinar si eran legibles, si estaban anotadas en los casilleros correspondientes, si había información innecesaria, o si habían anotado decimales donde requería un número entero. Otra tarea de la revisión requería que se anulaban trazando con dos líneas

cruzadas aquellas páginas de la cédula que estaban completamente en blanco. Esta última instrucción eliminaba la necesidad de que los digitadores tomaran tiempo para examinar todos los casilleros de las páginas en blanco a fin de cerciorarse que no hubiese anotación alguna, apresurando así el proceso de la entrada de datos. En fin, la revisión que se efectuaba no verificaba la veracidad de la información o su consistencia en vista de que se iba a hacer posteriormente por computadora.

Las instrucciones para la conversión, si bien se refería a la transformación de cantidades con decimales a enteros, también comprendía el redondeo o aproximación de cantidades y la conversión de cantidades anotadas en fracción a enteros. El último capítulo del manual contenía la mecánica operativa para la ejecución del trabajo. En él se acentuaba como se realizaría el trabajo. Cabe señalar que las instrucciones para las tareas de revisión, codificación y conversión indicaban que tenían que ejecutarse simultáneamente.

Manual de Crítica por Computadora

El Manual de Crítica por Computadora fue elaborado por partes, es decir, se redactaba el capítulo pertinente a la fase de la crítica que iba a ser ejecutada según lo programado. Como la crítica de los datos por computadora contó de cuatro fases, el manual consistió de cuatro capítulos. Estos fueron: (1) Análisis de Cobertura, (2) Chequeo y Proceso de Conversión, (3) Chequeo de Rango, y (4) Análisis de Consistencia. El manual se elaboró de esta forma ya que la información e ilustraciones necesarias para su desarrollo no estaban disponibles para la fecha que se inició. Vale decir que para la fase de chequeo de rangos, algunos de los límites correspondientes a los valores máximos y mínimos de ciertas variables no se establecieron hasta casi el final de dicha operación. Este fue el caso de la mayoría de las variables relacionadas con la explotación agropecuaria, por ejemplo, el rendimiento de los cultivos, límites de aplicación de diferentes insumos agrícolas por hectárea y cultivo, valor de venta de la producción de los diferentes cultivos, etc. Toda esta información fue obtenida y los límites establecidos de los resultados de la misma encuesta a medida que se iba realizando el trabajo. Una vez terminada la elaboración de todos los capítulos y las tareas respectivas, se encuadernaron e imprimieron los capítulos para que quedaran como documento de referencia.

El capítulo del análisis de cobertura comprendió las labores siguientes:

- Identificación de cédulas inválidas
- Cédulas omitidas
- Duplicidad de cédulas
- Análisis de la "Situación de la Cédula"
- Chequeo de códigos de fuente y respuestas
- Chequeo de datos obligatorios

Todas las instrucciones para este capítulo al igual que para los otros, contaban con ilustraciones y ejemplos. En efecto, algunas de las ilustraciones eran tomadas de los listados de salida de la computadora. Ambos ejemplos e ilustraciones proveían un entendimiento básico del trabajo para aquellos que tenían que desempeñarlo.

El capítulo sobre la verificación de la conversión recalcó la importancia de solucionar todos los errores detectados por la computadora que exigían intervención manual. Los errores podrían ser causados porque una cantidad o código había sido omitido o porque el código asignado estaba incorrecto. Las instrucciones indicaban, por consiguiente, que se investigaran los errores detectados y, de acuerdo con los procedimientos establecidos, se corrigieran debidamente.

El capítulo sobre el chequeo de rango contó de tres secciones: (1) finalidad del chequeo, (2) descripción de labores, y (3) procedimiento para corregir los errores. Como era fundamental que los críticos contaran con los límites máximos y mínimos establecidos para las variables en la cédula a manera que pudieran efectuar el trabajo, se adjuntó también el listado de los valores extremos permitidos. El listado de 45 páginas estaba compuesto de los valores predeterminados y de las siguientes 9 tablas:

- (1) Rendimiento de cada cultivo
- (2) Precio de venta por kilo de cada cultivo
- (3) Gastos de semilla por hectárea
- (4) Gastos de abono orgánico
- (5) Gastos de abono químico y pesticidas
- (6) Jornal hombre por hectárea
- (7) Pago de jornal y gasto en animales
- (8) Pago de alquiler de vehículos para transporte y compra de cajones, java, canasta, etc.
- (9) Animales por especie en el día de la entrevista, animales nacidos, perdidos, sacrificados, comprados, y vendidos

Para asegurar que los críticos interpretaran bien el listado, se incluyeron ejemplos con sus respectivas ilustraciones. El listado de valores máximo y mínimo, como se mencionó previamente, se estableció por partes. Ya establecidos los valores máximo y mínimo servirán como indicadores de referencia para cualquier encuesta que se realice en el futuro.

El último capítulo, Análisis de Consistencia, detalló los procedimientos para la fase final de la crítica por computadora la cual consistió en corregir los errores detectados al relacionar la información de dos o más variables "predefinidas". Para corregir los errores, los críticos tenían que conocer y comprender las especificaciones que se desarrollaron y las cuales fueron utilizadas por Informática para elaborar el programa de consistencia. Como las especificaciones de consistencia fueron elaboradas utilizando símbolos matemáticos, surgió la necesidad de explicarlos detalladamente. También se tuvo que explicar la lógica utilizada en la elaboración de las especificaciones, a saber, las correcciones automáticas, los mensajes de error, los pases y

las restricciones. Para esta tarea, los críticos tenían que utilizar las especificaciones y el listado de errores. A ese efecto, se incluyó como anexo al capítulo las 53 páginas de especificaciones que tenían alrededor de 250 pautas de consistencia.

Manual de Entrada de Datos

Como el personal de digitación que se empleó para la ENAHR no fue contratado e inexperto, sino se utilizó el personal de planta del INE, las instrucciones que se elaboraron fueron concisas. La Dirección General de Informática del INE tuvo a su cargo la elaboración del manual y tomó en cuenta las características del diseño de las cédulas y del equipo de máquinas e instrumentos destinados a la digitación de la información contenida en las cédulas diligenciadas.

En el manual se señaló el método para digitar los datos que consistía fundamentalmente en digitar, para cada pregunta de la cédula, el código de fuente seguido de la respuesta numérica contenida en cada casillero. Igualmente, se indicó el procedimiento para efectuar la digitación, basado en el "Formato del Registro Creado" mediante el cual se precisaron los dígitos para:

- Identificación de la cédula - con un total de 8 dígitos correspondientes al estrato, sub-estrato, segmento, vivienda muestral y hogar.
- Datos de la cédula - con un total de 13 dígitos para cada casillero, los 4 primeros para el código de fuente y los 9 restantes para el valor de la respuesta anotada en el casillero.

En el manual se incluyó también instrucciones específicas para datos con enteros y decimales, para anotaciones hechas con bolígrafo de diferente color y en preguntas cuyos códigos de fuente habían sido generados en la etapa de revisión, conversión y codificación.

CAPITULO VII

PRUEBA DE LA OPERACION DE REGISTRO Y DE LA CEDULA

En este capítulo se presenta todo el proceso de la prueba de campo incluyendo la capacitación del personal de campo, la ejecución del trabajo y la evaluación de los resultados. Cabe señalar que el propósito principal de una prueba de campo es el detectar oportunamente posibles errores en los documentos que se han elaborado para aplicar las medidas correctivas que sean necesarias. Otra finalidad de la prueba fue precisar los tiempos de duración del registro de viviendas y de las entrevistas, y el tiempo promedio de desplazamiento de un hogar a otro.

Capacitación de los Registradores y Encuestadores

Para la prueba que se realizó en las tres regiones naturales del país entre el 29 de febrero y 6 de marzo de 1984 fue necesario capacitar a 31 personas seleccionadas por los Directores de las Oficinas de Estadística del Ministerio de Agricultura en los departamentos en que se realizó la prueba de campo, es decir, de Lambayeque, Junín y Huánuco. De las personas capacitadas, sólo quince fueron seleccionados para la prueba. El número de instructores fue de 8 de los cuales 7 eran funcionarios de planta o contratados del INE y uno era de la OSE que había sido destacado al INE para las labores de la ENAHR.

La capacitación teórica se realizó en Lima y la práctica de campo en el centro poblado de Lurín a 15 km. al sur de Lima. La duración de la capacitación fue de 6 días, del 20 al 25 de febrero de 1984, y su horario de 8:00 a 13:00 y de 15:00 a 18:00 horas. Sin embargo, durante los primeros 4 días de la capacitación los instructores se vieron obligados a alargar el horario hasta las 19:00 horas para cubrir todos los temas programados. De los 6 días, 4 se utilizaron para exponer la parte teórica, 1 para la práctica, y el último, para la evaluación escrita y clausura del seminario. La selección de los 15 encuestadores y registradores se hizo en base a los resultados de la prueba escrita, la asistencia y la puntualidad de los participantes.

El centro poblado de Lurín fue escogido para la práctica de campo debido a su cercanía a Lima y por la diversidad de cultivos en esa zona. Para esta práctica se formaron equipos conformados por un supervisor y 4 encuestadores los cuales trabajaron en parejas. Los instructores acompañaron a los encuestadores a localizar, según el croquis, las manzanas designadas. Una vez localizadas, cada pareja efectuaba el registro de viviendas (cada participante registró una manzana). Posteriormente, cada pareja elegía 2 viviendas con productor agropercuario para encuestar. Mientras que uno actuaba como encuestador, el otro observaba y viceversa para la siguiente vivienda a encuestarse. Después de terminar, los participantes se reunieron con los instructores para la evaluación de la práctica.

Primera Prueba de Campo

La primera prueba a nivel nacional se realizó en las tres regiones naturales del país entre el 29 de febrero y el 6 de marzo de 1984. Durante esta prueba no solo se probó las cédulas sino también el Manual del Encuestador, el Manual del Registrador, el documento "Registro de Viviendas" y todos los procedimientos establecidos para la operación de campo de la ENAHR. Para la prueba se seleccionaron los centros poblados de Monsefú y Callanca en la Costa, Chupaca y Pincha en la Sierra, y Aucuyacu y Santa Lucía en la Selva. La selección se efectuó en base a la diversidad de cultivos en estas zonas y la facilidad para el desplazamiento del personal. También se tomó en consideración al alto número de viviendas que indicaron estar vinculadas a la agricultura durante el último Censo de Población y Vivienda en 1981.

En el área urbana, se realizó el reconocimiento de las manzanas que conformaban el segmento asignado para constatar si las manzanas habían sufrido alguna modificación en su estructura desde el último Censo de Población y Vivienda; en los casos en que se constató alguna modificación, se realizaban los cambios necesarios en el plano urbano. Seguidamente, se continuaba con el registro de las viviendas. Una vez terminado el registro, el cual se completó en un sólo día en el área urbana, el registrador entregaba a su supervisor el documento "Registro de Viviendas" para que seleccionara de éste, aleatoriamente, 10 viviendas para encuestar. La carga de trabajo programada para cada encuestador en el área urbana fue de dos entrevistas por día.

En el área rural, se efectuó el reconocimiento del Área de Empadronamiento Rural (AER) utilizando un croquis distrital como guía y un listado referencial de jefes de hogares (Doc. CPV-03.51) obtenido del Censo de Población y Vivienda de 1981. También se contó con el apoyo brindado por las autoridades locales para graficar los límites del segmento y la ubicación de algunas viviendas. Finalizado el reconocimiento, se inició el levantamiento del croquis del segmento rural. Según se visitaban las viviendas dentro del segmento, se iba marcando su localización dentro del croquis y registrándose en el documento "Registro de Viviendas." Como el tiempo para registrar todas las viviendas en un segmento rural se estimaba que iba a ser igual que el tiempo programado para la prueba, se decidió que mientras uno de los miembros de la brigada efectuaba el registro de las viviendas, el otro realizaba la encuesta. De hecho, para la prueba se asignaron dos personas para encuestar y una para registrar las viviendas en cada segmento del área rural. El encuestador a cargo de los hogares sin productor agropecuario tenía que realizar cuatro entrevistas por día, mientras que el de hogares con productor agropecuario hacía dos entrevistas diarias. Sin embargo, en el área rural, al igual que en el área urbana, se realizaron más entrevistas que las que habían sido programadas (ver Apéndice II).

Cinco de los instructores que participaron en la capacitación, al igual que varios técnicos de la OSE, cumplieron la función de supervisores. En calidad de observadores participaron dos asesores de BUCEN. Los vehículos para transportar el personal de campo a los segmentos donde se realizó el trabajo fueron proporcionados por las Oficinas Regionales de Estadística donde se encuentran ubicados los centros poblados seleccionados.

Para dar a conocer la ejecución del trabajo en Lambayeque, Junín y Huánuco, la OSE por intermedio de su personal especializado efectuó la difusión de la encuesta de la prueba de campo para lo cual preparó documentos, como cartillas y "cuñas." La difusión se inició en forma paralela a la prueba. Tanto en Monsefú (Costa), Huancayo (Sierra) y Tingo María (Selva) se efectuaron coordinaciones con las autoridades (Prefecto, Comandante de la Guardia Civil, Alcalde, Gobernador, etc.) para informarles sobre el desarrollo del trabajo. Igualmente, se hicieron anuncios por emisoras locales utilizando las "cuñas," y en Huancayo, se utilizó la televisión local. En la "Industria," periódico de Chiclayo, se publicaron notas de prensa sobre la prueba de campo. En todos los casos, los anuncios radiales y periodísticos tuvieron un impacto positivo en la población.

Evaluación de la Primera Prueba

Una vez finalizada la prueba de campo los supervisores de dicha operación se reunieron en Lima con los directivos del INE y de la OSE para redactar un informe de evaluación de la prueba. Este informe se elaboró en base a sus observaciones en el campo y los informes presentados por los registradores y encuestadores.

1. Evaluación de la Capacitación

El tiempo asignado para las instrucciones de temas específicos fue muy corto, por lo tanto, no se efectuaron las prácticas del llenado de las cédulas y del documento para el registro de viviendas.

2. Evaluación de la Operación del Registro de Viviendas

(a) Reconocimiento y Ubicación del Segmento: En el área urbana no se presentaron problemas significativos, sin embargo, en el área rural hubieron problemas relacionados con la delimitación de los segmentos. En algunos casos, por ejemplo en Callanca, el área de enumeración rural (AER) fue dividida en dos AER para el Censo de Población y Vivienda de 1981 porque el número de hogares en ese AER era más del doble del número promedio de hogares por AER, es decir, de 100 hogares; sin embargo, el límite divisor entre estos dos segmentos nunca fue precisado en el croquis distrital. La única manera de precisar el límite era refiriéndose al documento CPV-03.51 el cual contenía el nombre del centro poblado y los nombres de los jefes de hogar. Por consiguiente, aquellas viviendas ubicadas en la inmediación del límite de ambos segmentos y las cuales el nombre del jefe del hogar no constaba en el documento CPV-03.51, eran identificadas algunas veces en el segmento no correspondiente. Otro problema que se presentó fue en la Selva, en particular, en Santa Lucía. El croquis esquemático no se pudo elaborar por la excesiva vegetación que impedía la visibilidad.

(b) Registro de Viviendas: Se encontró que había un gran número de viviendas con ocupantes ausentes. También se detectó que las preguntas del documento "Registro de Viviendas" estaban redactadas a un nivel muy técnico y difícil de comprender por la población.

- (c) **Manual del Registrador:** El manual indicaba con claridad los procedimientos que el registrador debía seguir en el trabajo de campo en relación con la delimitación del segmento y el registro de viviendas.

3. Evaluación de las Entrevistas

- (a) **Mecánica para Iniciar la Entrevista:** Por no haberse instruido a los encuestadores en la forma como debían presentarse, ni en la mecánica para iniciar la entrevista, se observó que los encuestadores tenían dificultad para motivar a los informantes dado que no podían explicar claramente la finalidad de la visita ni los objetivos de la ENAHR.
- (b) **Número de Entrevistas:** La carga de trabajo asignada por encuestador de dos entrevistas en viviendas con productor agropecuario fueron relativamente holgadas, pues en todos los casos se cumplió con esta asignación.
- (c) **Funcionabilidad de las Cédulas:** La entrevista se hizo tediosa por el gran número de preguntas que tenía la cédula tipo 2, y se acrecentó aún más porque el encuestador hacía cálculos adicionales para llenar algunas preguntas sin intervención del informante. Esos momentos "vacíos" causaban desagrado e inquietud a los informantes, manifestando en repetidas ocasiones sus deseos de concluir la entrevista.

En la cédula tipo 2 el encuestador tenía que elaborar dos dibujos con la asistencia del productor. Uno de los dibujos exigía información puntual del mes de febrero de 1984, mientras que el otro requería información del período que se fijó para la prueba piloto, situación que propició mucha confusión en el traslado de datos a las preguntas de la sección correspondiente. Otro problema observado fueron los espacios designados para registrar las respuestas los cuales resultaron inapropiados por su reducido tamaño. Las preguntas relacionadas a la producción agrícola destinada al autoconsumo en el hogar y para la alimentación de los animales fueron difíciles de contestar. La sección sobre el inventario de los animales en la explotación era difícil de diligenciar por su diseño. Por último, el número de líneas para anotar el nombre y apellidos de todos los miembros del hogar fue insuficiente. Ambas cédulas tenían espacio para anotar hasta un máximo de 7 personas; en la prueba se dieron casos de hogares con 18 miembros.

- (d) **Manual del Encuestador:** El manual carecía de instrucciones básicas que le permitiera al encuestador desarrollar su trabajo debidamente. Entre las instrucciones que faltaban estaban la de ubicación de la vivienda seleccionada, la selección del informante, las medidas a tomarse en caso de rechazo de la entrevista, ausencia del informante y las viviendas desocupadas. Tampoco existía instrucciones para la revisión y devolución de los documentos de trabajo. Asimismo los encuestadores señalaron que el tamaño del manual no era adecuado y que se hacía difícil leer las instrucciones.

4. Evaluación de la Actividad de Comunicaciones

Debido a que el traslado del personal de difusión a las áreas donde se llevó a cabo la prueba piloto se produjo en forma paralela al del personal de campo, la campaña de difusión se inició en forma casi simultánea con el trabajo de campo. Consecuentemente, en un primer momento los informantes no tuvieron conocimiento de que se estaba realizando la prueba de la ENAHR. Más tarde y mientras duraba todavía la prueba, la campaña se dió con intensidad y precisamente en aquellas emisoras de sintonía en las áreas seleccionadas, lo que aseguró su audición.

Finalmente, se puede manifestar que la prueba alcanzó sus objetivos puesto que pudieron detectarse una serie de problemas en los documentos que se probaron, lo que dió lugar a que se hicieran los ajustes necesarios antes de iniciarse la operación de campo.

Cambios Debido a la Primera Prueba

Debido a los resultados de la prueba de campo llevada a cabo en los meses de febrero y marzo de 1984 se efectuaron los siguientes ajustes:

Para la capacitación de los registradores, se extendió el tiempo asignado para su desarrollo a una semana y para la capacitación de los encuestadores a dos semanas. También se programaron prácticas simuladas al término de las explicaciones de los temas de cada sección.

Para el registro de viviendas, se rediseño el documento "Registro de Viviendas" para facilitar su diligenciamiento y ampliar la información requerida para seleccionar la muestra. Por esta razón fue necesario elaborar nuevamente el manual, ampliándolo e introduciendo los ajustes necesarios. También se introdujeron instrucciones para solucionar el problema de las viviendas que al momento de la visita no tenían ocupantes presentes y para aquellas que estaban desocupadas.

En las cédulas, se mejoró su diseño sustancialmente. Por ejemplo, se utilizaron varios tipos de letras para las preguntas, indicaciones, advertencias, llamadas, etc. Además se introdujo el uso de color para sombrear las preguntas, dejando en blanco los espacios para la anotación de las respuestas. Las preguntas se modificaron para que fueran comprendidas fácilmente por los informantes. Se eliminó uno de los dos dibujos requeridos durante la prueba, y varias secciones y preguntas fueron reordenadas para mejorar el flujo de la entrevista. El número de líneas para anotar los nombres de los miembros del hogar fue incrementado de 7 a 22, al igual que se amplió los espacios donde se registrarían las respuestas. Para efectuar los ajustes debido a la inflación en el país, se añadieron preguntas sobre la fecha en que ocurrían las transacciones.

El Manual del Encuestador se elaboró nuevamente con el objeto de introducir los cambios efectuados en las cédulas. Su tamaño fue ampliado para facilitar su lectura. Se introdujo un capítulo de normas y técnicas de la entrevista y otro que trataba de la revisión y devolución de las cédulas diligenciadas. También se elaboraron ejemplos típicos relacionados con las regiones naturales del país.

Para la actividad de comunicaciones, la OSE preparó almanaques para el año 1985 los cuales se obsequiaron a los informantes. Se prepararon "cuñas" tanto en castellano como en Quechua para ser propaladas por las emisoras locales del país. Asimismo, se prepararon oficios de presentación para los hogares seleccionados a participar en la encuesta.

Segunda Prueba de Campo

La segunda prueba de campo se realizó del 28 de mayo al 01 de junio de 1984 en el departamento de Junín, en los distritos de Tarma, Sacsamarca, Acobamba y Pícoy en la Sierra y Chanchamayo en la Selva. El propósito de la prueba fue determinar la funcionabilidad de la cédula tipo 2, la cual había sido sustancialmente modificada en base a los resultados de la prueba anterior. Principalmente, se quiso determinar si era posible recoger información sobre los cultivos por campaña, la cantidad de cosechas por cultivo y los gastos asociados con cada uno. La selección de las localidades para la prueba se efectuó en razón de que en Tarma existe una diversificación de cultivos de corto período vegetativo, por ejemplo, las hortalizas, obteniéndose por año hasta más de cinco cosechas, y en Chanchamayo predominan los cultivos permanentes de gran rentabilidad. También se quiso investigar la factibilidad de preguntar el mes de siembra, cosecha y venta de cada cultivo al igual que el mes asociado con cada transacción a fin de ajustar los ingresos del hogar por inflación. Otro objetivo de la prueba fue el de medir el tiempo que tomaba diligenciar la cédula tipo 2 ya que existía una gran preocupación por el número de preguntas.

Los especialistas del INE, capacitaron insitu a 6 técnicos de la OSE de los cuales 4 actuaron como encuestadores y 2 como supervisores. Esta capacitación que se realizó del 24 al 26 de mayo fue prácticamente como un refrescamiento y afianzamiento, ya que el personal había participado en la prueba anterior. Además del personal de la OSE, 2 especialistas del INE participaron como supervisores formando junto con los de la OSE, 4 equipos de 2 personas cada uno y llegando a entrevistar en total 36 hogares.

Resultado y Evaluación

La prueba demostró que sí se podía recoger información sobre los cultivos por campaña, y que el promedio de cultivos por explotación fue de siete. También la prueba reveló que tomaba aproximadamente dos horas y media diligenciar la cédula cuando el número de cultivos era 6 o más. Sin embargo, es importante señalar que la prueba se realizó en una zona del país donde el número de cultivos por explotación es relativamente alto comparado con otras zonas y, por lo tanto, se dedujo que el tiempo requerido para diligenciar la cédula tipo 2 sería menor. El porcentaje de casilleros con respuestas fue bajo según los resultados de la prueba. La cédula con menos información tenía el 4.6% de los casilleros llenos y la más alta tenía 25%, siendo el promedio de 10.4%. Se había estimado en base a las experiencias con encuestas similares en otros países que las cédulas tuvieran entre el 15% y el 20% de los casilleros con información.

Para la prueba no fueron muchas las explotaciones que tenían facilidades de crédito y asistencia técnica; sólo 12 de las explotaciones, o sea el 33% tenían crédito y 3, 6 8.3%, recibían asistencia técnica. Los resultados

demostraron también que habfan cultivos donde los gastos fueron mayores que los ingresos. En conclusión, la prueba indicó que la cédula tal como había sido rediseñada funcionaba y que no era necesario modificarla. Los pocos problemas que se observaron tenían que ver mas bien con la capacitación de los encuestadores que con la cédula en sí. De hecho, para la capacitación del personal de campo se recomendó que se incluyeran ejemplos que reflejaran la realidad del trabajo a emprenderse; es decir, que se ofrecieran varios ejemplos de situaciones con diferentes cultivos para así darle un mejor conocimiento a los encuestadores.

