

5110471  
PN-AAP-783/62  
LEN-34258

PROYECTO DE APOYO AL SISTEMA DE  
PLANEAMIENTO REGIONAL  
PARA EL DESARROLLO RURAL  
CONVENIO AID - MPC No. 511 - 0471

63

METODOS DE RECOLECCION DE DATOS  
PARA LA PLANIFICACION REGIONAL

Enrique Gómez D'Angelo  
Walter Gómez D'Angelo  
Asesores USAID/MPC  
Septiembre de 1981.

## INDICE

	Página
1. Introducción.....	1
2. Diseño de la recolección de datos.....	6
2.1 Especificación de datos.....	6
2.2 Proceso de especificación de datos.....	9
2.3 Diseño de la muestra.....	11
2.3.1 Tipos de error.....	13
2.3.2 Métodos probabilísticos y no-probabilísticos de la selección de muestras.....	14
2.3.3 Técnicas aleatorias para la selección de muestras.....	16
2.3.4 Tamaño de la muestra.....	19
2.3.5 Diseño de la muestra para las comparaciones...	21
2.4 Diseño de los instrumentos.....	24
2.4.1 Rememoración.....	24
2.4.2 Mantenimiento de registros.....	27
2.4.3 Medición y observación directa.....	28
2.4.4 Experiencia lograda en las CDD's.....	28
2.5 Verificación de campo.....	29
2.5.1 Verificación de la muestra.....	29
2.5.2 Verificación del instrumento.....	30
3. Trabajo de campo.....	31
3.1 Selección y entrenamiento de los encuestadores.....	31
3.2 Supervisión.....	32
3.2.1 Rechazos o ausencias de unidades de observación.....	34
3.2.2 Deficiencia en la delimitación de áreas.....	36
4. Análisis.....	36
4.1 Entrada y limpieza de datos.....	36
4.2 Procesamiento de datos.....	39
4.2.1 Programas de verificación.....	40

11

	Página
4.3	Análisis estadístico..... 41
4.3.1	Comparaciones..... 41
4.3.2	Etapas en la formulación de comparaciones..... 43
4.4	Redacción del informe..... 43

#### Anexo 1

##### Tamaño de la muestra

1.	Especificación del universo muestral..... 1
2.	Tamaño del universo..... 1
3.	Muestreo aleatorio simple..... 2
4.	Muestreo por áreas..... 5

#### Anexo 2

##### Tests estadísticos

1.	Comparación simple entre dos grupos..... 1
2.	Comparaciones múltiples..... 5

## I. Introducción

Las Corporaciones Departamentales de Desarrollo (CDD's) y el Ministerio de Planeamiento y Coordinación (MPC) son responsables de señalar prioridades de inversión que permitan alcanzar los objetivos del desarrollo con la mayor eficacia posible en el uso de los escasos recursos de que dispone el país.

En una primera etapa, el señalar prioridades está dirigido principalmente a las inversiones que realizan las CDD's mismas. Posteriormente, el señalamiento de prioridades podrá extenderse mediante recomendaciones, incentivos o restricciones legales, etc, a fin de orientar las inversiones que realizan otras instituciones públicas o privadas.

Este proceso de toma de decisiones respecto a donde se debe orientar el uso de los recursos disponibles en cada región es responsabilidad primaria de las oficinas de planificación de las CDD's quienes deben presentar a sus ejecutivos de alto nivel la información que se requiere para la asignación de los fondos de que dispone la Corporación y la orientación que se da a otros organismos regionales.

Una vez que las CDD's han fijado metas y objetivos que se desean alcanzar en el proceso de desarrollo, y los han especificado claramente en términos cuantificables de ingreso, empleo, educación, salud, nutrición y vivienda, es función de los planificadores responder a la pregunta de: ¿cómo podemos alcanzar estos objetivos empleando los recursos de que disponemos, con la mayor eficiencia posible?

Al responder a esta interrogante los planificadores tropiezan con dos obstáculos importantes: primero, no es suficiente ni apropiado el emitir recomendaciones simplistas basadas únicamente en criterios subjetivos u

opiniones personales de técnicos que conocen la región. En segundo lugar, es necesario dar estas respuestas en forma oportuna ya que el proceso de inversión está limitado por fechas y plazos dentro de los cuales los ejecutivos de alto nivel deben tomar decisiones tanto en las CDD's como en el MPC.

Una planificación regional basada únicamente en el criterio subjetivo de los funcionarios de planificación resulta en una identificación de proyectos que depende en gran parte de la experiencia previa y la especialización académica de dichos funcionarios. Es así que los ingenieros civiles tienden a recomendar proyectos de vinculación caminera, los agrónomos proyectos de riego o investigación agrícola, los sociólogos proyectos de nutrición, etc. Siguiendo este camino, las ideas de proyectos se convierten en estudios de pre-factibilidad y factibilidad realizados en base a proyecciones "de escritorio" respecto a que cultivos producirán los campesinos que reciben riego, rendimientos a esperarse, precios futuros, número de campesinos que recibirán los beneficios de un proyecto de investigación y extensión agrícola, y naturaleza y alcance de dichos beneficios. Si estas proyecciones se basan sólo en la opinión del especialista que elabora el estudio, los resultados que se proyectan tienen un alto grado de incertidumbre. La experiencia observada en las CDD's muestra que estos estudios casi invariablemente arrojan resultados favorables al proyecto estudiado. Una vez demostrada la factibilidad económica de un proyecto, su ejecución final depende únicamente de la disponibilidad de fondos, que en muchos casos provienen de organismos internacionales de desarrollo. Los criterios empleados por estos organismos para la asignación de fondos, no responden necesariamente a las metas de desarrollo fijadas por el MPC y las CDD's. De esta forma, se llega a una situación en la cual los proyectos ejecutados no están dirigidos a las metas establecidas por las CDD's y el MPC, y tampoco son necesariamente los medios más eficientes para lograr las metas propuestas, con los recursos de que se dispone.

La metodología de planificación regional que proponemos tiene la capacidad de superar estos obstáculos y su característica más importante es que se basa en información y evidencias "experimentales". Su herramienta principal es la recolección de datos de campo cuyo grado de exactitud y representatividad se puede establecer a través de la estadística. De esta manera, los planificadores pueden presentar diferentes alternativas de inversión cuya relativa eficiencia en el uso de los recursos, para alcanzar las metas deseadas, se establece cuantitativamente. Esta misma disponibilidad de información fidedigna, obtenida mediante métodos objetivos, permite a las CDD's y al MPC adoptar posiciones más sólidas frente a otros organismos nacionales e internacionales a fin de dirigir adecuadamente sus inversiones en la región.

El uso de encuestas por muestreo permite obtener la información necesaria, cuya validez se establece estadísticamente, a un costo bajo y en un tiempo relativamente corto. El procesamiento de datos por computador es recomendado cuando se deben calcular variables que se definen a partir de los datos originales recabados en la encuesta, por ej. ingreso neto de la granja, etc.

Un supuesto implícito en esta filosofía de planificación, basada en información experimental, es que existen experiencias ya realizadas en la ejecución de ideas, proyectos o acciones, que se pueden extrapolar a la región o departamento, por ejemplo, crédito, apertura o mejoramiento de caminos, construcción de escuelas, riego, etc. En la práctica se observa que es muy difícil o prácticamente imposible encontrar casos en que no exista ninguna experiencia similar que permita estimar el impacto esperado de un proyecto que se desea evaluar a priori.

El proceso de recolección de datos que aquí se presenta consta de tres fases principales: diseño de la recolección de datos, trabajo de campo y

análisis. A su vez, cada una de estas fases se subdivide en actividades y sub-actividades o tareas. El Cuadro 1 muestra el esquema de las fases, actividades y tareas a realizarse en la recolección de datos.

Cada una de estas etapas es posible realizarla con diferentes niveles de complejidad que van desde los abordajes más simples que requieren menos esfuerzo y tiempo, a los más elaborados cuya mayor profundidad y uso más intenso de recursos resultan en una mayor riqueza analítica, mayor precisión estadística y mayor detalle en las conclusiones. Esto no significa que en cada etapa o tarea del proceso de recolección de datos se vaya a recomendar necesariamente el nivel más alto de complejidad. La experiencia observada en esfuerzos de análisis realizados en otros países de América Latina y Bolivia, y los realizados durante nuestro trabajo en este proyecto nos permite emitir recomendaciones respecto al método que debe emplearse en cada una de las etapas del proceso de recolección de datos.

CUADRO 1

FASES, ACTIVIDADES Y TAREAS DE LA RECOLECCION DE DATOS PARA LA  
PLANIFICACION REGIONAL

---

1. Diseño de la Recolección de Datos
  - Especificación de Datos
  - Diseño de la Muestra
  - Diseño de los Instrumentos
  - Verificación de Campo
  
2. Trabajo de Campo
  - Selección y Entrenamiento de los Encuestadores
  - Supervisión
  
3. Análisis
  - Entrada y Limpieza de Datos
  - Procesamiento de Datos
  - Análisis Estadístico
  - Redacción del Informe

## 2. Diseño de la Recolección de Datos

### 2.1 Especificación de Datos

Este es el primer paso en el proceso de recolección de datos, y su influencia en otras etapas posteriores, (tipo de muestreo y análisis de los datos) obliga al planificador a estudiar el asunto en tres planos diferentes en forma simultanea. Estos tres planos son:

#### a) Muestreo y recolección de datos:

Cuál es la dificultad de obtener el dato, tipo de muestreo a utilizar, tamaño de la muestra, factibilidad logística de la muestra que se pretende.

#### b) Análisis e interpretación de los resultados

Cómo se analizaría el dato, qué tipo de clasificación se emplearía, cuál sería su poder explicativo del punto de vista estadístico.

#### c) Experiencia en la materia

Cuán importante es este dato para la comprensión del sistema que se estudia. Qué relación tiene este dato con intentos de introducir mejoras en dicho sistema. Cuan importantes son las hipótesis operacionales que este dato verificaría.

Cualquiera de nosotros puede hacer referencia a esfuerzos de investigación realizados en las CDD's o el MPC que han quedado

truncos o, si se han terminado, ha sido con gran atraso, elevado costo y poca utilidad en la planificación. En todos estos casos se puede observar que los primeros errores cometidos se remontan a esta etapa. El investigador sin experiencia se lanza a requerir demasiada información que luego no es posible obtener, es muy difícil de procesar, y luego no se necesita emplear en el análisis.

La situación ideal y la que recomendamos es la de contar con un planificador que tiene experiencia en las tres áreas: ha realizado investigaciones de campo por muestreo, está familiarizado con el análisis estadístico a efectuarse, y tiene experiencia en la materia de análisis (agricultura, desarrollo rural, etc).

Aunque es relativamente difícil encontrar planificadores con experiencia en estas tres áreas, el personal de planificación que participa en este seminario y luego interviene en la aplicación de la metodología de planificación regional en sus respectivas Corporaciones, completará con su experiencia práctica el entrenamiento necesario para llevar a cabo la "especificación de datos" en forma apropiada.

A fin de dar mayor claridad al problema de asignación de personal para la especificación de datos examinaremos algunas alternativas de solución:

i. Equipo multidisciplinario para la especificación de datos.

El abordaje de "equipo multidisciplinario" consiste en organizar un equipo de especialistas donde generalmente se requieren por lo menos cuatro especialistas: un especialista en muestreo, un especialista en recopilación

de datos de campo, un analista estadístico, y un especialista en la materia de estudio que comprende el sistema rural bajo análisis.

