



# **GUIA PARA LA EVALUACION DE INDICADORES SOCIALES DE LAS ENCUESTAS DE HOGARES**

Lima, Noviembre 2001

## **DIRECCION Y SUPERVISION**

Rosario Aquije Valdez  
*Directora Técnica de Demografía e Indicadores Sociales*

Genara Rivera Araujo  
*Directora Adjunta de Demografía e Indicadores Sociales*

## **RESPONSABLES DEL ESTUDIO**

Econ. Rofilia Ramírez Ramírez  
Ing. Estad. Herman Edgar Castillo Ramón

## **SOPORTE INFORMATICO**

Sr. Walter Ayala Godiño

---

Preparado:	Dirección Técnica de Demografía e Indicadores Sociales (DES)
Impreso:	Talleres de la Oficina Técnica de Administración (OTA) del Instituto Nacional de Estadística e Informática
Diagramación:	Centro de Edición del INEI
Tiraje:	150 ejemplares
Nº de Orden:	693-OTA-INEI
Dispositivo Legal Nº:	150113-2001-4490

# PRESENTACION

Evaluar un indicador social implica revisar los procesos seguidos para determinar su valor numérico. Estos procesos van desde la conceptualización del indicador, el diseño y la formulación de las preguntas del cuestionario, las definiciones operacionales de las variables y la revisión de la metodología de otros países. En esta orientación la guía metodológica "Evaluación de Indicadores Sociales de las Encuestas de Hogares", explica la naturaleza de la revisión conceptual y describe los instrumentos para analizar los conceptos (Diagramas Conceptuales) y los procesos (Diagramas de Procesos) que intervienen en la generación de un indicador. La revisión culmina con la consistencia externa, es decir, cuando se compara el indicador calculado con otros. La combinación de estos instrumentos facilita la visión del objeto social en estudio así como la incorporación de medidas de control en cada una de las etapas de la revisión.

La guía contiene además la explicación de las técnicas estadísticas para el análisis de datos, que son enfocadas desde una perspectiva que facilita la descripción objetiva y comparable del indicador. Un indicador social resume las múltiples dimensiones de la realidad, lo que hace indispensable el empleo de las técnicas multivariadas, para evaluar la consistencia de los resultados obtenidos, tal como se muestra en el presente documento.

En el marco de transparencia metodológica, el Instituto Nacional de Estadística e Informática (INEI) pone al alcance de investigadores y usuarios la presente guía, esperando contribuir a una mejor evaluación de los indicadores sociales aplicando eficientemente los criterios y herramientas estadísticas y demográficas.

Lima, Noviembre 2001

**Gilberto Moncada Vigo**  
Jefe del INEI

# CONTENIDO

---

## PRESENTACION

<b>I. OBJETIVOS DEL MANUAL .....</b>	<b>7</b>
<b>II. INSTRUMENTOS Y PROCEDIMIENTOS PARA LA REVISIÓN CONCEPTUAL .....</b>	<b>9</b>
2.1 Elaborar mapas conceptuales .....	10
2.2 Revisar los conceptos locales .....	10
2.3 Revisar los conceptos de otros países .....	12
<b>III. INSTRUMENTOS Y PROCEDIMIENTOS PARA LA EVALUACIÓN DE LOS INDICADORES .....</b>	<b>13</b>
3.1 Elaborar diagramas de Procesos .....	13
3.2 Revisar las escalas de medición .....	14
3.3 Evaluar las definiciones operacionales .....	16
3.4 Comparar con las formas de medición de otros países .....	17
3.5 Analizar las ventajas y desventajas del indicador .....	18
<b>IV. TECNICAS ESTADISTICAS Y DEMOGRÁFICAS PARA EL ANALISIS DE LOS DATOS .....</b>	<b>21</b>
4.1 Realizar el análisis estadístico y demográfico .....	23
4.2 Evaluar la consistencia de la información .....	37
4.3 Efectuar el análisis multivariado .....	40
<b>V. IDENTIFICAR LOS ERRORES NO MUESTRALES MAS FRECUENTES .....</b>	<b>45</b>
5.1 Evaluar los errores no muestrales .....	45
5.2 Formas de minimizar los errores no muestrales .....	46
<b>VI. ERRORES MAS FRECUENTES EN LA MEDICION E INTERPRETACION .....</b>	<b>49</b>
<b>VII. RECOMENDACIONES PARA FORMULAR INDICADORES SOCIALES .....</b>	<b>51</b>
<b>REFERENCIAS BIBLIOGRAFICAS .....</b>	<b>53</b>
<b>ANEXO: CUADROS RESUMEN .....</b>	<b>55</b>

---



## CAPITULO I

### OBJETIVOS DEL MANUAL

#### 1. OBJETIVO GENERAL

necesarias para evaluar la calidad de los datos provenientes de las encuestas.

##### 1.1 Elaborar un documento guía para la evaluación de los Indicadores Sociales

Para la evaluación de los Indicadores Sociales se requiere el conocimiento conceptual del fenómeno en estudio y el uso apropiado de las técnicas estadísticas y demográficas de análisis de datos. También son necesarias aplicar los diagramas de conceptos asociados al indicador así como de las técnicas de interpretación de los gráficos de distribución de las variables. Estas herramientas se encuentran dispersas debiendo consultarse muchas fuentes para entenderlas, lo cual dificulta su uso integral y oportuno. En esta medida la presente guía reúne las principales herramientas y explica su uso con ejemplos prácticos de fácil entendimiento.

##### 2.2 Proponer criterios para formular alternativas útiles de medición

Tan importante como determinar la exactitud del indicador es encontrar su utilidad. Desde que los indicadores se ajustan a las convenciones aprobadas por la persona o institución que los formula es importante hacer explícitos los criterios empleados en su formulación de tal manera que puedan deducirse sus ventajas y limitaciones, y con ello apreciar su utilidad. Por ello, en este documento se presentan las condiciones para formular indicadores confiables..

##### 2.3 Brindar recomendaciones para mejorar la presentación e interpretación de los Indicadores Sociales obtenidos de las encuestas de hogares

En la medida que se conozca la fórmula de cálculo del indicador, será un procedimiento relativamente sencillo determinar su valor, encontrándose que las mayores dificultades se presentan al momento de presentar los resultados e interpretar su valor final. Por ello en el presente documento se brindan un conjunto de recomendaciones que esperamos permita a la comunidad de investigadores y usuarios en general superar estas limitaciones.

#### 2. OBJETIVOS ESPECIFICOS

##### 2.1 Proporcionar un diseño metodológico para evaluar la calidad de los datos.

Los datos son los insumos para calcular el indicador, por lo cual la calidad de los primeros determina el valor explicativo y predictivo del indicador social. Por ello en el documento se explican las técnicas estadísticas y herramientas analíticas



## CAPITULO II

### INSTRUMENTOS Y PROCEDIMIENTOS PARA LA REVISIÓN CONCEPTUAL

La revisión metodológica de los indicadores sociales<sup>1</sup> exige evaluar la correspondencia entre la definición del indicador y su operacionalización. Así, el conocimiento del fenómeno social adquirido en las etapas de la investigación respondiendo a las siguientes interrogantes: ¿Qué? ¿Cómo? ¿Dónde? ¿Cuándo? ¿De Dónde? ¿Porqué? debe tener relación con su expresión operativa. Probablemente en la operacionalización del concepto han intervenido dos condiciones: **claridad** en la formulación y **especificidad** en términos del tipo de respuesta a las preguntas formuladas.

La tarea de revisión se extiende además al conocimiento de los diseños de investigación que han dado lugar a la formulación de los indicadores. Los diseños más conocidos son: el Exploratorio,

el Descriptivo y el Explicativo. En el primer caso se trata de estudios nuevos donde no hay teorías precisas y los modelos no están explícitamente formulados. En el caso de las investigaciones descriptivas éstas son más específicas y organizadas que las exploratorias y su interés está más enfocado en las propiedades del objeto, como ocurre en el caso del diagnóstico. Las investigaciones explicativas cumplen un doble propósito, la verificación lógica y la verificación empírica. La primera verifica la consistencia entre las proposiciones que integran el sistema, mientras que la segunda busca la correspondencia entre la teoría y la realidad. El conocimiento de estas premisas nos debe llevar a formular la siguiente pregunta ¿Cómo empezar la revisión metodológica?

*Para que la revisión metodológica siga los procedimientos científicos, debe comenzar por la evaluación de los conceptos. Muchas veces el concepto obedece a convenciones y directivas determinadas de común acuerdo con otros países u organismos internacionales y de esta manera implícitamente el indicador también lo estará.*

En toda revisión conceptual se recomienda tener una visión global de las variables relacionadas con el fenómeno social. El instrumento más apropiado para

presentar las interrelaciones entre los conceptos es el mapa conceptual, cuya elaboración pasa a ser el primer paso en la evaluación del indicador.

<sup>1</sup> Un indicador social, es una medida resumen que describe un fenómeno social para su posterior análisis, comparabilidad y toma de decisiones. Por ejemplo la tasa de analfabetismo, la tasa de deserción escolar, la tasa de desempleo, el porcentaje de niños con diarrea, entre otros.



## 2.1 ELABORAR MAPAS CONCEPTUALES

**Definición.-** El Mapa Conceptual, es una representación gráfica de las relaciones significativas entre conceptos relacionados con el fenómeno social, en forma de

proposiciones. Sus principales componentes son: los términos conceptuales, los conectores y las proposiciones. Los términos conceptuales van encerrados en una elipse, para diferenciarlos del resto.

*El mapa conceptual es un conjunto de proposiciones representadas mediante términos conceptuales y conectores. Entre estos elementos deben existir jerarquías, y como toda definición responde a una visión del investigador, implican una selección de contenidos, los mismos que deben ser presentados de tal manera que faciliten el impacto visual.*

**Utilidad:** Los mapas conceptuales permiten la presentación de los conceptos sociales de un modo estructurado y de fácil comprensión. También facilitan la comparación de los diferentes puntos de vista en torno a un tema, ya que cada investigador diseña un mapa conceptual de acuerdo a su visión particular. Como instrumento de representación puede ser utilizado en todas las etapas de la revisión (revisión conceptual, evaluación de la consistencia, elaboración de modelos, entre otras).

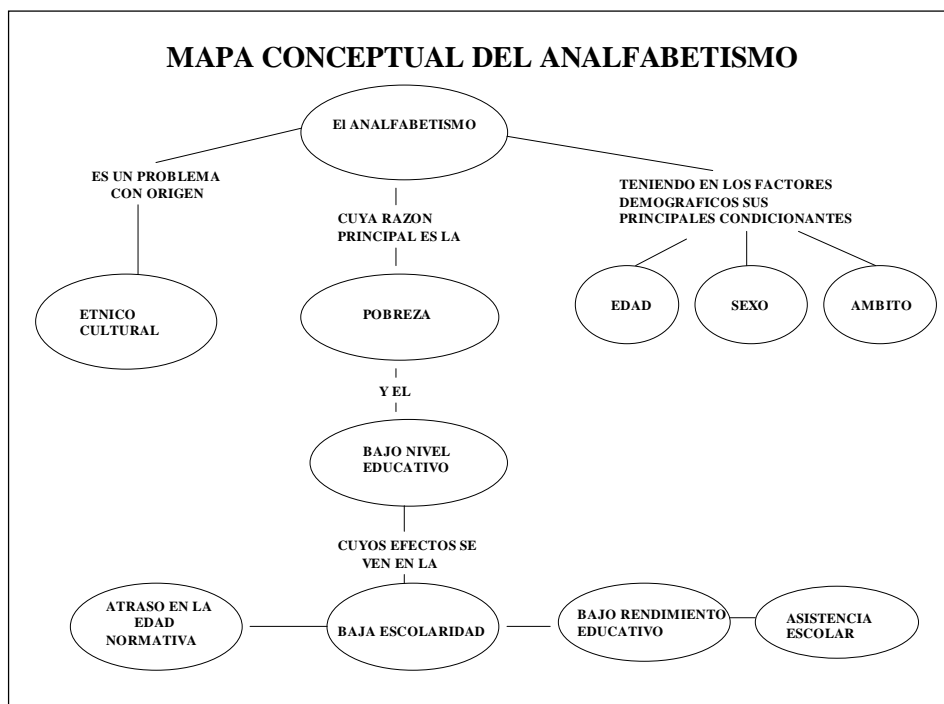
**Ejemplo:** Se toma el Analfabetismo para desarrollar un mapa conceptual. El analfabetismo puede definirse como un problema social complejo de origen étnico, muy relacionado con la pobreza y la ausencia de nivel educativo. Además está altamente concentrado en la población rural femenina adulta mayor. Las

estadísticas educacionales: asistencia escolar, deserción escolar, bajo rendimiento y atraso en la edad normativa reflejan los efectos en el sistema educativo de este fenómeno social. En el mapa conceptual siguiente se resumen las ideas expuestas: (ver gráfico página siguiente)

## 2.2 REVISAR LOS CONCEPTOS LOCALES

Formulada la visión del problema y representada en un mapa conceptual, el siguiente paso consiste en la búsqueda de los documentos y las fuentes de información local a partir de las cuales se ha generado el indicador. Se debe tener presente que además es necesario verificar que la operacionalización del concepto haya sido la misma en todo el período de análisis del indicador.

*Para una recopilación sistemática, se recomienda elaborar fichas bibliográficas y de contenido, de tal forma que se cuente con una relación de autores y definiciones. En esta revisión deben distinguirse los conceptos que han sido operacionalizados, de los no operacionalizados, ya que estos últimos si bien son teóricos y carecen de una expresión de cálculo, pueden sugerir nuevos aportes en la medición.*



De otro lado la operacionalización del concepto asociado al indicador se realiza con datos que proceden generalmente de los censos y las encuestas.

Mientras que los censos abarcan un conteo total de la población, las encuestas sólo cubren una parte de la población en estudio por lo que los indicadores obtenidos llevan asociados sesgos de cobertura. En este caso las estimaciones puntuales suelen brindarse acompañadas del respectivo intervalo de confianza dentro del cual se espera esté contenido el verdadero valor del parámetro.