CAPITULO VIII

CAPACITACION Y TRABAJO DEL PERSONAL DE CAMPO

La OSE fue la responsable del trabajo de campo para la ENAHR el cual constó de dos etapas: (1) el registro de viviendas en los segmentos seleccionados, y (2) el diligenciamiento de las cédulas en las viviendas seleccionadas. Para cada una de las etapas se realizó un curso de capacitación a fin de que los procedimientos a seguirse durante el desarrollo del trabajo fueran de conocimiento del personal encargado de realizar estas tareas, asegurando la aplicación uniforme de las normas establecidas.

Capacitación de Registradores

El curso de capacitación para la ejecución del registro de viviendas estuvo a cargo del personal del INE y de la OSE. También participó en calidad de observador e instructor un funcionario del BUCEN. El personal del INE se encargó de impartir las instrucciones teóricas y de la realización de las prácticas, y el personal de la OSE se responsabilizó de la organización del evento. Un total de 148 funcionarios de la OSE y las Regiones Agrarias fueron capacitados de los cuales 115 eran registradores y jefes de brigadas, 23 jefes departamentales y 10 observadores. El número de instructores fue de 7 incluyendo 3 cartógrafos.

Para el desarrollo del curso, que se realizó entre el 2 y 7 de julio de 1984, se eligió las instalaciones del Centro Vacacional de Huampaní ubicado a unos 50 kilómetros al este de Lima. Se eligió este sitio por reunir las condiciones apropiadas para el alojamiento y alimentación de los participantes, y contar con aulas en número suficiente y un auditorio que permitiera realizar sesiones conjuntas. Teniendo alojado a todos los participantes en un mismo local garantizaba la asistencia al curso y ofrecía también la oportunidad para que los participantes e instructores se reunieran por las noches, si era necesario, para aclarar cualquier tema discutido en clase. Sin embargo, la razón principal para centralizar la capacitación fue que no había suficientes instructores para descentralizar el evento a nivel de región natural. Específicamente, el presupuesto de la encuesta no permitía contratar más técnicos de los que fueron contratados.

Para el dictado del curso se organizaron tres grupos formados en función a la región geográfica en que el personal de campo efectuaría su trabajo, a saber, Costa, Sierra y Selva. Esta organización se planteó con la finalidad de lograr una mejor apreciación y conocimiento de las condiciones típicas de cada región.

Los temas tratados fueron:

- Plan del Curso de Capacitación
- Sistema Estadístico Nacional y su Programa de Encuestas
- Objetivo de la Encuesta Nacional de Hogares Rurales
- Diseño Muestral utilizado en la ENAHR
- Exposición del Manual del Registrador (Tratado por temas según capítulos)
- Realización de una Práctica de Campo
- Discusión y Evaluación sobre la Prueba
- Exposición del Manual del Jefe de Brigada

El tiempo programado para las exposiciones de los temas fue de ocho horas diarias (8:30 - 12:30 y 14:00 - 18:00 horas) y se desarrollaron en cuatro días. La prueba de campo, a su vez, tomó un día y medio y se realizó en los lugares cercanos a Huampaní. A los encargados de efectuar el registro se les sometió a una prueba de conocimientos mientras que al personal pre-seleccionado para desempeñarse como jefes de brigadas y departamentales se les impartió las instrucciones contenidas en el Manual del Jefe de Brigada.

Aunque el nivel de instrucción fue bueno y todos los temas importantes fueron cubiertos satisfactoriamente, la metodología empleada por los instructores no guardó uniformidad lo que produjo algunas contradicciones y confusión entre ellos. Este impase sin embargo fue superado antes de la conclusión del curso. Por otra parte, la participación de los cartógrafos permitió una mejor preparación en el manejo y elaboración de documentos cartográficos. Ellos cubrieron las sesiones pertinentes al uso de planos durante el recorrido de los segmentos seleccionados y el dibujo del croquis. Para los ejercicios prácticos del curso de capacitación, los cartógrafos también brindaron su apoyo, delineando en los planos censales los segmentos adyacentes a Huampaní. El ejercicio práctico en el área urbana tuvo una duración de medio día y en el área rural de un día. Los participantes fueron divididos en grupos, cada uno conformado de cuatro registradores y un jefe de brigada. Aunque el tiempo programado no fue suficiente para registrar todas las viviendas, los registradores hicieron un buen trabajo, identificando los límites del área de trabajo asignada, dibujando el croquis y diligenciando el documento "Registro de Viviendas" en parte del segmento.

Finalmente, vale decir que el personal capacitado demostró estar sumamente interesado por los temas expuestos y, en repetidas ocasiones, efectuó aportes valiosos durante el dictado de las clases teóricas así como en las prácticas, dado su experiencia en recolección de información estadística en el Sector.

Registro de Viviendas

La operación de campo para el registro de viviendas se ejecutó entre 18 de julio y el 31 de agosto de 1984 y contó con la participación de 137 personas pertenecientes a las 23 Regiones Agrarias, formando 32 equipos de trabajo. Cada equipo fue encabezado por un jefe de brigada y contó con el apoyo de los jefes departamentales. El número de registradores ascendió a 91, de donde se deduce que cada equipo contaba con un promedio de 3 registradores. Para aquellos departamentos donde el número de registradores era menor de 3, se decidió que el jefe departamental asumiera las labores del jefe de brigada (ver Apéndice 12). Teóricamente hubiera sido conveniente que cada jefe de brigada contara con mayor número de registradores, pero las condiciones geográficas y la dispersión de la muestra demostró que el número de registradores por brigada fue el adecuado. La realización del trabajo de campo contó también con la presencia de 16 técnicos, 11 de la OSE y 5 del INE, quienes desempeñaron el rol de supervisores nacionales y apoyaron decididamente en el desarrollo del trabajo.

Para la movilización del personal de campo dentro de sus respectivas regiones se contó con 31 vehículos entre camionetas, automóviles y motocicletas que fueron puestos a disposición de la encuesta por las diferentes Regiones

Agrarias a tiempo completo. La OSE apoyó con 4 vehículos en los departamentos de Ancash, Ayacucho, Apurímac y Cusco.

El material para la operación de campo fue proporcionado a los jefes departamentales al finalizar el curso de capacitación, lo cual les obligó a permanecer dos días más de lo programado en Lima para efectuar la revisión de la documentación recibida. La demora en la entrega del material se debió a 3 factores: primero, una huelga de tres semanas de duración en el sector público que retrasó la preparación del material; segundo, la descomposición de la máquina fotocopidora trajo como consecuencia que el trabajo de reproducción de croquis y otros documentos se terminara apresuradamente durante el período de capacitación; y por último, el cambio del diseño muestral acordado por el INE y la OSE pero no por los asesores de BUCEN y el cual motivó a que se regresara nuevamente al diseño original.

El material cartográfico demostró tener algunas deficiencias en cuanto a los límites de los segmentos y ubicación de los centros poblados en el ámbito de las áreas de trabajo seleccionadas. En algunos segmentos se encontraron significativas diferencias entre el número de viviendas consignadas en el listado correspondiente según el último Censo de Población y Vivienda realizado en el país y el número real empadronado en el registro de viviendas. En la mayoría de estos casos, el número de viviendas registradas era mayor a lo indicado por los datos censales. No obstante, el registro de viviendas fue realizado en forma satisfactoria debido a la experiencia del personal de campo. Los factores limitantes tales como: la falta de vehículos motorizados, falta de acémilas, caminos de difícil acceso y lugares donde no se expendía alimentos fueron superados por los registradores. El trabajo de campo desarrollado por los registradores en los segmentos seleccionados para Ayacucho, Huancavelica y Apurímac merece un reconocimiento especial por cuanto dicha zona fue declarada por el gobierno en "estado de emergencia" debido a las actividades de grupos sediciosos, y aún cuando se conocía la peligrosidad de esas zonas, se llevó a cabo el registro, quizás no con todo el éxito deseado pero sí respecto al logro de las metas trazadas en forma significativa.

Capacitación de los Instructores

A fin de mejorar el nivel de instrucción para el curso de capacitación de los encuestadores, se adoptaron algunas medidas correctivas tendientes a superar los problemas y deficiencias encontrados durante la capacitación de los registradores. Es así que se convino realizar un seminario taller con el personal que iba a desempeñarse como instructores en el curso de capacitación de los encuestadores. Además de capacitar a los instructores, otros objetivos del seminario taller fueron: (1) uniformizar los criterios técnicos que se emplearían durante la capacitación de los encuestadores; y (2) definir la mecánica de trabajo que debía utilizarse para una mejor comprensión de los temas.

El seminario taller se desarrolló del 10 al 15 de setiembre de 1984 en el mismo lugar que se realizó la capacitación de los registradores, es decir, en el Centro Vacacional de Huampaní. El personal administrativo de la OSE tuvo a su cargo la organización del evento. Un total de 16 personas

fueron capacitadas como instructores de los cuales 3 eran de la sede central de la OSE, 10 eran de las regiones agrarias, y 3 eran del INE. En calidad de instructores para este seminario taller participaron un funcionario del INE, de la OSE y del BUCEN. El horario previsto para la exposición de los temas fue de 8:00 a.m. a 6:00 p.m. incluyendo 2 horas para el almuerzo; sin embargo, en 4 de los 5 días el número de horas de clases excedió en 2 horas lo programado, terminándose a las 8:00 p.m.

Los participantes fueron sometidos a un programa de capacitación intensivo presentado en tres etapas. La primera etapa consistió en una revisión y análisis de la cédula tipo 2 el cual se efectuó sección por sección. Estas secciones de trabajo se desarrollaron con la presencia de todos los participantes a través de la modalidad de mesa redonda; es decir, cada uno de los participantes rotativamente dió lectura a una parte del Manual del Encuestador. Conforme se fue desarrollando la lectura, se explicó los aspectos que se estimaron problemáticos o de diversa interpretación. La duración de esta primera etapa fue de dos días. La segunda etapa que tuvo un período de duración igual a la primera, fue mas bien un afianzamiento y especialización en los aspectos técnicos. También fue en ella donde se designaron los instructores para el curso de capacitación de encuestadores y jefes de brigadas y departamentales. La última etapa fue una práctica de campo de un día de duración, la cual se realizó en los alrededores de Huampaní. Para la práctica, a cada participante se le asignó un hogar rural para entrevistar. Estos hogares fueron seleccionados por la OSE en base a la predisposición del jefe del hogar a cooperar. También el personal administrativo de la OSE hizo todas las coordinaciones con las autoridades locales para obtener la colaboración y garantías necesarias. Concluida la práctica, los participantes se reunieron para efectuar una evaluación del trabajo y formular recomendaciones para el curso que ellos tenían que dictar.

Capacitación de los Encuestadores, Jefes de Brigadas y Departamentales

Una vez concluida la capacitación de los instructores, se dió inicio al curso de capacitación del personal de campo. Al igual que los cursos anteriores, éste se realizó en el Centro Vacacional de Huampaní del 17 al 28 de setiembre de 1984. El número de participantes fue de 148 distribuido de la siguiente manera: 24 jefes departamentales, 23 jefes de brigadas y 101 encuestadores (ver Apéndice 13). El horario del curso fue de 8:30 a.m. a 5:30 p.m. de lunes a viernes incluyendo las 2 horas del almuerzo; el día sábado el horario fue de 8:30 a.m. a 12:30 p.m.

El curso se desarrolló satisfactoriamente puesto que se adoptaron muchas medidas tendiente a superar los problemas administrativos encontrados en el curso de capacitación de los registradores, como por ejemplo, el uso de ayuda de audio-visuales (duplicados de las páginas de la cédula en láminas transparentes para usarlas con un proyector). También para hacer el curso más comprensivo y asimilable, los participantes fueron agrupados en tres aulas, una por cada región natural y cada una de ellas contando con 5 instructores para aproximadamente 50 participantes. Solamente para las ceremonias de inauguración y clausura se reunieron los tres grupos.

El dictado del curso se desarrolló mediante exposiciones de temas, poniendo bastante énfasis en ejemplos que reflejaban las realidades de las tres

regiones naturales. Para complementar y ampliar los conocimientos teóricos, se desarrollaron dos tipos de prácticas: simuladas y de campo. Las prácticas simuladas consistieron en diligenciar, en clase, cada sección de la cédula tipo 2. Para este ejercicio, el instructor hacía el papel de informante mientras que los participantes tomaban turno para desempeñarse como encuestadores. La práctica de campo, por otra parte, les ofreció a los participantes la oportunidad de desenvolverse como encuestadores bajo condiciones reales. El propósito de la práctica era que los participantes efectuaran una aplicación de los conocimientos adquiridos a través de las exposiciones teóricas y para eso se efectuaron entrevistas en hogares rurales del departamento de Lima. Para las aulas de Costa, Sierra y Selva, se seleccionaron respectivamente las localidades de Barranca, Huacho y Huaral. La práctica de campo tomó 2 días, utilizando los participantes el primer día para hacer un reconocimiento de las viviendas asignadas y el segundo para diligenciar la cédula. Los instructores se desenvolvieron como jefes de brigadas, supervisando así toda la operación. También se les exigió que evaluaran las cédulas diligenciadas del personal componente de su grupo para determinar los errores e inconsistencias más significativas a fin de discutirlos en clase. Más de 150 cédulas fueron diligenciadas de las cuales 50 fueron posteriormente seleccionadas para probar el sistema de procesamiento.

Además de capacitar conjuntamente a los encuestadores, jefes de brigadas y departamentales, se realizó por separado un seminario de un día de duración para los jefes de brigadas. El énfasis del seminario fue en las obligaciones y deberes de los jefes de brigada durante la operación de campo, por ejemplo, observación y control de las entrevistas, crítica en el campo de las cédulas diligenciadas, solución de problemas encontrados por los encuestadores, etc. El personal que participó en este seminario fue pre-seleccionado por los Directores de las Regiones Agrarias del Ministerio de Agricultura por sus años de servicios. Obviamente, hubo quejas por partes de los encuestadores, los cuales manifestaron que muchos de ellos sabían más sobre el trabajo que las personas que fueron designadas como jefes de brigada. La observación de los encuestadores fue comunicada al Director General de la OSE para el análisis de la situación y, de ser el caso, dispusiera las acciones correctivas. Desafortunadamente el Director General de la OSE no tenía la autoridad para seleccionar a otras personas como jefes de brigada ya que los Directores de Regiones Agrarias no están dentro de su ámbito administrativo.

Organización del Trabajo de Campo

De las 148 personas capacitadas, 147 participaron¹ en el trabajo de la operación de campo mediante la siguiente organización: a nivel de departamento se designó al Director de Estadística como jefe departamental de la ENAHR; éstos a su vez, designaron entre su personal a los jefes de brigada. El 95% de los encuestadores fue seleccionado del personal que participó en la operación del registro de viviendas. Del personal capacitado de la Sede Central de la OSE se seleccionó a los supervisores nacionales y a los encuestadores suplentes

¹/ El departamento de Madre de Dios no salió seleccionado en la muestra, por lo tanto, el director de estadística de dicho departamento no tuvo que desempeñarse como jefe departamental.

los mismos que en algunos casos tuvieron que desplazarse a determinados departamentos con el objeto de reemplazar a los titulares. Además de 23 jefes departamentales, 23 jefes de brigadas y 101 encuestadores, participaron 44 personas que dieron apoyo logístico en la movilización del personal de campo.

El requerimiento de personal por regiones naturales fue el siguiente:

Región Geográfica	Total	TECNICOS				Auxiliar
		Encuestadores		Jefes de Brigada	Jefes Dept.	Choferes
		Efectivo	Suplente			
Total Nacional	147	97	4	23	23	44
Costa	50	33	-	7	10	16
Sierra	73	48	3	12	10	20
Selva	24	16	1	4	3	8

Diligenciamiento de las Cédulas

La operación de campo se programó para que comenzara el 3 de octubre de 1984, es decir, apenas unos cinco días de finalizar la capacitación del personal que tendría a su cargo esta actividad. La conclusión de las actividades estaba prevista para el 11 de noviembre de ese mismo año; sin embargo, debido a un clima laboral inestable (huelgas en el sector público), se modificaron las fechas de acuerdo a las circunstancias laborales en cada departamento. El trabajo de campo en realidad se inició el 5 de octubre en muchos de los segmentos y se concluyó con la recopilación de la información el 29 de diciembre.

De los 419 segmentos programados, 408 fueron trabajados. Los 11 segmentos restantes no se trabajaron debido a problemas socio-políticos. Es importante mencionar que en algunas zonas de los departamentos de Ayacucho, Apurímac, Huancavelica, Cuzco, San Martín y Huánuco, los encuestadores trataron de realizar el trabajo recibiendo en repetidas ocasiones el apoyo del Ejército Peruano. Este último apoyó en forma de transporte vía helicóptero a los encuestadores a las zonas convulsionadas, pero los resultados fueron negativos por la ausencia de los pobladores en esas zonas rurales.

Al concluir el diligenciamiento de un segmento, se enviaban las cédulas a la OSE en Lima para su posterior procesamiento en el INE. Para el mes de enero de 1985, 6,812 cédulas habían sido recibidas en el INE, de las cuales 655 representaban casos de no-entrevistas por diferentes razones. Sin embargo, durante el chequeo de cobertura en las oficinas del INE, se detectó que faltaban dos segmentos del departamento de Amazonas y uno de San Martín. Estos segmentos fueron nuevamente encuestados en el mes de marzo de 1985, resultando en un total de 6,069 cédulas diligenciadas. De este total, 4,146 cédulas fueron de tipo 2 y 1,923 de tipo 1.

El tiempo promedio por entrevista (diligenciamiento de la cédula tipo 2) según un estudio conducido durante la operación de campo en los departamentos de

Lima, Ica, Cuzco y San Martín, fue de una hora y 52 minutos. La sección que tomó más tiempo en diligenciar fue la Sección II - Producción Agrícola y su Destino con un tiempo promedio de 25 minutos, seguido por la sección III, Gasto de Insumos para Cultivos con un promedio de 21 minutos.

Control de Calidad en el Campo

1. Registro de Viviendas

El control de calidad durante la operación de campo del registro de viviendas estaba bajo la responsabilidad del jefe de brigada. El jefe de brigada tenía que acompañar a uno de sus registradores para observar el trabajo y, progresivamente, visitar a los demás a fin de constatar que cada uno de ellos efectuara el trabajo según las instrucciones del Manual del Registrador. Específicamente, el jefe de brigada acompañaba al registrador a 20 viviendas como simple observador, sin intervenir directamente hasta después de completada la entrevista. Una vez finalizada la entrevista, el jefe de brigada le señalaba al registrador las fallas en que había incurrido y lo instruía a fin de ayudarlo a mejorar su trabajo.

Otra medida que se tomó para mejorar la calidad del trabajo fue la revisión por el jefe de brigada del documento HR-06, Registro de Viviendas una vez que el registrador terminaba un segmento. Si durante la revisión del documento el jefe de brigada hallaba errores que no podía corregir, le devolvía el documento al registrador y le ordenaba que regresara a la vivienda correspondiente para resolver el error.

Además del control de calidad que efectuaba el jefe de brigada, el registrador a su vez tenía el deber de verificar todo su trabajo. Esto incluía la revisión diaria del documento HR-06, Registro de Viviendas y, al finalizar el segmento, la revisión de los planos urbanos o croquis graficados para el área rural según el caso.

El trabajo de campo del registro de viviendas fue realizado en forma satisfactoria. Los errores que se detectaron fueron pocos y tenían que ver más bien con factores ajenos al trabajo de los registradores.

2. Diligenciamiento de las Cédulas

La responsabilidad principal de revisar todo el trabajo de los encuestadores recaía en los jefes de brigada. Ellos tenían que verificar que las cédulas hubiesen sido diligenciadas de acuerdo a las normas establecidas en el Manual del Encuestador. Esta revisión consistía en verificar que se hallan formulado todas las preguntas y que las respuestas se hallan anotado en forma clara, completa y correcta. En tal sentido, el jefe de brigada tenía que realizar la crítica y consistencia del 100% de las cédulas diligenciadas por sus encuestadores durante las dos primeras semanas de trabajo. Después de la segunda semana, el jefe de brigada tenía que continuar revisando las cédulas pero a razón de dos cédulas por segmento, a no ser que el número de errores detectados hubiese superado el límite establecido.

Otra medida tomada para asegurar la validez de la información fue la re-entrevista. El jefe de brigada tenía la obligación de visitar nuevamente

a dos viviendas por segmento que habian sido encuestadas y hacerle unas preguntas breves a los jefes de hogares para evaluar la calidad de la información recogida por los encuestadores.

Los problemas que se detectaron durante el procesamiento de la información indican claramente que la supervisión en el campo no fue buena. Se encontraron errores que pudieron haberse evitados si los jefes de brigada hubiesen efectuado la crítica de las cédulas diligenciadas como se les indicó en el Manual del Jefe de Brigada. Afortunadamente, los errores se pudieron corregir automáticamente durante el procesamiento por computadora.

CAPITULO IX

CAPACITACION Y TRABAJO DEL PERSONAL DE OFICINA

En este capítulo se describe las tres fases del trabajo que efectuaron los crítico-codificadores durante el procesamiento de los datos de la ENAHR. La primer fase, antes de que la información recopilada en las cédulas fuese digitada, consistió de tres tareas distintas pero trabajadas en forma paralela. Estas tareas efectuadas, por el personal encargado, en las cédulas diligenciadas fueron: 1) revisión de la información, 2) codificación de respuestas literales contenidas en algunas preguntas, y 3) conversión de cantidades anotadas con decimales a enteros. Posteriormente a la digitación, se realizó la segunda fase que comprendió las siguientes tareas: 1) análisis de cobertura, 2) chequeo y proceso de los datos de conversión, 3) chequeo de rangos, y 4) análisis de consistencia. La última fase del trabajo fue la revisión y análisis de las tabulaciones. Sin embargo, antes de describir cada una de estas fases primero se detalla como se seleccionó y capacitó el personal encargado de las tareas mencionadas.

Selección y Capacitación de los Crítico-Codificadores

Para seleccionar a los crítico-codificadores que desempeñarían las tareas anteriormente mencionadas, se efectuó una evaluación de los postulantes a través del análisis de su curriculum vitae, así como una entrevista personal. De todos los postulantes se seleccionaron 55, los cuales participaron en el curso de capacitación para crítico-codificadores. Los objetivos del curso fueron:

- 1) Seleccionar al personal que se encargaría de desarrollar la primera fase del trabajo, es decir, las tareas de revisión, codificación y conversión.
- 2) Capacitar al personal en los procedimientos y reglas establecidas para la primera fase del trabajo.
- 3) Adiestrar al personal en el manejo de las cédulas y documentos auxiliares.

La responsabilidad del curso en cuanto a su coordinación, conducción y control estuvo a cargo del personal permanente y contratado del INE, habiendo participado como instructores en total cinco técnicos. El curso que se desarrolló en Lima tuvo una duración de seis días, y las clases se dictaron de 8:30 a 12:30 horas en las mañanas y de 14:00 a 17:00 horas en las tardes. La duración de cada clase fue de 50 minutos. La metodología empleada en el dictado de clases consistió en exposiciones teóricas complementadas por ejercicios prácticos.

Concluido el curso de capacitación se evaluó a los participantes y se seleccionó a los primeros 25 de acuerdo a su calificación. Cabe señalar que 9 de los 25 crítico-codificadores renunciaron a sus cargos por razones económicas antes de finalizar la segunda fase del trabajo, lo que prolongó más de lo programado el tiempo establecido para la validación de los datos.