La dificultad y desventaja seria del abordaje de equipo es que ninguno de los integrantes por sí solo es capaz de ver el cuadro completo y en consecuencia ninguno es capaz de emitir los juicios delicados acerca de la importancia relativa, y precisión relativa de la obtención de datos. El estadístico, incapaz de juzgar la relativa importancia de diferentes asuntos, buscará inevitablemente tests estadísticos cuyo aporte analítico será únicamente marginal en muchos casos, y explorará avenidas analíticas que no son realmente centrales al problema crítico estudiado. El profesional con experiencia en la materia, dado que no visualiza los asuntos estadísticos no formulará los preguntas del modo correcto para llegar a datos o análisis pertinentes. Los expertos en muestreo y recopilación de datos estarán esperando la especificación de datos para seguir adelante con su trabajo; pero, sin poder ver el cuadro total analítico, poco podrán contribuir para el proceso interactivo de especificación de datos en el que los costos relativos, exactitud y manejo de la recolección de datos deberían ser consideraciones permanentes. El resultado del abordaje del equipo multidisciplinario es la fragmentación. La "simultaneidad" del proceso se pierde cuando cada uno trata de hacer sólo lo que le corresponde.

ii. Entrenamiento de especialistas en la materia de recopilación de datos y metodología de análisis.

Cuando no se cuenta con un planificador con experiencia en las tres áreas, una posible solución consiste en tomar a un profesional con experiencia en una o dos de las áreas para luego darle un entrenamiento intensivo en el área o las áreas que no conoce.

Con frecuencia cuando se enfrenta esta decisión, las instituciones que se embarcan en la Planificación Regional, cometen el error de escoger analistas con habilidad en recolección de datos y análisis estadístico esperando poder "ponerlos al día", en un corto plazo, en la materia de desarrollo rural que se analiza. En este caso se cree erróneamente que es más fácil darle a un estadístico suficiente conocimiento de Proyectos de Desarrollo Rural, que darle a una persona con experiencia en proyectos de Desarrollo Rural un conocimiento suficiente de estadística. En casi todos los casos lo opuesto es cierto; para fines de análisis del desarrollo generalmente es más fácil aumentar las habilidades estadísticas de un especialista en desarrollo rural que aumentar el conocimiento de desarrollo rural en un especialista en estadística.

2.2 Proceso de Especificación de Datos

El trabajo de especificación de datos se inicia elaborando un listado de la información que se requiere. Este listado va de lo general a lo particular; por ejemplo, la encuesta de agricultores

del Departamento de Oruro tocaba los siguientes puntos en su listado inicial:

1. Ingreso: Producción, Gastos, Ventas, Ingresos extraprediales
2. Tamaño de la finca
3. Inventario de bienes
4. Demografía: Educación, nutrición, salud
5. Restricciones identificadas.

Luego se completa el listado incluyendo los detalles de cada rubro. por ejemplo:

2. Tamaño de la finca:
  - hectáreas totales
  - hectáreas cultivadas con productos temporales
  - hectáreas cultivadas con productos permanentes
  - hectáreas de pastos
  - hectáreas en descanso
  - hectáreas en barbecho

Por un proceso iterativo que considera la dificultad relativa de obtener los datos y su importancia en el análisis, se llega a una lista final de los datos a obtenerse en la investigación.

Una vez aclarado que datos son de interés, es necesario diseñar los cuadros de salida que se desean una vez que los datos se hayan procesado. Más adelante se explicará porqué es necesario diseñar, al menos en forma preliminar, los cuadros de salida.

El diseño se inicia con un listado de los cuadros deseados indicando únicamente el título. Luego se procede al diseño del formato con sub-títulos para cada columna y cada fila que debe contener el cuadro. Es decir, se precisa que tipo de estratificaciones y sub-estratificaciones se aplicarán a los datos que se obtendrán de la investigación muestral.

### 2.3 Diseño de la Muestra

El muestreo es una herramienta estadística que permite inferir características de un grupo sin necesidad de estudiar a todo el grupo sino sólo a una fracción relativamente pequeña que se denomina muestra. A partir de los datos de la muestra es posible calcular con un cierto grado de probabilidad de error qué valores de un parámetro cualquiera se pueden asignar al total del grupo o universo. Por ejemplo, a partir de la muestra de 287 campesinos realizada en tres provincias de Chuquisaca se puede concluir con un 80% de certeza que el ingreso medio del agricultor en esa región está entre \$0.9.227 y 13.587.-

El grado de precisión que se obtiene con una muestra depende más que todo del número de observaciones o tamaño absoluto de la muestra y no de la proporción de las observaciones efectuadas en relación al tamaño del universo o grupo total que se estudió. Por ejemplo, se necesitaría aproximadamente el mismo número de observaciones para estimar con la misma precisión la estatura promedio de los niños de cierta edad de un colegio de La Paz o del país entero.

Si se maneja la muestra adecuadamente, es posible obtener información estadística confiable que atañe a grupos muy numerosos, por medio de encuestas que incluyen sólo a un número reducido de

personas de ese universo.

Por ejemplo, supongamos que se desea estimar mediante una muestra el ingreso medio de los agricultores de una región. Si se desea estimar la verdadera media con una precisión tal que sólo existe una probabilidad de 5% de que el error se exceda de \$b. 1.000.--, el Cuadro 2 indica el tamaño que debe tener la muestra si el número de familias de la región oscila entre 50 y 10.000.000, si el ingreso medio es de \$b. 10.000.-- y la desviación típica es de \$b. 3.162.

CUADRO 2

NUMERO DE OBSERVACIONES NECESARIAS PARA UN PRECISION DADA

(E= 1.000, a = 5%, S= 3.162)

Tamaño de la Población	Tamaño de la Muestra	n/N
N	n	%
50	22	19,56
100	29	29,00
1.000	38	3,80
10.000	40	0,40
100.000	40	0,04
1.000.000	40	0,004
10.000.000	40	0,0004

---

(Para el cálculo de n, ver el Anexo 1)

### 2.3.1 Tipos de Error

se llama error muestral al error que resulta de haber entrevistado sólo a un sub-grupo de la población total. Pero, el error puede venir de otras fuentes adicionales: la persona entrevistada que no entiende la pregunta, o no responde la verdad, el entrevistador que formula mal una pregunta o transcribe erróneamente una respuesta, etc. A este tipo de errores se les conoce como errores no muestrales.

En la práctica de la recolección de datos se diseña la muestra para que el error muestral no sea superior a un 10 por ciento, excediéndose este margen de error sólo en una de cada 10 investigaciones realizadas. (90% de probabilidades de que el parámetro muestral no difiere del parámetro de la población total en más del 10%). En cambio el error no-muestral puede ser mayor debido a las posibles fuentes de error arriba mencionadas, que en Bolivia se agravan por problemas lingüísticos y culturales.

La presencia de estas dificultades parece quitarle toda validez a la investigación que utiliza una muestra como fuente de información ya que además del error muestral del orden del 10% se prevé un error no muestral cuya magnitud puede ser aun mayor. Sin embargo, si examinamos la información que se requiere en el análisis de la planificación regional se constata que no sólo los valores absolutos los que interesan al planificador sino los valores relativos que se emplean para hacer comparaciones. Por ejemplo, no nos interesa conocer el valor del ingreso del campesino en el oeste de Oruro per se sino su relación con el observado en el este o en el sur del mismo departamento. O en forma similar, nos interesará conocer la diferencia en el ingreso de los campesinos con crédito y

sin crédito agrícola en el departamento de Chuquisaca. En todos estos casos en que se hacen comparaciones es irrelevante para fines de la investigación si los campesinos tienden o no a ocultar sus ingresos en el curso de la entrevistas ya que, este error que afecta a todas las observaciones, no afecta a las diferencias observadas entre los sub-grupos que se forman dentro de la muestra para fines de análisis. Por lo tanto, los esfuerzos en una investigación deben dirigirse en dos direcciones: primero, obtener una muestra representativa de la población total, a fin de reducir al mínimo el error muestral; segundo, si hay errores no muestrales, reducirlos en lo posible poniendo énfasis en la consistencia, de modo que el error no muestral no sea sesgado en favor o en perjuicio de ningún sub-grupo dentro de la muestra. En secciones posteriores estudiaremos como controlar el error no-muestral. El error muestral se reduce mediante una muestra adecuada y es el tópico de la siguiente sección.

### 2.3.2 Métodos Probabilísticos y No-probabilísticos de Selección de la Muestra

Se dice que una muestra ha sido seleccionada empleando procedimientos probabilísticos cuando la probabilidad de seleccionar a un miembro cualquiera es conocida. Sólo de esta manera se pueden lograr muestras representativas del grupo total.

En la práctica, se emplean métodos no probabilísticos con frecuencia. Por ejemplo, en una visita de campo los funcionarios o investigadores se detienen en forma caprichosa para visitar a algunos campesinos. Aunque este procedimiento pueda parecer aleatorio a primera vista, no lo es. Se visita generalmente a los campesinos más próximos al camino, o conocidos por el extensionista

o dirigente campesino que hace de guía. Warwick y Lininger presentan este ejemplo:

"Un ejemplo mostrará el peligro de usar procedimientos muestrales no-probabilísticos, o de mezclarlos con procedimientos probabilísticos. Un estudio de mujeres se inició con una selección de hogares en el área del estudio. Sin embargo, apareció un sesgo al abandonar los procedimientos probabilísticos en la selección de las entrevistadas. Si las mujeres de los hogares seleccionados no estaban en casa cuando se presentó el entrevistador, fueron reemplazadas por una vecina. A primera vista, éste puede parecer un procedimiento perfectamente razonable, especialmente en un estudio con un presupuesto limitado y un plazo de terminación. Sin embargo, los resultados mostraron un sesgo marcado en los respondientes tomados para las entrevistas. Las mujeres que trabajan y las que son socialmente activas fueron seriamente sub-representadas en un método muestral que favorecía a las que se quedan en casa durante el día. Es así que una desviación aparentemente inocente de los procedimientos probabilísticos fue muy perjudicial para los resultados del estudio"<sup>1</sup>.

Cuando la probabilidad de selección es igual para todas las observaciones del grupo total (universo), la muestra se llama "auto ponderada" y sus resultados pueden ser directamente extrapolados al grupo total. Con frecuencia, las muestras son auto-ponderadas a nivel de un sub-grupo de la muestra, pero no lo son a nivel de la muestra total. Por ejemplo, para la encuesta agropecuaria de Oruro, se dividió el Departamento en tres regiones y se seleccionaron aleatoriamente 42 entrevistas en la zona este, 34 en la sud-oeste y 34 en la nor-oeste, o sea un total de 110 entrevistas en todo el Departamento. Estimamos que existen 18.400 familias en la zona este, 5.900 en el sud-oeste, y 16.700 en la

-----  
1. Donald Warwick y Charles Lininger. "El Estudio Muestral: Teoría y Práctica", McGraw Hill Book Co., New York, 1975, Pág. 70.

zona nor-oeste.<sup>1</sup> Los datos calculados de la muestra de cada zona, para la familia promedio, pueden ser extrapolados para el total de esa región multiplicando ese promedio por el número de familias de la región. Sin embargo si queremos calcular parámetros para el total del Departamento será necesario ponderar el promedio de cada región por el número de familias estimadas para la misma. En consecuencia, el promedio de las 110 entrevistas no es directamente representativo del Departamento.

### 2.3.3 Técnicas Aleatorias para la Selección de Muestras

Existen dos posibles situaciones para la selección de una muestra: aquellas donde se tiene acceso a una lista completa de las unidades de observación (granjas, hogares, empresas, etc.) y otras en que no se dispone de una lista completa. Empezaremos analizando el primer caso.