**Ejemplo:** La mayoría de los conceptos asociados al analfabetismo en el Perú, se refieren a este fenómeno social como la incapacidad para leer y escribir, considerando analfabeta a toda persona de 15 años y más que no lee ni escribe. Este

concepto actualmente se utiliza para calcular la tasa de analfabetismo. Otro concepto asociado es el analfabetismo funcional, considerándose en esta categoría a las personas que en algún momento aprendieron a leer y escribir pero que por diferentes razones muchas de ellas ligadas a la propia pobreza, perdieron esas destrezas. Otras investigaciones consideran analfabeta a aquel individuo que no conoce ni usa programas informáticos básicos.

Este ejemplo muestra los diferentes conceptos asociados al mismo fenómeno social.

Revisar los conceptos locales comprende además indagar por las fuentes de datos, el modo de obtención del indicador: tipo de preguntas formuladas en el cuestionario, fórmula (algoritmo de

cálculo), factores de expansión para determinar las cifras globales, manuales del usuario, entre otros.

### 2.3 REVISAR LOS CONCEPTOS DE OTROS PAÍSES

La búsqueda de los conceptos asociados al indicador en otros países debe centrarse en aquellos con características socioeconómicas, culturales y geográficas similares a la nuestra, de tal manera que pueda establecerse una correspondencia entre los términos encontrados. De este modo se facilitará la comparabilidad internacional y la armonización de conceptos y metodologías.

**Ejemplo:** La revisión de los conceptos asociados al analfabetismo se inicia indagando por las definiciones y las fuentes de datos en algunos países del continente como Argentina, Chile, Ecuador y Paraguay encontrando que todos se rigen por los principios de la UNESCO, lo que facilita su comparación. Esto permite el desarrollo de políticas y estrategias comunes en el campo Educativo.

De otro lado en los países mencionados el indicador proviene de los Censos de Población y Vivienda y de las Encuestas de Hogares. Las definiciones adoptadas por otros países son:

**Chile.** El Instituto Nacional de Estadística (INE) de Chile define la población alfabeta como aquella población que posee la capacidad de leer y escribir, medida en términos de algún estándar, mientras que el alfabeto funcional se relaciona con el estándar de poder funcionar adecuadamente en la sociedad. La tasa de analfabetismo se define como el porcentaje de la población de 15 años y más, incapaz de leer y escribir un párrafo breve.

**Argentina.** El Instituto de Estadística y Censos (INDEC) define la población analfabeta como aquellas personas que declaran no saben leer o escribir respecto a la población total de 15 años y mas.

**Ecuador.** El Instituto Nacional de Estadística del Ecuador (INEC) define la población analfabeta como aquellas personas que no saben leer o escribir o que solo leen o solo escriben. Mientras que la categoría analfabeto funcional se determina mediante la "prueba de medición de logros". Así recoge la recomendación de la UNESCO de considerar analfabetos funcionales a todas las personas que tienen tres años o menos de educación primaria, ya que se presupone que las destrezas de lectura, escritura y aritmética básica requieren, en general, de una mayor escolarización.

## CAPITULO III

### INSTRUMENTOS Y PROCEDIMIENTOS PARA LA EVALUACIÓN DE LOS INDICADORES

Los múltiples instrumentos estadísticos y demográficos que intervienen en la evaluación del indicador hacen necesario la elaboración de un diagrama en que se representen todos los procesos a seguir. De este modo se tendrá mayor control en todas las etapas de evaluación. Asimismo

debe identificarse las diferentes escalas de medición de las variables que intervienen en el cálculo del indicador y las definiciones operacionales (algoritmo). Finalmente se recomienda evaluar las ventajas y desventajas del indicador.

*Para evaluar la forma de cálculo del indicador se requiere elaborar un diseño de evaluación, identificar las escalas de medición de las variables que intervienen en el cálculo y un análisis de sus ventajas y desventajas.*

#### 3.1 Elaborar Diagramas de Procesos.

El instrumento principal de evaluación es el **diagrama de procesos**, que permite seguir un orden y control de cada uno de

los procesos, facilitando la descomposición de la Evaluación del Indicador en etapas con la posibilidad de generar medidas de control, en cada una de ellas.

*El diagrama de proceso es un recurso gráfico propio de la ingeniería y muy útil en la elaboración de Sistemas de Información. Con este recurso se identifican las actividades secuenciales para el proceso de evaluación, desde el nivel primario (datos, variables) hasta su consolidación en un indicador.*

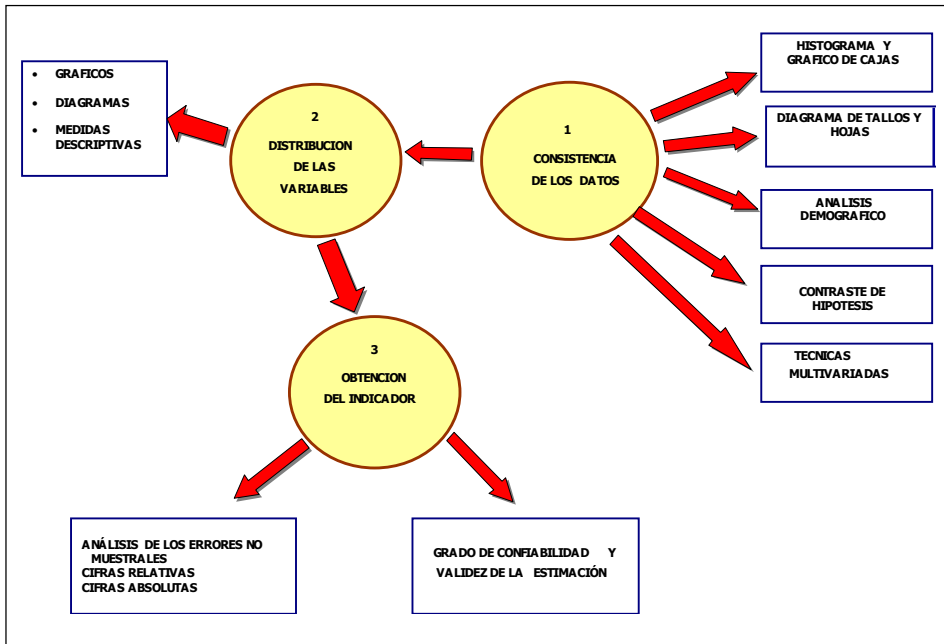
En el flujo de datos se representan las entidades (objetos que generan datos) y los procesos o actividades, en forma secuencial y separados por flechas. El flujo de información se representa con figuras geométricas. Los **círculos** representan los procesos o fórmulas, los **rectángulos** son las entidades o técnicas de análisis y las **flechas** permiten conocer la secuencia seguida en el análisis.

**Utilidad:** El Diagrama de Procesos permite **representar el diseño de revisión del indicador** (descomponer en etapas todo el proceso seguido) y los flujos. De este modo pueden incorporarse medidas de control en cada etapa.

**Ejemplo:** La evaluación de un indicador social en términos generales puede descomponerse en las siguientes etapas:

la consistencia de datos, el análisis estadístico de las variables (distribución de sus valores) y la consistencia del indicador obtenido.

Estas etapas se describen en el siguiente diagrama de procesos:



### 3.2 Revisar las Escalas de Medición.

Elaborado el diseño de evaluación mediante el diagrama de procesos, el siguiente paso es la revisión de las escalas de medición de las variables, a fin de

reconocer las unidades de medida en que están representadas las diferentes manifestaciones del fenómeno social, a fin de aplicar las **técnicas estadísticas más apropiadas**.

*Revisar las escalas de medición implica reconocer en las variables, las características de orden, distancia y origen que le confieren los números asociados. Así las técnicas de evaluación utilizadas serán las más apropiadas.*

El interés analítico del investigador sobre las variables en estudio determina los **tipos de escalas** a utilizar. Ejemplo: la "edad" puede representarse en tres escalas diferentes de medición: en edades simples,

en grupos quinquenales, o en poblaciones especiales de estudio como niños, jóvenes, adultos y adultos mayores. De este modo la variable "edad" puede medirse en una escala de medición interválica, ordinal o

nominal. Los cuestionarios de las encuestas contienen un conjunto de variables expresadas en los diferentes tipos de escalas las cuales pueden adecuarse al objetivo y utilidad de la investigación, a este proceso se le conoce como recodificación.

Las variables pueden representarse en las siguientes escalas:

**1. La escala nominal.** Es el nivel más elemental de clasificación de las variables u objetos en estudio que consiste en asociar números a cada una de las categorías de la variable. Los números sirven simplemente para distinguir entre las diferentes categorías de una variable. Las técnicas estadísticas utilizadas con más frecuencia para este tipo de escalas son: la distribución de frecuencias, el histograma y la moda. Ejemplo: la variable sexo tiene dos categorías: hombre y mujer, que en la ENAHO se representa con números: **hombre = 1** y **mujer = 2**. En este caso el número solo **sustituye** el nombre de la categoría. En la Encuesta Demográfica y de Salud Familiar (ENDES) la variable nominal "**unión conyugal**" tiene asociada la pregunta ¿Actualmente esta Ud. Casada o conviviendo? Sus categorías son: "si, actualmente casada"; "si, conviviendo", y la última "no, en unión", que se sustituyen por los números 1, 2 y 3, respectivamente.

**2. La escala ordinal.** Permite clasificar los objetos, hechos o fenómenos en forma jerárquica. Esta escala de medición facilita el ordenamiento de las diferentes categorías de la variable según su magnitud o importancia relativa, pues cada categoría no representa la misma cantidad de la variable. La escala ordinal carece de las propiedades de distancia y origen natural.

Ejemplo: El Nivel Educativo Aprobado por una persona, se expresa en solo una de las categorías posibles de esta variable: inicial, primaria, secundaria o superior.

Código	Nivel Educativo Aprobado
1	Inicial
2	Primaria
3	Secundaria
4	Superior No Univers.
5	Superior Universitaria

En la tabla se aprecia el ordenamiento de las categorías de la variable nivel educativo aprobado, debiendo tener cuidado al interpretar estos resultados, por ejemplo estar en un nivel superior implica haber pasado por los niveles anteriores, y por tanto existe un ordenamiento implícito pero se desconoce la cantidad de diferencia entre las categorías de la variables (magnitud de conocimientos). La mediana y la moda, son las medidas estadísticas más apropiadas para este tipo de escalas.

El cuestionario de la ENDES contiene algunas preguntas que incluyen variables medidas en escala ordinal, por ejemplo: " En promedio ¿Cuánto de los gastos de su hogar **se pagan** con lo que **usted gana**: casi nada, menos de la mitad, la mitad, más de la mitad, todo? " En esta pregunta es implícito el ordenamiento de las posibles respuestas de la variable: ingresos de la mujer medido a través de los gastos de su hogar.

En ninguno de los ejemplos citados se incluye como categoría de la variable el "cero", lo que nos lleva a presentar la siguiente escala.

**3. La escala ordinal con origen natural.** Característico de las variables ordinales que tienen además la propiedad de origen natural (incluye el cero) pero carecen de distancia. Ejemplo: El Índice de Desarrollo Educativo de la Niñez, se expresa en un puntaje (calificación), que está entre 0 y 20. Este puntaje permite configurar un ordenamiento de las unidades de análisis. En este caso el cero necesariamente tiene que ser el origen de la escala, mientras que el 20 significa un desarrollo educativo completo. Tener un puntaje de 10 no implica tener la mitad del desarrollo educativo, el número expresa simplemente una ubicación. En la ENDES encontramos la siguiente pregunta ¿Con qué frecuencia Ud. habló con su esposo compañero acerca de la planificación familiar en los últimos 12 meses? Y sus posibles respuestas son: Nunca, Algunas veces, Muy a menudo. En este caso la variable: habla con su esposo acerca de la planificación familiar esta expresada en una escala ordinal con origen natural por cuanto sus categorías pueden ordenarse e iniciarse con cero.

**4. La escala de intervalo.** Llamada también interválica, es aquella que agrupa a las variables cuantitativas, con las propiedades de orden, distancia y un origen no natural. Ejemplo: La edad de las personas, se expresa en días, meses o años. En este caso, tiene sentido afirmar que un individuo tiene el doble de la edad que otro, de modo que puede aplicarse

el criterio de distancia y por tanto, calcularse estadísticas como el promedio y la desviación estándar. Por ejemplo en la ENDES, mediante la pregunta ¿Cuántos meses de embarazo tiene? Se estudia la variable tiempo de embarazo, y sus posibles respuestas son expresadas en meses.

**5. La escala de proporción o razón**  
Este tipo de escala constituye el nivel más alto de medición para las variables cuantitativas. Contiene las características de una escala de intervalo con la ventaja adicional de poseer el cero absoluto. Es importante mencionar que el cero representa la nulidad o ausencia de lo que se estudia. Ejemplo: la proporción de analfabetos, reúne las dos propiedades mencionadas: origen natural (cero en el valor de la variable implica que no hay analfabetos) y distancia (la proporción de analfabetos hombres es la tercera parte de las mujeres). En el cuestionario de la ENAHO, se incluye la pregunta ¿Cuántas horas trabaja a la semana? Y sus respuestas se expresan en horas. En caso de no haber trabajado en el período de referencia, el cero expresa la propiedad de no haber trabajado y a su vez es el origen natural de la escala.

### 3.3 EVALUAR LAS DEFINICIONES OPERACIONALES.

Los indicadores de uso más frecuente en la investigación social son las tasas y porcentajes, por lo cual en la presente guía se presentan sus definiciones:

*Revisar las definiciones operacionales implica: analizar la fórmula utilizada para obtener el indicador y evaluar el algoritmo para operacionalizar el concepto*

**Porcentaje.-** Proporción del total expresado en unidades porcentuales. Tanto el numerador como el denominador están referidos al mismo tipo de información.

La siguiente fórmula resume la definición:

$$\text{Porcentaje} = \frac{\text{Frecuencia}}{\text{Total}} \times 100$$

Ejemplo: el porcentaje de analfabetos adultos mayores, se determina calculando el número de analfabetos mayores de 65 años entre el número total de analfabetos. Dicho cociente se multiplica por 100. Se

puede apreciar que tanto el numerador como el denominador son de la misma naturaleza.

**Tasa.-** La palabra tasa se emplea generalmente para designar la frecuencia relativa con que un hecho o suceso se presenta dentro de una población. Todas las tasas son razones o proporciones. El numerador y el denominador no necesariamente son de la misma naturaleza. La base numérica estándar puede ser 10, 100 o 1000, y con ello eliminamos los decimales para facilitar la interpretación y una más rápida comprensión.