Revisión de la Información

La tarea de revisión se efectuó a fin de verificar que las anotaciones contenidas en las preguntas de las cédulas diligenciadas fueron hechas de acuerdo a las instrucciones contenidas en el Manual del Encuestador. Para cada cédula la tarea de revisión consistió en:

- aclarar respuestas de difícil entendimiento;
- anular información innecesaria;
- eliminar información anotada en casilleros inapropiados y transferirlas a los casilleros, cuadros o renglones correspondientes;
- convertir respuestas literales a números arábigos en los casilleros que requerían cifras;
- verificar que el número de alternativas marcadas en las preguntas precodificadas estuviese de acuerdo a las instrucciones impartidas para su diligenciamiento;
- corregir los casilleros correspondientes a cantidades que traían más de un número anotado;
- anular páginas en blanco;
- comparar la anotación en el cuadro "Situación de la Cédula" con la información contenida en la cédula para verificar que estuviese correcta;
- eliminar decimales en cantidades referidas a soles; y
- comprobar que aquellas respuestas que deberían ingresar siempre con dos decimales lo estuviesen.

De los problemas que se detectaron durante la revisión de las cédulas, los que más sobresalieron fueron aquellos relacionados a la legibilidad de la información existente y los referentes a la situación de la cédula. Con respecto al primero, la mayor frecuencia se presentó en los casilleros donde se anotaron unidades de superficie y de peso. Sobre el segundo, no hubieron problemas determinando si una cédula estaba diligenciada íntegramente, pero sí cuando estaba diligenciada parcialmente. Inicialmente, se consideró "parcialmente" cuando faltaba información en algunas preguntas que deberían haberse contestado; sin embargo, el volumen de cédulas "parcialmente" demostró ser demasiado elevado, lo que dio lugar a que se cambiara el concepto para cuando la cédula carecía de información en una sección como mínimo.

Codificación

Esta tarea tuvo como finalidad la asignación de códigos a la información literal contenida en algunas preguntas de las cédulas diligenciadas con el objeto de facilitar el procesamiento posterior de los datos por computadora. Las respuestas que se codificaron fueron aquellas relacionadas a los nombres de: la ubicación geográfica, unidad de medida de la superficie, de los cultivos, ocupación y actividad económica de los miembros del hogar.

En líneas generales, no hubieron problemas que no pudieron superarse. Con respecto a la codificación de la unidad de medida de la superficie, se encontraron unidades que no figuraban en la relación de códigos. Estas fueron "Saco" para la zona de Junín con un equivalente de 0.06 hectáreas y "Taclla" para la zona del Cusco con un equivalente de 0.05 hectáreas. Ambas medidas fueron añadidas a la relación, así como sus equivalentes.

Con respecto a los cultivos tampoco se tuvieron problemas, tan sólo por las denominaciones propias que dan a los cultivos en algunas regiones; tal es el caso del zapallito italiano al cual le denominan "shupi" y a la arracacha que le denominan "virraca". Afortunadamente, los pocos casos que se dieron fueron solucionados por los mismos crítico-codificadores ya que algunos de ellos procedían de esas regiones donde se denominan a ciertos cultivos con nombres diferentes a los que se incluyeron en la relación.

La codificación de las preguntas sobre ocupación y actividad económica presentaron problemas en cuanto a la ambigüedad de las respuestas. Por ejemplo, se dieron casos en que la respuesta para ocupación era "empleado" o "agricultor" y, para rama de actividad, "empresa" o "fábrica". Para estos casos se tuvo que hacer un análisis minucioso de la cédula, tomando en cuenta la ubicación geográfica, el ingreso de los miembros del hogar e inclusive otras características similares de cédulas diligenciadas dentro del mismo segmento.

Conversión

La tarea de conversión consistió en transformar cantidades anotadas con decimales a enteros, a excepción de las cantidades referidas a superficie. Esta tarea se desarrolló sin ningún contratiempo. Todas las instrucciones para llevar a cabo la tarea de conversión estaban detalladas en el Manual de Codificación y Conversión. También, a cada crítico-codificador se le proporcionó una calculadora para minimizar los errores de cálculo y acortar el tiempo de proceso.

Análisis de Cobertura

El análisis de cobertura fue la primera tarea de la fase de Crítica por Computadora que tuvo por finalidad analizar y corregir los errores detectados por la computadora relacionados con la identificación y estructura de las cédulas; por ejemplo, omisión, duplicidad y situación de la cédula, así como los errores referentes a datos obligatorios, duplicidad e invalidez de códigos de fuente.

El análisis de cobertura fue realizado por cinco equipos de trabajo conformado cada uno por un jefe de equipo y cuatro crítico-codificadores; a su vez, los equipos estaban bajo la supervisión del Jefe de Crítica por Computadora quien estaba encargado de revisar las correcciones hechas por los equipos. Las correcciones fueron hechas por medio de un programa de actualización basado en los listados emitidos por la computadora y en las hojas de corrección, las cuales fueron:

- Documento HR-31, "Hoja de Anulación de Cédulas"
- Documento HR-32, "Hoja de Anulación de Celdas"
- Documento HR 33, "Hoja de Modificaciones."

Por otro lado, se utilizó el Documento HR-07, "Listado de las Viviendas Seleccionadas" para generar el Archivo Maestro donde se encuentran listadas las 6,996 viviendas seleccionadas para la encuesta.

Durante el análisis de cobertura se detectó que el número total de cédulas recibidas del campo fue de 6,916. Esta cifra fue más baja de la que se esperaba recibir, ya que el número de viviendas seleccionadas para la encuesta fue de 6,996 y cada vivienda, a su vez, podía contener más de un hogar. La diferencia entre las dos cifras, sin embargo, se debió a que en varios de los segmentos en las zonas de emergencia (con problemas sociopolíticos) no se realizó la encuesta. El Apéndice 14 muestra el porcentaje de viviendas sin encuestar por estrato. Los estratos de Selva Alta Centro, Selva Alta Urbana y Selva Alta Sur, correspondiente en su mayoría a los departamentos con problemas sociopolíticos, fueron los que tuvieron las tasas más alta de viviendas sin encuestar.

De las 6,916 cédulas recibidas del campo, se eliminaron 184 cédulas por las siguientes razones:

- la vivienda fue encuestada en cédulas de tipo 1 y tipo 2, teniéndose que anular una de ellas según correspondía;
- por ser una cédula adicional, es decir, que el encuestador utilizó más de una cédula para anotar la información, mayormente, en las preguntas relacionadas con la producción de cultivos; y
- por no ser una vivienda seleccionada.

De las 6,732 cédulas restantes, 6,069 cédulas integra y parcialmente diligenciadas fueron utilizadas para la generación del archivo de datos reformateado. De estas 6,069 cédulas, 1,923 cédulas fueron de tipo 1 y 4,146 de tipo 2.

Chequeo y Proceso de Conversión

Para tabular cantidades con diferentes unidades de medida fue necesario convertirlas a una unidad uniforme o estándar. Esta operación se efectuó por medio de la computadora y consistió en convertir las diferentes unidades de medida de superficie, producción agrícola y producción pecuaria a hectáreas, kilogramos o litros según el caso. Sin embargo, antes de iniciar la operación de conversión, la computadora verificó que todos los códigos correspondientes a las unidades de medida y los datos de cantidades estuviesen completos; de no ser así, la computadora emitía un listado de errores, el cual se le entregaba a los crítico-codificadores para que determinaran las fuentes de error y efectuaran las correcciones necesarias.

Chequeo de Rangos

Todos los valores de las variables contenidas en las cédulas fueron verificados por la computadora para determinar si se encontraban dentro de los límites establecidos; de lo contrario, la computadora emitía un listado de errores que era analizado por los crítico-codificadores. El chequeo de rangos

comprendió la examinación de tres clases de respuestas: las precodificadas, las cuantitativas abiertas y las literales que fueron codificadas en la oficina.

Con respecto a las respuestas precodificadas, los errores que se detectaron mayormente estaban relacionados a la aceptación de "blanco" como respuesta. Cuando se establecieron los límites, inadvertidamente, no se incluyó el valor de "blanco" en las especificaciones como valor aceptable para algunas variables, lo que resultó que la computadora detectara estos valores como errores. Este problema fue corregido modificando el programa de chequeo de rangos para las variables con dicho problema.

Los errores relacionados con las respuestas cuantitativas abiertas fueron más problemáticos, demandando un lapso de tiempo mayor al establecido para su corrección. La sección sobre Producción Agrícola y su Destino de la cédula tipo 2 presentó la mayor frecuencia de errores. Aquellas cédulas con valores identificados por haberse sobrepasado de sus límites fueron analizadas por los crítico-codificadores a fin de solucionar los errores. De no encontrarse error con la anotación efectuada por el encuestador o con la digitación de la información, o si la respuesta era consistente con otra información en la cédula, se aceptaba el valor ajustando los límites del programa para esa variable. Cabe señalar que los límites inferior y superior para las preguntas con respuestas cuantitativas abiertas, básicamente, fueron establecidos utilizando la información de la propia encuesta (ver Documento No. 3, Capítulo IV).

Para el último grupo, respuestas literales codificadas en la oficina, los errores detectados fueron pocos y en su mayoría relacionados a la digitación. Por ejemplo, 11,181 códigos de cultivos fueron verificados, de los cuales 964 ó 8.6% estaban incorrectos.

Análisis de Consistencia

La tarea de análisis de consistencia fue la última de la segunda fase del procesamiento de datos en la cual se dejó la información libre de errores para ingresar a la fase de tabulación. La tarea propiamente dicha tuvo la finalidad de comprobar que la información de dos o más variable afines estuviesen correcta; de no ser así, la computadora o corregía los errores automáticamente basado en los algoritmos establecidos, o emitía un mensaje de error en forma de listado con toda la información correspondiente. La corrección de estos últimos fue hecha por los crítico-codificadores que se valían del listado de errores, las cédulas identificadas con errores y las especificaciones de consistencia para solucionar los errores. Los errores que no podían resolverse, después de analizar la información anotada en la cédula y el dibujo de la explotación agropecuaria (cédula tipo 2), eran marcados para imputación.

Cabe señalar que el programa de consistencia se llevó a cabo en dos partes: A y B, en vista del gran tamaño que representaba el programa elaborado para esta tarea. El número de vueltas necesarias para llegar a un estado de corrección óptimo fue de cinco. Durante la primera vuelta, tanto en la parte A como la B, los casos que se presentaron con mayor frecuencia fueron los siguientes:

Número de la Especificación	Número de Casos Comprobados	Número de Casos con Errores	Porcentaje Errados
1 A	53	30	56.60
16 B	26	26	100.00
29 B	106	55	51.89
29 C	69	36	52.17
48 B	7,652	2,363	30.88
52 B	1,187	386	32.52
56 C	2,515	957	38.05
89 A	187	90	48.13
216 C	37	14	37.84

Los errores detectados se deben a varias causas, entre éstas se encuentran los errores de campo, de digitación, de especificación, de programación, etc. Por lo tanto, durante la primera vuelta del programa de consistencia los errores detectados presentan una combinación de errores. Esto explica la alta incidencia de casos errados en cada una de las especificaciones enumeradas en el cuadro anterior. Los errores se corrigieron en su totalidad después de analizar el origen de cada uno de ellos, cosa que no se hubiera podido hacer si los errores solamente hubieran provenido del trabajo de campo. Es importante mencionar también que durante la primera vuelta del programa de consistencia (partes A y B), se efectuaron en los datos de la ENAHR un total de 757,025 comprobaciones, de las cuales 10,226 ó 1.35% estaban erradas. Este porcentaje es sumamente bajo para una encuesta de la índole de la ENAHR.

Como se mencionó anteriormente, los casos que no pudieron corregirse se imputaron automáticamente con valores promedio calculado de la información contenida en las cédulas del dominio donde se efectuó la imputación. A continuación se presentan las variables que fueron imputadas con valores promedio.

Variable	Número de Casos Comprobados	Número de Casos Imputados	Porcentaje
LIR A+11	11,148	5	0.04
LIR B+8	10,428	599	5.74
LIR C+6	57,932	21	0.04
LIR C+9	57,932	10	0.02
LIR D+10	57,932	1	0.00
LIR E+9	12,208	72	0.59
CF* 2005	4,132	30	0.73

*CF = Código de Fuente

Análisis de Tabulaciones

La última fase del trabajo ejecutado por los crítico-codificadores fue el análisis de las tabulaciones. Para esta tarea se conformaron tres equipos de

los cuales dos de estos contaban con un coordinador (jefe de equipo) y tres analistas (crítico-codificadores); el otro equipo fue conformado por un coordinador y cuatro analistas. Los jefes de equipo, a su vez, estaban bajo la supervisión de un Coordinador General quien fue el responsable de todo el trabajo. En conjunto, ellos tuvieron la responsabilidad de revisar los 153 cuadros programados que reflejan los resultados definitivos de la ENAHR. La tarea en sí consistió en:

- revisar la parte literal de cada cuadro para comprobar que reflejara el contenido del cuadro; y
- comprobar que los datos en los cuadros se tabularon de acuerdo a las especificaciones.

Para facilitar el análisis y evitar errores que podían ser introducidos en el proceso de revisión, se decidió trabajar primero con un archivo de prueba compuesto de 120 cédulas (13 de tipo 1 y 107 de tipo 2). La información de las 120 cédulas fue tabulada por la computadora y presentada en forma de cuadros de salida. Al mismo tiempo, la información fue también tabulada manualmente a fin de compararla con la producida por la computadora. Las discrepancias entre los resultados obtenidos manualmente y automáticamente fueron investigadas y resueltas por los analistas y jefes de equipo en coordinación con los programadores. Los errores detectados podían ser de tres tipos: 1) de programación, 2) de tabulación manual, o 3) de estructura. Con respecto a este último, los analistas tuvieron que consistenciar la información de los cuadros que contenía la misma variable, ya que se dieron casos en que una variable se había especificado en distintas maneras.

El número total de cuadros revisados durante esta tarea ascendió a 159 aunque en realidad el plan de tabulaciones consta de 153 cuadros. La diferencia consistió en que hubieron cuadros que tuvieron que dividirse por problemas de espacio, mientras que otros se tuvieron que combinar.

CAPITULO X

PROCESAMIENTO DE DATOS

El desarrollo del sistema de procesamiento de los datos de la ENAHR y el procesamiento mismo, se llevó a cabo en el INE, utilizando una computadora WANG VS-100. El desarrollo de especificaciones, programación, prueba y documentación del sistema estuvo a cargo de un analista de sistemas y tres programadores contratados por el INE para trabajar en todas las actividades de computación de la ENAHR, con apoyo de técnicos de la Oficina del Censo de los Estados Unidos de América. Para la depuración o "limpieza" de los datos se utilizó CONCOR, un paquete generalizado para la crítica de datos de encuestas o censos. Para la tabulación de los 159 tabulados programados para la ENAHR, se utilizó CENTS 4, un paquete generalizado de tabulación; y para el cálculo de variancias, se utilizó el paquete SUPER CARP (Cluster Analysis and Regression Program), desarrollado en la universidad de Iowa State en los EE. UU. Además de estos paquetes de programas generalizados, se desarrollaron varios programas en COBOL con propósitos específicos; por ejemplo, los de generar el archivo de las viviendas seleccionadas, efectuar el chequeo de cobertura y reformatear el archivo de datos, producir listados de máximos y mínimos, actualizar el archivo reformateado, etc.

Las principales etapas que se llevaron a cabo durante el procesamiento por computadora de los datos de la ENAHR fueron: 1) entrada de los datos o digitación, 2) cobertura y generación del archivo de datos reformateado, 3) conversión a unidades de medida estándar, 4) chequeo de valores válidos y consistencia, 5) imputación, 6) agrupación, cálculo de ingresos y adiciones al archivo, 7) generación del archivo de cultivos, 8) generación de tabulados, y 9) estimados de variancia (ver Apéndice 15).

Entrada de los Datos

La digitación de las 6,732 cédulas se realizó en el INE utilizando equipo de digitación Inforex. Especificaciones detalladas (ver Documento No. 3, Capítulo II) basadas en el formato de la cédula y especialmente en el esquema de códigos de fuente, fueron desarrolladas para capacitar al personal de digitación. Con el formato de códigos de fuente cada respuesta en la cédula fue identificada por un código de fuente numérico. Durante la digitación sólo se digitaron las respuestas y sus respectivos códigos de fuente. Este método de códigos de fuente resultó muy efectivo, reduciendo el tamaño de cada registro digitado, y minimizando el número de errores de digitación. La operación de digitación se llevó a cabo en forma continua, comenzando con la llegada de las primeras cédulas del campo entregadas al INE por la OSE. La digitación fue verificada inicialmente en un 100 por ciento y fué gradualmente disminuyendo a un 20 por ciento a medida que el personal de digitación se familiarizaba con la cédula. Las cédulas fueron digitadas por agrupación de cédulas de segmentos completos. Una vez digitados varios segmentos, estos eran grabados en cinta magnética y pasados a la próxima etapa del procesamiento.

Chequeo de Cobertura y Generación del Archivo de Datos

Esta etapa se realizó también en forma continua, procesando grupos de varios segmentos a medida que estos eran concluidos en la etapa de digitación. El programa de cobertura y generación del archivo de datos, codificado en COBOL, tuvo como función principal verificar que la identificación de cada cédula fuese válida y, una vez verificado esto, reformatar el archivo según el tipo de cédula en trece tipos de registros (ver Documento No. 4, "Guía del Usuario") para su posterior procesamiento. Como entrada a este programa, se procesó el archivo de cédulas digitadas y el archivo maestro; como salida, se procesó el archivo de cédulas reformateado (ver Apéndice 15).

El archivo de cédulas digitadas estaba compuesto por la información de la cédula, digitada en registros que contenían la identificación de la cédula y una serie de valores numéricos que representaban el casillero de la cédula (código de fuente) y su respectivo valor. Puesto que una cédula estaba compuesta por más de un registro, este archivo de cédulas digitadas fue clasificado utilizando los campos de identificación de la cédula como clave. El archivo maestro fue creado digitando el "Listado de Viviendas Seleccionadas" (documento MR-07) que contenía la información de todas las viviendas que fueron incluidas en la muestra. Posteriormente, este archivo maestro fue actualizado adicionando hogares correspondientes a viviendas con más de un hogar. De esta manera, cada registro en el archivo maestro contenía la identificación de todos los hogares, además de otros campos para controlar la situación de cada cédula durante su procesamiento.

Durante este proceso se buscaba una igualdad entre la identificación en el archivo de cédulas digitadas y el archivo maestro. Esto aseguró que cada registro perteneciente a una cédula se generara una sola vez con una identificación válida. Una vez verificada la identificación, se procesaban todos los registros con una misma identificación, creando así registros reformateados para cada cédula. Todas las identificaciones inválidas u omitidas eran listadas e investigadas por el personal de oficina para su solución.

Durante la creación del archivo de cédulas reformateado, se realizó una verificación preliminar de todos los códigos de fuente y sus valores asociados. También durante esta etapa, y para facilitar el proceso posterior, se convirtieron todos los campos de ingreso a miles de soles. Problemas como códigos de fuente duplicados, valores inválidos, errores en la estructura de la cédula, etc., eran listados. Estos listados de errores eran revisados por el personal de oficina a fin de resolver las discrepancias, recurriendo en la mayoría de los casos a la cédula misma para su verificación. El documento "Hoja de Modificaciones" fue utilizado para corregir valores en el archivo de cédulas reformateado. Estas hojas de corrección eran llenadas con la identificación de la cédula y una serie de códigos de fuente y valores asociados utilizados para la actualización. Una vez llenadas, eran digitadas y luego procesadas por un programa de actualización en COBOL, que aplicaba las correcciones al archivo de cédulas reformateado.

Este proceso de cobertura y generación del archivo de datos se realizó en tres vueltas, estando sujeto al flujo de segmentos digitados, que a su vez

reflejar no-entrevistas de hogares y de segmentos no entrevistados a causa de problemas sociopolíticos. En consecuencia el archivo de datos quedó listo para ser tabulado.

Tabulaciones

Un total de 159 cuadros fueron programados utilizando el paquete generalizado de tabulación CENTS 4. Especificaciones para los 159 cuadros fueron desarrolladas en el INE (ver Documento No. 3, Capítulo XI). La programación de los cuadros comenzó en el mes de mayo a cargo de un programador, y a medida que el procesamiento de los datos lo permitía, se incorporaban otros programadores. Para comprobar que los cuadros eran programados de acuerdo a las especificaciones se preparó un archivo de prueba compuesto de 120 cédulas, de las cuales 107 eran con productor agropecuario (tipo 2) y 13 eran hogares sin productor agropecuario (tipo 1). Esta etapa de prueba se dividió en dos fases: (1) cuadros pasados por el archivo de prueba con los datos no expandidos a fin de comprobar las especificaciones, la programación, y (2) cuadros pasados por el archivo de prueba con los datos expandidos, para verificar que los datos no se sobrepasaran de su máximo número de dígitos (overflow) y, para desarrollar una metodología que controlara el error de redondeo.

Al completarse las pruebas mencionadas y una vez verificados y aprobados los cuadros, se procedió a utilizar el archivo de datos reformateado para la obtención de los datos sin expandir y expandidos, tanto en cifras absolutas como relativas. Un problema significativo que surgió durante esta etapa fue que las variables de superficie e ingreso al ser expandidas se sobrepasaban (overflow) del límite de dígitos establecidos por el sistema CENTS 4. Para corregir este problema se desarrollaron procedimientos generales para la programación de los cuadros que contenían dichas variables. Asimismo, otro problema significativo fue las modificaciones a las especificaciones de los cuadros, tanto en formato como en lógica, durante la programación. Todo esto tuvo como resultado un mayor tiempo de programación y prueba de los cuadros al originalmente establecido.

Cálculo de Variancias

Las estimaciones de variancia se calcularon utilizando el paquete SUPER CARP (Cluster Analysis and Regression Program), un sistema diseñado por la Universidad de Iowa State para el análisis de datos de encuestas. Los análisis realizados fueron estimaciones de razón, estimaciones de totales, medias de subpoblación, totales de subpoblación y razones de subpoblación. Previo al proceso de SUPER CARP se generó un archivo recodificado para ser utilizado por SUPER CARP, basado en las especificaciones desarrolladas para el efecto (ver Documento No. 3, Capítulo X).

CAPITULO XI

EVALUACION Y EXACTITUD DE LOS DATOS

Tipos de Errores en una Encuesta por Muestreo

Los usuarios de los datos de una encuesta por muestreo como la ENAHR, deben tener en cuenta que los resultados contienen errores de muestreo provenientes de basar las estimaciones en una muestra probabilística de la población bajo estudio y errores ajenos al muestreo que se originaron por diferentes circunstancias en las varias etapas de ejecución de la encuesta.

En el caso de la ENAHR, el efecto de errores ajenos al muestreo en la exactitud de los datos puede ser más serio que el de los errores de muestreo para algunas características, especialmente en los dominios con segmentos seleccionados que no se pudieron trabajar a causa de problemas sociopolíticos.

Origen de los Errores de Muestreo

El error de muestreo de una estimación de la encuesta es una medida de la variabilidad entre las estimaciones de todas las muestras posibles (usando el mismo diseño y tamaño de muestra), y es por lo tanto una medida de la precisión en que una estimación de esta muestra específica se aproxima al promedio de todas las muestras posibles. El error de muestreo es estimado por el error estándar (la raíz cuadrada de la variancia) que sirve para medir parcialmente el efecto de ciertos errores ajenos al muestreo, como la variabilidad entre encuestadores y los errores introducidos en la codificación, crítica y tabulación de los datos. Sin embargo, no mide sesgos sistemáticos en los datos, y por lo tanto se debe trabajar con cautela en el análisis de ciertas estimaciones que son más sujetos a sesgos.

Estos errores de muestreo no sólo dependen del tamaño de la muestra en cuanto al número de unidades de viviendas seleccionadas, sino también del número de conglomerados de donde se seleccionaron dichas unidades. Asimismo dichos errores de muestreo se ven afectados por los variables que se hayan utilizado en la estratificación.

Tal como se explicó en el Capítulo III, la población bajo estudio se dividió en 14 estratos según las regiones y zonas geográficas, y dentro de cada estrato se seleccionaron 30 conglomerados. Para cada uno de estos se escogieron en promedio 10 viviendas con productor y 10 viviendas sin productor, a excepción de los estratos urbanos donde se seleccionaron únicamente viviendas con productor, dando un total aproximado de 300 viviendas por subestrato señalado. Con este tamaño de muestra se cumplía con el margen de error especificado por los usuarios de 20% para cada estrato. Este error de muestreo está determinado básicamente por la variancia y por el nivel de confianza con el cual se desea trabajar en el análisis. La variancia es una medida estadística que permite determinar el grado de dispersión de la variable o la heterogeneidad u homogeneidad de los elementos en función de una característica dada. El nivel de confianza que va a ser usado en el análisis de las hipótesis será determinado por el usuario de la información, basado en las pruebas y niveles de significación estadístico requeridos.