#### i) Muestreo Aleatorio Simple

En base al listado de las unidades de observación se asigna un número a cada una de ellas, luego empleando una tabla de números aleatorios se seleccionan las unidades de observación que se incluirán en la muestra. Los números aleatorios pueden ser también obtenidos en algunas calculadoras electrónicas e incluso en las computadoras. Otra alternativa que también resulta satisfactoria es la de seleccionar aleatoriamente una observación y luego incluir en la muestra aquellas observaciones cuyos números están en un

---

1/ Suponiendo que cada segmento censal tiene 50 familias.

intervalo pre-determinado. Por ejemplo, si deseamos una muestra de 10 observaciones tomadas de un grupo de 1000, se puede seleccionar aleatoriamente un observación del grupo entre 1 y 100 (por ejemplo, sale el número 39) y luego se seleccionan las unidades siguientes en intervalos de 100 (las unidades de observación número 139, 239, 339, etc.). Este procedimiento se denomina muestreo sistemático.

ii) Muestreo en Etapas

Otra alternativa muy usada en la práctica consiste en dividir en sub-grupos a la población total de observaciones escogiendo aleatoriamente algunos de estos sub-grupos. Del listado de las unidades de observación de los sub-grupos escogidos se seleccionan finalmente las observaciones a ser incluídas en la muestra, de esta manera se requieren listados completos sólo para los sub-grupos pre-seleccionados. Este mismo proceso puede realizarse en más de dos etapas.

La división del grupo total en sub-grupos tiene aplicación práctica en dos casos: primero, cuando se realiza muestreo por áreas y segundo, cuando se estratifica la muestra.

a) Muestreo por Area

En este caso la división en sub-grupos es generalmente una división en áreas geográficas. Lo único que se requiere es la existencia de mapas con el detalle suficiente como para poder definir las divisiones. De esta manera se evita la necesidad de tener un listado completo del total de la población sino únicamente un listado que incluye a las unidades de observación localizadas en las áreas geográficas seleccionadas. Por

ejemplo, en la encuesta agropecuaria de CORDEOR se seleccionaron un total de 10 segmentos censales (del Censo Nacional de Población y Vivienda 1976) de un total de 820 que tenía el departamento. Luego se realizó un listado completo de los jefes de familia con vivienda permanente en cada uno de los 10 segmentos seleccionados y finalmente se seleccionaron aleatoriamente las familias de cada segmento para ser entrevistadas.

b) Estratificación de la Muestra

En algunos casos, se desea conocer los parámetros que corresponden a sub-grupos de la población. Por ejemplo, el ingreso de los campesinos con menos de 5 hectáreas de terreno. En estos casos se divide a la población en sub-grupos de acuerdo al parámetro pertinente, para luego proseguir con el proceso de muestreo sólo en el sub-grupo que interesa. La dificultad de este procedimiento es que requiere contar con un listado completo de la población con información sobre el parámetro por el cual se estratifica la muestra. Es decir, en el caso de los agricultores con menos de 5 hectáreas de terreno, se necesitaría un listado completo de todos los campesinos indicando el número de hectáreas de terreno que poseen, para luego seleccionar aleatoriamente a los que se ha de entrevistar de entre los que tiene menos de 5 hectáreas de terreno.

En la práctica, este tipo de estratificación no es recomendable por su alto costo. Si se desea, de todas maneras, información respecto a algunos sub-grupos de la población, se puede recurrir a otros mecanismos que explicaremos en la sección 4.3.

#### 2.3.4. Tamaño de la Muestra <sup>1/</sup>

El tamaño de la muestra está relacionado con tres factores que son: el margen de error muestral aceptable, la probabilidad de que el error sobrepase ese margen, y el grado de variabilidad que tiene el parámetro en la población que se estudia.

Es importante recalcar que, normalmente, el tamaño de la muestra no está relacionado con el tamaño del grupo total que se estudia.

De los tres factores arriba mencionados, los dos primeros son estipulados por el investigador siendo el tercero una característica del parámetro que se mide y de la población, característica que se mide con la desviación típica.

En la práctica, la desviación típica no se conoce en el momento de iniciar una investigación. Por esta razón cualquier cálculo que se haga del tamaño de la muestra es generalmente solo una estimación a priori.

La pregunta importante aquí es cuál es el tamaño que debe tener una muestra para proporcionar una precisión estadísticamente aceptable respecto al error muestral. No es posible dar una respuesta simple que sea válida para todos los casos pero sí es posible establecer un tamaño mínimo por debajo del cual es difícil obtener estimados aceptables. Este número es 30 observaciones, es decir 30 para cada sub-grupo para el que se desee calcular el parámetro. Por ejemplo, si desee tener un parámetro representativo del agricultor medio del Departamento de Oruro, necesitaré un mínimo de 30 entrevistas. Si requiero ese parámetro para tres regiones del Departamento el

---

<sup>1/</sup> Ver el Anexo 1 para un análisis más formal

mínimo se eleva a 90 entrevistas. Si además deseo estratificar a los campesinos en 5 sub-grupos de acuerdo a la tierra que poseen, en cada una de las tres regiones, necesito un total de 15 muestras, o sea 450 observaciones.

Por esta razón en el momento de hacer la especificación de datos es necesario establecer qué tipo de estratificación se pretende aplicar a los datos ya que ésto afecta el diseño de la muestra, como se ha demostrado.

Las muestras con menos de 30 observaciones resultan con niveles de confianza muy bajos, además de incluir características estadísticas indeseables. Warwick y Lininger indican que:

"Es importante observar que, excepto para la muestras pequeñas de 30 o menos, el atributo que se estima (como por ejemplo, el ingreso) no necesita estar distribuído normalmente en la población para que la distribución de la medias muestrales se aproxime a la curva normal. Esto significa que no es necesario que el ingreso esté distribuído normalmente en la población para que la distribución de las medias muestrales del ingreso se aproxime a la curva normal."

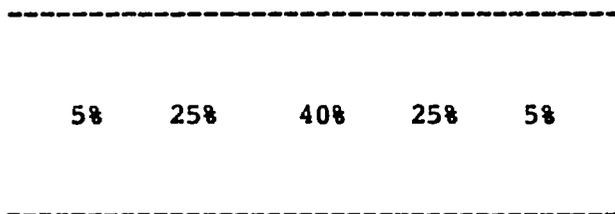
A fin de mantenerse por encima del mínimo de 30 y tomando en cuenta que algunas entrevistas son rechazadas en el campo o durante la limpieza de datos, es conveniente tomar un número de 40 observaciones aproximadamente para cada sub-grupo en la muestra. Por ejemplo, en la encuesta agropecuaria de Oruro se diseñó la muestra con 36 observaciones en las zonas nor y sud oeste y 48 en el este. Al terminar el análisis estos números se redujeron a 42 en el este, 32 en el sud-oeste, y 33 en el nor-oeste.

### 2.3.5 Diseño de la Muestra para las Comparaciones

El instrumento más poderoso con que cuenta la estadística para el análisis de resultados es la simple comparación entre sub-grupos o estratos que difieren entre sí en alguna característica que interesa al analista, por ejemplo, granjas pequeñas, medianas, y grandes; agricultores con riego y sin riego; granjas antes y después de la apertura de un camino, etc.

#### i) Encuestas Simples de Sección Cruzada

Son encuestas que se realizan en un momento en el tiempo. Una vez llevada a cabo la encuesta se estratifican las observaciones de la muestra según la característica que se desea analizar. Por ejemplo, la encuesta agropecuaria de CORDEOR estratifica a los agricultores en tres sub-grupos según el "Tamaño de la Explotación". La encuesta agropecuaria realizada en CORDECH ordena a los agricultores de la muestra según su ingreso para luego ignorar el 5% de las observaciones en ambos extremos y el 40% que está en el medio, quedando dos sub-grupos de 25% entre los cuales se efectúan comparaciones.



En este caso, para tener 40 observaciones en cada uno de los dos sub-grupos que se toman para comparar agricultores de "altos

ingresos" con los de "bajos ingresos" el tamaño de la muestra debía ser de por lo menos 160 observaciones.

i1) Encuestas con Grupo de Control

Son encuestas en que dos o más muestras seleccionadas reciben el mismo cuestionario. El objetivo es analizar relaciones causa-efecto en condiciones cuasi-experimentales. La situación ideal es que los dos grupos sean idénticos en todo excepto por la característica que se desea estudiar. Por ejemplo, la encuesta de crédito de CORDECH seleccionó en forma aleatoria una muestra de 41 agricultores que eran prestatarios del Banco Agrícola, luego se seleccionó un segundo grupo de control, de agricultores de la zona que no eran prestatarios del BAB. Sin embargo, este segundo grupo no fue seleccionado al azar, sino que se buscó un "par" con características similares, para cada uno de los agricultores "con crédito". Al grupo de agricultores "con crédito" se lo denomina grupo experimental, y al segundo grupo formado por sus pares, grupo de control.

Al seleccionar el grupo de control se debe tener mucho cuidado al decidir que características deben ser similares en los pares. En el caso de CORDECH nos preguntamos ¿Con qué parámetros puede el crédito tener una relación causa-efecto?: empleo de insumos mejorados, rendimientos, combinaciones de cultivos, ingreso, etc. Al seleccionar pares se aceptaba que tuvieran divergencias en esas características, en cambio, se exigía que los pares fuesen vecinos próximos (para captar así factores climáticos, suelos, acceso a mercados) y muy similares en características sin relación directa de causa efecto con el crédito agrícola como educación del jefe de familia, fuerza laboral de la familia, y tamaño de la explotación.

Se debe evitar el incluir demasiadas características en los requisitos de similitud entre pares ya que así se corre el peligro de eliminar las diferencias mismas que el análisis pretende capturar comparando el grupo experimental con el grupo de control.

iii) Encuestas Longitudinales Simples

En este caso se encuesta al mismo grupo dos o más veces a lo largo del tiempo. Este tipo de encuestas son de utilidad para establecer comparaciones "antes" y "después" de llevarse a cabo una intervención que afecta al grupo estudiado y cuyo efecto se desea evaluar. Un ejemplo sería un grupo de agricultores antes y después de la construcción de un sistema de riego o la recepción de crédito, etc.

Este tipo de comparaciones tiene una dificultad seria. A menudo se presentan otros cambios exógenos en el tiempo que influyen en los resultados del experimento. Por ejemplo, cambios en los precios de los productos, cambios en los precios de los insumos, factores climáticos (heladas, granizadas, etc) y otros. La presencia de estos factores exógenos hace muy difícil el establecer relaciones causa-efecto para el parámetro que se desea examinar. Por ejemplo, una elevación en el ingreso de un grupo de campesinos, percibido en el lapso de un período de tres años durante el cual han entrado en un programa de crédito agrícola. Si el ingreso aumentó, qué porción de ese incremento se debe al crédito recibido?, y qué porción corresponde a simples cambios climáticos, de precios, etc.?

iv) Encuestas Longitudinales con Grupo de Control

A fin de solucionar las dificultades que se presentan en las

encuestas longitudinales aplicadas a un grupo experimental en el tiempo, se incluye un grupo de control seleccionado en forma de pares (ver sección 2.3.5. ii) que también se encuesta a lo largo del tiempo. De esta manera se combina el análisis longitudinal con el grupo de control, lográndose así un análisis estadístico más riguroso que permite percibir con mayor precisión las relaciones causa-efecto de los parámetros que se estudian.

## 2.4 Diseño de los Instrumentos

Existen diferentes maneras de obtener la información que se requiere, una vez que las unidades de observación ya han sido seleccionadas. Es posible clasificar estas maneras en tres grupos: rememoración de la persona que se entrevista, registros correspondientes a la unidad de observación, y medición directa.

### 2.4.1 Rememoración

Aquí se recurre a la memoria del entrevistado para obtener información. Al diseñar un instrumento o formulario que requiere respuestas proporcionadas por una persona se debe tener en cuenta que las personas generalmente recuerdan eventos y no cifras aisladas y que la precisión de la memoria depende de tres factores: número de eventos que se deben recordar, tiempo transcurrido desde el evento al momento de la entrevista, y finalmente importancia del evento.