La siguiente fórmula resume la definición:

$$\text{Tasa} = \frac{\text{Frecuencia de la variable cuya variación se quiere saber}}{\text{Frecuencia de la variable de referencia}} \times \text{Base Numérica}$$

**Ejemplo:** La Tasa de Analfabetismo se calcula como una relación entre el número de personas mayores de 15 años que no

leen ni escriben, entre el número total de personas mayores de 15 años. El resultado se multiplica por 100.

$$\text{Tasa de Analfabetismo} = \frac{\text{Población de 15 y más años que no sabe leer ni escribir}}{\text{Población total de 15 y más años}} \times 100$$

### 3.4 COMPARAR CON LAS FORMAS DE MEDICIÓN DE OTROS PAÍSES

En esta etapa se requiere comparar los indicadores calculados con los de otros

países, cuyas características socioeconómicas, culturales y geográficas son similares a la nuestra, de tal manera que la visión de la realidad sea semejante a la nuestra.



Ejemplo: La tasa de analfabetismo en otros países del continente se calcula de la siguiente forma:

1. En México se calcula la tasa de alfabetismo como la relación entre la población de 15 años y más que sabe leer y escribir entre la población total en dicho grupo de edad. El analfabetismo se determina como complemento.

2. En Chile se mide el analfabetismo, aunque la tasa tiene una forma de cálculo un tanto diferente, por cuanto a la población proyectada analfabeta de 15 y más años se le resta el número de promovidos por el programa de alfabetización y esta diferencia se divide entre el total de la población proyectada en dicho grupo de edad.

**3.5 ANALIZAR LAS VENTAJAS Y DESVENTAJAS DEL INDICADOR**

En esta parte deben evaluarse las ventajas y desventajas de la medición de las variables mediante tasas y porcentajes.

Entre las ventajas tenemos:

- Las tasas y porcentajes son fáciles de construir
- Las tasas y porcentajes son fáciles de operar
- Cuando las tasas se presentan combinadas con otras variables permiten una mejor evaluación de la situación.

Algunas de las desventajas son:

- Los porcentajes solamente permiten representar una ordenación de las unidades de análisis
- Los porcentajes y tasas esconden variaciones en los cambios absolutos, ya que las cifras relativas pueden variar en un sentido diferente a la variación de las cifras absolutas. Por ejemplo:

Población Total (miles)	Población Analfabeta (miles)	Tasa de Analfabetismo
20 000	2 000	10,0
20 200	2 000	9,9

Mientras que las cifras relativas (tasa de analfabetismo) muestran que la tasa de analfabetismo ha disminuido, las cifras absolutas (población analfabeta) no se han modificado.

**Ejemplo:** En relación a la tasa de analfabetismo encontramos, las ventajas y desventajas siguientes:

**Ventajas :**

1. Es posible presentar el indicador anualmente.
2. El indicador se obtiene fácilmente a partir de la combinación de dos preguntas en la Encuesta de Hogares, ¿Sabe leer y escribir? , y Nivel Educativo Alcanzado. Esto no

- excluye la posibilidad de generar algunas preguntas filtro adicionales.
3. Puede calcularse a partir de otras encuestas y así evaluar la consistencia de la estimación hallada.
  4. Es posible brindar las estimaciones puntuales del indicador, con su correspondiente intervalo de confianza.
  5. Es de fácil interpretación y de amplia aceptación por la comunidad de investigadores.

**Desventajas:**

1. Puede ocasionar interpretaciones incorrectas cuando junto al indicador calculado no se incorporan los valores absolutos.
2. La cobertura limitada de las encuestas por muestreo afecta más a los indicadores expresados como tasas (analfabetismo), generándose los errores de sub-cobertura en las estimaciones.
3. Existe una alta probabilidad de encontrar sesgos de respuesta, por cuanto el indicador se calcula a partir de las respuestas del informante a la pregunta sabe leer y escribir y nivel educativo. Estos sesgos pueden controlarse con el desarrollo de métodos para comprobar la veracidad de las respuestas.

4. A fin de realizar un seguimiento de la evolución del indicador se requiere una medición longitudinal y por tanto la aplicación de Encuestas Panel.

**Ejemplo:** Las ventajas y desventajas de la tasa de desempleo, pueden analizarse de la siguiente forma:

**Ventajas :**

1. Es de fácil obtención porque es un cociente entre dos variables cuantitativas: número de personas desempleadas y Población Económicamente Activa (PEA).
2. Las estimaciones de la PEA pueden controlarse por el factor de expansión.

**Desventajas:**

1. Considera las personas que en la fecha de la encuesta pertenecen a la PEA, existiendo la probabilidad que en fecha posteriores algunas personas salgan o se incorporen a ella, por lo que puede subvalorarse o sobrevalorarse la estimación de la tasa de desempleo.
2. El desempleo es una variable dinámica y la medición es un momento específico.



## CAPITULO IV

### TECNICAS ESTADISTICAS Y DEMOGRAFICAS PARA EL ANALISIS DE DATOS

Los **datos** son insumos para la obtención del indicador, por tanto además de conocer su naturaleza, forma de obtención y procedencia, es importante aplicar algunas herramientas analíticas estadísticas y demográficas de control a fin de obtener estimaciones de calidad.

Estos datos pueden proceder de un conteo total de la población en estudio, en cuyo caso el indicador calculado a partir de ellos se denomina **parámetro**, cuyos sesgos de cobertura son mínimos. También los datos pueden provenir de una muestra. En este caso se generan **estimaciones** del indicador y llevan implícitos sesgos de cobertura, que se controlan con el diseño muestral. En estadística se define el sesgo para un indicador como la diferencia entre

el valor del parámetro (obtenido del conteo total de la población) y la estimación (obtenida de una parte de la población).

Otra fuente de datos son los registros administrativos. En ellos las entidades responsables como el Ministerio de Educación, Salud, Registros Públicos, entre otros, acopian la información del evento en formatos y fichas. Los errores mas frecuentes en este caso, se originan como consecuencia de una contabilidad mal llevada (formularios no actualizados, omisión de registros), carencia de procesos de actualización, entre otros, haciendo necesario la generación de mecanismos para su control y entrega oportuna.

*Para evaluar los indicadores deben generarse técnicas de control de datos para contrastar las hipótesis relacionadas con la variable de interés sin que intervengan factores ajenos a la medición.*

Las técnicas de control hacen posible la observación sistemática y el análisis de los experimentos sociales permitiendo comparar objetivamente los datos obtenidos. En esta comparación debe tenerse en cuenta la temporalidad del indicador, es decir, la información debe corresponder a los mismos períodos de tiempo y los métodos de obtención de los datos también deben ser similares (encuestas con características similares).

Además tener presente las diferencias existentes cuando los datos provienen de encuestas paneles.

Cuando se tienen los datos dispuestos en una serie temporal, pueden calcularse las **diferencias simples**, entre las tasas o porcentajes calculadas. Si estas diferencias simples se dividen por el valor del indicador en el año inicial se tienen las **tasas de cambio** para la serie.

Una medida que resulta de dividir las diferencias simples entre el complemento del indicador en el año inicial (100 - valor inicial) se denomina **índice de efectividad**.<sup>2</sup>

**Ejemplo:** La tabla siguiente muestra las tasas de analfabetismo calculadas para diferentes años:

Año	Tasas Calculadas	Diferencias Simples	Tasas de Cambio (%)	Índice de Efectividad (%)
1996	10.6			
1998	11.3	0.7	6.6	0.78
2000	10.7	-0.6	-5.3	-0.67

**Interpretación:** Las diferencias simples muestran la variación absoluta en las tasas, en este caso corresponde un incremento de más de medio punto porcentual en el año 1998 con respecto al año 1996 y una caída de más de medio punto porcentual del año 2000 con respecto a 1998.

Los diferentes valores obtenidos implican que existen diferentes formas de medir los cambios temporales del indicador, debiendo escogerse la más adecuada al tipo de análisis que se pretenda.

Las tasas de cambio muestran que en el año 1998 la tasa de analfabetismo aumentó en 6.6% con respecto a 1996, mientras que en el año 2000 la tasa disminuyó en 5.3% con relación a 1998.

**La utilidad del índice de efectividad:** Su utilidad se aprecia mejor cuando se necesita evaluar la eficiencia de alguna técnica, programa social o tratamiento (método de aprendizaje, tratamiento de salud, etc.).

El índice de efectividad señala que en 1998 la tasa aumentó en 0.78% con relación a su variación potencial del año 1996 y en el 2000 cayó en 0.67% con relación al cambio potencial de 1998.

En el cuadro adjunto se muestran los resultados de la aplicación de un Programa Social en dos zonas diferentes, encontrándose los siguientes resultados :

	Grupo Experimental	Grupo de Control
Tratamiento	Si	No
Situación Inicial	15	75
Situación Final	30	85

<sup>2</sup> El índice de efectividad es una medida propuesta por Hovland, C en su documento "A base line for the measurement of percentage change".

Al evaluar los cambios entre la situación final y la inicial, para el grupo experimental y el de control encontramos las tasas de cambio de 100% y 13% respectivamente. Estas medidas prescinden del cambio potencial. El grupo experimental parte de 15 y su incremento potencial es 85, mientras que el grupo de control parte de 75, y entonces su incremento potencial es menor (25). De esta manera al calcular el índice de efectividad encontramos que en el primer caso el cambio es sólo de 17.6%, mientras que en el segundo caso es de 40%.

Otro procedimiento de control para evaluar los datos dispuestos en series temporales es el **Índice de Relativos**. Este índice se obtiene al dividir dos categorías relacionadas de la variable y comparar el resultado para otros años, si este cociente se mantiene constante, entonces es una señal de la consistencia del indicador. Se asume que el fenómeno social en este período no se altera drásticamente.

**Ejemplo:** En el cuadro se muestran cifras hipotéticas desagregadas por sexo para el período 1998 - 2000

	Año 1998	Año 1999	Año 2000
Tasa	8	10	14
Hombre	4	6	7
Mujer	12	18	20
Índice de Relativos	0.33	0.33	0.35

**Interpretación:** A pesar de los incrementos de las tasas (total y por sexo), el valor del índice de relativos se mantiene constante a lo largo del período 98-2000, de modo que los cambios en el volumen no han afectado las diferencias por sexo.

**Utilidad:** El índice de relativos se construye a partir de una serie temporal en que los datos han sido desagregado por principales características (sexo, área, grupos de edad, entre otros), permite apreciar que a pesar de las variaciones en el comportamiento del indicador a través del tiempo, la **composición estructural** (Relación mujer/hombre en el ejemplo), permanece constante.

y demográfico de las variables mediante diagramas y gráficos y la generación de estadísticas de tendencia central y dispersión, a fin de hallar la distribución de los valores de las variables y detectar la probable presencia de valores discordantes. En este análisis se emplean herramientas más estandarizadas que las presentadas anteriormente. Así, se utilizan el promedio, la moda y la mediana entre las medidas de tendencia central. Entre las medidas de dispersión se recurren a: la varianza, la desviación estándar y el coeficiente de variabilidad, entre otras. La composición de la población por edad y sexo se analiza mediante las pirámides poblacionales.

#### 4.1 REALIZAR EL ANÁLISIS ESTADÍSTICO Y DEMOGRÁFICO

La evaluación de los datos debe complementarse con un análisis estadístico

Los procesos de inferencia, por los cuales se extienden los resultados de la muestra a la población exigen el cumplimiento de ciertos requisitos, que se explican en las

siguientes secciones con bastante detalle, tanto para la construcción de los intervalos de confianza como la elaboración de las pruebas de hipótesis.

*Las principales herramientas del análisis estadístico gráfico-descriptivo son: el histograma, el diagrama de tallos y hojas y el gráfico de cajas.*

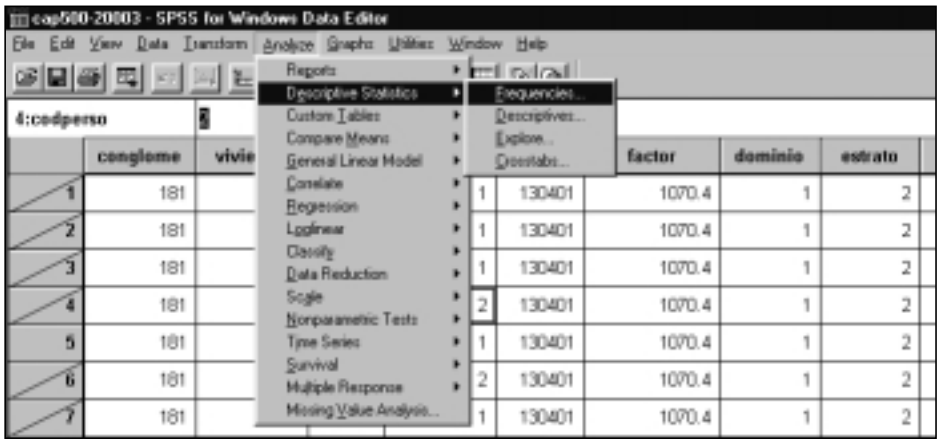
Con el software SPSS se generan las estadísticas mencionadas, para lo cual debe ubicar la ventana Analyze del menú principal.

**4.1.1 Tabla de Frecuencias**

Es un modo de agrupar los datos en una tabla. En esta tabla se muestran los valores de la variable en estudio, tanto en

frecuencias simples como relativas. El término frecuencia indica el número de veces que se repiten los valores de la variable. La variable puede haberse medido en cualquiera de las escalas mencionadas.

**Modo de Obtención:** En el software SPSS, la tabla de frecuencias se obtiene del modo siguiente:



**Utilidad:** Permite mostrar los valores para las diferentes categorías de la variable. Además en esta tabla se muestran los valores perdidos "missing" (NEP) cuyo número debe ser el menor posible de modo que no se invaliden los procesos de inferencia.

**Ejemplo practico:** El módulo de Empleo de la ENAHO, incluye la pregunta "¿En su trabajo, negocio o empresa incluyendo usted cuántos laboraron?". La siguiente tabla de distribución de frecuencias resume los resultados.