Estimaciones del Error Estándar y Cómo Deben ser Usados

Fue escogido un grupo representativo de estimaciones importantes de cada sección del Plan de Tabulaciones de la encuesta para calcular los errores estándar correspondientes usando el paquete SUPER CARP. Estas estimaciones incluyen totales, razones y proporciones a diferentes niveles de desagregación, los mismos que son presentados en el Documento No. 5 "Resultados Definitivos" con las variancias, errores estándar y coeficientes de variación (c.v.) correspondientes. El c.v. o error estándar relativo es definido como el error estándar dividido por el valor esperado del estimado. Examinando los c.v. de las diferentes estimaciones, o a diferentes niveles de desagregación, se puede comparar las precisiones relativas de estas estimaciones.

Una manera sencilla de interpretar el error estándar es en término de intervalos de confianza. Este se define como un intervalo estimado que incluye el promedio de las estimaciones de todas las muestras posibles con una asignada probabilidad de confianza. Asumiendo una distribución normal, el intervalo de un error estándar por debajo del valor estimado y un error estándar por encima del mismo valor estimado tiene una probabilidad del 68% de incluir el promedio de las estimaciones de todas las muestras posibles. De manera similar se puede obtener los intervalos de confianza de 90%, 95% y 99% para cualquier estimado de la encuesta R (total, media, razón, proporción) de la siguiente manera:

$$R \pm t \text{ s.e. (R)}$$

donde:

- t = 1.64 para un intervalo de confianza de 90%
- t = 1.96 para un intervalo de confianza de 95%
- t = 2.58 para un intervalo de confianza de 99%

s.e. (R) = error estándar de la estimación R

Fuentes de Error Ajeno al Muestreo

1. Error de cobertura

El marco muestral usado para ENAHR incluye cada área de empadronamiento del censo (urbano y rural) en los 24 dominios geográficos definidos para la encuesta. Dado que estas áreas de empadronamiento fueron definidas de tal manera que cubrían el área rural completa del país y también el área de los centros poblados con una población de por lo menos 2,000 habitantes y con un mínimo de 20% de viviendas con productor agropecuario, cada vivienda ubicada dentro de estas áreas, teóricamente, tenía una probabilidad mayor que cero de ser incluida en la muestra. Sin embargo, es posible que hayan sido omitidas en la operación del registro aquellas viviendas inaccesibles y difíciles de ubicar, resultando así en una falta de cobertura. También pueden contribuir al sesgo de cobertura los límites deficientemente definidos; especialmente este problema se presenta en las áreas rurales donde a veces los límites son definidos por líneas imaginarias. En algunos segmentos en la muestra hubo una discrepancia

entre el listado del censo y el registro de viviendas de la encuesta; a veces eso fue debido a deficiencias en la cobertura del censo o cambios durante el período entre el censo y el registro de la encuesta. Comparando los resultados del registro de viviendas de ENAHIR con las cifras del censo aparentemente el registro en general tiene una tasa de cobertura más alta que el censo. Los errores de cobertura pueden ser positivos o negativos, pero el sesgo neto de cobertura es normalmente negativo indicando falta de cobertura. Considerando que no se encontraron mayores problemas de cobertura en la operación del registro de la encuesta, el sesgo correspondiente debe ser mínimo.

También hay un sesgo potencial en el marco muestral urbano dado que la definición de áreas urbanas a incluirse fue basado en la información del censo, en cuanto al porcentaje de viviendas con productor. Solamente los centros poblados urbanos con población de 2,000 habitantes y más, y con por lo menos el 20% de viviendas con productor agropecuario fueron incluidos en los dominios urbanos definidos para la encuesta. La definición de vivienda con productor basado en los datos del censo (viviendas con por lo menos un trabajador independiente dedicado a la agricultura como actividad principal) no corresponde exactamente a la definición de la encuesta (viviendas con tierras cuya superficie era de un mínimo de 0.10 hectáreas o con un número mínimo de animales). Comparando el número de viviendas con productor en segmentos seleccionados obtenido del censo y del registro de la encuesta, se encontró que las cifras de la encuesta eran generalmente más altas, indicando que la definición del censo es más restringida. En este caso, pueden haber algunos centros poblados urbanos excluidos del marco muestral porque no cumplen con los requisitos de la definición basados en los datos del censo, pero que podían tener por lo menos el 20% de viviendas con productor agropecuario basado en la definición de la encuesta. Sin embargo, dado que el límite de 20% fue más o menos arbitrario, éste puede ser considerado un problema conceptual en la interpretación de los resultados de los dominios urbanos.

Probablemente el sesgo más serio relacionado a problemas de cobertura es el que se origina como consecuencia de no haber podido registrar ni encuestar segmentos enteros debido a problemas sociopolíticos y de seguridad. Este punto es tratado seguidamente bajo el rubro "no-entrevistas". Para dichos segmentos omitidos, un ajuste de no-entrevista sencillo fue efectuado, pero el sesgo no cuantificado correspondiente puede ser serio.

2. No-entrevistas

La encuesta fue afectada por dos tipos de no-entrevistas: hogares no entrevistados como consecuencia de que el informante rehusó la entrevista o que nunca se le pudo entrevistar por ausencia, enfermedad, viaje, etc.; y no-entrevista de segmentos completos que no pudieron ser trabajados a causa de problemas sociopolíticos y de narcotráfico. Aunque un ajuste de no-entrevista fue hecho a los factores de expansión para cada clase de no-entrevista, este ajuste fue basado en el supuesto de que los hogares no-entrevistados eran similares a los entrevistados. En el caso de segmentos seleccionados que no pudieron ser trabajados íntegramente, el

sesgo potencial de no-entrevistas es más serio, ya que los problemas sociopolíticos que impidieron su accesibilidad pueden resultar también en otras diferencias significativas entre los hogares no-entrevistados y los entrevistados en otros segmentos. Debido a que el sesgo puede ser fuerte en los dominios afectados, los resultados correspondientes deben ser usados con cautela.

3. Error de Respuesta

El error de respuesta se refiere a la variabilidad de respuestas obtenidas a cierta pregunta si se parte de la premisa que los mismos hogares fuesen entrevistados nuevamente e independientemente de la encuesta y si se utiliza el mismo período de referencia. La diferencia entre el valor esperado de las respuestas de todas las repeticiones posibles de la encuesta y el valor verdadero para esa característica es el sesgo de respuesta. En la ENAHR una de las mayores fuentes de error de respuesta es el error de memoria, debido a que se les pidió a los productores que recordaran datos cuantitativos como ingresos recibidos y gastos agropecuarios incurridos durante el Año Agrícola 83-84. La mayoría de los agricultores, sin embargo, no llevan una contabilidad de los ingresos y los gastos de la explotación agropecuaria.

Los errores de respuesta también pueden ser causados por fallas ya sean en la forma de formular la pregunta, o en su interpretación por el informante dando motivo a una respuesta incorrecta. Asimismo puede suceder que la pregunta impresa en la cédula no ha sido correctamente redactada. Para cierto tipo de información, como ingreso, algunos informantes pueden dar deliberadamente una información falsa, no importando cuantas veces se les asegure de que los datos serán para fines estadísticos solamente. También puede haber deshonestidad por parte de los encuestadores cuando llenan las cédulas con datos ficticios.

En el caso de ENAHR se trató de minimizar esos errores con un buen programa de motivación y un adecuado curso de capacitación para el personal de campo. Para reducir el sesgo de respuestas omitidas o inválidas, se usó imputación dinámica o por promedio. Sin embargo, queda un sesgo no cuantificado, los mismos que pueden ser medidos por estudios especiales como reentrevistas o diseño de muestras interpenetrantes que permiten determinar el efecto del encuestador en los resultados. Tales estudios no fueron posibles efectuarlos debido a los altos costos, pero se hizo una evaluación general de los errores ajenos al muestreo.

4. Errores Introducidos en la Codificación, Entrada y Procesamiento de Datos

Otra fuente de error es la que se origina una vez que las cédulas llegan a la oficina, es decir, durante las operaciones de codificación, crítica y entrada de datos. Para reducir este tipo de error, se incluyeron algunos procedimientos de control de calidad, tal como verificación del 100% de la entrada de datos y la revisión de la crítica y codificación.

Evaluación de los Errores Ajenos al Muestreo

Aunque ciertos tipos de errores ajenos al muestreo no pueden ser detectados ni medidos sin un costoso programa de reentrevista o validación, es posible obtener una indicación general de la calidad de los datos examinando los resultados de la crítica. Los procedimientos para la crítica e imputación para la encuesta consisten en cambiar el valor en un registro cuando se encuentra inconsistente o fuera de rango. Una alta tasa de cambios para una celda particular indicaría que la calidad de los datos para esa celda es pobre y tales datos deben ser usados con cautela.

Básicamente habría tres maneras en que el valor para una cédula podría ser cambiado por encontrarse invalidado: una manera fue por los críticos de oficina llenando una "Hoja de Modificaciones" y especificando el valor válido en el caso de poderse determinar dicho valor. La segunda manera fue una imputación automática durante la crítica de consistencia ejecutado principalmente para preguntas de filtro con respuesta "sí" o "no". Una tercera forma fue una imputación dinámica o por promedio a cédulas para las cuales no se podía determinar el valor válido. Dichas variables fueron identificadas con el signo de "\$" durante la fase de crítica de rango y consistencia.

Análisis de Errores ajenos al Muestro

Para el propósito del análisis de errores, la tasa de error para una celda de la cédula fue definida como el número total de errores para esa celda dividido entre el número de casos donde la celda tuvo un valor mayor de cero; es decir, el número de observaciones (base). El número de errores para cada celda fue obtenido del número total de cambios efectuados manualmente durante la crítica de consistencia y documentados en las hojas de modificaciones, más el número total de cambios efectuados automáticamente por el programa CONCOR. Hubieron casos donde algunas celdas fueron cambiadas más de una vez durante la crítica de consistencia, lo cual causó que se contaran más de una vez esos errores y aumentara la tasa de error. Este procedimiento de estimación principalmente fue utilizado en el análisis de errores dadas las características de la crítica de consistencia por computadora. Considerando que los cambios múltiples a una celda, por lo general, aseveran menos confianza en la calidad de los datos correspondientes, parece razonable, sin embargo, incluir dichos cambios en la tasa de error. Por otra parte, es también interesante examinar la tasa de error que se obtiene de contar una sola vez cada celda errada sin importar el número de veces que haya sido modificada. En este caso las tasas de error pueden calcularse comparando los datos previos a la etapa de depuración con los datos depurados para determinar el número de celdas que fueron cambiadas.

La tasa de error total para las celdas de la cédula fue de 6.1 por ciento, lo cual es bastante razonable dado la naturaleza de la ENAHR. Para la Encuesta de Costos de Producción llevada a cabo en 1976 en la República Dominicana y la cual se asemeja bastante a la ENAHR, la tasa de error total fue de un 6.02 por ciento. Analizando las tasas de error en ambas encuestas se puede deducir que la ENAHR probablemente es la que tiene la tasa de error menor, ya que en la encuesta de la República Dominicana los errores para cada celda solamente se contaron una vez, sin importar el número de veces que la celda se había

modificado, mientras que para la ENAHR se contaron todos los cambios efectuados en una misma celda. Por lo tanto, la calidad de los datos, en general, aparenta ser bastante buena aunque la tasa de error varía considerablemente de acuerdo a cada celda, lo que indica que se debe tener cuidado cuando se utilizan aquellas celdas cuyas tasas de error son superiores a un 20%. El Cuadro No. 1 demuestra la distribución de frecuencias de las 404 celdas distintivas de la cédula según las tasas de error. Se puede observar en este cuadro que más del 50 por ciento de las celdas de la cédula tienen una tasa de error de menos del 5 por ciento, y solamente 17.1 por ciento tienen una tasa de error de 20 por ciento o más.

Cuadro No. 1 DISTRIBUCION DE FRECUENCIAS DE LAS CELDAS DE LA CEDULA SEGUN TASA DE ERROR

Número de Celdas	Tasas de Error (Porcentaje)								
	Total	0-1.9	2.0-4.9	5.0-9.0	10.0-14.9	15.0-19.9	20.0-29.9	30.0-49.9	50.0+
Número de Celdas	404	112	96	85	26	16	17	9	43
Porcentaje de Celdas	100.0	27.7	23.8	21.0	6.4	4.0	4.2	2.2	10.6

El Cuadro No. 2 presenta una lista de las variables que resultaron ser problemáticas, o sea, las 69 celdas de las cédulas tipo 1 y 2 que tuvieron una tasa de error de por lo menos un 20 por ciento. El Cuadro también contiene para cada celda identificada el número de correcciones automáticas al igual que las efectuadas manualmente, el número total de errores, el número de observaciones (base) y la tasa de error para cada una.

Cuadro No. 2 NUMERO DE CORRECCIONES AUTOMATICAS Y MANUALES, NUMERO TOTAL DE ERRORES, NUMERO DE OBSERVACIONES (BASE) Y TASA DE ERROR SEGUN EL NUMERO DE CADA CELDA

Número de Celda o LIR	Número de Correcciones Automáticas	Número de Correcciones Manuales	Número Total de Errores	Base	Tasa de Error (%)
0070 & 3912	-	4	4	15	26.7
0071 & 3913	-	4	4	15	26.7
0072 & 3914	-	4	4	15	26.7
0073 & 3915	-	6	6	15	40.0
0074 & 3916	-	4	4	15	26.7
0075 & 3917	-	4	4	2	100.0
0076 & 3918	-	4	4	2	100.0
0077 & 3919	-	4	4	2	100.0
0078 & 3920	-	4	4	2	100.0
0079 & 3921	-	7	7	2	100.0
0090 & 3932	-	3	3	13	23.1
0092 & 3934	-	4	4	10	40.0
0093 & 3935	-	2	2	10	20.0

Número de Celda o LIR	Número de Correcciones Automáticas	Número de Correcciones Manuales	Número Total de Errores	Base	Tasa de Error (%)
0094 & 3936	-	2	2	4	50.0
0095 & 3937	-	2	2	4	50.0
0096 & 3938	-	2	2	4	50.0
0097 & 3939	-	1	1	1	100.0
0098 & 3940	-	1	1	1	100.0
0099 & 3941	-	1	1	1	100.0
0100 & 3942	-	1	1	1	100.0
0101 & 3943	-	1	1	1	100.0
0104	-	85	85	72	100.0
A+7	-	171	171	140	100.0
A+8	-	98	98	183	53.6
A+25A	290	21	311	398	78.1
A+26A	-	80	80	146	54.8
A+27A	-	40	40	172	23.3
A+32A	-	7	7	34	20.6
A+34A	-	3	3	1	100.0
A+35A	-	7	7	1	100.0
A+35B	-	24	24	28	85.7
B+2	-	2,849	2,849	8,463	33.7
B+5	332	2,156	2,488	1,706	100.0
B+6	18	2,115	2,133	8,175	26.1
B+8	-	1,563	1,563	2,891	54.1
D	1,241	331	1,572	2,192	71.7
E+1	54	258	312	187	100.0
E+3	-	35	35	156	22.4
E+9	-	103	103	180	57.2
F	1,687	538	2,225	1,362	100.0
1938	175	96	271	184	100.0
1947	83	98	181	279	64.9
1951	78	95	173	167	100.0
2005	-	66	66	197	33.5
2006	-	55	55	197	27.9
2007	-	91	91	197	46.2
G+2	-	61	61	227	26.9
2202	35	37	72	282	25.5
2203	25	28	53	66	80.3
2212	-	2	2	9	22.2
2220	27	23	50	142	35.2
2221	24	41	65	37	100.0
2223	10	19	29	129	22.5
2224	-	2	2	2	100.0
2225	-	2	2	1	100.0
2226	-	1	1	4	25.0
2227	-	3	3	1	100.0
2228	-	5	5	1	100.0
2282	187	11	198	636	31.1
2283	1,994	367	2,361	4,146	56.9
2284	-	633	633	1,271	49.8

Número de Celda o LIR	Número de Correcciones Automáticas	Número de Correcciones Manuales	Número Total de Errores	Base	Tasa de Error (%)
2285	-	668	668	791	84.5
2292	-	4	4	12	33.3
2295	-	275	275	135	100.0
2296	-	17	17	6	100.0
2297	-	43	43	12	100.0
2298	-	59	59	23	100.0
J+26	-	15	15	64	23.4
J+28	-	413	413	520	79.4

Los resultados de la depuración de los datos fueron analizados para así poder comprender la naturaleza de los problemas relacionados con estas variables. A continuación se presentan las observaciones correspondientes a las celdas identificadas en el Cuadro No. 2.

1. La mayoría de las celdas identificadas tienen una base relativamente baja (en algunos casos hasta 1 ó 2 observaciones). Este es el caso de las celdas 0058 a 0101 correspondientes a la compra y venta de tierra. En algunos casos, el número de errores es mayor que su base debido a las modificaciones múltiples en estas celdas; en estos casos se presenta la tasa de error como de un 100 por ciento. Dado el número tan bajo de observaciones para compra y venta de tierra y sus altas tasas de error, no se pueden utilizar, inclusive a nivel nacional, los estimados derivados de estas variables.
2. La celda 0104 está relacionada con aquellos hogares con productor agropecuario que efectuaron todas las actividades correspondientes a la explotación en sociedad con otros (en total se presentaron 72 casos). La pregunta en sí investiga el porcentaje de las ganancias o pérdidas derivadas de la explotación que le correspondió al hogar. Dado a que se efectuaron más de 85 cambios para esta celda, se puede deducir que los productores no comprendieron bien la pregunta y, al mismo tiempo, los encuestadores no indagaron lo suficiente para poder obtener una respuesta adecuada. Por lo tanto, la información obtenida para esta celda es de carácter dudoso, sin embargo, dado el número tan bajo de observaciones, no va a tener un impacto significativo sobre la calidad de los datos de ingresos. Las celdas correspondientes a las LIR (A+7) y (A+8) también estaban relacionadas a la explotación de la tierra en sociedad y tienen problemas similares a los de la celda 0104.
3. Las LIR (A+25A), (A+26A), (A+27A) y la (A+34A) están relacionadas a la venta de cultivos a crédito. La calidad de la información obtenida para estas LIR parece que no es buena porque hubo mal entendimiento de las preguntas por parte de los informantes y porque los encuestadores no indagaron lo necesario a fin de obtener respuestas precisas. También puede ser que el diseño de esta sección de la cédula no fuese el más

adecuado para obtener este tipo de información debido a que el espacio designado para anotar la respuesta fue dividido en dos, acomodando de esta manera si la venta del cultivo fue a crédito o al contado. Por lo tanto, la información obtenida se debe utilizar con mucha cautela. El número de ventas a crédito, sin embargo, fue relativamente bajo (aproximadamente 7 por ciento del total de todas las ventas de cultivos), lo que implica que el impacto de estas variables sobre la calidad de los datos correspondientes a la venta de cultivos es mínimo. En el caso de las LIR (A+35A) y (A+35B) correspondientes a las ventas de cultivos a "otros", hubo solamente una observación, por lo que algunos de los errores probablemente se produjeron a causa de la reclasificación de ciertas ventas de cultivos bajo una de las categorías indicadas.

4. En el caso de la LIR (B+2) correspondiente al mes de siembra, la fuente de error posiblemente se debió a problemas de "recordación" por parte de los productores debido a que la siembra podía haber ocurrido hasta 18 meses antes del día de la entrevista. Dada la gran cantidad de imputaciones para esta LIR, la información de ésta se debe utilizar con cuidado.
5. La LIR (B+5) y la LIR (B+6) corresponden al tipo de semilla utilizada. El número tan alto de cambios efectuados para la LIR (B+5) (No utilizó semilla) se debió a que los informantes contestaron que no habían utilizado semillas cuando en realidad era necesario utilizarlas. Esto indica falta de sondeo o indagación por parte del encuestador.
6. En el caso de la LIR D (vacas y cabras ordeñadas), 1,241 casos fueron automáticamente cambiados de "2" (No) a blanco. De acuerdo a las instrucciones específicas en la cédula, si en la LIR C para vacas o cabras se marcaba un "2" (No), entonces no había que hacer las preguntas 75 a 84 y, por lo tanto, la LIR D debía estar en blanco. El número tan alto de imputaciones a blanco obviamente indica que los encuestadores no siguieron las instrucciones. Este error no tiene un impacto directo en la calidad de la información.
7. Hubo mucha inconsistencia entre la información de la LIR (E+1) y la LIR (E+3) correspondiente al número de semanas que se produjo queso y mantequilla en la explotación, resultando en un número alto de modificaciones, especialmente en la LIR (E+1). Aparentemente, el encuestador no indagó lo suficiente para resolver las inconsistencias. En el caso de la LIR (E+9), correspondiente a los gastos relacionados con la producción de queso y mantequilla, muchos de los informantes aparentemente no pudieron proveer esta información, resultando en un alto porcentaje de imputación. Otro problema con estos datos fue que el manual de los encuestadores indicaba que la leche producida en la explotación y destinada a queso o mantequilla para consumo del hogar se anotara en la LIR (D+11) que correspondía a la cantidad de leche diaria que en promedio se utilizaba para el consumo del hogar, excluyéndose entonces de la sección de "Productos Derivados de la Leche." Dados estos problemas y el reducido número de observaciones, se considera que los datos obtenidos no son confiables.
8. El problema con la LIR F (producción de lana o fibra) es similar al de la LIR D, y requirió que 1,687 casos fueran cambiados de "2" (No) a blanco por no haberse seguido las instrucciones de los "pases" en la cédula.

9. En el caso de la celda 1938, solamente habían 184 hogares con productor agropecuario que informaron tener gallinas las cuales no produjeron huevos. Como el número de cambios documentados para esta celda fue de 271, aparentemente hubieron cédulas donde se marcó "1" en la celda 1938 (indicando que no tuvieron producción de huevos), pero que posteriormente fueron cambiadas debido a las anotaciones en celdas relacionadas. Como ésta fue una pregunta "filtro," no afecta directamente la calidad de la información en esta sección. Los casos de las celdas 1947 y 1951 fueron similares al de la 1938.
10. Las celdas 2005, 2006 y 2007 están relacionadas al alquiler de tierras. Dado el bajo número de observaciones (197) y la tasa de error relativamente alta en cada una de estas celdas (respectivamente el 33.5 por ciento, 27.9 por ciento y 46.2 por ciento), los datos obtenidos no son muy confiables.
11. Para la LIR (G+2), correspondiente a los gastos para producir y vender productos elaborados en el hogar, hubieron relativamente pocas observaciones (227) y la tasa de error fue de 26.9 por ciento, indicando que estos datos se deben utilizar con mucha cautela. El alto número de imputaciones puede deberse a que muchos productores no llevan una contabilidad o desconocen los gastos asociados con la producción de productos elaborados.
12. La información en la Sección VII, Asistencia Técnica, celdas 2200 a 2228, también parece que es de baja calidad, debido a las pocas observaciones y la alta tasa de error en cada una de las celdas. Las celdas 2224 a 2228 (recomendaciones) definitivamente no se deben utilizar en las tabulaciones o en el análisis, debido a que hubieron 9 observaciones en total y la mayoría de estas fueron cambiadas.
13. En el caso de la celda 2282 se imputó "2" (uso de vehículo ajeno) automáticamente cuando se detectó que en la sección de gastos de insumos para cultivos se habían incurrido gastos por alquiler de vehículo para el transporte de los cultivos. Se efectuaron 187 imputaciones debido a inconsistencias, lo que indicó una deficiencia por parte de los encuestados por no haber consistenciado la información en el campo.
14. La principal razón por las altas tasas de error para las celdas 2283 a 2285 (otra maquinaria o implemento agrícola) se debió a que se imputaron estas celdas si se había indicado uso de pesticidas en la LIR (B+14). La lógica detrás de esto es que no se podían aplicar pesticidas a no ser que se utilizara una mochila (aplicador).
15. Las celdas 2287 a 2294 correspondientes a asociaciones agrícolas contaron con pocas observaciones y muchos cambios, en particular, las preguntas relacionadas a la razón principal por dejar de pertenecer a la asociación (celdas 2295 a 2298). Consecuentemente, mucha cautela se debe tener al utilizar esta información, inclusive aquella que corresponde al nivel de la república.
16. LIR (J+26) correspondiente a ingresos por concepto de indemnizaciones, seguro de accidentes o vejez, herencias, etc., solamente tiene 64

observaciones con un total de 15 cambios, de manera que cualquier inferencia de ésta es muy limitada. Sin embargo, el impacto de ésta sobre el ingreso total del hogar es mínimo. En el caso de la LIR (J+28), valorización de la especie recibida, el problema es más serio ya que el número de observaciones es mayor (520) y la tasa de error es alta (79.4 por ciento). Obviamente, fue difícil para los encuestadores obtener buena información para esta variable ya que en muchos casos los informantes no podían valorizar las especies recibidas. Esto puede resultar en un sesgo significativo para los hogares "marginales" que dependen del intercambio de productos o mercancía para su sostén. Sin embargo, dado el valor relativamente bajo de ingreso en especie, el ingreso promedio de los hogares rurales no va a ser seriamente afectado.