Por esta razón es muy difícil pedirle, por ejemplo, a un agricultor que recuerde el número de días que trabajó durante el último año. Sin embargo, si trabajó sólo unos meses, y lo hizo en forma regular podría recordarlos. Por ejemplo, si trabajó durante 4 meses, 22

días por mes, todo lo que necesita es multiplicar 4 x 22 para tener el total. En este caso el número de eventos que ha tenido que recordar es de sólo dos.

Cada vez que se inserta un ítem de datos en un formulario o en el formato de un estudio de caso para su respuesta a partir de una rememoración, el que planifica el estudio debe identificar el número aproximado de ítems que será requerido, o el nivel de detalle que se tendrá que recordar. Si el número es alto, el ítem debe ser examinado cuidadosamente para ver si no se puede emplear una medición directa o hacer llevar un registro en lugar de emplear la rememoración.

En algunos casos la persona que diseña el formulario tiene la tentación de resumir las preguntas y presentarlas en forma de cuadros clasificados en columnas y filas de modo que cada casilla contiene un ítem de información. Si bien esto tiene la ventaja de emplear menos papel en el cuestionario, el resultado es que el encuestador muchas veces no formula las preguntas que requiere el cuadro. Por su parte, el supervisor, al revisar el formulario y encontrar un gran número de casillas en blanco, no podrá comprobar si las casillas están en blanco porque las preguntas no se plantearon o porque las respuestas eran cero o no correspondían.

Cuando se hacen preguntas de opinión u otras que tienen una respuesta no numérica, es conveniente escribir ya en el formulario las respuestas más frecuentes que se esperan, asignando a cada una un número de código y dejando uno o más códigos libres para registrar respuestas abiertas que no encajan en las respuestas cerradas pre-codificadas. Caso contrario, si se dejan completamente abiertas las respuestas, después de terminar el

trabajo de campo, el planificador deberá revisar las respuestas obtenidas para clasificarlas en grupos y asignar a cada grupo un código.

La distancia en el tiempo del evento a ser recordado y el tipo de dato solicitado afectarán la calidad de la respuesta. Si se considerara necesario tener estimados buenos de lo que hace la gente con su tiempo, podrá preguntarse sólo acerca de los últimos días si se ha de emplear el método de rememoración. Para obtener una visión del uso del tiempo a lo largo de un período de un año se necesitarán visitas frecuentes para obtener una rememoración razonable.

Los diferentes eventos entran en la memoria con diferente intensidad, y dicha intensidad depende de muchos factores culturales y económicos. Las muertes en la familia pueden violar la regla de rememoración "de eventos recientes"; si se pide al encuestado recordar el número de defunciones en la familia en los últimos diez años, la importancia del evento sobrepasará a la tendencia del tiempo a borrar la memoria. Es obvio que la fecha del contacto entre el entrevistado y los encuestadores es un factor que debe ser acomodado para reducir tanto el número de eventos a ser recordados como lo reciente de los eventos.

Aunque, por ejemplo, es imposible obtener datos de empleo en base a un rememoración anual, mediante visitas semanales o bisemanales puede reducir el tiempo transcurrido y el número de estos eventos, de forma que se puedan recordar los datos a ser obtenidos. Sin embargo, estas visitas múltiples son muy caras, y dado el presupuesto de la mayor parte de los esfuerzos del análisis de desarrollo rural en el Ministerio de Planeamiento y/o las

Corporaciones de Desarrollo Rural, es virtualmente imposible a cabo dichas investigaciones de visitas múltiples.

#### 2.4.2 Mantenimiento de Registros

Otro método para recopilar datos es obtenerlos de registros mantenidos por la empresa, hogar, u otra unidad de observación encuestada. Si estos registros se llevan normalmente, se facilitará el proceso, pero también surgirán problemas especiales. Si, por ejemplo se na de medir el ingreso, es importante que todos los hogares y empresas estén sujetos a las mismas convenciones de contabilidad, métodos de depreciación equivalentes, y las mismas definiciones de ingreso, etc. Es improbable que aunque existan registros disponibles éstos sean suficientemente similares en su estilo de contabilidad y conceptos como para ser empleados en el análisis de desarrollo rural.

Las dificultades que presentan los registros son numerosas: el analfabetismo limita radicalmente el tipo de personas entrevistadas con este sistema; el cumplimiento en los períodos entre una y otra entrevista es difícil de controlar. Hay algunas situaciones limitadas en que puede ser útil; por ejemplo, si el problema central es recordar una serie de eventos que son demasiado numerosos para ser cubiertos con una rememoración anual, los sistemas simples de mantener registros que escapan el problema de analfabetismo pueden ser empleados. Se puede entregar una pequeña caja para que el entrevistado la coloque en un lugar prominente en su hogar o negocio y deposite pequeños palitos (también suministrados por el encuestador) cada vez que trabaja un día casi entero en la granja o negocio.

### 2.4.3 Medición y Observación Directa

Los esfuerzos dirigidos al bienestar, con frecuencia tienen una mejora en la nutrición, como uno de los objetivos de control. La nutrición se puede medir ya sea como una mejora en la dieta, una mejora en la salud, en la forma de un mejor crecimiento físico o disminución de enfermedades.

El control del crecimiento físico es un ejemplo clásico de una situación donde la observación directa es el único método confiable para obtener datos exactos. La medición de estatura, peso y circunferencia del brazo son indicadores comunes del estado nutricional de los niños. Es razonablemente simple y barato obtener datos, pidiendo al personal de campo de la encuesta que pese y mida a los niños en los hogares entrevistados. La edad del niño es el único dato de rememoración necesario. Como la medición en los niños de más de 6 años de edad es sensible sólo a diferencias en años, la rememoración no representa un problema principal. Para los niños de menos de 5 años de edad, sin embargo, la edad mensual es necesaria y existe considerable evidencia de que la rememoración a este nivel de detalle es generalmente bastante pobre. Estos son ejemplos sencillos de técnicas de medición directa.

### 2.4.4 Experiencia Lograda en las CDD's

Los formularios de encuesta empleados por el proyecto en las CDD's han recurrido a la rememoración como camino primario para obtener información, excepto por la nutrición donde se empleó la medición directa del perímetro braquial de los niños. (No se midió el peso y la estatura por carecer de balanzas de campo adecuadas y la

necesidad de no alargar la entrevista excesivamente).

Debido a la dificultad de rememoración, no se incluyeron algunas preguntas como las relacionadas con el empleo de mano de obra familiar en las actividades agrícolas.

Además de tomar en cuenta los factores ya mencionados, al diseñar el formulario se debe emplear el lenguaje de las personas a entrevistarse de modo que el entrevistador formule las preguntas empleando las palabras del instrumento escrito.

Cuando se entrevista a personas que hablan otro lenguaje como Quechua o Aymara la solución ideal es el empleo de formularios escritos en esos idiomas. Por falta de recursos, en las encuestas realizadas recurrimos al uso de encuestadores bilingües con resultados aparentemente satisfactorios.

## 2.5 Verificación de Campo

### 2.5.1 Verificación de la Muestra

Antes de iniciar el trabajo de campo es conveniente verificar si la información en base a la cual se diseña la muestra y se seleccionan las unidades de observación es fidedigna. Por ejemplo, si se trata de un listado completo de pequeñas industrias y artesanos en una ciudad, es recomendable verificar al menos unos cuantos casos para ver si las listas recibidas son reales.

Cuando se aplica un muestreo por áreas se debe tener especial cuidado en dos puntos, primero, si los límites geográficos que aparecen en los mapas son identificables en el terreno, y en

segundo lugar, si las áreas geográficas contiene el número de unidades de observación que dicen tener.

En las encuestas realizadas en las CDD's se observó que los límites de los segmentos empleados (del Censo Nacional de Población y Vivienda 1976) eran bastante precisos pero no así el número de familias que debían contener. En muchos casos, la población de los segmentos resultó muy inferior a la establecida por el censo. Algunos segmentos que debían incluir unas 50 familias tenían solo 12 ó 15, sin existir señales de un drástico descenso de población en los últimos años que pudiese explicar una diferencia tan grande.

#### 2.5.2 Verificación del Instrumento

Una vez diseñado el formulario de la encuesta, es indispensable someterlo a una prueba de campo. En la práctica se observa que a pesar de la experiencia y el cuidado de los responsables del diseño del formulario, al utilizarlo en el campo se encontrarán detalles y deficiencias que deben ser corregidos antes de reproducir el instrumento para su utilización en las entrevistas.

En el momento de reproducir el formulario final que se empleará en la encuesta se debe proceder con la pre-codificación. Este procedimiento consiste en asignar un número a cada pieza de información que luego se extraerá del formulario. Por ejemplo:

4. ¿Cuál es la edad del Jefe de Familia?

---

35

\_\_\_\_\_ años

---

En este ejemplo, la información sobre la edad del jefe de familia se identificará por el campo número 35, que aparece en la esquina superior derecha de la casilla.

### 3. Trabajo de Campo

#### 3.1 Selección y Entrenamiento de los Encuestadores

Con el propósito de reducir en lo posible el error no-muestral se debe hacer un esfuerzo especial para estandarizar el proceso de recolección de datos, o en otras palabras, hacer que todas las entrevistas sean conducidas del mismo modo para que las diferencias en las respuestas se deban únicamente a diferencias entre los entrevistados y no provengan de otras causas. Existen tres vías para homogenizar este procedimiento: emplear un formulario escrito, entrenar al personal de campo y seleccionar adecuadamente el personal de campo. El primer punto ya fue elaborado en la sección anterior; ahora trataremos de los otros dos.

Para la selección de encuestadores es recomendable emplear personas que además de poder comunicarse en el lenguaje de las personas entrevistadas y tener suficiente nivel escolar como para leer el formulario y escribir correctamente las respuestas que reciben, no estén relacionadas profesionalmente con la material de la encuesta.

Por ejemplo, no es conveniente emplear agentes de extensión agrícola en una encuesta agropecuaria debido a las ideas preconcebidas que tienen en cuanto a las mismas preguntas que formulan. Estos preconceptos se traslucen en el tono o los términos empleados en la entrevista y llegan a influir en las respuestas recibidas.

Una vez seleccionados los entrevistadores es preciso entrenarlos en el uso del formulario. Se deben emplear uno o dos días de entrenamiento para dar a los encuestadores las definiciones de los conceptos que incluye el formulario y el modo como funciona. Una vez completado este entrenamiento "teórico" se debe dar a cada encuestador la oportunidad de realizar por lo menos dos entrevistas de entrenamiento en el campo, en presencia del entrenador o supervisor. Una vez realizadas las entrevistas de entrenamiento, se debe reunir a los encuestadores para que el grupo conozca las dificultades y dudas encontradas por cada uno y qué solución se les debe dar. En este sentido, es importante que cualquier interpretación que se dé a alguna pregunta sea homogénea para todos los encuestadores.

Normalmente el personal de campo es contratado sólo temporalmente para efectuar un trabajo específico. Es importante que el tipo de pago además de un monto fijo por día, incluya un monto por encuesta realizada y aceptada posteriormente por el supervisor. De esta forma es posible controlar la calidad del trabajo de campo realizado por los encuestadores.

El costo del trabajo de campo incluyendo la impresión de formularios, supervisión, transporte, etc. por entrevista completada varía según los casos. (Fue de \$b. 500 por entrevista en la encuesta agropecuaria de Oruro, y \$b. 180 por entrevista en la encuesta agropecuaria de Chuquisaca).