**En su trabajo negocio o empresa incluyendose Ud.laboraron:**

		Frequency	Percent	Valid Percent
Valid	Menos de 100 personas?	9821296	85.9	86.7
	De 100 a 499 personas?	315617	2.8	2.8
	De 500 y mas personas?	1191350	10.4	10.5
	Total	11328263	99.1	100.0
Missing	9 NEP (*)	4823	.0	
	System	100335	.9	
	Total	105158	.9	
Total		11433421	100.0	

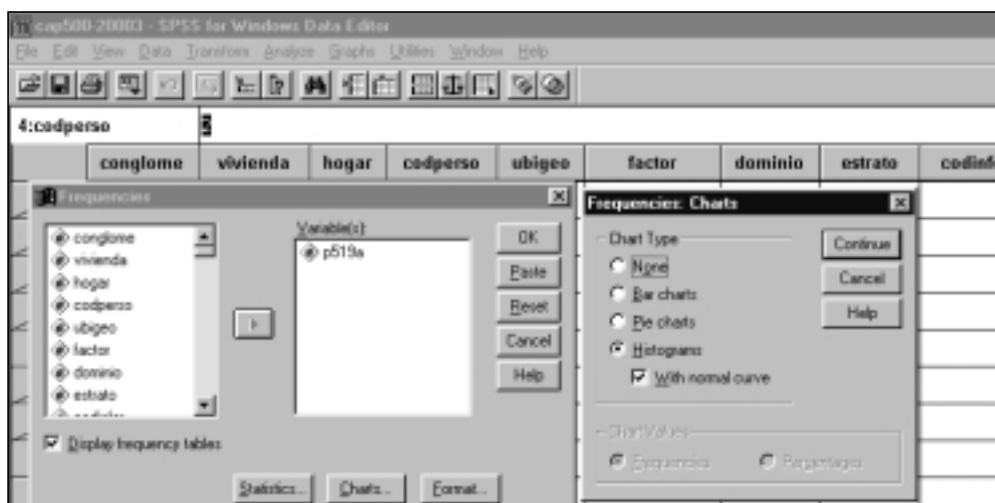
**Análisis e Interpretación:** Se observa que la mayor frecuencia de casos se da entre las empresas que laboran con menos de 100 personas, mientras que es menos probable encontrar personas laborando en empresas de 100 a 499 personas.

fenómeno estudiado queda representado por una serie de rectángulos semejantes a los del diagrama de barras; sin embargo las barras de histograma se colocan sólo verticalmente y deben ir uno al lado de las otras sin que haya un espacio que las separe. La base de cada rectángulo es la amplitud de la clase de la variable correspondiente.

**4.1.2 Histograma**

Es la representación gráfica de la tabla de frecuencias. Muestra las distribuciones de frecuencias absolutas o relativas de datos expresadas en cualquiera de las escalas mencionadas en la sección anterior. El

**Modo de Obtención:** Puede obtenerse en el SPSS, ingresando a la opción Analyze del menú principal, y seleccionando las ventanas que a continuación se muestran:





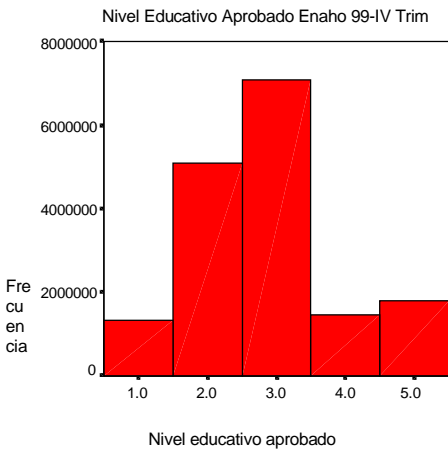
**Utilidad:** Permite comparar visualmente dos distribuciones de efectos diferentes y deducir la forma de la distribución de los valores, conocer los valores centrales y la dispersión entre las categorías de la variable en estudio.

Este tipo de herramienta se aplica frecuentemente a las variables de tipo ordinal y de intervalo.

**Explicación:** En el histograma se representan en el eje horizontal las

categorías de la variable mientras que en el eje vertical quedan representadas las frecuencias correspondientes a cada categoría. La barras que se levantan por encima del eje horizontal representan el número de casos que hay en cada categoría.

**Ejemplo practico:** El gráfico representa la distribución de frecuencias del Nivel Educativo de la población de 15 años y más (ENAH0 1999-IV Trimestre).



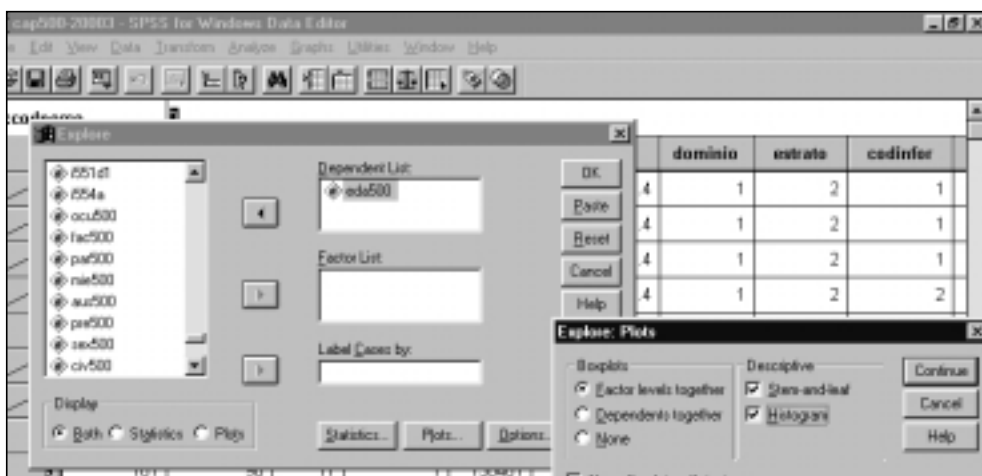
Codigo	Nivel Educativo Aprobado
1	Inicial
2	Primaria
3	Secundaria
4	Superior No Univers.
5	Superior Universitaria

**Análisis e Interpretación:** Se observa la mayor frecuencia de casos en el nivel de secundaria (moda de la distribución). La relación entre el nivel mas frecuente (Secundaria) y el menos frecuente (Inicial) es aproximadamente cuatro a uno, lo cual implica una alta dispersión. La distribución del histograma muestra que el nivel educativo predominante en la población del Perú, es secundaria. Por otro lado, su forma es muy semejante a la distribución normal (curva de Gauss), lo cual otorga mayor validez a las inferencias que posteriormente se hagan.

**4.1.2 Diagrama de Tallos y Hojas**

Es un procedimiento semi-gráfico (tabular y gráfico) de presentar la información. Los datos se disponen en un gráfico semejante a un árbol, facilitando su observación.

**Modo de Obtención:** Puede obtenerse en el SPSS ubicando la opción Analyze del menú principal, luego debe ubicar Explore, en Descriptive seleccionar Stem and leaf y el Box Plot (gráfico de cajas):



**Utilidad:** Permite detectar los valores extremos y la forma como están distribuidos los datos. Esta herramienta se aplica generalmente a las variables de tipo cuantitativo e interválicas.

**Ejemplo practico:** La variable edad de la población de 15 años y más, se compone de dos dígitos, el primer dígito es la cifra de las decenas mientras que el segundo corresponde a las unidades. Así en el árbol, el **tallo** representa las decenas, mientras que las **hojas** las unidades. De este modo cada tallo, según el dígito inicial conforma una clase, la cual se compone de un conjunto de hojas. El número de **hojas** es la frecuencia de dicha clase. El diagrama de tallos y hojas que se presenta, corresponde a las edades de la población de 15 años y más de la ENAHO 1999-I Trimestre.

**Explicación:** La longitud de cada fila corresponde al número de casos que hay dentro del intervalo.

Cada caso es representado dentro de la fila con un valor numérico que

corresponde al valor observado. En el diagrama el tallo lo constituye el primer dígito (base), que para el ejemplo es (1, 2, ....8) y las hojas corresponden a las unidades (5,6,7,8,9).

**Análisis e Interpretación:** En la primera fila se tienen 3,471 casos que caen dentro del intervalo de los valores 15-15...16...16...17...17...18...19...(15 a 19 años)

Cada hoja "Each-leaf 78": significa que el programa SPSS ha determinado que cada hoja sea equivalente a 78 casos: Así en la primera fila del diagrama encontramos 702 casos de edades de 15 años, 702 de 16 años, 702 de 17 años, 700 de 18 años y 546 de 19 años

El diagrama muestra que la distribución de las edades de la población de 15 años y más está concentrada en los grupos de menor edad (15-19 años). Se han encontrado 161 casos de personas que declaran tener 85 años y más de edad.

Frecuencia Tallo & Hoja  
 No de Casos Base  
 3471.00 1 . 555555556666666677777777888888889999999  
 2996.00 2 . 00000000111111122222223333334444444  
 2464.00 2 . 55555566666677777788888899999  
 2136.00 3 . 000000111122222333334444  
 2088.00 3 . 5555566666777788888899999  
 1658.00 4 . 00000111222223333444  
 1351.00 4 . 5556667778888999  
 1153.00 5 . 00001122233444  
 885.00 5 . 555667788899  
 812.00 6 . 0001223344  
 628.00 6 . 55567889  
 464.00 7 . 001234  
 328.00 7 . 5688  
 166.00 8 . 08  
 161.00 Extremes (>=85)  
 Stem width: 10  
 Each leaf: 78 case(s)

**Comparación entre la Tabla de Frecuencias y el Diagrama de Hojas y Tallos**

En el siguiente ejemplo se tienen las edades de los desempleados representados en dos formas distintas. Como tabla de frecuencias:

Edad	Frecuencia	Porcentaje
14	13	1.4
15	20	2.2
16	27	3.0
17	32	3.6
18	53	5.9
19	56	6.2
20	43	4.8
...	...	...
60	5	0.6
61	4	0.4
63	3	0.3
64	5	0.6
...	...	...
80	1	0.1
<b>Total</b>	<b>897</b>	<b>100.0</b>

Y también como un diagrama de tallos y hojas:

**Gráfico de Tallos y Hojas: Edad de los Desempleados**

Frequency Stem & Leaf  
 13.00 1 . 444  
 188.00 1 . 555566667777778888888899999999999  
 189.00 2 . 000000001111112222223333344444444  
 141.00 2 . 555555666666777778888899999  
 92.00 3 . 000001112222333344  
 74.00 3 . 55566677888999  
 50.00 4 . 0011222334  
 39.00 4 . 556778899  
 37.00 5 . 0112334  
 33.00 5 . 5667889  
 17.00 6 . 0134  
 24.00 Extremes (>=65)  
 Stem width: 10  
 Each leaf: 5 case(s)

Como puede apreciarse, tanto en la tabla como en el diagramas existen 13 casos de desempleados con edad de 14 años, 188 casos con edades entre 15 y 19 años, y así sucesivamente. El número total de casos es de 897. La combinación de un tallo con una hoja representa aproximadamente 5 casos (each leaf: 5 case (s)). El número exacto de casos dentro de cada categoría de edad lo proporciona la columna "frequency", cuya suma arroja el total de casos.

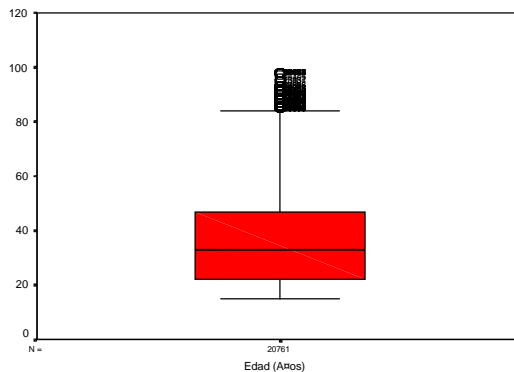
#### 4.1.3 Grafico de Cajas (Box-Plot)

Es otro modo de presentar en forma resumida el conjunto de datos. Nos da una idea del grado de concentración de los datos. La caja se conforma a partir de los

cuartiles inferior y superior (percentil 25 y 75 respectivamente) y la mediana (percentil 50). Esta caja rectangular esta alineada ya sea horizontal o verticalmente y se extiende desde el cuartil inferior al superior siendo atravesada de un lado a otro por la mediana. A partir de los extremos de la caja se extienden líneas ("bigotes") hasta los valores mínimo y máximo.

**Utilidad:** Permite deducir la dispersión de los valores de la variable y su grado de simetría.

**Ejemplo:** El gráfico de cajas corresponde a la variable Población de 15 años y más (I Trimestre ENAHO 99).



**Análisis e Interpretación:** La distribución de la población de 15 años y más es simétrica porque la mediana ocupa el centro de la caja, por encima del límite máximo se observan un conjunto de

valores extremos. (población con edades superiores a los 85 años).

#### 4.1.4 Pirámides de Población

*La Pirámide Demográfica es un recurso gráfico para representar la estructura de la población por edad y sexo en un período específico que permite comparar la población obtenida de la Encuesta con la población estimada en base a proyecciones del Censo.*

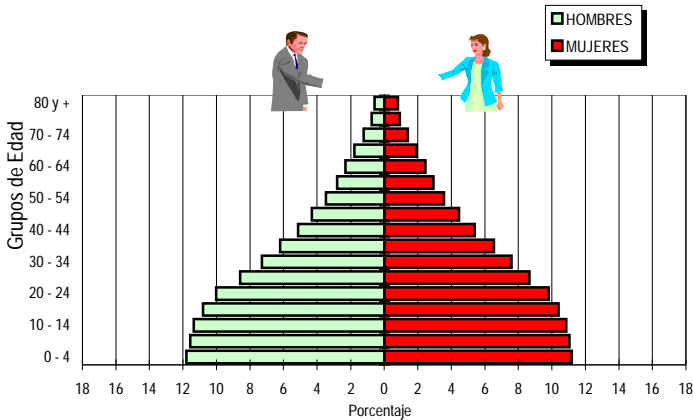
Una pirámide muestra gráficamente la distribución por edad y sexo de la población en un periodo específico. En el eje vertical se indican los grupos de edad, desde los de menor edad que están en la base, hasta los adultos mayores que están en la cima, mientras que en el eje horizontal se indica el porcentaje de cada grupo de edad respecto del total. De este modo podemos comparar las proyecciones poblacionales y asegurar que cada grupo poblacional este presente en la muestra.