La conclusión derivada de este análisis de errores fue que la calidad de los datos en general es bastante buena, con una tasa de error de solamente 6.1 por ciento. Los datos para celdas individuales con una alta tasa de error como las indicadas en el Cuadro No. 2, por otro lado, se deben usar con mucha precaución. El número de observaciones para muchas de estas celdas es bajo, lo cual limita las inferencias que se pueden hacer de estos datos.

CAPITULO XII

RECOMENDACIONES PARA ENCUESTAS SIMILARES EN EL FUTURO

En este capítulo se presentan las recomendaciones, producto de la experiencia tenida con la ENAHR, que son necesarias para el planeamiento y ejecución de encuestas similares que se realicen en el futuro. Es necesario enfatizar que estas recomendaciones sólo constituyen elementos a ser tenidos en cuenta al considerar las circunstancias particulares que involucra cada encuesta que se realiza.

Dirección y Administración

Para el desarrollo de encuestas en que la responsabilidad es compartida, como en el caso de la ENAHR, es necesario conformar un "comité técnico de la encuesta" con representantes de los organismos encargados de la ejecución de la encuesta que tenga como función resolver los problemas técnicos y aprobar, en última instancia, los documentos básicos. Obviamente, los representantes deben tener poder de decisión pues también el comité técnico sería encargado de aprobar los métodos y procedimientos a utilizarse en todas las actividades de la encuesta a partir de la aprobación de los documentos básicos. El comité técnico debe funcionar permanentemente brindando apoyo a la actividad en ejecución y evaluando periódicamente el cumplimiento de las metas programadas a fin de tomar las medidas correctivas convenientes. En este sentido, el comité técnico también debe tener ingerencia en los aspectos presupuestales.

Es importante también que los organismos encargados de la ejecución de la encuesta cuenten oportunamente con las asignaciones presupuestales programadas para la ejecución de las tareas, ya sea con fondos de transferencia externa (préstamos, donaciones, etc.) o de la contrapartida nacional.

Para el desarrollo armónico de las actividades de una encuesta se debe establecer una permanente coordinación entre el personal responsable del diseño muestral, el personal encargado de las tareas de campo y los encargados del procesamiento automático de los datos pues existen reglas, normas, procedimientos, limitaciones, etc. que necesariamente deben tenerse en cuenta.

Es imprescindible que los analistas principales de los datos participen integralmente en el planeamiento de la encuesta y de los objetivos, así como los conceptos de la encuesta, se definan en forma clara y precisa.

En lo que concierne a las actividades administrativas, es necesario que no sean absorbidas por los técnicos. Los problemas de carácter económico, como son las remuneraciones, y cualquier influencia sindical, deben ser encarados por un personal administrativo.

Diseño Muestral

En términos generales, el diseño de la muestra de la ENAHR en función a las limitaciones presupuestales, los requerimientos de los usuarios y la información disponible, se consideró adecuado. Sin embargo, para optimizar la relación costo/beneficio se debe tener en cuenta las siguientes sugerencias:

Primero, considerando que en el país se ejecuta periódicamente encuestas por muestreo, es recomendable elaborar un marco de muestreo "maestro" que permita seleccionar muestras a bajo costo. Dicho marco "maestro", además de las variables que tipifican las viviendas en el área urbana, debe incluir variables de caracterización del área rural similares a las que se especificaron en el documento HR-06, "Registro de Viviendas".

Segundo, es importante que se realice el control de calidad de los archivos que se generan para la elaboración de los marcos de muestra en razón de que para la ENAHR se comprobó un significativo número de centros poblados omitidos, lo cual demandó un fuerte trabajo adicional para completar dicho marco.

Tercero, es necesario que los croquis distritales a nivel rural incluyan una delimitación adecuada para los diferentes AER que los conforman.

Diseño de Cédulas

Para el caso específico de la ENAHR las cédulas elaboradas pueden señalarse como adecuadas ya que facilitaron la labor del encuestador; no obstante, es conveniente tener presente que las cédulas con un número grande de preguntas dan lugar al cansancio del informante como del encuestador, lo que genera en muchos casos la recopilación de información de baja calidad, sobre todo para aquellas variables complejas.

Operación de Campo

De acuerdo a la experiencia de la encuesta es recomendable capacitar un mayor número de encuestadores de reserva, ya que por diversas causas tuvo que reemplazarse a varios de los titulares y no se contó con el número de encuestadores suplentes. Esta situación obligó a efectuar una capacitación en forma acelerada del personal que iba a reemplazar a los titulares.

Es importante que para otras encuestas en el futuro se vele por el fiel cumplimiento de las instrucciones impartidas en los manuales correspondientes, especialmente en la ejecución de las labores de campo. La supervisión de las tareas de campo debe hacerse bajo normas previamente establecidas que respondan a un "plan de supervisión". En este sentido, los supervisores deben verificar el correcto diligenciamiento de las cédulas y documentos auxiliares garantizando que la información tenga un mínimo nivel de consistencia. La supervisión debe ser muy exhaustiva al inicio del trabajo de los encuestadores a fin de controlar y detectar los errores para tomar las medidas correctivas pertinentes.

Procesamiento de Datos

En lo referente a las actividades del procesamiento de datos es conveniente que las especificaciones de validación de los datos se definan con la debida anticipación, sobre todo aquellas que tienen mayor tiempo de duración, como son las del proceso de establecer los valores mínimos y máximos de las variables, la de consistencia y la de tabulación. Analizando el procesamiento de la ENAHR se puede resumir las siguientes recomendaciones:

- Para establecer los rangos de las variables estudiadas, es decir, los límites mínimos y máximos, primero se deben visitar todos los organismos

públicos y privados que recopilan información a través de encuestas u otro tipo de investigación a fin de obtener datos que permitan fijar dichos límites. Entre las posibles fuentes de información se encuentran el Ministerio de Agricultura, el Banco Agrario, asociaciones de agricultores, universidades, empresas privadas, etc.

- Las especificaciones para la crítica de consistencia deben ser revisadas y aprobadas por los analistas previa a su programación para así reducir las modificaciones a las especificaciones una vez programadas. Asimismo, los programas basados en las especificaciones deben ser probados exhaustivamente a fin de detectar y corregir errores de programación o de especificación. De esta manera, se minimizan las modificaciones a los programas durante la validación de la información.
- Las especificaciones de ingresos, precio promedio, ajuste por inflación, y la de los cuadros de salida deben ser elaboradas con la participación de los analistas usuarios de la información.
- Se recomienda que en el futuro se tabule para cada casillero en la cédula una distribución de frecuencia inmediatamente después que se haya concluido la actividad de la entrada de datos. El plan preliminar de tabulación entonces puede finalizarse por los analistas una vez determinado el nivel de desagregación más bajo para el cual se puede obtener resultados razonables para cada variable, basado en el número de observaciones que tenga un valor mayor que cero en el casillero correspondiente en la cédula. De esta manera, se puede evitar el prolongado proceso de ajustar los niveles de desagregación de los cuadros una vez que hayan sido tabulados.
- Es conveniente que el sistema (paquete) de programación CENTS 4 se modifique para trabajar hasta con 15 dígitos. La actual limitación de nueve dígitos del CENTS 4 hace que se pierda precisión en algunos casos; no obstante, esta limitación, en balance, es superada por otras ventajas que resultan conveniente. El CONCOR version 2.3, por otro lado, resultó útil por sus características tales como las de corrección automática, informes comprensivos de validación y documentación del sistema. Finalmente, en el análisis de variancia se utilizó el paquete SUPER CARP el cual fue un aporte significativo ya que, al igual que CONCOR y CENTS 4, permitió que esta actividad se desarrollara con más eficiencia.

APENDICE 1

DEPARTAMENTOS QUE CONFORMAN CADA DOMINIO DE ESTUDIO*

1 Costa Urbana (con productor agropecuario)

Tumbes	Ancash	Moquegua
Piura	Lima	Tacna
Lambayeque	Ica	Cajamarca
La Libertad	Arequipa	

2 Sierra Urbana (con productor agropecuario)

Piura	Lambayeque	Junín
Puno	Moquegua	Huancavelica
Arequipa	Huánuco	Amazonas
Lima	Pasco	Tacna
Ayacucho	Cajamarca	Apurímac
La Libertad	Cusco	

3 Selva Alta Urbana (con productor agropecuario)

Cajamarca	Huánuco	Ayacucho
Amazonas	Pasco	Cusco
San Martín	Junín	Puno

4 Selva Baja Urbana (con producto agropecuario)

Loreto	Ucayali	Junín
Amazonas	Huánuco	Cusco
San Martín	Pasco	Madre de Dios

5-6 Costa Norte Rural (con y sin productor agropecuario)

Tumbes	Cajamarca
Piura	La Libertad
Lambayeque	

7-8 Costa Centro Rural (con y sin productor agropecuario)

Ancash	Ica
Lima	Prov. Const. del Callao

9-10 Costa Sur Rural (con y sin productor agropecuario)

Arequipa
Moquegua
Tacna

- 11-12 Sierra Norte Rural (con y sin productor agropecuario)
- | | |
|------------|-------------|
| Piura | Amazonas |
| Lambayeque | La Libertad |
| Cajamarca | |
- 13-14 Sierra Centro Rural (con y sin productor agropecuario)
- | | |
|---------|--------------|
| Ancash | Pasco |
| Lima | Junín |
| Ica | Huancavelica |
| Huánuco | |
- 15-16 Sierra Sur Rural (con y sin productor agropecuario)
- | | |
|----------|----------|
| Arequipa | Puno |
| Apurímac | Moquegua |
| Cusco | Tacna |
- 17-18 Selva Alta Norte Rural (con y sin productor agropecuario)
- | | |
|-----------|-------------|
| Amazonas | San Martín |
| Cajamarca | La Libertad |
- 19-20 Selva Alta Centro Rural (con y sin productor agropecuario)
- | | |
|---------|----------|
| Huánuco | Junín |
| Pasco | Ayacucho |
- 21-22 Selva Alta Sur Rural (con y sin productor agropecuario)
- | |
|-------|
| Cusco |
| Puno |
- 23-24 Selva Baja Rural (con y sin productor agropecuario)
- | | | |
|---------|---------------|----------|
| Loreto | San Martín | Amazonas |
| Junín | Madre de Dios | Cusco |
| Huánuco | Ucayali | Pasco |

* Todos los departamentos que se mencionan son incluidos en cada dominio en forma parcial. Por lo general, un departamento incluye mas de una región geográfica.

Afijación de la Muestra

<u>Dominio</u>	<u>Número de Segmentos Muestreados</u>	<u>Número de Viviendas Muestreadas</u>
Costa Urbana Hogares Agrícolas	30	300
Sierra Urbana Hogares Agrícolas	30	300
Selva Alta Urbana Hogares Agrícolas	30	300
Selva Baja Urbana Hogares Agrícolas	30	300
Costa Norte Rural Hogares Agrícolas	30	300
Costa Norte Rural Hogares No Agrícolas	30	300
Costa Central Rural Hogares Agrícolas	30	300
Costa Central Rural Hogares No Agrícolas	30	300
Costa Sur Rural Hogares Agrícolas	30	300
Costa Sur Rural Hogares No Agrícolas	30	300
Sierra Norte Rural Hogares Agrícolas	30	300
Sierra Norte Rural Hogares No Agrícolas	30	300
Sierra Central Rural Hogares Agrícolas	30	300
Sierra Central Rural Hogares No Agrícolas	30	300
Sierra Sur Rural Hogares Agrícolas	30	300
Sierra Sur Rural Hogares No Agrícolas	30	300
Selva Alta Norte Rural Hogares Agrícolas	30	300
Selva Alta Norte Rural Hogares No Agrícolas	30	300
Selva Alta Central Rural Hogares Agrícolas	30	300
Selva Alta Central Rural Hogares No Agrícolas	30	300
Selva Alta Sur Rural Hogares Agrícolas	30	300
Selva Alta Sur Rural Hogares No Agrícolas	30	300
Selva Baja Rural Hogares Agrícolas	30	300
Selva Baja Rural Hogares No Agrícolas	30	300

88

NUMERO DE SEGMENTOS Y VIVIENDAS
POR DOMINIO

APENDICE 2

TOTAL**420****7,200**

APENDICE 3

FORMULAS USADAS PARA CALCULAR LA VARIANCIA (σ_h^2) Y EL COEFICIENTE DE CORRELACION (δ_h) TOMANDO COMO BASE LOS DATOS DEL CENSO DE 1981

$$\sigma_h^2 = \frac{\sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} x_{hij}^2}{N_h} - \left(\frac{\sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} x_{hij}}{N_h} \right)^2$$

$$\delta_h = \frac{\frac{1}{M_h} \sum_{i=1}^{M_h} \left(\frac{\sum_{j=1}^{N_{hi}} x_{hij}}{N_{hi}} \right) - \left(\frac{\sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} x_{hij}}{M_h N_h} \right)^2}{\left(\frac{N_h - M_h}{N_h} \right) \left[\frac{\sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} x_{hij}^2}{N_h} - \left(\frac{\sum_{i=1}^{M_h} \sum_{j=1}^{N_{hi}} x_{hij}}{N_h} \right)^2 \right]} - \frac{M_h}{N_h - M_h}$$

Donde:

- N_h = Número total de viviendas empadronadas en el Censo de 1981 con Cédula 2 (con datos de ingreso) en el estrato h.
- M_h = Número de UPM (conglomerados) en el estrato h.
- N_{hi} = Número total de viviendas empadronadas en el Censo con Cédula 2 en el UPM_i del estrato h.
- X_{hij} = Ingreso de la vivienda "j" empadronada con Cédula 2 en el UPM_i del estrato h.

APENDICE 4

RANGO DE VALORES OPTIMOS DE \bar{n}_h PARA CADA DOMINIO, DADOS
LOS ESTIMADOS DE LAS TASAS DE COSTO Y LOS COEFICIENTES DE
CORRELACION INTRACLASE (δ_h)

DOMINIO	$\frac{C_o^2}{CC_2}$	$\frac{C_1}{C_2}$	δ_h 1/	Rango de \bar{n}_h 2/
1. Costa Urbana C.P.	0.44	2.83	.11	3-6-11
2. Sierra Urbana C.P.	7.52	2.83	.21	3-5-10
3. Selva Alta Urbana C.P.	4.00	2.83	.01; (.04)	7-14-26
4. Selva Baja Urbana C.P.	25.58	2.83	.12	7-12-24
5. Costa Norte Rural C.P.	0.13	6.23	.56; (.50)	2-3-5
6. Costa Norte Rural S.P.	0.13	6.23	.06	6-11-21
7. Costa Centro Rural C.P.	0.13	6.23	.56; (.50)	2-3-5
8. Costa Centro Rural S.P.	0.13	6.23	.04	6-11-21
9. Costa Sur Rural C.P.	0.13	6.23	.12	4-7-12
10. Costa Sur Rural S.P.	0.13	6.23	.06	6-11-21
11. Sierra Norte Rural C.P.	2.78	6.23	.28	3-5-10
12. Sierra Norte Rural S.P.	2.78	6.23	.21	3-6-11
13. Sierra Centro Rural C.P.	2.78	6.23	.22	3-6-11
14. Sierra Centro Rural S.P.	2.78	6.23	.05	8-15-29
15. Sierra Sur Rural C.P.	2.78	6.23	1.01; (.50)	2-3-5
16. Sierra Sur Rural S.P.	2.78	6.23	.80; (.50)	2-3-5
17. Selva Alta Norte Rural C.P.	1.37	6.23	.32	3-6-10
18. Selva Alta Norte Rural S.P.	1.37	6.23	-.17; (.04)	7-13-26
19. Selva Alta Centro Rural C.P.	1.37	6.23	.35	3-6-10
20. Selva Alta Centro Rural S.P.	1.37	6.23	.17	4-7-13
21. Selva Alta Sur Rural C.P.	1.37	6.23	.35	3-6-10
22. Selva Alta Sur Rural S.P.	1.37	6.23	.69; (.50)	2-3-5
23. Selva Baja Rural C.P.	23.88	6.23	.94; (.50)	2-5-10
24. Selva Baja Rural S.P.	23.88	6.23	-.23; (.04)	12-23-40

1/ δ_h fue calculado basado en la fórmula que se presenta en el Apéndice 3.
Los valores entre paréntesis fueron usados para la determinación de \bar{n}_h

2/ El número central es el \bar{n}_h óptimo; el primer y último número especifica el
rango de valores de \bar{n}_h cuyo resultado en la variancia estaría dentro del 10%
del óptimo.

APENDICE 5

ESTIMACION DEL C.V. PARA EL INGRESO PROMEDIO POR VIVIENDA QUE RESULTO DEL DISEÑO PRELIMINAR DE LA MUESTRA Y NUMERO DE CONGLOMERADOS NECESARIOS PARA OBTENER UN C.V. DE 20% POR DOMINIO

DOMINIO	Ingreso (\bar{x}_h) Promedio por Vivienda (miles de S/.)	σ_h^2	$\delta_h \frac{1}{h}$	CV _x	\hat{m}_h
1. Costa Urbana C.P.	60	28.40	.11	.01	1
2. Sierra Urbana C.P.	40	933.69	.21	.07	5
3. Selva Alta Urbana C.P.	72	551.19	.31; (.04)	.02	1
4. Selva Baja Urbana C.P.	73	851.95	.12	.03	1
5. Costa Norte Rural C.P.	47	853.60	.56; (.50)	.09	6
6. Costa Norte Rural S.P.	62	490.87	.06	.03	1
7. Costa Centro Rural C.P.	66	735.82	.56; (.50)	.06	3
8. Costa Centro Rural S.P.	121	660.75	.04	.01	1
9. Costa Sur Rural C.P.	126	967.53	.12	.02	1
10. Costa Sur Rural S.P.	80	746.53	.06	.02	1
11. Sierra Norte Rural C.P.	23	229.58	.28	.07	4
12. Sierra Norte Rural S.P.	46	485.03	.21	.05	2
13. Sierra Centro Rural C.P.	25	520.26	.22	.09	7
14. Sierra Centro Rural S.P.	62	868.15	.05	.03	1
15. Sierra Sur Rural C.P.	19	690.75	1.01; (.50)	.11	26
16. Sierra Sur Rural S.P.	46	237.85	.80; (.50)	.05	2
17. Selva Alta Norte Rural C.P.	43	117.17	.32	.03	1
18. Selva Alta Norte Rural S.P.	67	858.54	-.17; (.04)	.03	1
19. Selva Alta Centro Rural C.P.	47	612.72	.35	.06	3
20. Selva Alta Centro Rural S.P.	84	804.69	.17	.03	1
21. Selva Alta Sur Rural C.P.	40	576.26	.15	.07	4
22. Selva Alta Sur Rural S.P.	70	308.93	.69; (.50)	.03	1
23. Selva Baja Rural C.P.	50	914.95	.94; (.50)	.08	6
24. Selva Baja Rural S.P.	112	855.81	-.23; (.04)	.02	1

1/ δ_h fue calculado basado en la fórmula especificada en el Apéndice 3. Los valores que están entre paréntesis fueron usados para estimar el C.V. \bar{x} y \hat{m}_h

APENDICE 6

COEFICIENTES DE LA FUNCION COSTO* Y TASAS PARA CADA GRUPO DE ESTRATOS

GRUPO	C_0	C_1	C_2	C	$\frac{C_0^2}{CC_2}$	$\frac{C_1}{C_2}$
1. Costa Rural	406.68	117.60	28.50	45,492.10	0.13	6.23
2. Costa Urbana	406.68	80.70	28.50	13,198.50	0.44	2.83
3. Sierra Rural	2,229.52	177.60	28.50	62,785.10	2.78	6.23
4. Sierra Urbana	2,229.52	80.70	28.50	23,182.60	7.52	2.83
5. Selva Alta Rural	1,474.22	177.60	28.50	55,619.70	1.37	6.23
6. Selva Alta Urbana	1,474.22	80.70	28.50	19,045.60	4.00	2.83
7. Selva Baja Rural	5,458.58	177.60	28.50	43,775.90	23.88	6.23
8. Selva Baja Urbana	5,458.58	80.70	28.50	40,868.90	25.58	2.83

* Obtenidos de la función Costo $C = C_0/m + C_1m + C_2m^2$. Una explicación de la función costo y estimación de coeficientes se presentan en el Apéndice 7.