### 3.2. Supervisión

La supervisión constituye un punto clave e incluso es quizás el punto más crítico para el éxito de un trabajo de recolección de

datos. Una vez que los formularios están siendo procesados, poco es lo que se puede hacer para enmendar errores. Los errores deben ser corregidos en el trabajo de campo y no en la oficina, y la responsabilidad de hacerlo recae sobre el supervisor.

La relación ideal es de un supervisor por cada tres o cuatro encuestadores, pero en ningún caso más de cinco. El supervisor debe revisar cada uno de los formularios lo antes posible y no esperar uno o dos días para hacerlo.

Contrariamente a lo que se dijo para los encuestadores, es necesario que el supervisor tenga conocimiento de la materia que trata la encuesta. En una encuesta agropecuaria, por ejemplo, no se requiere que el supervisor sea un agrónomo, pero debe conocer el tipo de vida de los agricultores, rendimientos posibles en la zona, precios aproximados, niveles de salarios, cultivos, etc. El supervisor debe ser capaz de captar todas las contradicciones o absurdos recibidos en una encuesta para luego enviar nuevamente al encuestador para corregir o completar la entrevista.

Es importante que los encuestadores estén conscientes de que se les exigirá retornar a completar o corregir una entrevista aunque el entrevistado esté lejos. Sobre todo durante los primeros días el supervisor no debe tener reparos en enviar nuevamente a los encuestadores por lo menos en un 30% de los casos, o sea una de cada tres entrevistas. Al final de la jornada se debe aprovechar de recibir comentarios de las dificultades encontradas, cuando el grupo está reunido. De esta manera se da mayor fuerza a la homogeneización del procedimiento que siguen los encuestadores en la realización de las entrevistas.

El supervisor debe también verificar que las respuestas recibidas en el formulario sean legibles y puedan ser interpretadas, ésto es particularmente importante en el caso de unidades de superficie, peso, y volumen que manejan los agricultores. En el momento de procesar la información, será necesario que se emplee una misma unidad en todos los casos. Por ejemplo, cuando se toman cifras de tamaño de la explotación y extensión cultivada, todas deben estar expresadas en hectáreas, lo mismo con la producción, que puede ser expresada, por ejemplo, en kilogramos. Dado que el agricultor entrevistado en muchos casos maneja distintas unidades de área, peso, y volumen, el entrevistador debe anotar la respuesta en dichas unidades, dejando al supervisor la responsabilidad de realizar los cálculos de transformación a unidades standard. Es obvio que cuando se manejan unidades cuyo valor no está bien definido y cambia según la zona o el producto al que se aplica, el encuestador deberá indicar en el formulario mismo cual es su equivalencia con unidades standard.

Existen dos asuntos sobre los cuales debe el supervisor tomar decisiones en el campo: rechazo o ausencia de unidades de observación seleccionadas para la muestra, y deficiencias en la delimitación de segmentos o áreas.

### 3.2.1 Rechazos o Ausencias de Unidades de Observación

El peligro que existe aquí no está en la pérdida de tiempo o el incremento de costo de la investigación debido a rechazos o ausencia de las personas a entrevistar, sino en el hecho de que el fenómeno se presente en forma no aleatoria. Por ejemplo, podría suceder que todos los campesinos que también se dedican a la compra y venta de animales, o tienen tierras en un proyecto de colonización

estén ausentes en determinadas fechas, en cuyo caso una encuesta realizada durante su ausencia no tendrá la representatividad debida.

En caso de haber una ausencia el supervisor debe averiguar si la ausencia es permanente o temporal. Si es permanente, se puede reemplazar tomando a otra persona (seleccionada al azar, por supuesto). Si la ausencia es temporal, es recomendable regresar para realizar la entrevista aunque el costo de dicha entrevista adicional sea elevado.

El supervisor también es responsable de definir, en los casos en que haya duda, si una entrevista determinada corresponde o no corresponde a la definición de la unidad de observación. Por ejemplo, en la encuesta agropecuaria del Departamento de Oruro se encontró algunas familias campesinas que tienen su vivienda permanente en un sitio, pero también habitaciones y pequeños corrales construidos en otros puntos situados en segmentos censales contiguos. En estos casos, al efectuar el listado del segmento se tomó como referencia la ubicación de la vivienda en la que habitan la mayor parte del tiempo.

En cuanto a los recnazos, a fin de evitarlos se debe tomar contacto con las autoridades locales y comunales antes de iniciar el trabajo de campo, para explicar el tipo de encuesta y los objetivos que tiene. De la misma forma, al contactar con las personas a ser entrevistadas el encuestador debe dar una explicación clara y plausible del motivo de la investigación indicando qué institución lo realiza. La experiencia en un alto número de investigaciones realizadas en América Latina nos indica que el porcentaje de recnazos no pasa del 5% normalmente.

Si el número de rechazos es mayor al 5%, la causa estará seguramente en un formulario con preguntas mal estructuradas, o en que la institución que realiza la investigación no es bien vista por los entrevistados, o que la explicación de los motivos de la investigación no es plausible.

### 3.2.2 Deficiencias en la Delimitación de Areas

A pesar de haberse verificado si los límites geográficos de los mapas son identificables en el campo, en el curso del trabajo se encuentran casi siempre algunos errores. Por ejemplo, después de identificar los límites de un segmento censal (ríos, caminos y sendas, cumbres de montañas, etc) se encuentra que alguna comunidad que aparece dentro del segmento en el mapa, en realidad está fuera de él, o que al contrario, el segmento incluye comunidades que en el mapa aparecen fuera de él. En este caso es necesario decidir entre dos alternativas: dar prioridad a los límites indicados en el mapa (segmento cerrado) o al listado de las comunidades que el mapa muestra dentro del segmento (segmento abierto). No importa que alternativa se escoja, lo importante es ser consistente en todos los casos para mantener la aleatoriedad de la muestra.

## 4. Análisis

### 4.1 Entrada y Limpieza de Datos

Es conveniente asignar un número a cada trozo de información de campo que se ha de extraer de la encuesta. Por ejemplo, si la encuesta tiene información de edad, sexo y nivel escolar del jefe de familia, se requieren 3 campos (edad, sexo, educación). Si esta información abarca a todos los miembros de la familia, se debe

estimar el número máximo de miembros que pueden haber en una familia. Si éste número es de diez miembros, se requieren 30 campos para registrar su información de edad, sexo, y educación. Las familias con menos de diez miembros tendrán vacíos los campos correspondientes.

No toda la información que contiene el formulario debe ser utilizada y procesada. Por ejemplo, se preguntan los gastos efectuados en la compra de cada tipo de herramientas, y el formulario incluye un listado completo de herramientas. Sin embargo, lo que interesa puede ser el gasto total que se reduce a ocupar sólo un campo.

Existen muchas alternativas distintas para la introducción de los datos en el computador; aquí mencionaremos únicamente las tres alternativas que son más frecuentes en Bolivia.

El modo más tradicional emplea tarjetas IBM en cuyo caso es necesario definir a qué campo se asignan las columnas de que dispone cada tarjeta. Los computadores modernos emplean también un sistema equivalente que introduce los datos a una cinta magnética en lugar de la tarjeta. Esta variante puede incluir un programa que presenta en una pantalla las mismas preguntas que aparecen en el formulario, de esta manera se facilita la labor del operador que transcribe los datos de los formularios de encuesta al computador.

Cuando los formularios obtenidos contienen datos en casi todos sus campos estos dos métodos de transcripción son adecuados, pero si la mayor parte de los campos están vacíos el operador empleará mucho tiempo. En este caso es más conveniente adoptar un método de transcripción que requiere que el operador marque sólo los campos

que tienen datos introduciendo para cada uno el número de campo y el dato correspondiente.

Durante la transcripción es posible que el operador cometa errores. Algunos de estos errores puede percibirlos el operador mismo al introducir los datos (o las tarjetas) en el computador gracias a los programas de verificación que explicaremos en la sección 4.2.1. Sin embargo para eliminar errores en esta etapa es necesario realizar dos transcripciones independientes que luego se comparan, o en su defecto comparar cada uno de los formularios con el listado inicial de los campos obtenido del computador.

Una vez transferida la información a un computador, o manualmente a un cuadro, se inicia la limpieza examinando un listado donde cada campo ocupa una columna y cada entrevista una fila. De esta manera una simple revisión visual permite identificar respuestas absurdas o inaceptables que deberán corregirse confrontandose con el formulario de la entrevista respectiva.

Una vez corregidos todos los errores de transcripción que se han identificado, se encuentran a menudo incongruencias que aparecen en el formulario mismo y no fueron percibidas por el supervisor. Si estas dificultades son serias, se hace necesario descartar toda la entrevista para fines de análisis. Cualquier corrección en base al promedio de las otras entrevistas u otro procedimiento similar no es recomendable ya que afecta la validez estadística de los resultados.

Un posible camino, que ha dado buenos resultados en la encuesta agropecuaria de Chuquisaca consiste en calcular un parámetro o variable para cada encuesta que sea importante para el análisis que

se realiza (por ejemplo, ingreso de la granja) y luego descartar todas las entrevistas en los extremos (5% a cada lado) en base al argumento de que aquellos que indican un ingreso demasiado alto o demasiado bajo se deben a un error, o son casos atípicos. El error en cuestión puede ser una unidad de observación incorrectamente seleccionada (una familia que no es campesina), respuestas mal registradas, o simplemente entrevistados que mienten a sabiendas.

A partir de los datos primarios de una encuesta es posible calcular otros parámetros que resultan de una combinación de ellos (suma o multiplicación de campos). A estos parámetros calculados se les llama variables. Por ejemplo, si sumamos todos los campos que contienen cifras de gastos de producción, se obtiene una variable que corresponde a "Gastos Totales de Producción", etc.

El mismo examen visual efectuado previamente con los campos es aplicable a las variables para continuar con el procedimiento de limpieza de datos.

#### 4.2 Procesamiento de Datos

Algunas investigaciones no requieren el cálculo de muchas variables a partir de los campos; si además, no se trata de una encuesta con muchas unidades de observación (no sobrepasa unas 150), el procesamiento y limpieza de datos puede realizarse en forma manual; caso contrario es recomendable emplear un computador.

Para solicitar los servicios de un centro de computación es preciso definir el trabajo que se desea. El primer paso es la definición de los campos de la encuesta. Si el formulario empleado estaba pre-configurado solo se requiere indicar para cada campo lo siguiente:

1. Si es literal o numérico (si la información que contiene son letras o números)
2. Cuántas letras o números puede contener como máximo.

Si el campo es numérico se debe añadir:

1. Rangos máximos y mínimos
2. Si lleva decimales, y cuantos

El segundo paso es el de la definición de variables que se desean calcular a partir de los campos. Por ejemplo, se calculará el ingreso de la finca sumando los campos correspondientes al valor de la producción de cultivos, valor de la producción animal, valor de los productos elaborados y procesados, etc.

Finalmente, es necesario preparar un diseño de los cuadros de salida deseados. Aquí se debe precisar qué columnas y qué filas tiene cada cuadro o lo que es lo mismo, qué estratificaciones se desea y para qué campos o variables.

#### 4.2.1 Programas de Verificación

Acompañando a la definición de variables se pueden incluir algunas definiciones de relaciones entre campos y variables para verificar la consistencia de los datos. Por ejemplo, se acostumbra a incluir en el formulario la superficie total y la superficie clasificada según uso de la tierra (cultivos, pastos, barbecho, descanso, etc). La suma de los campos de uso de la tierra debe ser siempre igual al campo de superficie total.

También se puede especificar para cada campo un rango de valores posibles precisando límites superiores e inferiores.