**Utilidad:** La pirámide al presentar el número proporcional de hombres y mujeres en cada grupo de edad, nos da la idea de las características de la población.

**Ejemplo:** La pirámide siguiente representa la Estructura de Población por Edad y Sexo para 1999, según datos correspondientes a las proyecciones de población.

**Análisis e Interpretación:** Puede notarse el engrosamiento de la pirámide conforme se desciende en los grupos de edad, síntoma característico de una población expansiva y con elevada tasa de crecimiento poblacional. La amplia base de la pirámide alude a una población joven, es decir con una mayor población entre las edades de 0-24 años de edad.

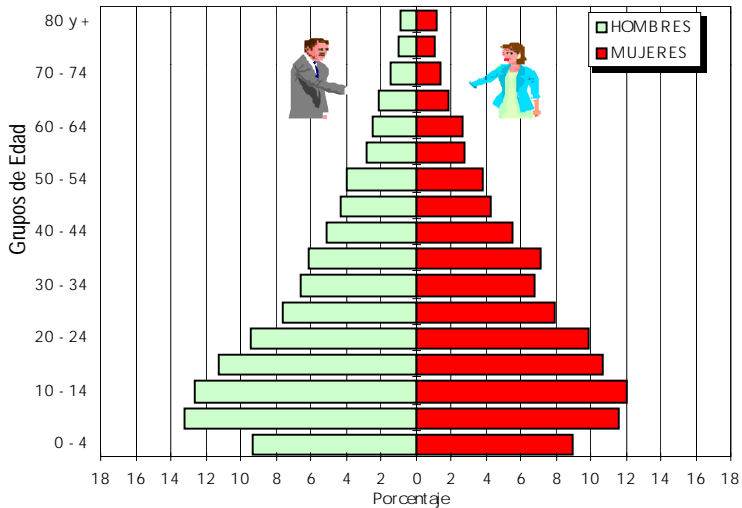
PERU: ESTRUCTURA DE LA POBLACION POR GRUPOS DE EDAD Y SEXO: 1999-PROYECCIONES



**La utilidad analítica:** su utilidad analítica se da cuando se compara la estructura de población por grupo de edad y sexo de las Proyecciones con la estructura de población determinada a partir de la Encuesta de Hogares. En las pirámides mostradas puede apreciarse la composición poblacional diferente. En la

encuesta los grupos quinquenales de la base de la pirámide son menores que los correspondientes a la pirámide de las proyecciones de población. Ello indicaría que el recojo de datos no está registrando los niños menores de cuatro años. Hay que tener presente que este análisis debe ser realizado para periodos iguales de tiempo.

**ESTRUCTURA DE LA POBLACION POR GRUPOS DE EDAD Y SEXO: 1999-I - ENCUESTA DE HOGARES**



**4.1.5 Estadísticas de Tendencia Central y Dispersión**

Los gráficos brindan una descripción de los datos, permitiendo a partir de ellos deducir la forma de la distribución de los mismos, pero los **procesos de inferencia** no podrían desarrollarse sin las medidas estadísticas de tendencia central y dispersión. Estas medidas, que se determinan de acuerdo al tipo de variable, en el caso de las tasas y porcentajes (proporciones) tienen en la proporción

muestral la mejor medida de tendencia central, y en el caso de las medidas de dispersión tienen en la varianza muestral y el coeficiente de variabilidad, las medidas de dispersión de uso más frecuente.

Los indicadores presentados sintetizan el conjunto de datos constituyendo el paso previo a la generación de pruebas de hipótesis y construcción de los intervalos de confianza necesarios para generalizar los resultados a la población.

*Las estadísticas de tendencia central y dispersión son medidas numéricas resumen de los datos y por tanto indicadores que permiten analizar la calidad de los resultados y su generalización mediante las pruebas de hipótesis e intervalos de confianza.*

Estas medidas en el caso de las tasas y porcentajes estas medidas, tienen las siguientes formas de cálculo:

**4.1.5.1 Tendencia Central**

Las medidas de tendencia central son aquellas que además de ser

representativas del conjunto de datos nos indican el centro de la distribución. En el caso de las tasas y porcentajes, el mejor estimador -máximo verosímil- para la proporción de unidades de análisis con cierta característica (p), viene dado por la siguiente relación:

$$p = \frac{a}{n}$$

donde

- p : proporción de unidades de análisis con cierta característica
- a : número de unidades de análisis que tienen la característica
- n : número total de personas en la muestra

#### 4.1.5.2 Dispersión

Las medidas de dispersión son aquellas que nos indican la variabilidad de los datos, siendo las más usuales la varianza, la desviación estándar y el coeficiente de variación.

**La varianza**, para la proporción de unidades de análisis (p) con cierta característica (a) tiene la siguiente forma de cálculo:

$$\text{Var}(p) = \frac{(a/n)(1-a/n)}{n}$$

y la raíz cuadrada de la varianza se denomina desviación estándar o error estándar (DE).

**El Coeficiente de Variación (CV)**, relaciona las dos medidas anteriormente explicadas mediante la siguiente expresión:

$$\text{CV} = \frac{\text{DE}}{E(p)} \times 100$$

Así, el coeficiente de variación es un cociente que resulta de dividir el error estándar entre el valor esperado del estimador E(p) y multiplicar dicho cociente por 100. Es una medida de dispersión relativa, que normaliza la desviación estándar y hace posible comparar otras distribuciones de frecuencia expresadas en unidades diferentes. Su interpretación se explica en la siguiente sección.

Ejemplo: Para ilustrar la forma de cálculo e interpretación de la proporción, varianza y el coeficiente de variabilidad, se presenta el siguiente ejemplo en el cual se tiene una muestra aleatoria de 225 personas mayores de 15 años de las cuales 150 son analfabetos. Se necesita calcular la proporción de analfabetos, y una medida de variabilidad.

Utilizando el estimador máximo verosímil de la proporción y las formas de calculo presentadas, obtenemos el cuadro siguiente:

Tamaño de Muestra (n)	Numero de Analfabetos (a)	Proporción de Analfabetos E(p) = p = (a) / (n)	1 - p	Varianza Var (p)	Desviación Estándar (DE)	Coefficiente de Variación (CV)
225	150	0.7	0.3	0.0014	0.037	5.6%

**Interpretación:** La proporción de analfabetos es de 0.7. Desde que se trata de una muestra aleatoria este valor puede interpretarse como la proporción media (o esperada) de analfabetos en la población. La desviación estándar es de 0.037 puntos porcentuales y su coeficiente de variabilidad es 5.6 por ciento. La desviación estándar es pequeña, apenas el 5.6% del valor esperado de la media, por lo cual la estimación es confiable.

#### 4.1.6 Análisis de los Errores Muestrales

Los procesos de muestreo y la consiguiente generación de estimaciones de los indicadores llevan asociados sesgos por cuanto se trabaja con una parte de la población y no con la totalidad de los elementos. En este sentido se recomienda presentar las estimaciones del indicador acompañadas con sus respectivos errores muestrales.

*El error de muestreo o error estándar es la raíz cuadrada de la varianza del estimador, y sirve para determinar la precisión del indicador obtenido de la muestra. Cuanto más pequeño sea este valor la calidad del estimador es mejor.*

Además de los errores de muestreo, debe determinarse el coeficiente de variación, el efecto de diseño y los intervalos de confianza. Los dos primeros son medidas relacionadas con la calidad de las estimaciones, cuya fórmula de cálculo se explicó en la sección anterior. En cambio los intervalos de confianza son medidas de control, que establecen límites de una, dos o tres desviaciones estándar del parámetro, con una probabilidad de error de 0.1, 0.05 o 0.01 respectivamente.

El **coeficiente de variación**, se interpreta según el valor calculado esté ubicado en la siguiente escala: Si el coeficiente de variabilidad es 5% o menos, entonces las estimaciones son muy buenas. Si se encuentra en un rango de 5% a 10% las estimaciones son buenas. Entre 10% y 20% las estimaciones son aceptables. Si los coeficientes superan el 20% entonces las estimaciones no son confiables (solo referenciales).

Los **intervalos de confianza** se determinan a partir del error estándar y asumiendo un nivel de confianza de 95% (probabilidad de error de 0.05) se calculan, los límites superior e inferior. Su forma de cálculo se explica más detalladamente en la siguiente sección.

El **efecto de diseño**, se define como la razón de la varianza de la estimación obtenida a partir de la muestra más compleja a la varianza de la estimación obtenida a partir de una muestra aleatoria simple del mismo número de unidades. Así, cuanto más cerca de uno se encuentre el valor resultante de la división, implica que el diseño muestral ha contribuido a controlar los errores muestrales tan eficientemente como si lo habría hecho una muestra simple aleatoria y por tanto las estimaciones son de calidad.

Ejemplo: Continuando con el ejemplo anterior, en el cuadro adjunto se presentan los intervalos de confianza para la proporción calculada y se interpreta el coeficiente de variabilidad.



Proporción de Analfabetos	1.96 * (DE)	Límite Inferior de Confianza	Límite Superior de Confianza
0.7	0.07	0.63	0.77

La proporción de analfabetos está entre 0.63 y 0.77 con un 95% de confianza.

El coeficiente de variación, calculado anteriormente, fue de 5.6%, valor que se encuentra en el rango de las estimaciones buenas.

**Ejemplo:** La tasa de analfabetismo calculada, según la Encuesta Nacional de Hogares (ENAHOG), presenta los siguientes errores muestrales:

ERROR MUESTRAL DE LA TASA DE ANALFABETISMO, SEGÚN PRINCIPALES VARIABLES

Principales Variables	Tasa de Analfabetismo (%)	Error Standar (%)	Coeficiente Variación (%)	Intervalos de Confianza		Efecto Diseño
				Inferior (%)	Superior (%)	
PERU	11.2	0.3	2.54	10.6	11.7	4.2
Area Urbana	5.2	0.2	3.55	4.8	5.5	2.48
Area Rural	24.5	0.7	2.79	23.1	25.8	4.03

1/ Incluye la Provincia Constitucional del Callao  
 FUENTE: INEI- Encuesta Nacional de Hogares, Anual 1999

Las estimaciones de la tasa de analfabetismo correspondientes al área urbana y rural son buenas, según el coeficiente de variabilidad. El error estándar es pequeño lo que determina que los intervalos de confianza sean de amplitud reducida. Los valores del efecto de diseño indican que el diseño ha controlado mejor los errores muestrales en el área urbana a comparación del área rural.

**4.1.7 Estadística Inferencial**

Los indicadores determinados a partir del procedimiento de muestreo, además del error estándar llevan asociados una

distribución de probabilidad. En el caso de las estimaciones de tasas o porcentajes, su distribución debe compararse con la distribución normal. Para lograr este contraste el indicador calculado se transforma de tal manera que el nuevo valor tenga una distribución normal estándar. Los valores de esta distribución se extienden entre -4.5 y 4.5, tienen la propiedad de simetría y las probabilidades asociados a estos valores están completamente determinadas en tablas especialmente creadas (Tabla de Distribución Normal Estándar). De esta manera las inferencias estarán referidas ahora, a este nuevo valor del indicador.

*El proceso de inferencia estadística permite la extensión de los resultados a la población mediante las funciones de probabilidad determinadas a partir de los valores muestrales calculados.*

Las pruebas de hipótesis relativas a proporciones (tasas y porcentajes) son muy semejantes a las relacionadas con las medias, de una distribución continua. Si el tamaño de muestra es 30 o menos, el proceso de inferencia se basa en la distribución binomial. Si la muestra supera las 30 unidades de análisis, entonces los procesos de inferencia se basan en la distribución normal.

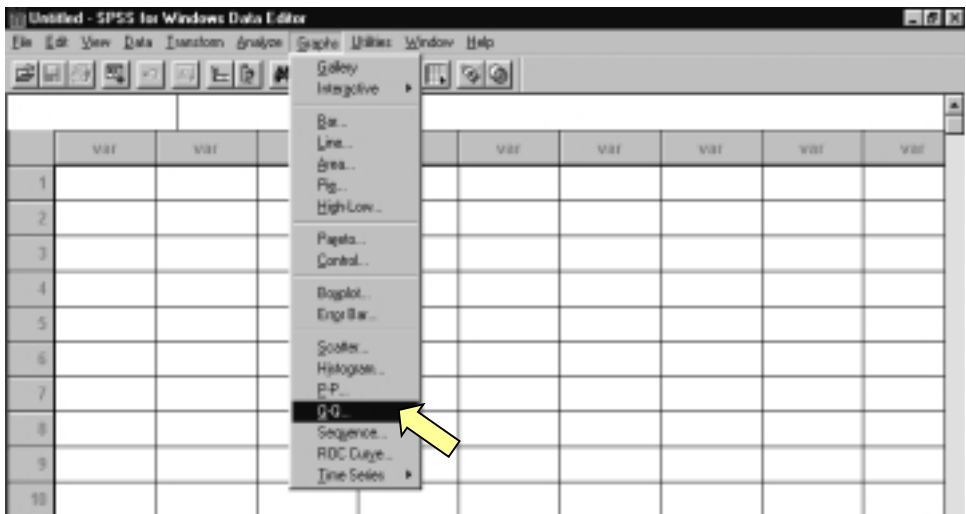
Las Encuestas de Hogares tienen un tamaño promedio superior a las 10,000 unidades, por lo que las variables en ellas contenidas, tienen las características de normalidad, favoreciendo los procesos de inferencia. El número de unidades se va reduciendo conforme se va desagregando los datos a nivel de departamentos, provincias y distritos, lo cual afecta las estimaciones.

De manera general, es importante comprobar si los datos proceden de una

distribución normal o cuasi-normal. Para verificar esta condición se recomienda utilizar en el SPSS, las siguientes pruebas:

1. Gráfico Normal QQ Plot. La normalidad queda determinada por la cercanía de los puntos a la recta.
2. Gráfico Detrended Normal Q-Q Plot. Se infiere normalidad si al observar los gráficos no se aprecia en su distribución un patrón definido, estando dispersos.
3. Prueba Kolmogorov-Smirnov. El cual arroja un coeficiente que en la medida de un menor valor es más evidente la normalidad.