APENDICE 7

Función Costo Para Encuesta Nacional de Hogares Rurales

Simbología

$$C = C_0 \sqrt{nA} + C_1 n + C_2 nm$$

C = Costo total excluyendo los costos generales

C_0 = Costo por kilómetro para viajar entre segmentos

C_1 = Costo de la selección y listado de las manzanas seleccionadas

C_2 = Costo por cuestionario, incluyendo el costo de la impresión, entrevista, supervisión de la entrevista, codificación y tabulación

\sqrt{nA} = Estimación de la distancia recorrida entre manzanas

A = Area del dominio

m = Número de segmentos en la muestra

n = Número promedio de viviendas en la muestra por manzana

Determinación de los Costos (en US\$)

C_0 (determinado por la distribución de los costos totales para transporte proporcional a \sqrt{nA} y un diferencial de costo/km de 1/3/5/6):

1. Costa: $C_0 = \$1.13$, $A = 129,522 \text{ km}^2$
2. Sierra: $C_0 = \$3.38$, $A = 435,100 \text{ km}^2$
3. Selva Alta: $C_0 = \$5.63$, $A = 68,566 \text{ km}^2$
4. Selva Baja: $C_0 = \$6.76$, $A = 556,739 \text{ km}^2$

Componentes de C_1 (costo por segmento)

1. Selección de segmentos \$28.00
2. Impresión de hoja de listado (10 hojas) 0.40
3. Fotocopias de croquis (2 hojas) y pre-listado (10 hojas)
 - a. Rural (incluyendo mecanografiar listado del censo) 2.40
 - b. Urbano 1.25

4. Listado de segmentos

a. Rural

- Viático --- Encuestador (5 días)	83.30
Jefe de Brigada (1 día)	24.00
Supervisor nacional (1/4 día)	7.30
Chofer (1/4 día)	4.20
- Transporte dentro del segmento (listado, encuesta)	28.00

b. Urbano

- Viático --- Encuestador (2 días)	33.30
Jefe de Brigada (1/4 día)	12.00
Supervisor nacional (1/8 día)	3.65
Chofer (1/8 día)	2.10

Componentes de C₂ (costo por entrevista)

1. Impresión de la Cédula	0.65
2. Encuesta	
- Viático --- Encuestador (1/2 día)	8.35
Jefe de Brigada (1/8 día)	3.00
Supervisor nacional (1/16 día)	1.85
Chofer (1/16 día)	1.05
3. Trabajo de Oficina	
- Crítica/codificación y supervisión	1.10
- Entrada de datos	3.35
- Procesamiento de datos	9.15

Función Costo por Grupo de Dominios:

1. Costa Rural

$$\begin{aligned} C &= 1.13 \sqrt{129,522n} + 177.60n + 28.50 \text{ nm} \\ &= 406.68\sqrt{n} + 177.60n + 28.50 \text{ nm} \end{aligned}$$

2. Costa Urbana

$$C = 406.68\sqrt{n} + 80.70n + 28.50\text{nm}$$

3. Sierra Rural

$$\begin{aligned} C &= 3.38 \sqrt{435,100n} + 177.60n + 28.50 \text{ nm} \\ &= 2229.52 \sqrt{n} + 177.60n + 28.50\text{nm} \end{aligned}$$

4. Sierra Urbana

$$C = 2229.52\sqrt{n} + 80.70n + 28.50\text{nm}$$

5. Selva Alta Rural

$$\begin{aligned} C &= 5.63 \sqrt{68,566n} + 177.60n + 28.50\text{nm} \\ &= 1474.22 \sqrt{n} + 177.60n + 28.50\text{nm} \end{aligned}$$

6. Selva Alta Urbana

$$C = 1474.22\sqrt{n} + 80.70n + 28.50\text{nm}$$

7. Selva Baja Rural

$$\begin{aligned} C &= 6.76 \sqrt{652.027n} + 177.60n + 28.50 \text{ nm} \\ &= 5458.58 \sqrt{n} \end{aligned}$$

8. Selva Baja Urbana

$$C = 5458.58 \sqrt{n} + 80.70n + 28.50\text{nm}$$

APENDICE 8

ESPECIFICACIONES DE SEGMENTOS PARA DEFINIR NUEVOS ESTRATOS REQUERIDOS EN LAS ESTIMACIONES DE VARIANCIAS

ESTRATO 01: COSTA URBANA

<u>NUEVO ESTRATO</u>	<u>SEGMENTOS</u>	<u>TIPO DE UPM</u>
001	01, 06, 07	AR (colapsado)
002	02, 03	AR
003	04, 05	AR
004	08, 09	NAR
005	10, 11	NAR
006	12, 13, 14	AR
007	15, 16	AR
008	17, 18	AR (colapsado)
009	19, 20	NAR
010	21, 24	NAR
011	22, 23	AR
012	25, 26	AR
013	27, 28	AR (colapsado)
014	29, 30	NAR

ESTRATO 02: SIERRA URBANA

015	01, 07	AR (colapsado)
016	03, 04	NAR
017	05, 06	NAR
018	08, 09	NAR
019	10, 11	NAR
020	12, 13	NAR
021	14, 15	NAR
022	16, 17	NAR
023	18, 19	NAR
024	20, 21	NAR
025	22, 23	NAR
026	24, 25	NAR
027	26, 27	NAR
028	28, 29, 30	NAR

ESTRATO 03: SELVA ALTA URBANA

029	01, 14	AR (colapsado)
030	02, 03	AR
031	04, 05	AR
032	06, 07, 08	AR
033	09, 10	AR
034	11, 12, 13	AR
035	15, 16, 17	AR
036	18, 19	AR

ESTRATO 03: SELVA ALTA URBANA

<u>NUEVO ESTRATO</u>	<u>SEGMENTOS</u>	<u>TIPO DE UPM</u>
037	20, 21	AR
038	22, 23	AR
039	24, 25, 28	AR (colapsado)
040	26, 27	AR
041	29, 30	NAR

ESTRATO 04: SELVA BAJA URBANA

042	01, 02	AR
043	03, 04, 05	AR
044	06, 07	AR
045	08, 20, 21	AR (colapsado)
046	09, 10	AR
047	11, 26, 27	NAR
048	12, 13, 14	AR
049	15, 16	AR
050	17, 18, 19	AR
051	22, 23	AR
052	24, 25, 30	AR (colapsado)
053	28, 29	AR

ESTRATO 05: COSTA NORTE RURAL

054	01, 06	NAR (CCPP)
055	02, 03	NAR (AER)
056	04, 05	NAR (AER)
057	07, 08	NAR (AER)
058	09, 10	NAR (AER)
059	11, 12	NAR (AER)
060	13, 14	NAR (CCPP)
061	15, 27	NAR (CCPP)
062	16, 17	NAR (AER)
063	18, 19	NAR (AER)
064	20, 21	NAR (AER)
065	22, 23	NAR (AER)
066	24, 25	NAR (AER)
067	26, 28	NAR (AER)
068	29, 30	NAR (AER)

ESTRATO 06: COSTA CENTRO RURAL

069	01, 18	NAR (CCPP)
070	04, 05, 06	NAR (AER)
071	07, 08	NAR (CCPP)
072	09, 10	NAR (AER)
073	11, 12	NAR (AER)

ESTRATO 06: COSTA CENTRO RURAL

<u>NUEVO ESTRATO</u>	<u>SEGMENTOS</u>	<u>TIPO DE UPM</u>
074	13, 14	NAR (AER)
075	15, 16	NAR (CCPP)
077	19, 20	NAR (AER)
078	21, 22	NAR (AER)
079	23, 24	NAR (AER)
080	25, 26	NAR (AER)
081	27, 28	NAR (AER)
082	29, 30	NAR (AER)

ESTRATO 07: COSTA SUR RURAL

083	01, 02	NAR (CCPP)
084	03, 23	NAR (CCPP)
085	05, 06	NAR (CCPP)
087	08, 09	NAR (AER)
088	10, 11	NAR (AER)
089	12, 13	NAR (AER)
090	14, 15	NAR (AER)
091	16, 17	NAR (AER)
092	18, 19	NAR (AER)
093	20, 21, 22	NAR (AER)
094	24, 25	NAR (AER)
095	26, 27, 28	NAR (AER)
096	29, 30	NAR (AER)

ESTRATO 08: SIERRA NORTE RURAL

097	01, 17, 18	NAR (CCPP)
098	02, 03	NAR (AER)
099	04, 05	NAR (AER)
100	06, 07	NAR (AER)
101	08, 09	NAR (AER)
102	10, 11	NAR (AER)
103	12, 13	NAR (AER)
104	14, 15, 16	NAR (AER)
105	19, 20	NAR (AER)
106	21, 22	NAR (AER)
107	23, 24, 25	NAR (AER)
108	26, 27, 28	NAR (AER)
109	29, 30	NAR (AER)

ESTRATO 09: SIERRA CENTRO RURAL

<u>NUEVO ESTRATO</u>	<u>SEGMENTOS</u>	<u>TIPO DE UPM</u>
110	01, 13	NAR (CCPP)
111	02, 03	NAR (AER)
112	04, 05, 06	NAR (AER)
113	07, 08	NAR (CCPP)
114	09, 10	NAR (AER)
115	11, 12	NAR (AER)
116	14, 15	NAR (AER)
117	16, 17	NAR (AER)
118	18, 19, 20	NAR (CCPP)
119	21, 22, 23	NAR (AER)
120	24, 27	NAR (CCPP)
121	25, 26	NAR (AER)
122	28, 29, 30	NAR (AER)

ESTRATO 10: SIERRA SUR RURAL

123	01, 08, 09	NAR (CCPP)
124	03, 04, 05	NAR (AER)
125	06, 07	NAR (AER)
126	10, 11	NAR (AER)
127	12, 13	NAR (AER)
128	14, 15	NAR (AER)
129	16, 19	NAR (AER)
130	17, 18	NAR (CCPP)
131	20, 21	NAR (AER)
132	22, 23	NAR (AER)
133	24, 25, 26	NAR (AER)
134	27, 28	NAR (AER)
135	29, 30	NAR (AER)

ESTRATO 11: SELVA ALTA NORTE

136	02, 11	NAR (CCPP)
137	03, 04	NAR (AER)
138	05, 06	NAR (AER)
139	07, 08	NAR (AER)
140	09, 10	NAR (AER)
141	12, 13	NAR (AER)
142	14, 15	NAR (AER)
143	16, 17	NAR (AER)
144	18, 19	NAR (AER)
145	20, 21	NAR (AER)
146	22, 23	NAR (AER)
147	24, 25	NAR (AER)
148	26, 27	NAR (AER)
149	29, 30	NAR (AER)

ESTRATO 12: SELVA ALTA CENTRO

<u>NUEVO ESTRATO</u>	<u>SEGMENTOS</u>	<u>TIPO DE UPM</u>
150	01, 10	NAR (CCPP)
151	02, 04	NAR (AER)
152	05, 06	NAR (AER)
153	07, 09	NAR (AER)
154	11, 12	NAR (AER)
155	13, 14	NAR (AER)
156	15, 16	NAR (AER)
157	17, 18	NAR (AER)
158	19, 20, 21	NAR (AER)
159	22, 23	NAR (AER)
160	24, 26	NAR (AER)

ESTRATO 13: SELVA ALTA SUR

161	01, 02	NAR (CCPP)
162	03, 22	NAR (CCPP)
163	04, 05, 06	NAR (AER)
164	08, 09	NAR (AER)
165	10, 11	NAR (AER)
166	12, 13	NAR (AER)
167	14, 15	NAR (AER)
168	16, 17	NAR (AER)
169	18, 19	NAR (AER)
170	20, 21	NAR (AER)
171	23, 24	NAR (AER)
172	25, 26	NAR (AER)
173	27, 28	NAR (AER)
174	29, 30	NAR (AER)

ESTRATO 14: SELVA BAJA RURAL

175	01, 04	NAR (CCPP)
176	02, 03	NAR (AER)
177	05, 06	NAR (AER)
178	07, 08	NAR (AER)
179	09, 10	NAR (AER)
180	11, 12, 13	NAR (AER)
181	14, 19	NAR (CCPP)
182	15, 16	NAR (AER)
183	17, 18	NAR (AER)
184	20, 21	NAR (AER)
185	22, 23	NAR (AER)
186	24, 25	NAR (AER)
187	26, 27	NAR (AER)
188	29, 30	NAR (AER)

DOC. HR-08
(19-9-84)

"AÑO DEL RESUCENTENARIO DEL BATAJO DEL ALMIRANTE GRAU"

INSTITUTO NACIONAL DE ESTADÍSTICA
DIRECCIÓN GENERAL DE CENSOS Y ENCUESTAS



MINISTERIO DE AGRICULTURA
OFICINA SECTORIAL DE ESTADÍSTICA

ENCUESTA NACIONAL DE HOGARES RURALES REGISTRO DE VIVIENDAS

Hoy ____ de ____ de 19__

A. UBICACIÓN GEOGRÁFICA

Distrito	Provincia	SOLO PARA EL ÁREA URBANA
Segmento	Distrito	
Decanato	Centro Poblado	Manzana

B. NOMBRE Y APELLIDOS DEL REGISTRADOR

C. FECHA DE

Inicio	Termino
--------	---------

N° de Orden de la Vivienda (1)	DIRECCIÓN DE LA VIVIENDA					Condición de Ocupación de la vivienda	¿Cuál es el nombre y apellidos del jefe de familia?	¿Algunas personas de esta vivienda trabajan en las tierras o crian animales por su cuenta? (8)	¿Cuánta tierra trabajan en total? (9)		¿Cuántos animales crían y de qué especies? (10)	¿Trabajan esas tierras lo crían los animales en sociedad con otro/a? (11)	¿El socio de más edad vive en esta vivienda? (12)	USO DE OFICINA		Observaciones (16)
	URBANA. Nombre de la calle, prón, avenida, pasaje, etc. RURAL. Descripción de la ubicación de la vivienda (Km., paraje, anexo, etc.) (2)	N° de la Puerta de Calle (3)	N° de Pisc. (4)	N° Interior o Letra (5)	¿Cuenta tierra trabajada desde el 11? Si es 1,000 m ² o más paise a el 12				Cantidad (13)	Unidad de Medida (14)				3 vacunos, 10 caprinos, 10 ovinos, 5 porcinos, 10 sucinos, 100 aves de corral, 100 cuyes, 100 conejos (15)	C	
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			
					1 <input type="radio"/> Ocupada 2 <input type="radio"/> Desocupada (Termino)			1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)			1 <input type="radio"/> Si - Siga 2 <input type="radio"/> No - Termine	1 <input type="radio"/> Si 2 <input type="radio"/> No (Termino)	1 <input type="radio"/> S 2 <input type="radio"/> No			

CENTROS POBLADOS (CCPP) SELECCIONADOS, NUMERO DE ENCUESTADORES Y NUMERO DE CEDULAS DILIGENCIADAS
PARA LA PRUEBA PILOTO - ENAHR

REGION	EN AREA URBANA				EN AREA RURAL					
	CON PRODUCTOR				CON PRODUCTOR				SIN PRODUCTOR	
DEPARTAMENTO	CCPP	No. de Encuestadores	No. de Cédulas a Diligenciar	CCPP	No. de Encuestadores	No. de Cédulas a Diligenciar	CCPP	No. de Encuestadores	No. de Cédulas a Diligenciar	
PROVINCIA										
DISTRITO										
<u>COSTA</u>										
Lambayeque										
CHICLAYO										
Monsefú	Monsefú	01	12	Cayanca	02	24	Cusupe	01	29	
<u>SIERRA</u>										
Junín										
Huancayo										
Chupaca	Chupaca	01	12	Pincha	02	24	Buenos Aires	01	24	
<u>SELVA</u>										
Huánuco										
Leoncio Prado										
J. Crespo y Castillo	Aucaycu	01	12	Santa Lucía	02	24	Sto. Domingo de Anda	01	24	
TOTAL		03	36		06	72		03	72	

REGION AGRARIA	SEGMENTOS			TECNICOS					PERSONAL AUXILIAR Choferes
	TOTAL	URBANOS	RURALES	TOTAL	Registra- dores	Suplentes	Jefe de Equipo	Jefe Depar- tamental	
TOTAL	420	120	300	137	89	02	23	23	44
I Tumbes	05	03	02	02	01	-	-	01	01
II Piura	26	08	18	08	05	-	02	01	02
III Lambayeque	17	10	07	03	02	-	-	01	01
IV La Libertad	18	04	14	06	04	-	01	01	02
V Ancash	14	02	12	07	04	01	01	01	02
VI Lima	22	03	19	06	04	-	01	01	02
VII Ica	08	-	08	03	02	-	-	01	01
VIII Arequipa	28	04	24	09	06	-	02	01	03
IX Moquegua	03	-	03	02	01	-	-	01	01
X Tacna	07	01	06	03	02	-	-	01	01
XI Cajamarca	36	07	29	11	07	-	03	01	03
XII Amazonas	12	-	12	05	03	-	01	01	02
XIII San Martín	54	40	14	12	08	01	02	01	04
XIV Huánuco	14	02	12	05	03	-	01	01	02
XV Pasco	08	03	05	03	02	-	-	01	01
XVI Junín	29	08	21	09	06	-	02	01	03
XVII Huancavelica	04	01	03	03	02	-	-	01	01
XVIII Ayacucho	15	03	12	06	04	-	01	01	02
XIX Apurímac	06	01	05	03	02	-	-	01	01
XX Cusco	38	06	32	11	08	-	02	01	03
XXI Puno	23	-	23	08	05	-	02	01	02
XXII Loreto	24	11	13	09	06	-	02	01	03
XXIII Ucayali	09	03	06	03	02	-	-	01	01
XXIV M. de Dios	-	-	-	-	-	-	-	-	-

APENDICE 13

DISTRIBUCION DEL PERSONAL TECNICO Y AUXILIAR POR REGIONES AGRARIAS

REGION AGRARIA		TECNICOS					Auxiliar
		TOTAL	Efectivos	Suplentes	Jefe de Brigada	Jefe Departamental	Chofer
TOTAL		147	97	4	23	23	44
I	Tumbes	2	1	-	-	1	1
II	Piura	8	5	-	2	1	2
III	Lambayeque	4	3	-	-	1	1
IV	La Libertad	6	4	-	1	1	2
V	Ancash	6	4	-	1	1	2
VI	Lima	7	5	-	1	1	2
VII	Ica	3	2	-	-	1	1
VIII	Arequipa	9	6	-	2	1	3
IX	Moquegua	2	1	-	-	1	1
X	Tacna	3	2	-	-	1	1
XI	Cajamarca	13	8	1	3	1	3
XII	Amazonas	6	4	-	1	1	2
XIII	San Martín	12	8	1	2	1	4
XIV	Huánuco	6	4	-	1	1	2
XV	Pasco	3	2	-	-	1	1
XVI	Junín	11	7	1	2	1	3
XVII	Huancavelica	3	2	-	-	1	1
XVIII	Ayacucho	6	4	-	1	1	2
XIX	Apurímac	3	2	-	-	1	1
XX	Cusco	13	9	1	2	1	3
XXI	Puno	9	6	-	2	1	2
XXII	Loreto	9	6	-	2	1	3
XXIII	Ucayali	3	2	-	-	1	1
XXIV	Madre de Dios	-	-	-	-	-	-

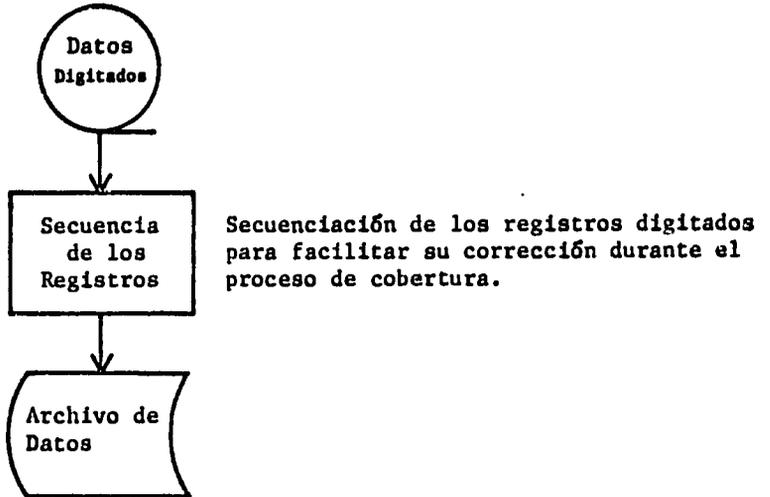
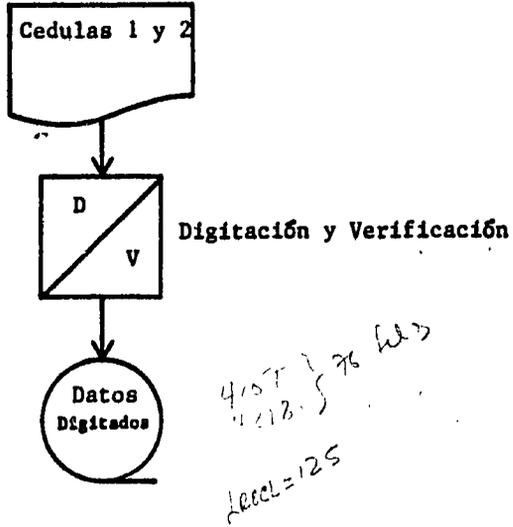
Escrato	Número de Viviendas Seleccionadas	Cédulas Diligenciadas Integramente	Cédulas Diligenciadas Parcialmente	Número de Viviendas Sin Encuestar	Porcentaje de Viviendas sin Encuestar	
01	Costa Urbana	300	275	-	25	8.3
02	Sierra Urbana	319	267	9	46	14.4
03	Selva Alta Urbana	300	241	2	57	19.0
04	Selva Baja Urbana	325	268	3	56	17.2
05	Costa Norte Rural	610	579	12	25	4.1
06	Costa Centro Rural	607	542	29	49	8.1
07	Costa Sur Rural	621	584	1	42	6.8
08	Sierra Norte Rural	480	423	3	54	11.2
09	Sierra Centro Rural	566	505	8	55	9.7
10	Sierra Sur Rural	537	451	8	82	15.2
11	Selva Alta Norte	608	522	1	86	14.1
12	Selva Alta Centro	683	455	1	234	34.3
13	Selva Alta Sur	514	417	5	92	17.9
14	Selva Baja Rural	526	456	2	76	14.4
Total		6,996	5,985	84	979	14.0