#### 4.3 Análisis Estadístico

El análisis que se realiza para la planificación regional examina el impacto potencial de proyectos, programas, acciones, y actividades que tienen lugar especialmente en el área rural. Este tipo de análisis suele ser de dos clases, el primero un esfuerzo exploratorio general, y el segundo una evaluación del impacto de una acción o intervención específica.

Para estos dos tipos de análisis proponemos un método que se apoya fundamentalmente en comparaciones para llegar a conclusiones a partir de los datos de campo obtenidos mediante el muestreo.

##### 4.3.1 Comparaciones

Las comparaciones experimentales son casi instintivas para cualquier adulto, están en el fondo del proceso de aprendizaje humano.

Comparamos el pan con levadura con el que no tiene levadura y aprendemos de la experiencia observada cómo hacer un buen pan. El poder de las simples comparaciones ha sido oscurecido por los abundantes y complejos procedimientos analíticos y estadísticos que han evolucionado como un modo de hacer comparaciones más precisas, más exactas, y más amplias. El método que proponemos está basado en una simplificación, manteniendo lo esencial y más necesario de estos procedimientos estadísticos e ignorando el resto. Lo esencial del método de análisis está en obtener datos de alta

calidad y estadísticamente confiables estructurados con vista a la realización de comparaciones entre diferentes grupos.

La estructuración de comparaciones para verificar hipótesis de políticas, para descubrir y comprender cuál es el efecto de ciertas intervenciones o acciones y explorar como mejorar su funcionamiento, no es un proceso fácil. Existe una interacción entre lo que ya conocemos y qué preguntas haríamos y qué hipótesis verificaríamos. Cada pieza nueva de información, cada hipótesis verificada, sugieren una nueva luz con la cual interpretar la siguiente hipótesis, un nuevo modo de mirar el problema, un modo distinto de hacer las próximas preguntas, o incluso nuevas preguntas a ser formuladas. El análisis del desarrollo rural, si es efectivo, debe tener un modo de tratar con esta característica de interrelación en espiral de la naturaleza del análisis. No es un proceso simple lineal o de un sólo intento en el cual un sólo conjunto de hipótesis son formuladas y verificadas. Dicho abordaje sería difícil y extremadamente costoso, exigiendo grandes esfuerzos de análisis para poder clarificar áreas relativamente simples de desarrollo de políticas.

El método de comparaciones experimentales es factible de ser aplicado incluso cuando no hay un experimento explícito. La variación en las condiciones del mundo real crea cuasi-experimentos que un proceso de recopilación de datos cuidadosamente estructurados capturará para el análisis. Cuando el esfuerzo de análisis na de evaluar el impacto de una intervención específica, la estructuración de estos grupos de comparación será más formal, requiriendo grupos de control antes y después de la intervención, etc. Cuando el análisis del desarrollo rural es exploratorio, aunque algunas de las comparaciones serán estructuradas en la etapa

de especificación de datos, la mayor parte será formulada durante la etapa de análisis, de modo que los datos deben ser recopilados con vista a suministrar una muestra con suficiente amplitud de tamaño y contenido cuantitativo para permitir una flexibilidad razonable. Es por ésto que se pone tanto énfasis en este documento en la etapa de especificación de datos. El analista en la especificación de datos debe ser capaz de visualizar los tipos de comparaciones que se desean hacer durante el análisis aunque no siempre pueda visualizar con anticipación cada una de las comparaciones mismas que eventualmente se realizarán durante el proceso de análisis.

#### 4.3.2 Etapas en la Formulación de Comparaciones

El método de comparación es casi intuitivo y sigue un procedimiento simple de 4 pasos:

- a) 1ro. Agrupar los datos para clarificar el asunto, responder a la pregunta, o verificar la hipótesis.
- b) 2do. Sub-agrupar los datos para controlar por distorsiones en atribución.
- c) 3ro. Examinar las diferencias entre los grupos buscando cifras con un significado para el análisis del desarrollo rural.
- d) 4to. Estimar la significancia estadística de las diferencias observadas entre los grupos.

El primer paso involucra el diseño de grupos en los datos para verificar una importante hipótesis del análisis del desarrollo rural o responder una interrogante importante de políticas. Al proseguir el análisis hay un proceso de realimentación en el cual se llega a una delimitación de líneas importantes de investigación y una clarificación de los problemas y posibles soluciones. Un ejemplo podría ser la pregunta: ¿hasta qué punto la educación afecta el ingreso? Si los datos contienen un buen indicador o escala de educación, podríamos agrupar a las personas (y por lo tanto los datos respecto a ellas) en 4 grupos iguales o cuartiles, según su nivel escolar. Con los datos estratificados en estos 4 grupos se podrían estimar los indicadores de ingreso para cada grupo. El primer paso es la formulación de un grupo que se dirigirá a un asunto, responderá a una interrogante, verificará una hipótesis. La relación entre lo que estamos preguntando y la estructuración de grupos es vital.

El segundo paso involucra la construcción de sub-grupos o grupos cruzados para controlar la atribución. Si por ejemplo, sospechamos que la edad está confundiendo nuestros resultados en la comparación anterior (es decir sospechamos que la educación puede afectar el nivel de ingresos en las personas más jóvenes que no han tenido actividades educacionales completas) deseamos estratificar los grupos por edad de modo que tengamos 8 grupos (4 grupos de educación y 2 de edades para cada grupo de educación). De esta forma, se hace posible verificar si el nivel escolar afecta los ingresos de personas que están dentro de un mismo grupo de edad.

En esta etapa todo lo que se debe decir es que los grupos simples y sub-grupos son el método que proponemos tanto para estructurar las comparaciones como para controlar la causa y las distorsiones de

causa y atribución múltiples. Gracias a la capacidad de reagrupar los datos para diferentes asuntos y para controlar las diferentes fuentes potenciales de causación se aumenta grandemente la capacidad analítica del método de comparaciones. Simplemente extendiendo el ejemplo altamente simplificado que se dió arriba (edad y nivel de ingreso) uno puede ver que el análisis del desarrollo rural puede llegar a ser una estructura lógica muy intrincada.

La tercera etapa involucra el examen de las implicaciones de las diferencias observadas entre los grupos y sub-grupos en la formulación de políticas. Este es el enlace analítico más importante, es aquí donde el analista debe dar un significado más sustantivo a las cifras y las diferencias en las cifras. Debe relacionar el concepto en los datos con asuntos de política o asuntos de hipótesis y debe elaborar sobre lo que ésto significa para mejorar las intervenciones del Ministerio de Planificación y/o las Corporaciones Departamentales de Desarrollo. Se deben formular dos preguntas en esta etapa:

-¿Cuáles son las consecuencias de lo que se encontró? y

-¿Cuán importante es la consecuencia?

El arte del análisis para la planificación regional es el proceso de evaluar la importancia relativa. Las magnitudes observadas en las diferencias entre los subgrupos estructurados son su herramienta principal en determinar la importancia de lo que se descubre. Existe una debilidad común en los analistas estadísticos y es que ignoran la importancia del tamaño de una diferencia

observada debido a su interés para explorar la significancia estadística de la diferencia. Las magnitudes en estas diferencias observadas entre los grupos estructurados le dicen al analista cuán importante es lo encontrado para fines de políticas, le permiten empezar a catalogar el número masivo de diferentes items en el análisis de acuerdo a su importancia relativa. Esta catalogación no es un proceso sistemático, sin embargo, ya que el número de interrelaciones posibles es infinito; lo que se va logrando es una expansión gradual de la comprensión del sistema y cómo las intervenciones pueden afectarle. Las magnitudes son la llave para mantener una visión clara y prevenir una confusión interminable y la falta de resultados concretos de políticas que resultan de ello.

El cuarto paso es estimar la significancia estadística de las diferencias observadas. La significancia estadística de una diferencia observada entre los grupos nos dice que la diferencia observada no es simplemente debida a la casualidad. La significancia estadística estima la probabilidad de que dicha diferencia se mantendría si repetimos la recopilación de datos con toda la población y no sólo con un pequeño grupo o muestra de la población. Hay muchas técnicas estadísticas diferentes para medir la significancia estadística, creemos que sería de gran beneficio usar un número reducido de estas técnicas y hemos encontrado en la práctica que la metodología que proponemos puede casi siempre ser conducida aceptablemente sólo con el test estadístico "t" para comparaciones simples entre dos grupos, y  $\chi^2$  (chi-cuadrado) para grupos múltiples.

Para tener éxito, el método propuesto depende más de la habilidad del analista para construir adecuadamente la lógica de las comparaciones y controlar la atribución por sub-grupos que de la

manipulación estadística.

La estadística principal que se computa para cada uno de los grupos y sub-grupos es la media aritmética, el promedio común. La razón para esta decisión es que la media aritmética mide la tendencia central y la tendencia central es lo que el análisis de políticas está buscando. Si de hecho no hay una tendencia central poco es lo que el análisis puede hacer para manejar los resultados de los números. Para investigar hasta qué punto la tendencia central es característica de los datos, la técnica más útil y manejable es un histograma visual o diagrama de dispersión que muestra los valores mismos de todas las observaciones en un cuadro. Estos histogramas o diagramas de dispersión sugerirán como sub-agrupar los datos de modo que la media aritmética sea una medida aceptable de la tendencia central.

#### 4.4 Redacción del Informe

El resultado final de un esfuerzo de recolección de datos es un informe escrito y publicado. Este informe debe indicar cuál fue la muestra, como se diseñó, cuál fue su universo, cuántas observaciones tiene, dónde se realizó, a cuánto ascendieron los costos de trabajo de campo, qué formulario se empleó, cuáles fueron las principales dificultades encontradas, cómo se definieron las variables del análisis, qué cuadros de salida se obtuvieron, si se hizo algún análisis estadístico, y finalmente debe incluir una interpretación o análisis de los resultados.

Es importante precisar todos estos detalles a fin de que la persona que lee el análisis de los resultados sepa el grado de validez que tienen los datos obtenidos y las consecuencias o recomendaciones de

políticas que de ellos se derivan.

Con frecuencia el planificador, que ha sufrido contratiempos y atrasos en las primeras etapas del proceso de recolección de datos, se ve obligado a terminar el trabajo dentro de un plazo establecido. Esto da como resultado informes finales que constan únicamente de cuadros y cuadros con un análisis de una o dos páginas. La utilidad del trabajo se ve seriamente reducida ya que el personal ejecutivo, (presidentes y gerentes de las CDD's, Directores Generales y Sub-Secretarios del MPC) no dispone del tiempo necesario o no tiene el nivel técnico especializado como para interpretar los resultados así presentados y basarse en ellos para la toma de decisiones.

En otros casos la o las personas que realizaron la investigación ya no están presentes en la institución, en cuyo caso los resultados y recomendaciones del análisis son totalmente ignorados si no están acompañados de un documento que explica todo el detalle y método seguidos en la recolección de la información.

El esfuerzo de redacción del informe está ligado al de análisis en forma iterativa. El analista no completa todo el análisis (comparaciones, sub-división en grupos, tests estadísticos) antes de iniciar el informe final. Al contrario, el analista debe empezar a escribir tan pronto tiene algunos resultados. Durante este proceso de redacción el analista encontrará nuevas interrogantes, asuntos que aclarar, comparaciones adicionales, estratificaciones nuevas, etc., que le obligarán continuamente a pasar de la redacción al análisis y a la interpretación de los resultados obtenidos.

En otras palabras, el proceso mismo de la redacción es el modo más eficaz de completar adecuadamente el análisis e interpretación de los resultados.

La publicación del informe final también permite que el planificador intercambie ideas con otras personas, colegas, funcionarios de otras instituciones, personal ejecutivo, etc. En algunos casos este proceso de intercambio de ideas es tan fructífero que justifica la publicación de un informe posterior que presenta en forma resumida las conclusiones adicionales más importantes a las que se ha llegado.