Para efectuar estas pruebas, ubicar en el menú de barras del SPSS la siguiente ventana:



### 4.1.8 Prueba de Hipótesis

*Los posibles valores del indicador son contrastados mediante la conformación de una hipótesis nula la cual se acepta o rechaza en base a los resultados muestrales con una probabilidad de error determinada a priori.*

La prueba de hipótesis consiste en formular una hipótesis en torno al valor del indicador, a la que se denomina hipótesis nula quedando latente la opuesta a ésta a la que se denomina hipótesis alternante. En base a los resultados muestrales y fijada la probabilidad de error (probabilidad de rechazo de la hipótesis nula siendo esta cierta) se acepta o rechaza la hipótesis formulada. Los siguientes pasos, describen todo el proceso de inferencia:

1. Se fija un margen de error generalmente 1%, 5% ó 10%.
2. La forma del indicador determina la selección de la estadística de prueba, cuya distribución de probabilidad estará en función del tamaño de la muestra. Por ejemplo: la tasa de analfabetismo tiene en la proporción muestral su estadística de prueba y el tamaño de muestra de la Encuesta de Hogares, determina que la distribución de probabilidad es normal.
3. En la distribución de probabilidad normal, los márgenes de error determinan las regiones críticas de aceptación y rechazo de la Hipótesis Nula. Los límites de las regiones críticas vienen dados por los valores de las tablas de la distribución de probabilidad. Por ejemplo en el caso de la distribución normal se ubican estos valores en la Tabla de Distribución Normal Estándar.
4. La estadística de prueba se compara con los límites de las regiones, determinados a partir de la tabla, si es mayor que estos límites, entonces se rechaza la hipótesis nula. El SPSS muestra una probabilidad, la cual en caso de ser menor que la fijada a priori, también conduce al rechazo de la hipótesis nula.
5. La prueba concluye con el rechazo o aceptación de la hipótesis nula. En el lenguaje estadístico, si las pruebas determinan el rechazo de la hipótesis nula entonces se concluye la significación estadística de la estimación. Caso contrario se dirá que no hay evidencia suficiente para rechazarla.

**Ejemplo:** La función de distribución de probabilidades asociada a las proporciones para un tamaño de muestra grande es la normal estándar (Z). Dada la variable aleatoria p, esta se normaliza de la siguiente manera:

$$Z = \frac{p - P}{\text{Raiz } (p(1-p)/n)}$$

Donde:

- p : proporción de unidades de análisis con cierta característica hallados en la muestra de tamaño n
- P : proporción de unidades de análisis con cierta característica en la población

Si la Z calculada es superior a los límites tabulares, entonces se rechaza la hipótesis nula con un margen de error.

#### 4.1.9 Intervalos de Confianza

A fin de brindar confiabilidad estadística al indicador, se determinan los intervalos de confianza, los cuales se calculan a partir de la siguiente expresión:

$$P [ | p - P | < d ] = 1 - a$$

Donde:

- p : proporción de unidades de análisis con cierta característica hallados en la muestra de tamaño n
- P : proporción de unidades de análisis con cierta característica en la población
- d : máxima discrepancia aceptada
- a : probabilidad de error

Los intervalos de confianza (nivel de confianza del 5%) para las estimaciones halladas vienen dadas por las siguientes expresiones:

- **Limite superior de confianza:**

$$p + 1.96 \text{ raiz } (p (1-p) / n)$$

- **Limite inferior de confianza:**

$$p - 1.96 \text{ raiz } (p (1-p) / n)$$

#### 4.2 Evaluar la Consistencia de la Información

Las técnicas presentadas anteriormente han permitido evaluar la validez interna del indicador. Pero ¿Estamos seguros que la medida calculada reproduce el fenómeno social en estudio? Para dar una respuesta a ello, se hace necesario relacionar el indicador calculado con otro u otros, de tal manera que las diferentes categorías del indicador se relacionen con las categorías de otra variable. Este proceso de consistencia empieza por establecer una hipótesis nula (H<sub>0</sub>) relativa a la no existencia de una relación entre el indicador y la otra variable que a priori la consideramos relacionada. Luego de efectuar el cruce entre los dos indicadores, se genera la tabla de contingencia, punto de inicio del análisis.

*Una forma de evaluar la consistencia de un indicador es realizando una crosstabs (cruce), con otra variable altamente relacionada con la observada.*

La suma de los cuadrados de las diferencias entre los valores calculados del cruce de las variables y los valores de las frecuencias esperadas, determinan una estadística (Jí-cuadrado calculada). Este valor debe compararse con el Jí-Cuadrado tabular. Se infiere la existencia de la relación entre las variables, si el Jí-cuadrado calculado es mayor al Jí-Cuadrado tabular. Entonces se rechaza la hipótesis nula permitiendo afirmar que las variables no son independientes o están relacionadas.

La secuencia de pasos a seguir en esta prueba son:

1. Se fija un nivel de confianza, generalmente 1% , 5% o 10%.
2. Se escoge la estadística de prueba que según el tamaño de la muestra llevará una distribución de probabilidad determinada, en este caso la distribución Jí-Cuadrado.

3. Se compara la estadística calculada con la teórica, o se comparan las probabilidades calculadas con la fijada a priori. Si el valor calculado es mas grande que la estadística teórica o si la probabilidad calculada es menor a la teórica, entonces se rechaza la hipótesis nula.

**Aplicación conceptual:** El siguiente ejemplo corresponde a la tabla de contingencia generada a partir de dos métodos de enseñanza aplicados a un grupo de personas. Se quiere saber si hay relación entre el sexo y nivel educativo.

	Mejoraron	No Mejoraron	Total
Hombres	40	60	100
Mujeres	10	70	80
	50	130	180

Esta tabla de valores observados se contrasta con la tabla de valores esperados, los cuales se obtienen de la siguiente manera:

	Mejoraron	No Mejoraron
Hombres	$50 \cdot 100 / 180$	$130 \cdot 100 / 180$
Mujeres	$50 \cdot 80 / 180$	$130 \cdot 80 / 180$

Dando lugar a la siguiente tabla:

	Mejoraron	No Mejoraron
Hombres	27.78	72.22
Mujeres	22.22	57.78

Estos valores esperados son proporcionales al total de su fila y columna

En caso de no existir relación estos valores estarían muy cerca de los observados. Por ello, las diferencias se calculan mediante el algoritmo siguiente:

	Mejoraron	No Mejoraron
Hombres	$(40-27.78)^2/27.78$	$(60-72.22)^2/72.22$
Mujeres	$(10-22.22)^2/22.22$	$(70-57.78)^2/57.78$

Si se suman estos valores obtenemos el valor de la Ji-Cuadrado calculada, resultando: 16.7

Este valor calculado lo comparamos con el Ji-Cuadrado de tabla. Los grados de libertad para esta prueba están dados por la relación:  $(\text{fila}-1)(\text{columna}-1)$ . La probabilidad de error fijada es de 0.05.

Así se determina un grado de libertad y se ubica en la tabla con 0.05 de probabilidad un valor de 3.8. Como el Ji-Cuadrado calculado es mayor que el Ji-Cuadrado tabular, rechazamos la hipótesis que no hay asociación entre los métodos de enseñanza y el sexo de los participantes. Por lo tanto el sexo de los participantes influye en la captación de los métodos de enseñanza.

**Aplicación práctica:** El siguiente ejemplo se realiza con los datos de la Encuesta Nacional de Hogares, 2000 - III trimestre. Se formula la hipótesis nula (Ho):

el aprender algún oficio a través de la experiencia es independiente del sexo del entrevistado, asumiendo una probabilidad de error de 0.05.

**P592 Ha aprendido algún oficio a través de la experiencia en una empresa o taller? \*  
SEXO - Crosstabulation**

			SEXO		
			HOMBRE	MUJER	Total
P592 Ha aprendido algún oficio a través de la experiencia?	1 Si	Count	9233	5115	14348
		Expected Count	6960.6	7387.4	14348.0
	2 No	Count	7355	12490	19845
		Expected Count	9627.4	10217.6	19845.0
Total	Count		16588	17605	34193
	Expected Count		16588.0	17605.0	34193.0

La lectura de la tabla de contingencia muestra que del total de personas que aprendieron algún oficio a través de la experiencia, mas de la mitad son hombres, mientras que del total de personas que

no aprendieron algún oficio el 63% son mujeres.

Aplicando la prueba estadística de Ji-Cuadrado (Pearson Chi-Square) se tiene los siguientes resultados:

**Chi-Square Tests**

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	2482.558 <sup>b</sup>	1	.000		
Continuity Correction <sup>a</sup>	2481.466	1	.000		
Likelihood Ratio	2512.382	1	.000		
Fisher's Exact Test				.000	.000
Linear-by-Linear Association	2482.485	1	.000		
N of Valid Cases	34193				

a. Computed only for a 2x2 table

b. 0 cells (.0%) have expected count less than 5. The minimum expected count is 6960.62.

**Interpretación:** El Chi-Square Test arroja una probabilidad (Asymp. Sig. 2-sided) de 0.000, valor muy inferior a la probabilidad de error a priori fijada de 0.05. Por tal razón se rechaza la hipótesis nula. Entonces de los datos podemos inferir que el aprender un oficio depende del sexo de los entrevistados. Así, se concluye que los hombres aprenden más algún oficio a través de la experiencia que las mujeres.

### 4.3 Efectuar el Análisis Multivariado

La consistencia interna de los datos mediante las técnicas estadísticas presentadas anteriormente debe complementarse con el análisis multivariado. Este enfoque transversal permite analizar en forma simultánea las variables relacionadas con el indicador cuya consistencia se va determinar, mediante un modelo matemático.

*Una de las aplicaciones prácticas más importantes del análisis multivariado es que permite evaluar la consistencia del indicador mediante las diferentes técnicas de tratamiento simultáneo de las variables.*

Con el Análisis Multivariado se puede realizar un análisis simultáneo de las variables que configuran un fenómeno social, siendo sus usos más frecuentes: el Análisis de la Dependencia y el Análisis de la Interdependencia. En el primero se establece una relación de dependencia entre una (o varias) variable(s) con otra (u otras), encontrando sus aplicaciones más comunes en el análisis de regresión multivariado, el análisis de contingencia múltiple y el análisis discriminante. La interdependencia abarca desde la

independencia total hasta la colinealidad, es decir, cuando una de ellas es combinación lineal de las otras o es una función cualquiera de las demás variables, así se puede analizar tanto la interdependencia entre variables como entre individuos mediante el análisis factorial, el análisis de conglomerados o "cluster", el análisis de correlación canónica, el análisis de componentes principales y algunos métodos no paramétricos.

*La consistencia del indicador calculado puede determinarse a partir de su relación simultánea con otras variables explicativas del fenómeno social en estudio..*

El primer paso para efectuar el análisis multivariado es conformar la matriz de datos (n-filas y p-columnas). Las filas la

conforman las unidades de análisis y las columnas están conformadas por las variables.

**Ejemplos:** A partir de la ENAHO, se puede conformar muchas matrices de datos, de acuerdo a las diferentes entidades de la población objetivo, tal como a continuación se describe:

1. Unidad de Análisis: **hogar** (fila) **Variables** (columna): ingreso promedio del hogar, condición de pobreza, tamaño del hogar, gastos del hogar en salud, entre otras.
2. Unidad de Análisis: **vivienda** (fila) **Variables** (columna): tipo de pared, material predominante en los pisos,

número de cuartos, estado de la vivienda, etc.

3. Unidad de Análisis: **jefe de familia** (fila) **Variables** (columna): edad, sexo, ingreso mensual, nivel educativo, condición de actividad, estado civil, entre otras.

En la matriz de datos, las primeras columnas se usan para identificar las unidades de análisis y cada columna siguiente viene a ser una variable. A continuación se muestra una matriz de datos:

	confoocap	vivienda	hogar	parentesco	sexo	marital	confoocap
1	181	20	11	Jefe de Hogar	Mujer	Soltero	No Trabajé ni busqué Trabajo
2	181	19	13	Jefe de Hogar	Mujer	Viudo (a)	No Trabajé ni busqué Trabajo
3	181	19	23	Jefe de Hogar	Hombre	Casado (a)	No Trabajé ni busqué Trabajo
4	181	19	23	Esposa (o)	Mujer	Casado (a)	No Trabajé ni busqué Trabajo
5	181	19	33	Jefe de Hogar	Hombre	Casado (a)	Trabajé
6	775	10	11	Jefe de Hogar	Hombre	Convivient	No Trabajé ni busqué Trabajo
7	775	10	11	Esposa (o)	Mujer	Convivient	Trabajé
8	775	10	11	Hijo (a)	Mujer	.	.
9	775	10	11	Suegro	Mujer	Separado (a)	No Trabajé ni busqué Trabajo
10	775	10	11	Otros Parientes	Mujer	Soltero	No Trabajé ni busqué Trabajo
11	775	7	11	Jefe de Hogar	Mujer	Separado (a)	No Trabajé ni busqué Trabajo
12	775	7	11	Hijo (a)	Hombre	Casado (a)	Trabajé
13	775	7	11	Tierra/ fuera/Padres	Mujer	Casado (a)	No Trabajé ni busqué Trabajo
14	775	6	11	Jefe de Hogar	Hombre	Casado (a)	Trabajé
15	775	6	11	Esposa(o)	Mujer	Casado (a)	Trabajé

El siguiente paso consiste en reconocer la naturaleza de la relación entre las variables y su escala de medición, para proceder a seleccionar el método o tipo de análisis

que más se acomoda a los datos. Así tenemos que pueden aplicarse los siguientes métodos multivariantes:



## METODOS MULTIVARIANTES APLICADOS A TASAS Y PORCENTAJES

Porcentaje o Tasa =  $a/b$

### MODELOS DE INTERDEPENDENCIA

$f(a/b, X1, X2, \dots, Xn)$

### TIPOS DE ANALISIS

CONGLOMERADOS O CLUSTER  
FACTORIAL  
COMPONENTES PRINCIPALES  
ANALISIS DE CORRELACION  
NO PARAMETRICOS

### MODELOS DE DEPENDENCIA Y EXPLICATIVOS

$a/b = f(X1, X2, \dots, Xn)$

CON UNA VARIABLE EXPLICATIVA  
ANALISIS DE REGRESION SIMPLE  
ANALISIS DE REGRESION MULTIPLE  
CON MAS DE UNA VARIABLE  
ANALISIS DISCRIMINANTE  
ANALISIS DE CONTINGENCIA

**Aplicación conceptual** El siguiente es un ejemplo aplicado para la conformación de un modelo explicativo con más de una variable explicativa. En este caso la variable a explicar (dependiente) se expresa como un porcentaje o tasa y las variables explicativas son al menos de naturaleza ordinal. Las variables permiten **discriminar** entre las unidades de análisis, de tal manera que se conformen grupos

de unidades de análisis (hogar, individuos, etc.) de acuerdo a su mayor o menor grado de relación intervariables. Algunas variables son más influyentes que otras mostrando sus efectos cuando se les relaciona con otras. A este efecto conjunto se le denomina factor. Cuando estos factores se presentan como una combinación lineal se le denominará, **función discriminante**.