PORCENTAJE DE VIVIENDAS SIN ENCUESTAR

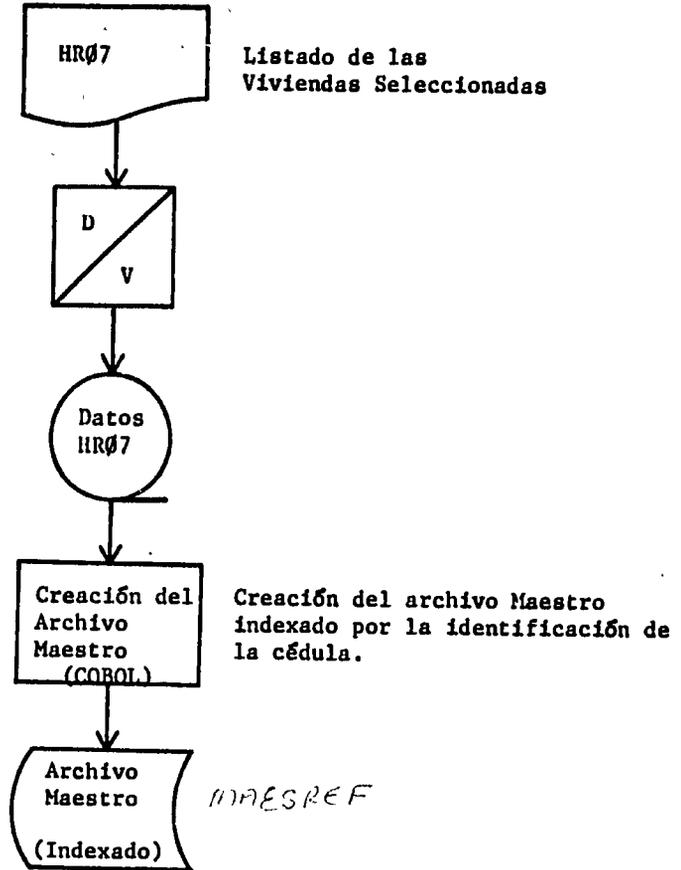
APENDICE 15

SISTEMA DE PROCESAMIENTO DE DATOS

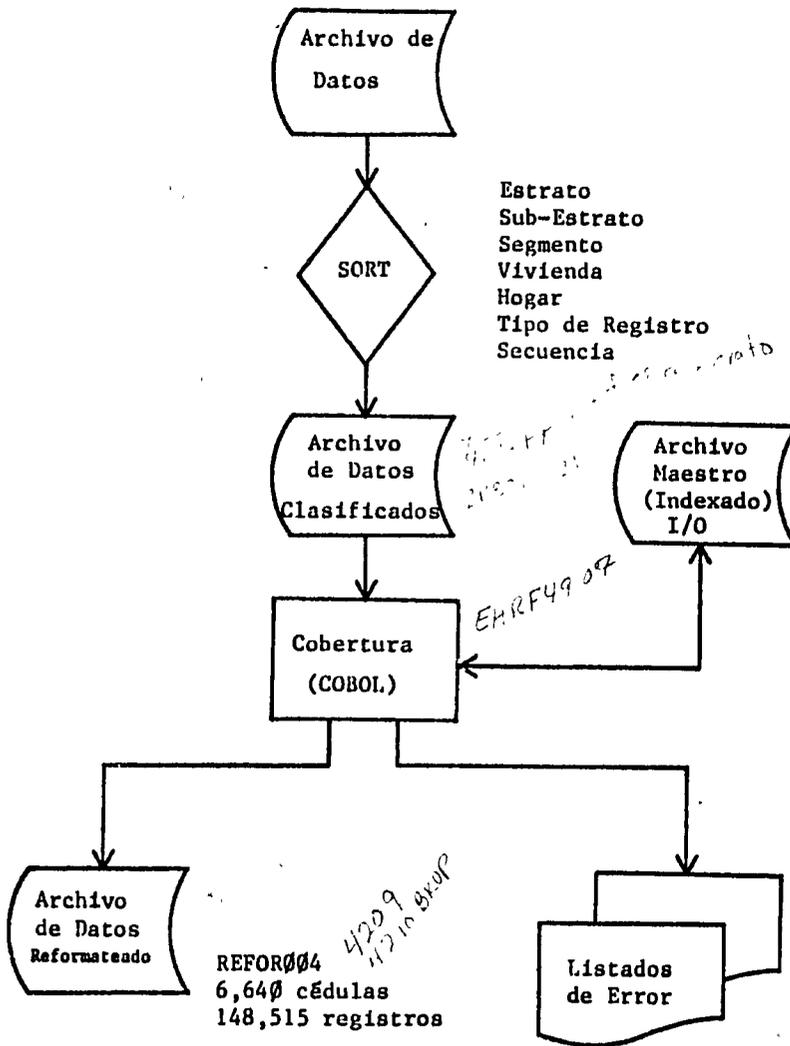
Proceso de Entrada de Datos



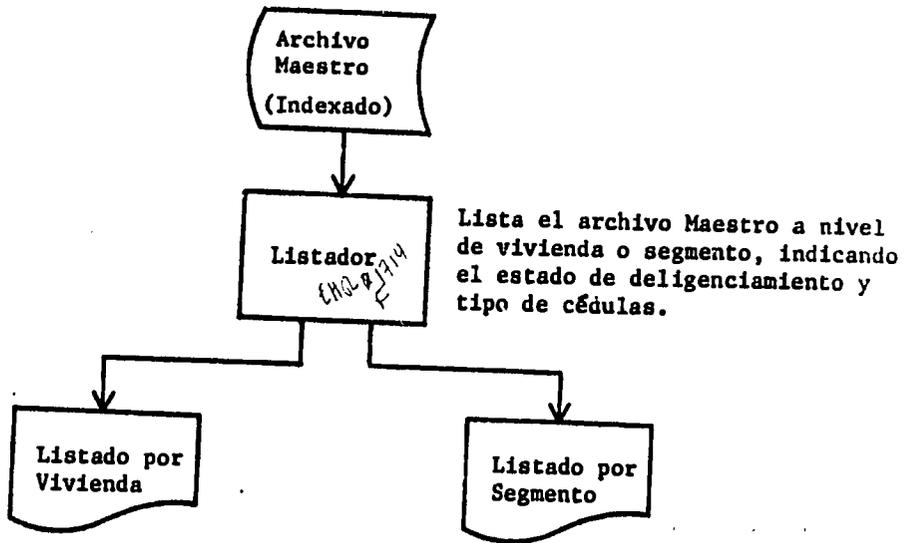
Generación del Archivo Maestro



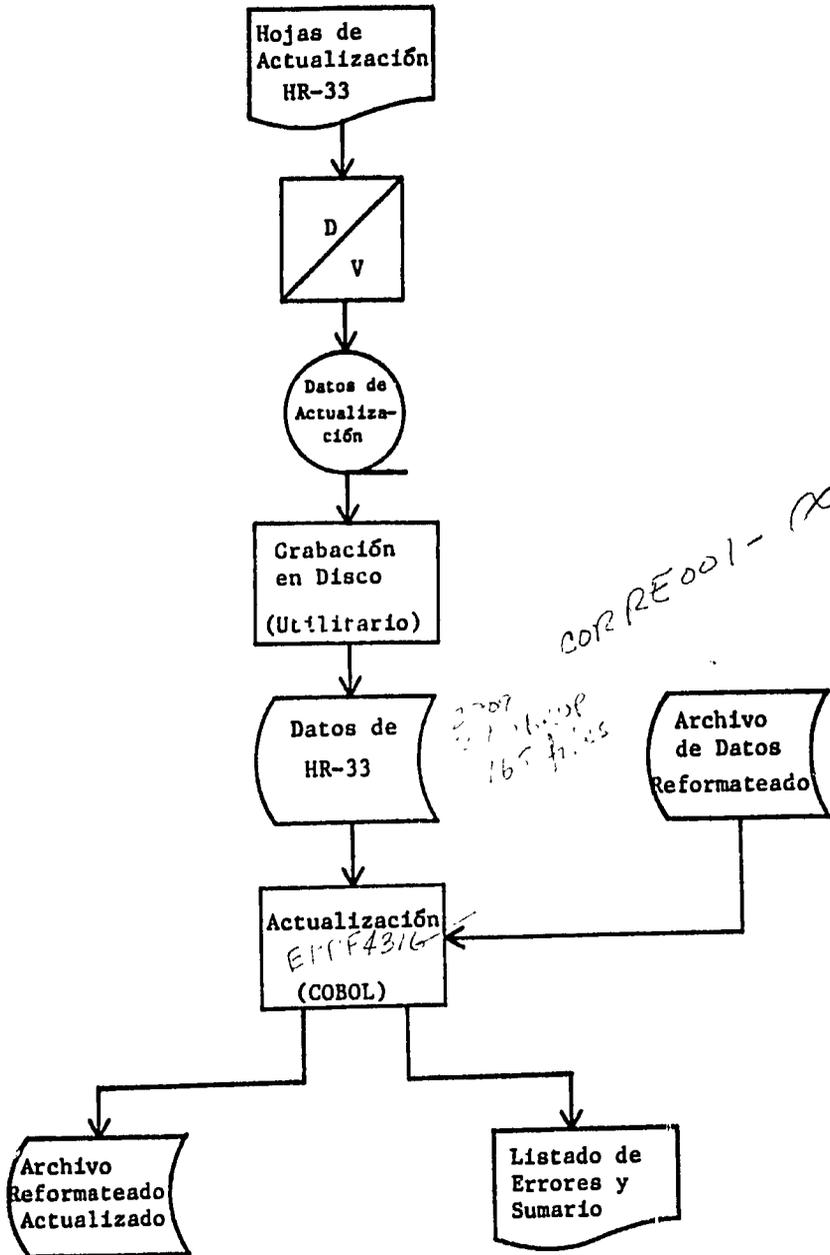
Proceso de Chequeo de Cobertura y Generación de Archivo de Datos



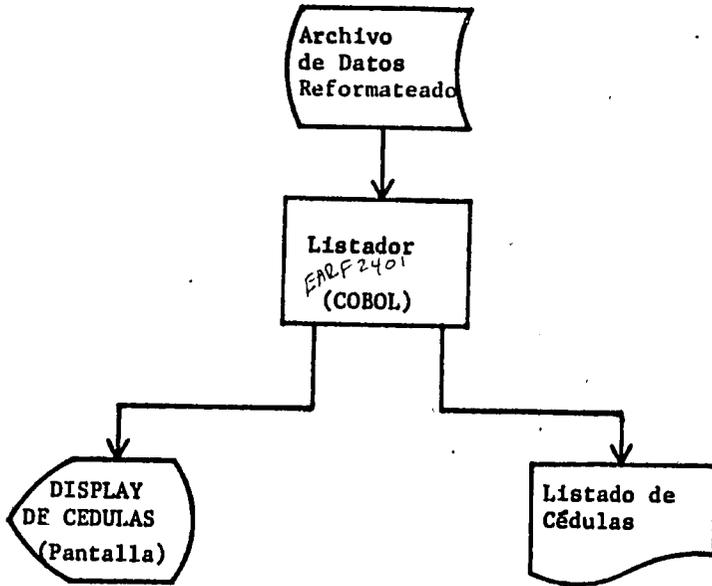
Proceso de Listar Información del Maestro



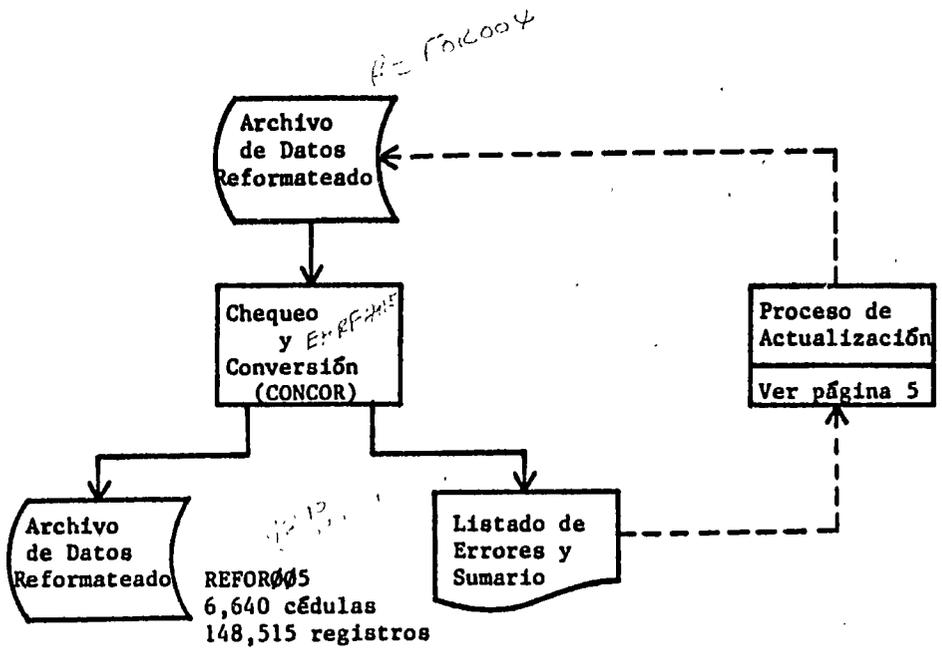
Proceso de Actualización



Proceso de Listado de las Cédulas

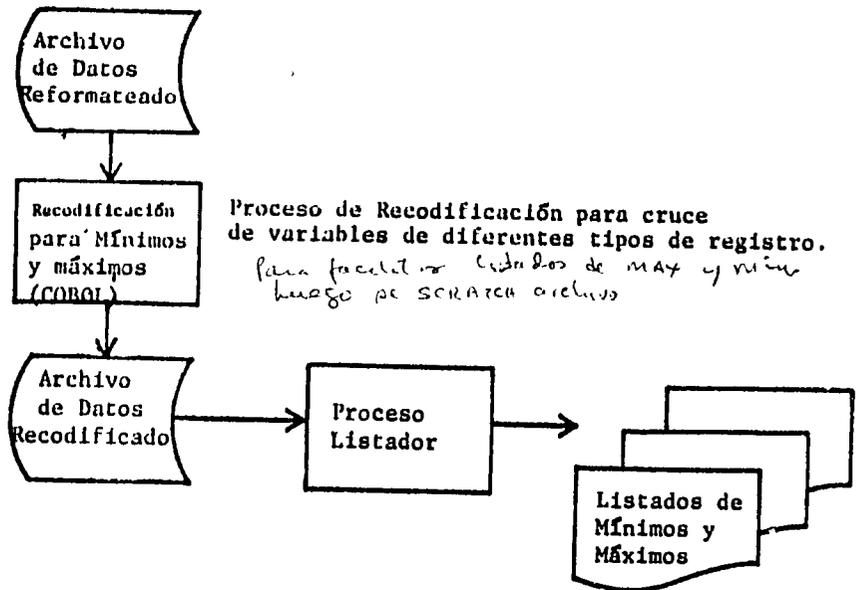
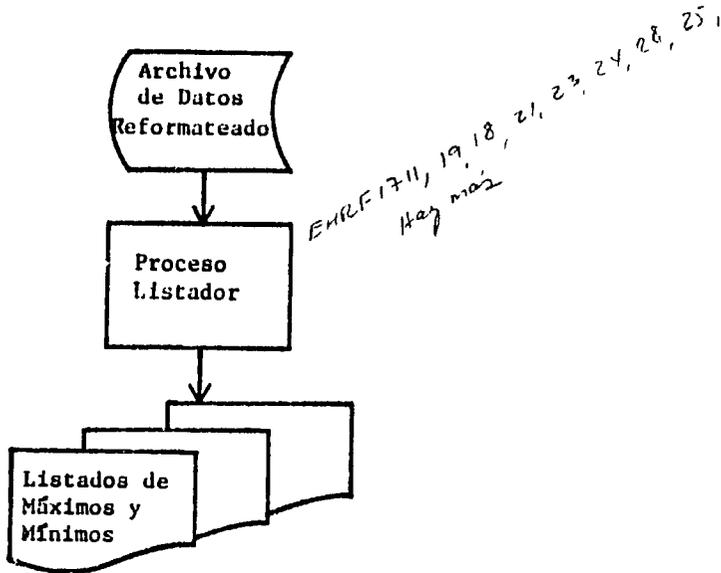


Proceso de Chequeo y Conversión de las Unidades de Medidas

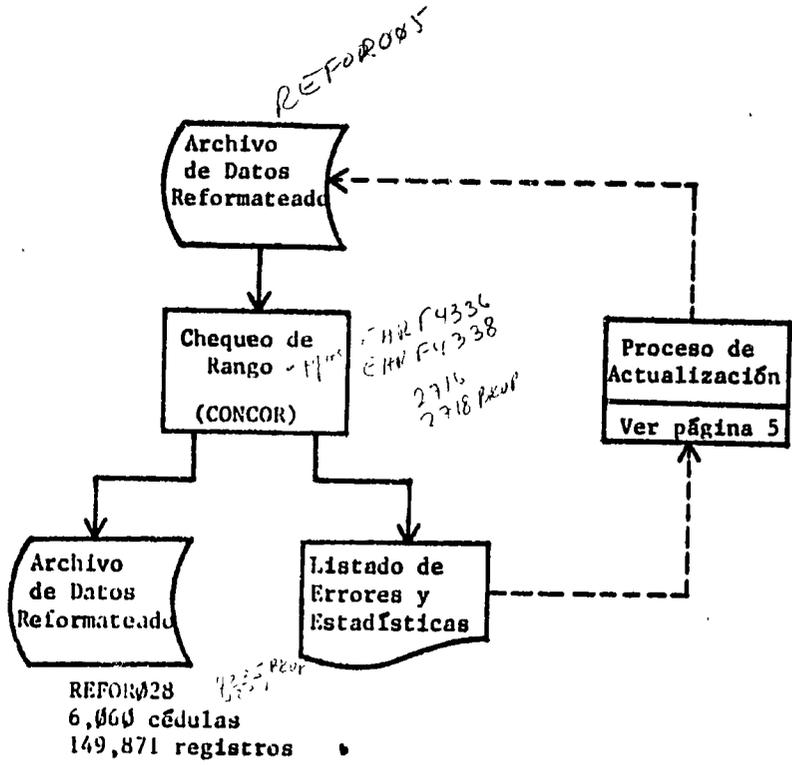


Proceso de Chequeo de Rango

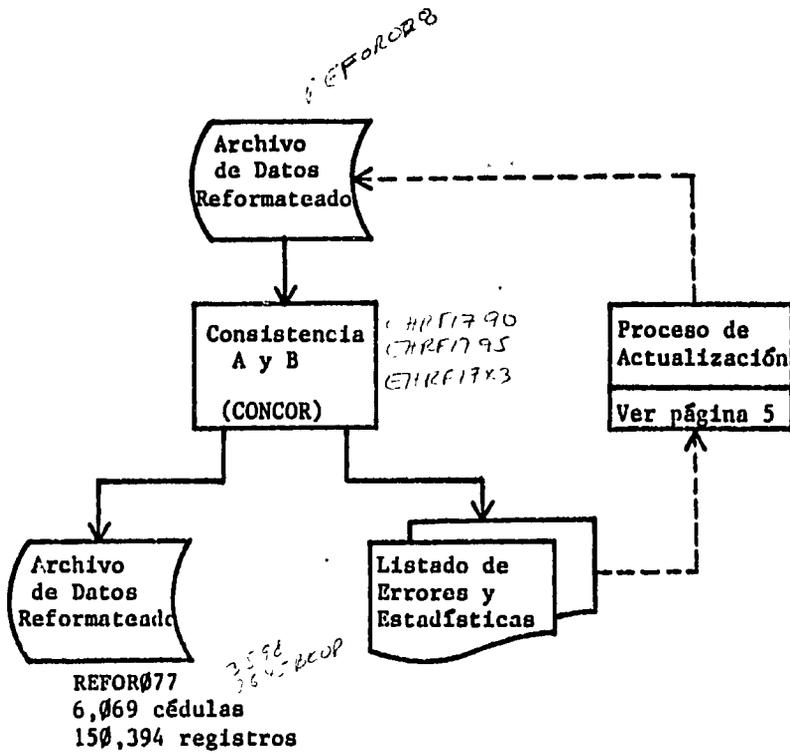
1. Proceso para obtención de Mínimos y Máximos por segmento



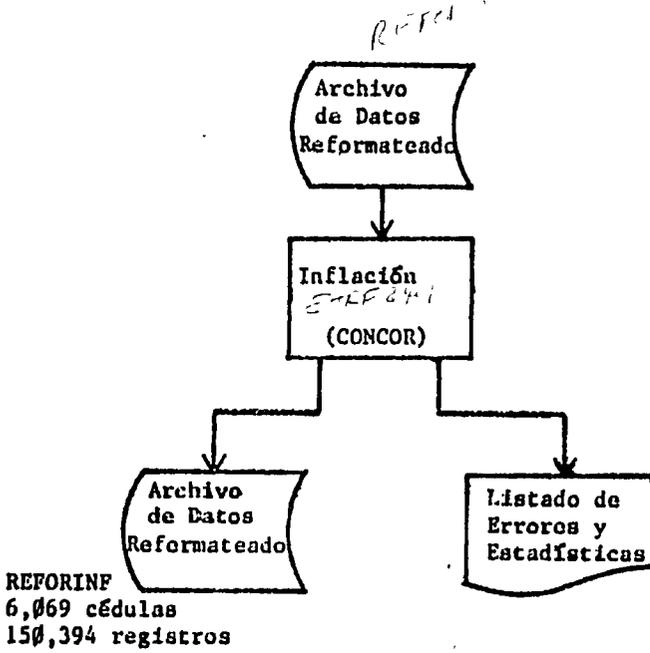
2. Chequeo de Rango



Proceso de Consistencia

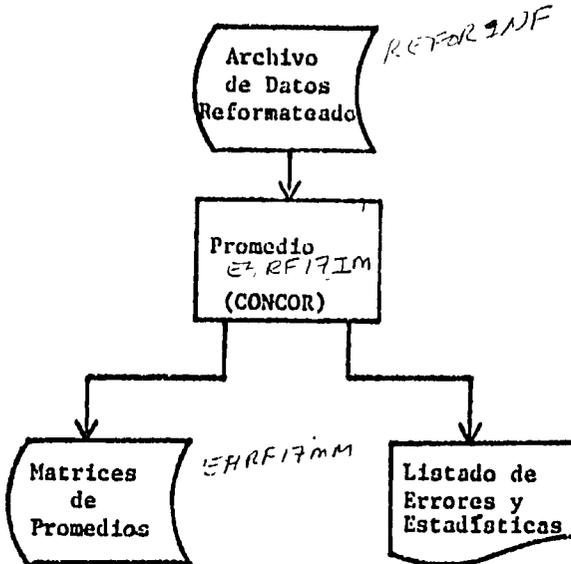


Proceso de Inflación

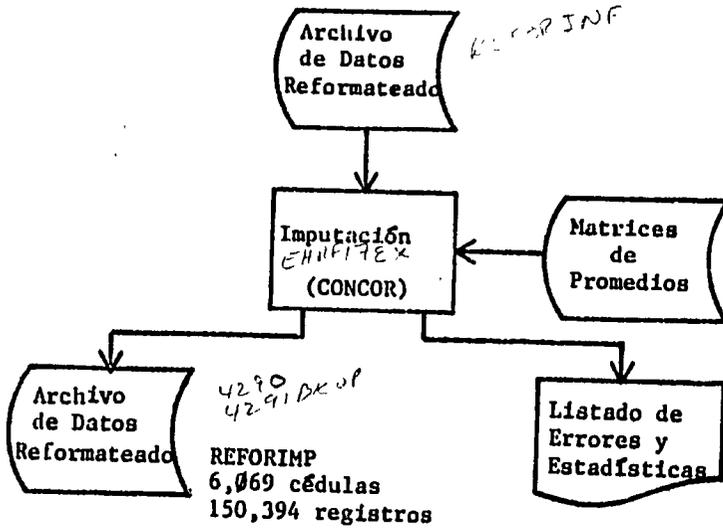


*364
259800*

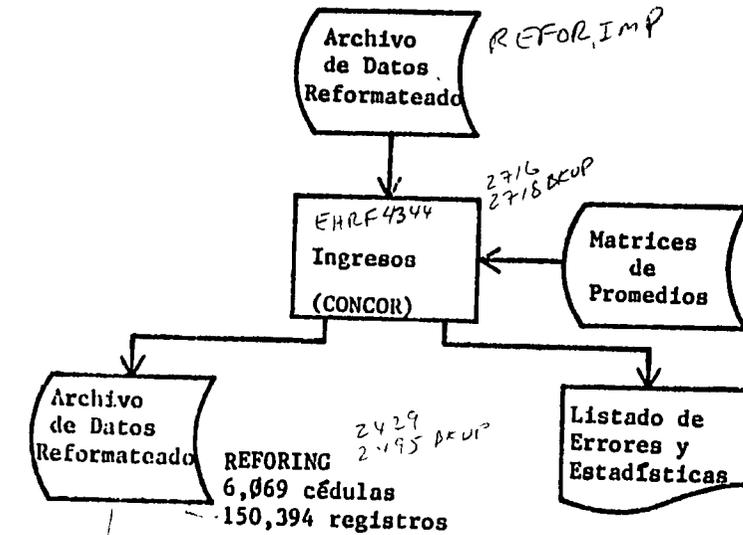
Proceso de Cálculo de Promedios



Proceso de Imputación con Promedios



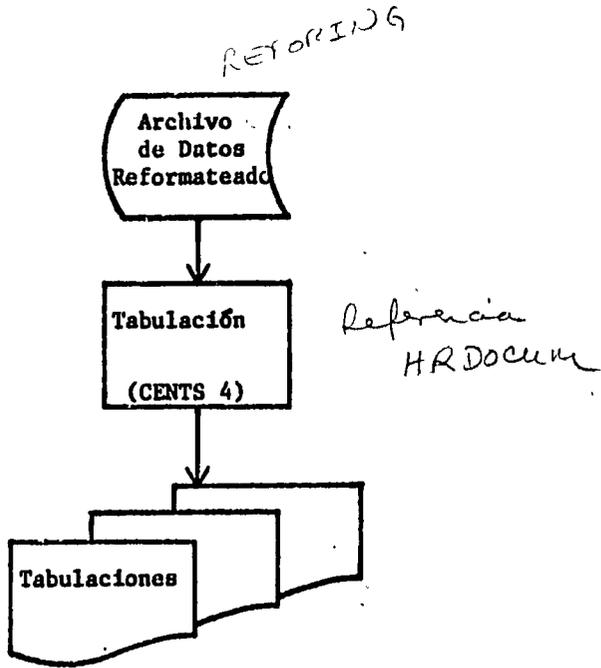
Proceso de Cálculo y Agrupación de Ingresos