ANEXO 1

TAMAÑO DE LA MUESTRA

La definición del tamaño óptimo de cada muestra depende de muy diversos factores. Este proceso de definición se desarrolla de la siguiente manera:

1. Especificación del universo muestral Para comenzar es preciso definir cuál es la población objeto del análisis. Si son los hogares campesinos de determinadas provincias de un departamento, habrá que establecer sin ambigüedades los requisitos que debe satisfacer un "hogar campesino".

Por ejemplo, todos los que viven en una casa, y comparten permanentemente los gastos de alimentación, son considerados como un hogar. No importando si en el hogar convive algún hijo con su cónyuge. Además, para que el hogar sea "campesino", deben poseer por lo menos, por ejemplo:

- Una hectárea de terreno de cultivo temporal, o
- Media hectárea de terreno de cultivo con riego, o
- Una vaca, o
- Una yunta, o
- Cinco ovejas, o
- Cinco llamas

Esta población objeto del análisis, constituye lo que se denomina universo de la muestra.

2. Tamaño del Universo Una vez establecido el universo de la muestra,



se considera el tamaño de este universo. En algunos casos, debido a que el universo es reducido, o gracias a que existe alguna institución relacionada con dicho universo (Por ej., el Banco Agrícola y los campesinos clientes del banco), es posible disponer de un listado completo del universo en cuestión. En otros casos, resultaría prohibitivo el costo de construir un listado completo del universo muestral.

Cuando se dispone de un listado, se aplica lo que se llama el "muestreo aleatorio simple", caso contrario, es preciso utilizar el "muestreo por áreas".

### 3. Muestreo aleatorio simple

Disponiendo de un listado del universo muestral, lo que debe determinarse es a cuántas unidades de este universo se requiere entrevistar. Para ello se calcula cuál es el número de sub-grupos o categorías en que se clasificará la muestra. Si simplemente queremos comparar campesinos con crédito vs campesinos sin crédito tenemos dos categorías. Si además vamos a clasificar a los campesinos en tres grupos según el tamaño de la explotación, ya se tienen 6 categorías.

Si tuviéramos, gracias a una muestra anterior, un estimado del valor de la desviación típica del atributo en estudio (ingreso, tamaño del área de cultivo, etc), para cada subgrupo del universo muestral, podría estimarse el tamaño que debe tener cada subgrupo para alcanzar una cierta precisión, con un determinado grado de confianza.

La fórmula para estimar este "n" mínimo es la siguiente:

$$n = \frac{k^2 v'}{E^2} \quad (1)$$

donde  $v'$  = varianza obtenida en la muestra anterior

$E$  = error admisible en el atributo que se estima

$k$  = coeficiente proporcional al grado de confianza "a"

La relación entre  $k$  y  $a$  es aproximadamente la siguiente:

a	80%	90%	95%	99%
k	1,3	1,6	2,0	2,6

Así, si quisiéramos estimar la estatura media de los conscriptos en Bolivia, con un error  $E = 5$  cm, con un grado de confianza de 90% en nuestra afirmación; y en una muestra anterior se obtuvo una  $v' = 100 \text{ cm}^2$ , tendríamos:

$$n = \frac{1,6^2 \times 100}{25} = 10,24$$

Si el "n" obtenido es mayor al 5% de "N", tamaño del universo muestral, la fórmula (1) se convierte en:

$$n = \frac{k^2 N v'}{NE^2 + k^2 v'} \quad (2)$$

Así, si en el ejemplo anterior  $N = 200$ , debería ajustarse el  $n$  a:

$$n = \frac{1,6^2 \times 200 \times 100}{200 \times 5^2 + 1,6^2 \times 100} = 9,74$$

En la práctica se tropieza con dos obstáculos para calcular  $n$  utilizando estas fórmulas.

- No se dispone de varianzas de muestras anteriores

52

- Se quieren analizar varios atributos

Por ello, se recurre a una simple receta que ha mostrado ser adecuada en la experiencia de campo, y se procuran 40 encuestas para cada categoría en que se piensa clasificar la muestra. Considerando que algunas encuestas serán rechazadas en el procesamiento, se prevé terminar con unas 35 encuestas por cada categoría. De esta forma, el tamaño de la muestra se calcula como:

$$n = 40 \times N_g$$

donde,  $N_g$  = número de categorías en que se pretende clasificar la muestra.

Así, para solo 2 categorías (con crédito vs sin crédito), se tendrá

$$n = 40 \times 2 = 80 \text{ encuestas.}$$

Para 6 categorías: tres tamaños de explotación X con/sin crédito, se tendrá:

$$n = 40 \times 6 = 240 \text{ encuestas}$$

### 3.1 Marco de la muestra

Una vez que se ha determinado el tamaño de la muestra, se procede a seleccionar las unidades que serán encuestadas. El proceso de selección debe ser aleatorio, es decir, cada unidad del universo muestral debe tener la misma probabilidad de ser seleccionada. El conjunto de unidades seleccionadas para la muestra constituye el

marco muestral.

Asignando a cada unidad del universo un número (de 1 a N), y seleccionando n números aleatorios entre 1 y N, se tendrá el marco de la muestra.

Alternativamente, se puede recurrir al muestreo sistemático. Se divide el universo muestral en n grupos de  $k_n$  unidades cada uno. Se selecciona al azar un número entre 1 y  $k_n$ , digamos  $k_j$ . Se toman en el marco muestral las unidades  $k_j, k_n + k_j, 2k_n + k_j, \dots, (n-1)k_n + k_j$ , es decir, a partir de la unidad  $k_j$  se toma una unidad cada  $k_n$  del listado total.

#### 4. Muestreo por Areas

En aquellos casos en que no se dispone de un listado completo de las unidades que constituyen el universo muestral, y resultaría costoso elaborar tal listado, la alternativa es recurrir al muestreo por áreas.

Se toma el universo muestral, y se lo divide en áreas con un tamaño similar de población, a las que se denomina unidades primarias. Se seleccionan al azar las unidades primarias donde se llegará con la encuesta.

En las unidades primarias seleccionadas se confecciona un listado de unidades muestrales, a las que se llaman unidades secundarias. De estos listados, se seleccionan al azar las unidades secundarias que conformarán la muestra o marco muestral.

Se tiene entonces que:

- N : No. de unidades primarias (UP's) en el universo
- $M_j$  : No. de unidades secundarias (US's) en la j unidad primaria.
- n : No. de unidades primarias incluidas en la muestra.
- $m_j$  : No. de unidades secundarias (US's) incluidas en la muestra, que pertenecen a la j unidad primaria.
- $\bar{m}$  : promedio aritmético de  $m_j$ .
- $n \bar{m}$  : tamaño de la muestra.

#### 4.1 Tamaño de la muestra

De la misma forma que en el muestreo aleatorio simple, si se dispone de datos obtenidos de una muestra por áreas preliminar, aplicados al universo en cuestión, es posible calcular el tamaño óptimo de la muestra mediante las siguientes fórmulas.

Si la función de costos es de la forma:

$$C_o = C_o + c_1 n \bar{m} + c_2 n$$

donde,  $C_o$  : Costos fijos

$c_1$  : Costo marginal por encuesta

$c_2$  : Costo marginal por UP incluida en la encuesta.

Se puede calcular aproximadamente el valor óptimo de  $m$  mediante la expresión:

$$\bar{m} \text{ óptimo} = \bar{m} = \sqrt{\frac{c_2}{c_1} \frac{\bar{m}^*}{\bar{m}^* S_D^2 / S_W^2 + 1}} \quad (3)$$

donde,  $\bar{m}^*$  :  $\bar{m}$  de la encuesta preliminar

$S_D^2$ : varianza "entre" áreas de la encuesta preliminar

$S_w^2$ : varianza interna de las áreas de la encuesta preliminar

Adoptando valores obtenidos de la encuesta por áreas realizada en Chuquisaca, se tiene por ejemplo:

$$c_2 = 4,500 \text{ \$b.}$$

$$c_1 = 150 \text{ \$b.}$$

$$\bar{m}^* = 47,8$$

$$S_b^2 = 17,80 \times 10^6$$

$$S_w^2 = 156,86 \times 10^6$$

con estos datos, y usando la expresión (3), se obtiene un  $\bar{m} = 18$

Es decir, si se quisiera realizar otra encuesta por áreas para determinar el promedio de ingresos de los campesinos del Noroeste de Chuquisaca, se debería procurar obtener 18 encuestas de cada área muestral.

Para determinar el número óptimo de áreas, se dispone de otra fórmula

$$n = \frac{k^2}{E^2} \frac{S^2}{\bar{m}} \left[ 1 + (\bar{m} - 1) r_{ci} \right] \quad (4)$$

donde: E: error aceptable

k: coeficiente del grado de confianza exigido

$S^2$ : varianza de la muestra preliminar

$r_{ci}$ : coeficiente de correlación intraclásica, obtenido de la muestra preliminar.

Adoptando siempre los valores obtenidos en Chuquisaca, se tiene:

$$S^2 = 169,8 \times 10^6$$

$$r_{ci} = 0,0855$$

y para  $k = 1,3$  y  $E = 1,500$

Sustituyendo estos valores en la expresión (4) se llega a un  $n$  óptimo de  $n=17$

### Recomendaciones

Ai igual que en el caso del muestreo aleatorio simple, se tropieza con dos obstáculos para la aplicación de estas fórmulas (de por sí ya bastante complejas): (a) Normalmente no se dispone de muestras por áreas realizadas anteriormente y (b) Se quieren analizar varios atributos.

Por tanto, es preferible recurrir a recetas aproximadas. Estas recetas indican:

1. El número "n" de áreas muestrales incluídas en la muestra debe oscilar entre 15 y 25 áreas.
2. El número promedio de encuestas por área ( $\bar{m}$ ) es más flexible, oscilando entre 10 y 40 encuestas por área.
3. El número total de encuestas ( $n \bar{m}$ ) se lo estima de acuerdo a la receta aplicada en el muestreo aleatorio simple. Esto es:

$$n \bar{m} = 40 \times N_g \quad (5)$$

4. Los valores finales de  $n$  y  $\bar{m}$ , dentro los rangos mencionados, deben establecerse considerando que:
  - Al aumentar  $n$  (y disminuir  $\bar{m}$ ), la precisión de la muestra aumenta
  - Al mismo tiempo, los costos aumentan.

Así si se quiere obtener una buena precisión, no importando mucho los costos, se toma por ejemplo un  $n$  alto y se procede a calcular:

$$\bar{m} = \frac{40 \times Ng}{n}$$

Si el  $\bar{m}$  así calculado resulta menor a 10, se adopta  $\bar{m} = 10$  por ser un mínimo.

En cambio, si la consideración más importante es minimizar costos, se puede adoptar un  $\bar{m} = 40$ , y calcular

$$n = \frac{40 \times Ng}{\bar{m}} = Ng$$

si el  $n$  así obtenido es inferior a 15, se adopta como  $n$  el valor mínimo de 15 áreas y se corrige  $\bar{m}$  para satisfacer la expresión (5).

#### Observación

Estas recomendaciones son válidas para casos en los que  $c_2/c_1$  es un valor alto (del orden 10 a 20). En casos excepcionales puede darse que  $c_2/c_1$  sea bajo (menor a 5) casos en los que habría que ajustar las recetas, aumentando  $n$  y disminuyendo  $\bar{m}$ .