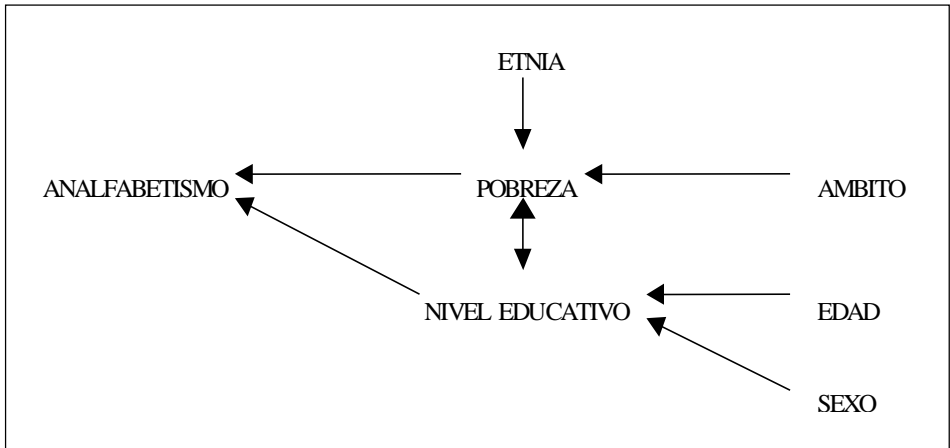
*La consistencia del indicador por el análisis discriminante implica la generación de un modelo analítico de contraste que permita clasificar las unidades de análisis de tal manera que se compare la nueva disposición con la determinada a priori por el indicador .*

**Aplicación práctica:** El siguiente diagrama causal, corresponde a un modelo explicativo del analfabetismo en el cual se representan las variables más explicativas del fenómeno en estudio. El análisis discriminante permite generar un modelo confirmatorio de la clasificación a priori de las unidades de análisis: alfabetos y analfabetos. De este modo es posible

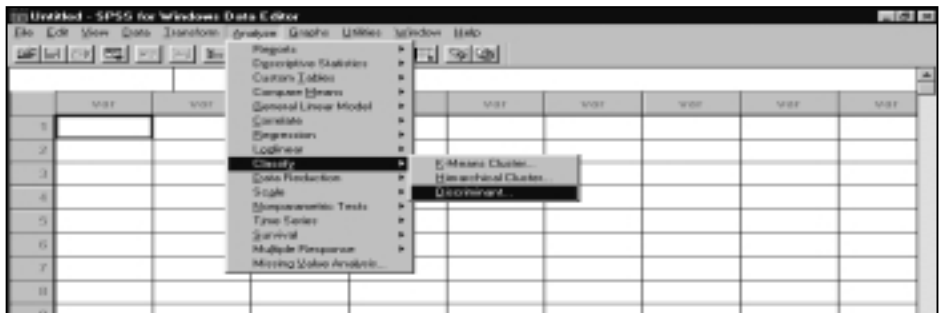
evaluar la consistencia de la clasificación a priori establecida.

Para efectuar el análisis discriminante siga los siguientes pasos:

1. Establezca un diagrama causal en el cual se representen las variables seleccionadas



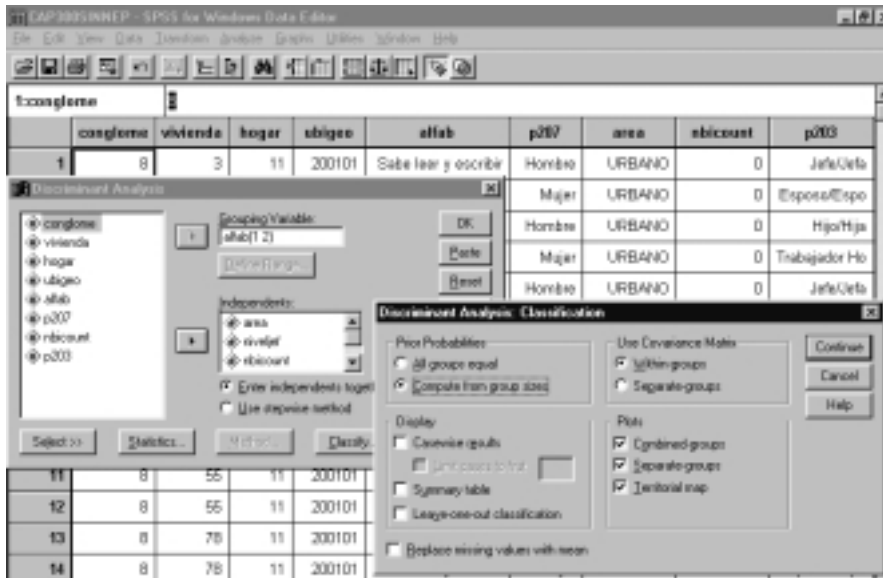
2. Identifique en la base de datos del SPSS las variables nominales y numéricas seleccionadas y conforme la matriz de datos con sus respectivas categorías y valores.



3. Luego, ingrese al modulo Classify del SPSS según el cuadro siguiente:

4. Tome como variable de agrupación el analfabetismo y como independientes todo el resto. El rango de la variable analfabetismo va de 1 a 2 ( alfabetos y analfabetos). Mantenemos la opción "Enter independents together". Seleccionamos

todos los estadísticos y damos "Continue". En la opción "Classify" dejamos "Compute from group sizes" y "Within groups", seleccionando todo lo demás a excepción "Limit cases to first", "Leave-one-out classification" y "Replace missing values with mean", y damos Continue. Se ejecuta el procedimiento con OK.



5. Los resultados se muestran en los cuadros de salida, teniendo en los Eigen values, Lambda de Wilks y el Análisis de Varianza algunas medidas de control de las estimaciones. Los coeficientes del modelo explicativo son mostrados en el cuadro de Coeficientes estandarizados de las Funciones Discriminantes, también la matriz de correlaciones entre las variables y las funciones discriminantes halladas es mostrada por el SPSS, luego, los coeficientes no estandarizados de las funciones discriminantes, que servirán para determinar los puntajes discriminantes para cada unidad de análisis (combinación lineal entre estos coeficientes no estandarizados y las variables independientes). Asimismo se muestran las tablas de clasificación de los individuos, el Mapa Territorial y la matriz con los resultados globales de la clasificación en cuya diagonal se muestra el porcentaje de casos correctamente clasificados.

**Comentario de los resultados:** las variables más discriminante de la condición

de alfabeto son aquellas con coeficientes más altos. Así, el nivel educativo, la edad, el área y el sexo son las variables que discriminan más entre los grupos de alfabetos y analfabetos.

Utilidad: El análisis discriminante permite determinar que las variables que influyen más en el valor del indicador son: el nivel educativo, la edad, el área y el sexo. Por lo que se recomienda mayor control sobre estas variables, en todas las etapas de la investigación desde el diseño de la pregunta, el recojo de información y el procesamiento.

Variables	Función 1
Nivel Educativo	-0.63
Edad en años cumplidos (años)	0.47
Sexo	0.2
Nivel Educativo del Jefe de Hogar	-0.17
Area	0.24
Quintiles de Ingreso Percápita (Población)	-0.05
Idioma Aprendido en la niñez	0.21
Constante	-0.4

## CAPITULO V

### IDENTIFICAR LOS ERRORES NO MUESTRALES MAS FRECUENTES

En las secciones precedentes se trató la evaluación de la consistencia del indicador por métodos estadísticos y demográficos estandarizados, y se presentó el análisis multivariado como una forma de evaluación transversal, estudiando la interrelación del indicador con otras

variables. Pero el indicador también puede verse afectado por factores ajenos al diseño de investigación y que no están bajo control. A estos factores causantes de las variaciones en el valor del indicador se les denomina errores no muestrales.

*Mientras que los errores muestrales son controlables por el diseño muestral y sus formas de cálculo están determinadas así como estandarizados los procedimientos para su identificación, los errores no muestrales comprenden otros factores causantes de las variaciones de los datos, cuya detección es complicada.*

#### 5.1 PRINCIPALES ERRORES NO MUESTRALES

Los principales errores no muestrales son:

- Error de Cobertura: Cuando la muestra no reproduce las características de la población
- Error en la Formulación de las Preguntas y en el riesgo de las respuestas: El orden en la formulación de las preguntas influye en la calidad de las respuestas
- Error de No Respuesta: El informante no responde a un módulo o una pregunta del cuestionario o esta ausente.
- Error de Respuesta en Contenido y Veracidad de la Información: El informante distorsiona la respuesta lo cual ocasiona inconsistencias en el llenado del cuestionario

- Error de Procesamiento o Digitación: Los códigos se atribuyen a otras preguntas ocasionando errores

**Ejemplo Aplicado:** A continuación se presentan los errores no muestrales, encontrados en el cálculo de la Tasa de Analfabetismo:

**Error de Cobertura.** Las muestras de la ENAHO correspondientes al I y IV Trimestre de 1,999 tienen una composición diferente; en el I Trimestre la proporción muestral fue de 32.9 en el área rural, mientras que en el IV Trimestre fue 36.7. En fenómenos sociales en los cuales la región es un factor influyente, los indicadores del IV Trimestre mostraran con mayor intensidad esas características en comparación a los del I Trimestre.

**Error en la Formulación de las Preguntas.** Cuando la pregunta para determinar la condición de alfabeto (¿Sabe leer y escribir?) se realiza antes de preguntar por el nivel de estudios, se pierde la posibilidad de establecer un filtro inicial por nivel educativo. Es mejor preguntar por la condición de alfabeto a aquellos, entrevistados que declaran no tener nivel educativo, con educación inicial o con primaria. Así tenemos que en el cuestionario de la ENAHO, para obtener el indicador relacionado al analfabetismo, se incluye las preguntas cerradas ¿Sabe leer y escribir? y ¿Qué nivel de estudios ha alcanzado? Siendo de fácil interpretación y permitiendo una inmediata codificación, por cuanto se han estructurado las categorías de las respuestas. Es influyente el orden de las preguntas en el cuestionario determinando la calidad de la información a obtener, al haberse comprobado que conviene empezar primero por la pregunta relativa al nivel educativo y luego la relacionada a la condición de alfabeto.

**Error de No Respuesta u Omisiones.** El porcentaje de omisiones del IV Trimestre del 99 fue de 1.8% mayor al registrado en la ENAHO 99 I Trimestre que fue de 0.1%. Esto probablemente tuvo que ver con la aplicación de una muestra panel en el IV Trimestre. Para estos trimestres la tasa de analfabetismo calculada difirió en casi 3 puntos porcentuales.

**Error de Respuesta Contenido y Veracidad de la Información.** Cuando las respuestas de las personas no son veraces. Por ejemplo a la pregunta sabe leer y escribir alguna personas responden afirmativamente sin saber leer realmente, o a la pregunta relacionada a los ingresos responden ocultando su verdadero nivel.

**Error de Procesamiento o Digitación.** Los programas de procesamiento tienen incorporados filtros para detectar inconsistencias en la digitación. En la actualidad, a la etapa de llenado del cuestionario le sigue el procesamiento, evitándose los errores que se pueden cometer en la etapa de crítica codificación.

## 5.2 RECOMENDACIONES PARA MINIMIZAR LOS ERRORES NO MUESTRALES

**Minimizar el Error de Cobertura.** Para minimizar el error de cobertura se requiere:

1. Ampliar el tamaño de muestra
2. Distribuir la muestra de acuerdo a la estructura poblacional por sexo, grupo de edad, ámbito y región
3. Ajustar los factores de expansión de acuerdo a la proyección de población
4. Probar el diseño muestral mediante algunas prepruebas específicas (diseño, cuestionario, procesamiento) para tener una idea de su dificultad y estimar el tiempo y el costo necesarios para el levantamiento de la información.
5. Estimar y controlar el porcentaje de omisión en cada una de las variables principales

**Minimizar el Error de Formulación de las preguntas.** Las principales maneras de minimizar el error de formulación de las preguntas son:

1. Redactar las preguntas de acuerdo a la población objetivo
2. Probar el cuestionario aplicando el mismo a todo tipo de entrevistados

3. En las pre-pruebas del cuestionario deben aplicarse preguntas abiertas de tal manera que se determinen y estandaricen las categorías apropiadas de la pregunta definitiva
4. En las pre-pruebas deben determinarse el orden más eficiente para disponer las preguntas abiertas y cerradas

**Minimizar el Error de no respuesta.**

Las principales maneras de minimizar el error de no respuesta de las preguntas son:

1. Capacitar mejor a las encuestadoras
2. Mantener un grupo estable de encuestadoras
3. Realizar campañas publicitarias difundiendo los estudios a realizar
4. En las pre-pruebas deben probarse todas las versiones posibles del cuestionario

**Minimizar el Error de respuesta.** Las principales maneras de minimizar el error de respuesta de las preguntas son:

1. Diseñar el cuestionario de tal manera que el orden de las preguntas facilite la respuesta veraz del informante
2. Realizar preguntas filtro
3. Elaborar modelos de simulación para el contraste de las respuestas

**Minimizar el Error de Digitación y**

**Procesamiento.** Las principales maneras de minimizar el error de Digitación y Procesamiento del cuestionario son:

1. Establecer programas informáticos de control para la entrada de datos
2. Realizar el procesamiento paralelo a la toma de datos
3. Obtener indicadores con resultados parciales
4. Generar listados de control del procesamiento
5. Seleccionar sub-muestras con la finalidad de verificar si la codificación automática responde a la codificación manual
6. Revisar y actualizar periódicamente las tablas de consistencia e imputaciones



## CAPITULO VI

### ERRORES MAS FRECUENTES EN LA MEDICION E INTERPRETACION

En esta sección se reúnen los errores mas frecuentes encontrados al brindar los resultados, las omisiones en el diseño de investigación, las interpretaciones equivocadas de las tasas y porcentajes, entre otros.