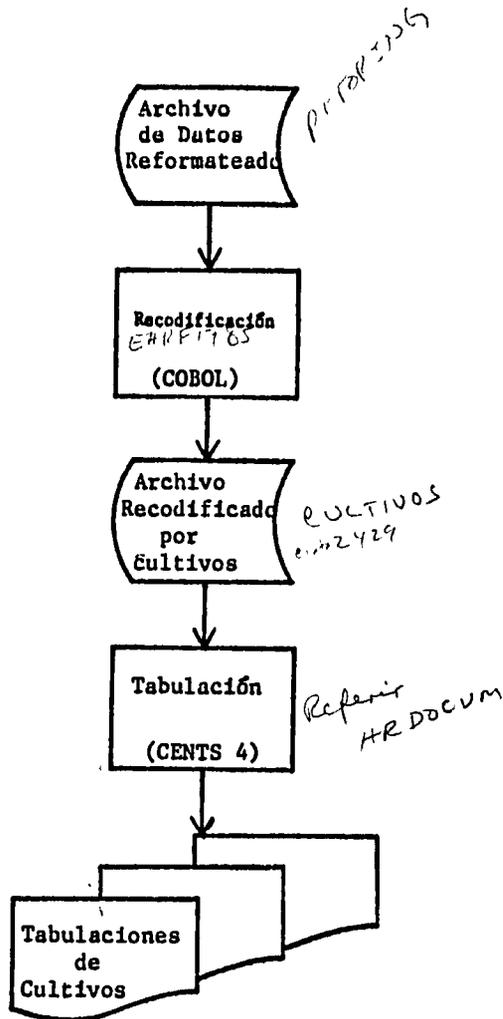
Cambios Seguros Dominio

REFORING
150,394

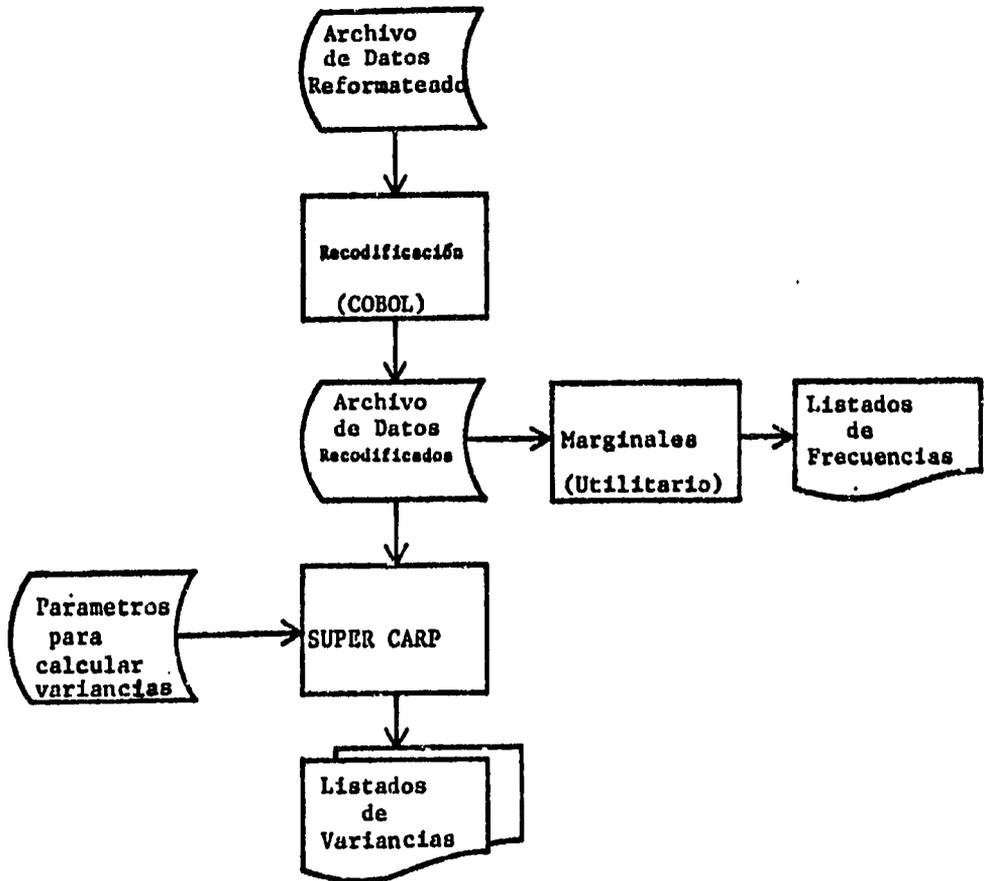
Proceso de Tabulación



Obtención de Tabulados de Cultivos



Proceso de Cálculo de Variancias



APENDICE 16

CALENDARIO DE PRINCIPALES ACTIVIDADES

Descripción de la Actividad	Periodo de Ejecución	
	Comienzo	Conclusión
1. <u>Planificación</u>		
a. Evaluación de las estadísticas agropecuarias.	7 mar. 83	25 mar. 83
b. Formulación del plan de implementación de la encuesta.	7 mar. 83	25 mar. 83
2. <u>Diseño y Selección de la Muestra</u>		
a. Familiarización con muestras similares y fuentes de información sobre el sector y el área rural.	1 ago. 83	19 ago. 83
b. Definición de los objetivos, el universo y la unidad de muestreo para la muestra.	1 ago. 83	19 ago. 83
c. Obtención de información y tabulados para el diseño de la muestra.	28 nov. 83	30 dic. 83
d. Obtención del marco muestral.	1 ene. 84	30 abr. 84
e. Fijación del diseño de la muestra y el diseño de una pequeña submuestra para estudiar la confiabilidad de los datos.	28 nov. 83	21 dic. 83
f. Documentación del diseño de la muestra y submuestra.	1 ene. 84	30 abr. 84
g. Identificación y listado de las UPM por dominio.	12 dic. 83	30 abr. 84
h. Desarrollo de especificaciones para la selección de los segmentos.	12 dic. 83	30 dic. 83
i. Coordinación de la selección manual de los segmentos (incluye selección de segmentos).	21 mar. 84	24 mayo 84

j.	Diseño de la hoja de registro:		
	Borrador	28 nov. 83	12 dic. 83
	Definitivo.	12 abr. 84	18 abr. 84
k.	Revisión de especificaciones para el registro de viviendas dentro de los segmentos seleccionados.	28 nov. 83	12 dic. 83
l.	Selección de muestra de viviendas.	20 ago. 84	15 set. 84
m.	Evaluación de la encuesta.	21 ene. 86	14 feb. 86
3.	<u>Elaboración del Plan de Tabulación</u>		
a.	Desarrollo de los cuadros.	1 ago. 83	24 oct. 83
b.	Revisión del plan de tabulación.	29 mar. 84	20 abr. 84
c.	Desarrollo de especificaciones de cuadros.	22 oct. 84	3 mayo 85
4.	<u>Diseño de la Cédula</u>		
a.	Conceptualización de la cédula incluyendo variables a investigarse y definiciones de términos.	1 ago. 83	24 oct. 83
b.	Diseño del primer borrador.	24 oct. 83	12 dic. 83
c.	Primera prueba en el campo.	14 dic. 83	14 dic. 83
d.	Evaluación y revisión del primer borrador.	19 dic. 83	10 feb. 83
e.	Segunda prueba de la cédula.	29 feb. 84	6 mar. 84
f.	Evaluación y revisión del segundo borrador.	8 mar. 84	18 mayo 84
g.	Tercera prueba de la cédula.	29 mayo 84	1 jun. 84
h.	Evaluación y revisión del tercer borrador.	8 jun. 84	18 jul. 84
5.	<u>Elaboración de Manuales</u>		
a.	Manual del Registrador	16 mayo 84	8 jun. 84
b.	Manual del Jefe de Brigada (Registro).	11 jun. 84	22 jun. 84

c. Manual del Encuestador: Para la prueba Definitivo	9 ene. 84 4 jun. 84	15 feb. 84 24 ago. 84
d. Manual del Jefe de Brigada (Operación de Campo)	30 jul. 84	31 ago. 84
e. Manual de Codificación y Conversión	25 set. 84	9 nov. 84
f. Manual de Crítica por Computadora	13 nov. 84	20 mayo 85
6. <u>Capacitación</u>		
a. Capacitación de 31 encuesta- dores (15 seleccionados) para la prueba del registro y el segundo borrador de las cédulas tipo 1 y 2.	20 feb. 84	25 feb. 84
b. Capacitación de 8 encuestadores para la prueba del tercer borrador de la cédula tipo 2.	27 mayo 84	28 mayo 84
c. Capacitación de 148 personas (jefes departamentales, jefes de brigadas y encuestadores) para el registro de viviendas.	2 jul. 84	7 jul. 84
d. Capacitación de 16 instructores.	10 set. 84	15 set. 84
e. Capacitación de 147 personas (jefes departamentales, jefes de brigadas y encuestadores) para las entrevistas.	17 set. 84	28 set. 84
f. Capacitación de 54 personas (25 seleccionados) para la codificación de las cédulas.	22 nov. 84	29 nov. 84
g. Capacitación de 20 digitadores para la entrada de datos.	10 dic. 84	14 dic. 84
h. Capacitación de 25 críticos para el análisis de cobertura.	8 ene. 85	8 ene. 85
i. Capacitación de 25 críticos para la crítica de rangos.	25 feb. 85	25 feb. 85
j. Capacitación de 25 críticos para la crítica de consistencia.	28 mayo 85	30 mayo 85

7. Campana de Difusion

a. Prueba de campo de feb./mar. 84:		
Preparación	5 dic. 83	24 feb. 84
Ejecución	28 feb. 85	6 mar. 84
b. Registro de viviendas.		
Preparación	1 jul. 84	15 jul. 84
Ejecución	17 jul. 84	31 ago. 84
c. Entrevistas		
Preparación	1 set. 84	29 set. 84
Ejecución	3 oct. 84	24 dic. 84

8. Operación de Campo

a. Ejecución del registro de viviendas.	18 jul. 84	31 ago. 84
b. Diligenciamiento de las cédulas.	4 oct. 84	29 dic. 84

9. Desarrollo de Especificaciones

a. Especificaciones de cálculo y agrupación de ingresos.	18 jun. 84	20 jul. 84
b. Valores máximos y mínimos para la crítica de rangos.	16 jul. 84	14 ago. 85
c. Especificaciones para la crítica de consistencia.	17 ago. 84	6 set. 85
d. Especificaciones para los cuadros de salida.	22 oct. 84	27 set. 85
e. Especificaciones de precio promedio para imputación.	18 mar. 85	21 mar. 85
f. Especificaciones para ajuste por inflación.	22 mar. 85	1 jun. 85
g. Especificaciones para la expansión de datos.	1 mayo 85	5 jun. 85
h. Especificaciones para el cálculo de varianzas.	1 jun. 85	30 set. 85
i. Especificaciones para el ajuste de "no entrevista".	1 dic. 84	28 feb. 85

- | | | | |
|-----|--|------------|------------|
| j. | Especificaciones para los segmentos no entrevistados. | 1 dic. 84 | 31 ene. 85 |
| 10. | <u>Revisión, Conversión y Codificación</u> | | |
| a. | Revisión de cédulas, conversión de diferentes unidades de medidas a una estándar y codificación. | 4 dic. 84 | 7 ene. 85 |
| 11. | <u>Desarrollo del Sistema de Procesamiento por Computadora</u> | | |
| a. | Diseño del sistema. | 28 nov. 83 | 9 dic. 83 |
| b. | Especificaciones para entrada de datos. | 22 oct. 84 | 26 oct. 84 |
| c. | Especificaciones para el programa de corrección. | 30 ene. 84 | 10 feb. 84 |
| d. | Especificaciones para el programa de listar. | 30 ene. 84 | 10 feb. 84 |
| e. | Especificaciones para generar archivos de cédulas. | 30 ene. 84 | 10 feb. 84 |
| 12. | <u>Desarrollo y Prueba de los Programas de Computadora</u> | | |
| a. | Programa de cobertura y generación del archivo de cédulas. | 9 feb. 84 | 24 mar. 84 |
| b. | Programa de corrección. | 9 feb. 84 | 24 mar. 84 |
| c. | Programa de listar. | 9 feb. 84 | 20 mar. 84 |
| d. | Instalación y seminario de CONCOR. | 21 mayo 84 | 8 jun. 84 |
| e. | Programa de rangos. | 10 jul. 84 | 28 nov. 84 |
| f. | Programa de consistencia. | 10 ago. 84 | 1 mar. 85 |
| g. | Programa de imputación. | 1 ago. 85 | 15 ago. 85 |
| h. | Programa de ponderación. | 5 ago. 85 | 10 ago. 85 |
| i. | Programa de cálculo y agrupación de ingresos. | 23 abr. 85 | 30 jul. 85 |
| j. | Programa para generar archivo de cultivos. | 1 ago. 84 | 8 ago. 84 |

k. Instalación y seminario de CENTS 4.	21 feb. 85	13 mar. 85
l. Programación de cuadros.	1 feb. 85	31 oct. 85
m. Instalación y seminario de SUPER CARP.	10 jun. 85	14 jun. 85
13. <u>Procesamiento de Datos por Computadora</u>		
a. Entrada de datos.	17 dic. 84	30 mar. 85
b. Generación de archivo maestro.	23 dic. 84	31 dic. 84
c. Cobertura y generación de archivo de cédulas.	14 ene. 85	15 abr. 85
d. Resolución de errores de cobertura.	14 ene. 85	12 abr. 85
e. Generación de archivo de "no entrevistado".	30 jun. 85	30 jun. 85
f. Conversión a unidad de medida estándar.	20 ene. 85	30 abr. 85
g. Resolución de errores de conversión.	15 feb. 85	22 feb. 85
h. Chequeo de rangos	4 mar. 85	30 ago. 85
i. Resolución de valores fuera de rangos.	18 feb. 85	5 jun. 85
j. Chequeo de consistencia.	5 jun. 85	9 oct. 85
k. Resolución de inconsistencias.	5 jun. 85	7 oct. 85
l. Asignación de ponderaciones.	16 oct. 85	16 oct. 85
m. Cálculo y agrupación de ingresos y precios.	15 oct. 85	15 oct. 85
n. Imputaciones al archivo de cédulas.	11 oct. 85	11 oct. 85
o. Generación de archivo de cultivos.	17 oct. 85	17 oct. 85

p. Generación de tabulaciones.	2 set. 85	28 feb. 86
q. Generación de variancias.	21 oct. 85	31 mar. 86
14. <u>Documentación</u>		
a. Documento de Conceptos y Definiciones.	27 mayo 85	8 nov. 85
b. Informe Metodológico	6 jun. 85	7 feb. 86
c. Documento de Especificaciones para el Procesamiento de Datos	6 ene. 86	7 feb. 86
d. Guía del Usuario.	1 ago. 85	7 feb. 86
e. Resultados Definitivos	2 set. 85	31 mar. 86

**FUNCIONARIOS Y PERSONAL TÉCNICO QUE PARTICIPO EN LA
ENCUESTA NACIONAL DE HOGARES RURALES**

1. DEL INSTITUTO NACIONAL DE ESTADISTICA

1.1 ALTA DIRECCION

Dra. GRACIELA FERNANDEZ BACA DE VALDEZ
Jefa del Instituto Nacional de Estadística

Estad. VICTOR LOZANO DIAZ
Director Técnico

1.2 DE LA DIRECCION GENERAL DE CENSOS Y ENCUESTAS

Ing. JORGE VANDERGHEN BRANIZZA
Director General de Censos y Encuestas
Jefe de la Actividad 4a: "Manejo de la Encuesta"
(Hasta Octubre de 1984)

I.O. ATILIO PIZARRO HUAMAN
Director General de Censos y Encuestas
Jefe de Actividad 4a: "Manejo de la Encuesta"
(Desde Noviembre de 1984)

Prof. LUIS APOLAYA INGUNZA
Director Ejecutivo
(Hasta Setiembre de 1984)

Ing. WALTER CAVERO DHAGA
Director Ejecutivo
(Desde Abril de 1985)

1.2.1 De la Dirección de Métodos y Muestreo

Estad. RAMON DE LA CRUZ YUPANQUI
Sub-Director de Muestreo y Análisis Estadístico

Eca. ALFREDO BALMACEDA CORREA
Estadístico Matemático

1.2.2 De la Dirección de Operaciones

Ing. AUGUSTO VILLON TORRES
Director
(Hasta Setiembre de 1985)

Ing. MANUEL CABRERA NEYRA
Generalista

Ing. RICARDO JARA RODRIGUEZ
Especialista

Eco. MARTHA VARGAS ARIZA
Especialista
(Hasta Junio de 1985)

Bach. WILLIAM LAZARO APOLAYA
Especialista
(Desde Julio de 1985)

1.2.3 De la Dirección de Cartografía

Géog. LUIS GUILLEN PASTUS
Director

Ing. ALDO ALAY MALDONADO
Jefe de la Unidad de Ejecución Cartográfica

Sr. GERARDO CUELLAR MAYHUA
Supervisión Cartográfica

Sr. FLORENCIO CUELLAR MAYHUA
Capacitación y Supervisión Cartográfica

1.3 OTRAS DEPENDENCIAS DEL INE

1.3.1 De la Dirección General de Informática

Lic. MARIO CAMARA FIGUEROA
Director General de Informática
(Hasta Junio de 1985)

Lic. CHERLY ORE DE CRUZADO
Director Ejecutivo

Lic. MARIO CASTILLO BARCENA
Analista PAD

Programadores (Procesamiento de Datos):

Bach. FLOR BUSTOS DIAZ

Bach. MARTHA RODRIGUEZ CALLE

Bach. ADRIANA CABANILLAS GAMBOA (Hasta ABR '85)

Bach. OSCAR PUICON MONTERO

Bach. FELIX DAVALOS TORRES

Sra. TERESA GASPAR VELASQUEZ

1.3.2 De la Oficina de Comunicación, Información e Impresiones

Ing. WALTER CAVERO DHAGA
Director General
(Hasta Marzo de 1985)

Sr. AUGUSTO CHERO ASCENCIOS
Director de Publicaciones e Impresiones:

2. DEL MINISTERIO DE AGRICULTURA

Ing. ALFREDO VALENCIA JUSTO
Director General de la Oficina Sectorial de Estadística

Ing. HUGO RAMOS BARRANCA
Director de Estadísticas Agropecuarias
Administrador de la Actividad: "Operación de Campo"

Srta. ROSA SAMANEZ CIACON
Encargada de la Compañía Publicitaria

Sr. ROHAN TOVAR PEREZ
Coordinador de la Capacitación

Instructores

Ing. JOSE KAWATA MAKABE
Ing. JESUS QUIÑONES GALLEGOS
Ing. SANTOS MAZA Y SILUPU
Ing. RAUL ANDRADE ORE
Bach. RODOLFO PEÑA ARCE
Eco. RAUL MERCADO RODRIGUEZ
Ing. ANDRES ISLA MALDONADO
Eco. JUAN SANTAMARIA ARANDA
Ing. MELCHOR VASQUEZ AREVALO

Supervisoras Nacionales:

Ing. FRANCISCO SIGUEÑAS ROBLES
Eco. JUAN SANTAMARIA ARANDA
Dem. FRANCISCO MORALES ANGELES
Eco. JAIME MUÑOZ CORREA
Eco. CESAR VARGAS CASTRO
Sr. ANTONIO ROBLES VILLAFUERTE
Bach. JOSE CHICOMA HUITMAN

Jefes Departamentales:

Téc.	VIRGILIO SOLIS GUTIERREZ	Tumbes
Ing.	NORBERTO GARCIA MIRANDA	Piura
Ing.	MANUEL HARO VERDE	La Libertad
Ing.	ANTONIO GONZALES ALVA	Ancash
Ing.	JUAN VERA HUARCAYA	Lima
Ing.	VICTOR ROIG POLAR	Arequipa
Bach.	OLGA MALDONADO CARRASCO	Moquegua
Est.	LUIS FERRUA VIVANCO	Tacna
Méd.	Vet. JUAN DIAZ BARCO	San Martín
Ing.	MARCO CENTURION LEON	Cajamarca
Ing.	JOSE OSORIO CHUQUILLANQUI	Huánuco
Ing.	LUIS ABREGU TELLO	Pasco
Ing.	RAUL ANDRADE ORE	Junín
Ing.	RAUL MERCADO RODRIGUEZ	Huancavelica
Ing.	BENILDE CERRON VEGA	Ayacucho
Ing.	MARIO PRADA TORRES	Apurímac
Ing.	RODOLFO NAJAR OLAGUIVEL	Cuzco
Ing.	EDILBERTO MANSILLA MAMANI	Puno
Ing.	ANDRES SILVA MALDONADO	Ucayali
Ing.	JESUS QUIÑONES GALLEGOS	Amazonas
Ing.	CESAR ZUMAETA TUESTA	Loreto
Ing.	GRIMANIEL BERNAL RUBIO	Lambayeque

3. DE ORGANISMOS INTERNACIONALES

3.1 AGENCIA PARA EL DESARROLLO INTERNACIONAL

Sr. DAVID FLOOD
Jefe del Proyecto

Sr. LOREN L. PARKS
Consultor

3.2 OFICINA DEL CENSO DE LOS ESTADOS UNIDOS DE NORTEAMERICA

Sr. MIGUEL CUEVAS
Jefe del Proyecto

Sr. DAVID MEGILL
Diseño y Selección de la Muestra

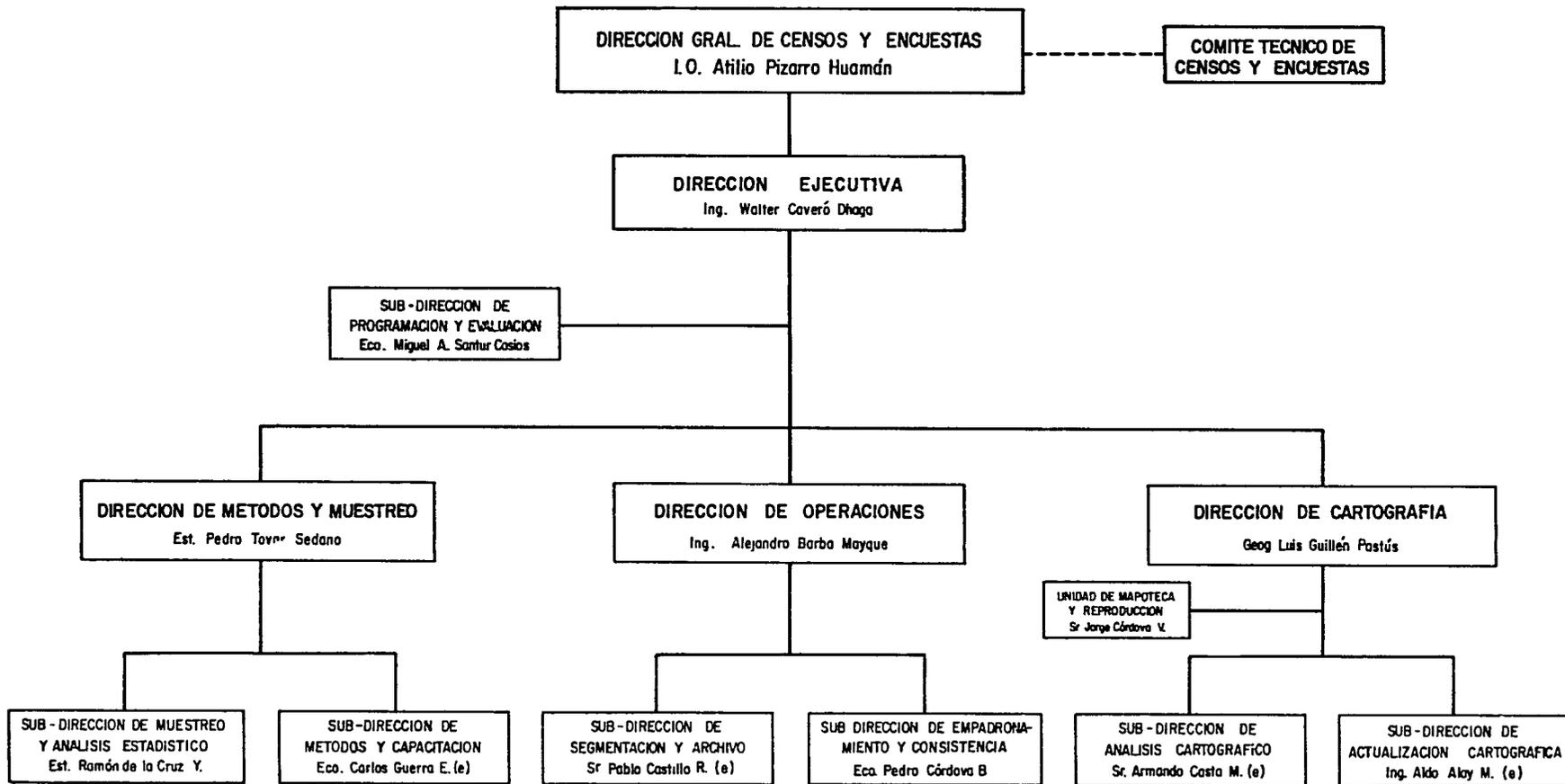
Sra. SANDRA ROWLAND
Diseño de Cédula y Especificaciones

Srta. EVELYN S. FODORKO
Elaboración del Formata de Cédulas

Sr. PETER J. VALENTINO
Elaboración de las Especificaciones de Ingreso

Sr. ENRIQUE GOMEZ
Procesamiento de Datos (Diseño y Supervisión)

Srta. VIVIAN TORO
Procesamiento de Tabulados



133