#### 4.2 Varianza de la media muestral

Las recomendaciones acerca del tamaño óptimo de la muestra pueden ser comprendidas mejor analizando la expresión para la varianza de la media muestral.

$$v_x = \frac{N-n}{N} \cdot \frac{s^2}{n \bar{m}} (1 + (\bar{m} - 1)r_{ci}) \quad (6)$$

Se observa en esta expresión.

a) Cuando "N" es mucho mayor a "n", el término (N-n)/N tiende a la unidad.

b) El término  $r_{ci}$  expresa la correlación intraclásica, esto es, el grado de homogeneidad existente dentro cada área muestral, con relación a la población total. En la encuesta por áreas realizada en el noroeste de Chuquisaca se obtuvo un  $r_{ci}$  de 8,6%. En la encuesta de los Valles del Sur realizada en 1977 en una región con áreas más heterogeneas entre sí, se obtuvo un  $r_{ci} = 35\%<sup>1</sup>$ .

La expresión para esta relación es:

$$r_{ci} = \frac{S_b^2 - S_w^2/\bar{m}}{S^2} \quad (7)$$

c) Cuando  $\bar{m}$  es mayor a 10,  $v_x$  es poco sensible a mayores aumentos de  $\bar{m}$ . Aumentar  $\bar{m}$  por encima de  $\bar{m} = 40$  no tiene prácticamente ninguna influencia en reducir  $v_x$ .

---

<sup>1</sup>/ 1978 Bolivian Socio-Economic Farm Survey: Sample Design, Don Larson (May 1978) United States Department of Commerce, Bureau of the Census, Washington D.C. 20235.

d)  $v_x$  es inversamente proporcional a  $n$ . Cualquier aumento en  $n$  redundará siempre en un menor valor de  $v_x$

#### 4.3 Planificación del Trabajo de Campo

En el caso de muestreo por áreas realizado en zonas rurales, lo más indicado es utilizar como áreas muestrales los segmentos censales del Censo de Población y Vivienda de 1976. Para estos segmentos censales se dispone de mapas razonablemente adecuados. Además, se puede obtener del INE datos sobre la población: hombres, mujeres, y hogares de cada segmento.

A fin de contar con áreas censales con similar número de hogares, se pueden juntar segmentos adyacentes de poca población, y dividir aquellos de mucha población. De esta forma se recomienda conformar áreas o unidades primarias entre las que las mayores diferencias en número de unidades secundarias sean inferiores al 100%.

Habiendo seleccionado al azar las unidades primarias de la muestra, y conociendo el número de unidades secundarias ( $M_j$ ) de cada una de ellas, se procede a estimar el porcentaje  $(m/M)_j$  de US's que deberán ser entrevistadas en cada UP para lograr el promedio ( $m$ ) que se pretende. Para ello recurrimos a la siguiente expresión:

$$(m/M)_j = \frac{n \bar{m}}{\text{SUM } M_j} \quad (8)$$

Una vez determinado este porcentaje  $(m/M)$ , es importante mantenerlo escrupulosamente igual en todas las UP's que se incluyen en la muestra. Solo así, todas las US's del universo muestral tendrán la

misma probabilidad de aparecer en la encuesta. Esta es una condición necesaria para que los promedios obtenidos de la encuesta sean representativos del universo muestral, sin tener que aplicar engorrosas ponderaciones. Las muestras que gozan de esta propiedad son denominadas autoponderadas.

ANEXO 2

TESTS ESTADISTICOS

1. Comparación Simple Entre Dos Grupos

Cuando se desea verificar estadísticamente si la media de un parámetro cualquiera es igual para dos grupos muestrales, se puede utilizar el estadístico  $t$ .

La hipótesis que se plantea es que los dos grupos provienen de la misma población. Si ese es el caso sus medias serían iguales:

Hipótesis:  $\mu_1 = \mu_2$

El estadístico  $t$  se calcula con la expresión:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{S_{\bar{x}_1 - \bar{x}_2}}$$

donde:  $\bar{x}_1$  : media muestral del grupo 1

$\bar{x}_2$  : media muestral del grupo 2

$S_{\bar{x}_1 - \bar{x}_2}$  : desviación típica de la diferencia de medias.

Por ejemplo, el Cuadro 2.1 presenta el ingreso total de dos muestras de familias campesinas obtenidas en la encuesta de crédito del departamento de Chuquisaca.

CUADRO 2.1

INGRESO TOTAL DE LOS CAMPESINOS CON CREDITO Y SIN CREDITO EN EL  
DEPARTAMENTO DE CHUQUISACA

(\$b. x 1.000)

CON CREDITO	SIN CREDITO
108,1	23,9
40,3	13,7
6,8	27,1
40,4	35,8
19,6	9,3
22,6	7,4
1,9	9,5
12,8	5,1
7,1	2,9
22,0	23,3
18,6	14,8
28,0	29,1
25,5	10,7
16,2	25,8
14,2	-2,2
76,2	10,1
29,5	15,1
13,8	9,4
27,1	6,7
52,7	38,2
35,6	27,1
35,4	28,2
4,4	29,4
16,0	5,6
29,3	30,0
16,5	39,3
37,4	60,1
65,3	13,3
24,8	31,7
39,9	24,0
32,0	10,2
22,1	15,0
72,3	16,0
35,9	11,5
34,5	26,2
36,9	7,3
25,4	26,6
75,8	99,7

El grupo 1 está formado por 38 campesinos que han recibido crédito agrícola. El grupo 2 consta de 38 campesinos que no han recibido crédito. Los primeros 38 fueron seleccionados aleatoriamente, los otros son sus pares.

Podemos calcular la media como sigue:

$$\bar{x}_1 = \frac{\sum x_{1j}}{n_1} = \frac{1222,9}{38} = 32,18 \quad (\$b \times 1000)$$

$$\bar{x}_2 = \frac{\sum x_{2j}}{n_2} = \frac{816,9}{38} = 21,50 \quad (\$b \times 1000)$$

La diferencia entre ambos promedios es de:

$$\bar{x}_1 - \bar{x}_2 = 10,68 \quad (\$b \times 1000)$$

Por lo tanto, el grupo de campesinos con crédito tiene ingresos totales anuales que exceden en \$b 10.680 al resto. Pero prosigamos con el test estadístico. Después de calcular  $\bar{x}_1$  y  $\bar{x}_2$ , es preciso calcular  $S_{\bar{x}_1 - \bar{x}_2}^2$ :

$$S_{\bar{x}_1 - \bar{x}_2}^2 = \frac{s^2}{n_1} + \frac{s^2}{n_2}$$

donde:  $n_1$  : número de observaciones del grupo 1

$n_2$  : número de observaciones del grupo 2

El término  $s^2$  se calcula con la expresión:

$$s^2 = \frac{(n_1 - 1) S_1^2 + (n_2 - 1) S_2^2}{n_1 + n_2 - 2}$$

64

donde  $s_1$  = desviación típica del grupo 1

$s_2$  = desviación típica del grupo 2

La desviación típica se calcula con la ecuación:

$$s_1^2 = \frac{\sum x_{1j}^2 - \frac{(\sum x_{1j})^2}{n_1}}{n_1 - 1}$$

$$s_1^2 = \frac{57975,69 - \frac{1222,9^2}{38}}{37} = 503,27$$

$$s_2^2 = \frac{29701,41 - \frac{816,9^2}{38}}{37} = 328,11$$

Ahora se puede calcular el valor de  $s^2$

$$s^2 = \frac{37 \times 503,27 + 37 \times 328,11}{74} = 415,69$$

Una vez calculado  $s^2$  se puede calcular el valor de  $s_{\bar{x}_1 - \bar{x}_2}^2$

$$s_{\bar{x}_1 - \bar{x}_2}^2 = \frac{415,69}{38} + \frac{415,69}{38} = 21,88$$

$$s_{\bar{x}_1 - \bar{x}_2} = \sqrt{21,88} = 4,68$$

Finalmente, se calcula el valor de  $t$  :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_{\bar{x}_1 - \bar{x}_2}} = \frac{10,68}{4,68} = 2,28$$

65

El valor crítico del estadístico  $t$ , para un grado de confianza de 95%, y con  $n_1 + n_2 - 2 = 74$  grados de libertad es de: (el valor es obtenido de la tabla de  $t$ )

$$t_{0,95,74} = 1,99$$

Como el valor calculado de  $t$  sobrepasa el valor crítico, la hipótesis es rechazada. Por lo tanto, se concluye que los dos grupos (campesinos con crédito y sin crédito) tienen un ingreso promedio significativamente distinto.

## 2. Comparaciones múltiples

Con frecuencia lo que se desea comparar no es simplemente el promedio de un grupo con el de otro sino una clasificación no numérica aplicada a las unidades de observación. Por ejemplo, si queremos comparar el nivel de instrucción de la población rural en las tres regiones (podríamos hacerlo para sólo dos regiones también) del departamento de Oruro, no tenemos para cada región una cifra promedio ( $\bar{x}$ ) para comparar, sino una estratificación con "numero de niños" en cada nivel de instrucción, y en cada región tal como aparece en el Cuadro 2.2.

La hipótesis a verificarse es: "la estratificación según nivel de instrucción no depende de la zona". Para esta verificación se emplea el estadístico Chi cuadrado que se calcula con la expresión:

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

donde:  $O_{ij}$  = número de observaciones en la columna  $i$ , fila  $j$   
 $E_{ij}$  = valor esperado en la columna  $i$ , fila  $j$ .  
 $p$  = número de columnas  
 $q$  = número de filas

66

CUADRO 2.2

Población de 5 y Más Años por Nivel de Instrucción,  
Encuesta Agropecuaria de Oruro

	Z O N A			Total
	<u>Este</u>	<u>Suj-Oeste</u>	<u>Nor-Oeste</u>	
Ninguno	37 (38.2)	27 (29.8)	36 (31.9)	100
Alfabetización	4 (5.7)	2 (4.5)	9 (4.8)	15
Básico	105 (105.5)	90 (82.3)	81 (88.1)	275
Intermedio	19 (20.6)	16 (16.1)	19 (17.2)	54
Medio u Otros	17 (11.9)	7 (9.2)	7 (9.9)	31
Total	182	142	152	476

---

Nota: En paréntesis están los valores esperados.

El valor esperado se calcula con la ecuación:

$$E_{ij} = \frac{N_i \times N_j}{N}$$

donde:  $N_i$  = número total de observaciones en la columna  $i$

$N_j$  = número total de observaciones en la fila  $j$

$N$  = número total de observaciones

Por ejemplo, en el Cuadro 2.2 se han calculado los valores esperados y aparecen entre paréntesis. El ejemplo tiene 3 columnas y 5 filas. El valor calculado para Chi cuadrado es de:

$$\chi^2 = 11,4$$

El valor crítico de  $\chi^2$  se obtiene de tablas. Para un 95 por ciento de confianza y  $(p-1) (q-1)$  grados de libertad el valor crítico es de:

$$\chi^2_{0,95;8} = 9,49$$

El valor calculado sobrepasa al valor crítico, por lo tanto, se rechaza la hipótesis. En consecuencia, la estratificación de la población de 5 y más años según nivel de instrucción es significativamente distinta en las tres zonas del departamento de Oruro.

58

### REFERENCIAS

- Samuel R. Daines, *An Overview of Economic & Data Analysis Techniques for Project Design & Evaluation*, Agency for International Development, Washington, D.C., 1977
- Samuel R. Daines, *Data Gathering & Statistical Methods for Rural Development Planning, (Preliminary Version) Proyecto de Planeamiento para el Desarrollo Rural: Manual de Métodos*, MPC-AID-PCI, 1981.
- Donald Warwick y Charles Lininger, *The Sample Survey: Theory and Practice*, McGraw Hill Book Company, New York, 1975.
- Morris H. Hansen, William N. Hurwitz, y William G. Madow, *Sample Survey Methods and Theory. Volume I, Methods and Applications*, John Wiley & Sons, Inc. 1953.
- Francisco Azorín Poch, *Curso de Muestreo y Aplicaciones*. Ed. Aguilar 1972.

(0188A)