**1. Cuando los porcentajes no van acompañados de las cifras absolutas pueden ocasionarse interpretaciones incorrectas.** Ejemplo: En 1,993 en base a los Censos de Población y Vivienda de 1993, la tasa de analfabetismo se calculó en 12.8% y se estimó el número de analfabetos en 1 millón 784 mil. En 1999 (I Trimestre), en base a la Encuesta de Hogares la tasa calculada fue de 9.7% y el número estimado de analfabetos fue de 1 millón 701 mil. Ello implica una disminución de la tasa en más de tres puntos porcentuales, que no refleja el cambio absoluto observado. Esto se explica porque las tasas calculadas en esos años se han estimado tomando como referencia bases diferentes, mientras que el grupo de referencia en los Censos se tenían 13 millones 938 mil personas de 15 y mas años, en el segundo caso la población de referencia estaba conformada por 17 millones 536 mil personas.

**2. En las interpretaciones se incurre frecuentemente en falacias.** El 12% de la población del país no sabe leer ni escribir. Es una falacia. La población de referencia para el cálculo de la tasa de analfabetismo está conformado por la población de 15 y más años, la cual

constituye solo el 70% de la población total.

**3. Los diseños de investigación experimental carecen de Grupo de Control.** Cuando se afirma que los programas de vacunación no han contribuido a una reducción significativa de la tasa de mortalidad infantil, se cae en una falacia. Los programas tienen que evaluarse asignando aleatoriamente éstos a dos grupos: uno experimental y otro de control. De tal manera que puedan medirse los cambios y el impacto, comparando las variaciones en uno u otro de los grupos.

**4. En la investigación experimental no se consideran modelos explicativos de contraste de los resultados.** Antes de la experimentación deben formularse las hipótesis y los modelos para el contraste de los resultados, de tal manera que las estimaciones sean confirmatorias de una situación a priori controlada.

**5. Las estimaciones no van acompañadas de los errores muestrales ni intervalos de confianza.** Los datos, insumos para el cálculo del indicador muchas veces son generados a partir de una Encuesta, la cual por ser una parte de la población, está sujeta a errores muestrales los cuales son controlados por el investigador y por tanto las estimaciones para todos los indicadores deben ir



acompañadas de sus respectivos errores muestrales e intervalos de confianza.

**6. En las Encuestas la unidad de análisis no coincide con la unidad informante.** Las Encuestas de Hogares tienen generalmente como unidad informante al jefe de familia el cual responde las preguntas del cuestionario relativas a su relación con el resto de integrantes del hogar, lo cual tiene cierto riesgo cuando responde las preguntas asociadas a otros miembros del hogar, no siendo tan veraces como se espera, en algunos casos.

**7. Las estimaciones generalmente se brindan para grandes agregados de población, no permitiendo la focalización del problema.** Las limitaciones presupuestales condicionan el

tamaño de muestra, no permitiendo desagregar los datos a unidades menores (departamentos, provincias, distritos y localidades) sin perder significación estadística. Así, la tasa de analfabetismo calculada con la Encuesta de Hogares del IV Trimestre del 99, no permite brindar estimaciones a nivel departamental.

**8. Comparación de Indicadores calculados con métodos de estimación diferente.** No deben compararse los indicadores calculados por procedimientos diferentes. Por ejemplo la tasa de analfabetismo calculada incluyendo las omisiones, difiere de la tasa calculada que no considera las omisiones. Por tanto debemos asegurarnos antes de efectuar las comparaciones que los métodos de obtención del indicador hayan sido los mismos.

## CAPITULO VII

### RECOMENDACIONES PARA FORMULAR INDICADORES SOCIALES

Para formular los indicadores sociales, deben tenerse en cuenta las siguientes condiciones:

modo la estimación carece de consistencia.

#### 1. Validez Interna:

- **Principio de Realidad.** El indicador tiene que mostrar lo que sucede en la realidad de otro modo no sería de utilidad ni aceptado por la comunidad de investigadores
- **Principio de Sencillez y Replica.** El indicador debe ser fácilmente calculable, pudiendo ser repetido sin dificultad las veces que se considere necesaria.

#### 2. Validez Externa:

- **Principio de no Contradicción.** El indicador calculado debe tener relación con otras medidas relacionadas al fenómeno, lo cual en estadística se denomina correlación, tanto serial como transversal, de otro

#### Ejemplos:

1. El Índice de Precios calculado con una base de datos desactualizada, no mostrará los cambios reales en los precios, generando desconfianza en la opinión pública.
2. La tasa de desempleo se calcula relacionando la población no ocupada entre la PEA. Esta medida es relativamente fácil de calcular, así como repetir los procesos para su obtención.
3. La disminución de la pobreza debe tener relación con la disminución de la tasa de analfabetismo cuando sus valores se comparan en una serie temporal. Asimismo la variación del Índice de Desarrollo Humano para un año determinado debe tener relación con el crecimiento económico observado.



## VI. REFERENCIAS BIBLIOGRAFICAS

1. **Hanz Zeisel,**  
*"Dívalo con Números"* -  
 Fondo de Cultura Económica, México 1962
  
2. **Sanchez Carrión, Juan Javier**  
*"Introducción a las Técnicas de Análisis Multivariantes Aplicadas a las Ciencias Sociales"*-  
 Centro de Investigaciones Sociológicas, Madrid, 1984
  
3. **Lazarfeld, Torgenson, Barton, Coleman y otros**  
*"Medición y Construcción de Índices"*-  
 Nueva Visión, Buenos Aires, 1971
  
4. **Babel , Earl**  
*"Métodos de Investigación por Encuesta"*-  
 Fondo de Cultura Económica, México 1988
  
5. **Pineda, Elia Beatriz**  
*"Metodología de la Investigación"*-  
 Organización Panamericana de la Salud, USA 1994
  
6. **Boggino, Norberto**  
*"Cómo Elaborar Mapas Conceptuales en la Escuela"* -  
 Serie Educación Homo Sapiens Ediciones, Buenos Aires, 1997



**CUADRO RESUMEN METODOLOGICO No 1**

**CRITERIOS DE EVALUACION DE INDICADORES SOCIALES ASOCIADOS A TASAS Y PORCENTAJES**

NIVEL DE LA INFORMACION	METODOS DE EVALUACION	PROCEDIMIENTOS	INTERPRETACION Y RESULTADOS
<b>DATOS</b>			
LOS DATOS SE DETERMINAN A PARTIR DE LA CONCEPTUALIZACION DEL PROBLEMA	DETERMINAR LA FRACCION DE MUESTREO	COCIENTE ENTRE EL TAMAÑO DE MUESTRA Y EL TOTAL DE POBLACION	LA MUESTRA ES REPRESENTATIVA DEL CONJUNTO DE DATOS
CENSOS (SESGO = 0)	DETERMINAR UN DISEÑO MUESTRAL	EVALUAR EL DISEÑO MUESTRAL COMPARANDO	LOS ERRORES MUESTRALES HAN SIDO MINIMIZADOS
ENCUESTAS (SESGO >0)	APROPIADO QUE PERMITA CONTROLAR	LOS ERRORES MUESTRALES DEL TIPO DE MUESTREO	POR EL DISEÑO MUESTRAL DE TAL MANERA QUE SON
REGISTROS ADMINISTRATIVOS (SESGO ?)	LOS ERRORES MUESTRALES	APLICADO CON UNA MUESTRA SIMPLE ALEATORIA	MENORES A LOS QUE SE HABRIAN OBTENIDO CON UNA MUESTRA SIMPLE ALEATORIA
<b>VARIABLES</b>			
SE GENERAN A PARTIR DE LA OPERACIONALIZACION DE LOS CONCEPTOS Y SON LAS QUE REFLEJAN LA PROPIEDAD LATENTE DEL FENOMENO SOCIAL	TABLA DE FRECUENCIAS O HISTOGRAMA	CONSTRUYA UNA TABLA DONDE SE MUESTRE LAS CATEGORIAS DE LAS VARIABLES Y SUS FRECUENCIAS ASOCIADAS	LOS DATOS SIGUEN UNA DISTRIBUCION NORMAL O CASI NORMAL. NO HAY VALORES EXTREMOS Y SE APRECIA LA MODA DE LA DISTRIBUCION
	DIAGRAMA DE TALLOS Y HOJAS	UTILICE EL SPSS-DESCRIPTIVE STATISTICS -EXPLORE	IDENTIFICAR LOS VALORES EXTREMOS EN EL GRAFICO
	DIAGRAMA DE CAJAS	UTILICE EL SPSS-DESCRIPTIVE STATISTICS -EXPLORE	CONFORMAR LIMITES SUPERIOR E INFERIOR DE CONTROL
	CRUCES DE VARIABLES	CONSTRUYA UNA TABLA DE CONTINGENCIA Y CALCULE LA ESTADISTICA JI-CUADRADO	SI SE RECHAZA LA HIPOTESIS NULA SE DETERMINA QUE LAS VARIABLES SON INDEPENDIENTES
LAS VARIABLES ASOCIADAS A LAS TASAS SON POR LO GENERAL CUANTITATIVAS Y SE EXPRESAN AL MENOS EN UNA ESCALA ORDINAL	ESTADISTICAS DE TENDENCIA CENTRAL Y DE DISPERSION	CALCULE EL INDICADOR Y LUEGO DETERMINE EL COEFICIENTE DE VARIABILIDAD	LAS ESTIMACIONES SON DE CALIDAD Y BRINDAN LA CONFIANZA ESTADISTICA SUFICIENTE

Continúa

**CUADRO RESUMEN METODOLOGICO No 1**  
**CRITERIOS DE EVALUACION DE INDICADORES SOCIALES ASOCIADOS A TASAS Y PORCENTAJES**

Conclusión

NIVEL DE LA INFORMACION	METODOS DE EVALUACION	PROCEDIMIENTOS	INTERPRETACION Y RESULTADOS
<b>INDICADORES</b> SE OBTIENEN COMO COMBINACION LINEAL DE LAS VARIABLES SELECCIONADAS EN EL CASO DE LAS TASAS Y PORCENTAJES PORCENTAJES SE OBTIENEN COMO UN COCIENTE	<b>METODOS DETERMINISTICOS</b>		
	ELABORAR DIAGRAMA DE PROCESOS	DESCOMPONGA EL DISEÑO DE EVALUACION EN ETAPAS MUTUAMENTE EXCLUYENTES	PROCESOS DE EVALUACION DESCRITOS Y REPRESENTADOS EN UN DIAGRAMA E IDENTIFICADO EL FLUJO DE DATOS
	REVISAR LOS CONCEPTOS OPERACIONALES	EVALUE LA OPERACIONALIZACION DE LOS CONCEPTOS	LA FORMA OPERATIVA SE ADECUA AL CONCEPTO
	ANALIZAR LAS VENTAJAS Y DESVENTAJAS	REALICE UN LISTADO DE VENTAJAS Y DESVENTAJAS	POTENCIAR LAS VENTAJAS Y CONTROLAR LAS DESVENTAJAS
	REVISAR LA FORMA DE MEDICION DE OTROS PAISES	EFFECTUE CONSULTAS BIBLIOGRAFICAS Y ELECTRONICAS DE OTRAS FORMAS DE MEDICION	ARMONIZACION Y ESTANDARIZACION DE CONCEPTOS
	APLICAR EL METODO DE LAS DIFERENCIAS SIMPLES TASAS DE CAMBIO Y EFECTIVIDAD	CALCULE LAS DIFERENCIAS SIMPLES, LAS TASAS DE CAMBIO Y LA EFECTIVIDAD	EVALUAR LAS VARIACIONES TEMPORALES ABSOLUTAS RELATIVAS Y POTENCIALES
	APLICAR EL METODO DE LAS DIFERENCIAS RELATIVAS	DISPONGA LOS INDICADORES EN UNA SERIE TEMPORAL Y CALCULAR LAS DIFERENCIAS RELATIVAS POR SEXO, EDAD, AMBITO Y REGION	LAS DIFERENCIAS RELATIVAS SE MANTIENEN CONSTANTES ENTONCES EL INDICADOR ES CONSISTENTE EN EL PERIODO
	CONSTRUIR PIRAMIDES DEMOGRAFICAS	UTILICE EL EXEL PARA SU CONSTRUCCION	LA POBLACION Y LA MUESTRA TIENEN ESTRUCTURA POBLACIONAL SEMEJANTE
	<b>METODOS PROBABILISTICOS</b>		
	REALIZAR LAS PRUEBAS DE CONTRASTE DE HIPOTESIS	FORMULE DOS HIPOTESIS UNA NULA Y OTRA ALTERNANTE REFERIDAS AL VALOR DEL INDICADOR LA HIPOTESIS NULA ESTA ASOCIADA AL SUPUESTO VALOR DEL INDICADOR	SE RECHAZA LA HIPOTESIS NULA PARA UN TAMAÑO DE MUESTRA DADO Y UNA PROBABILIDAD DE ERROR FIJADA DETERMINANDOSE QUE EL INDICADOR ES SIGNIFICATIVO O CUMPLE CIERTA CONDICION
ESTABLECER LOS INTERVALOS DE CONFIANZA	HALLE LOS LIMITES SUPERIOR E INFERIOR PARA EL INDICADOR CALCULADO	CON UN 95% DE CONFIANZA SE ESPERA QUE EL INDICADOR CALCULADO SE ENCUENTRE ENTRE CIERTOS LIMITES	
REALIZAR EL ANALISIS MULTIVARIADO	HAGA DIAGRAMAS CAUSALES Y UTILICE EL SPSS PARA HALLAR LA MATRIZ DE CORRELACION Y LOS PUNTAJES PARA CLASIFICAR LOS INDIVIDUOS	CONTRASTAR LOS POSTULADOS A PRIORI DEL FENOMENO SOCIAL CON LOS RESULTADOS DEL ANALISIS MULTIVARIADO Y FORMULAR MODELOS PREDICTIVOS	

CUADRO RESUMEN METODOLOGICO No 2  
DIAGRAMA DE LOS PROCESOS SEGUIDOS PARA LA REVISION DEL INDICADOR

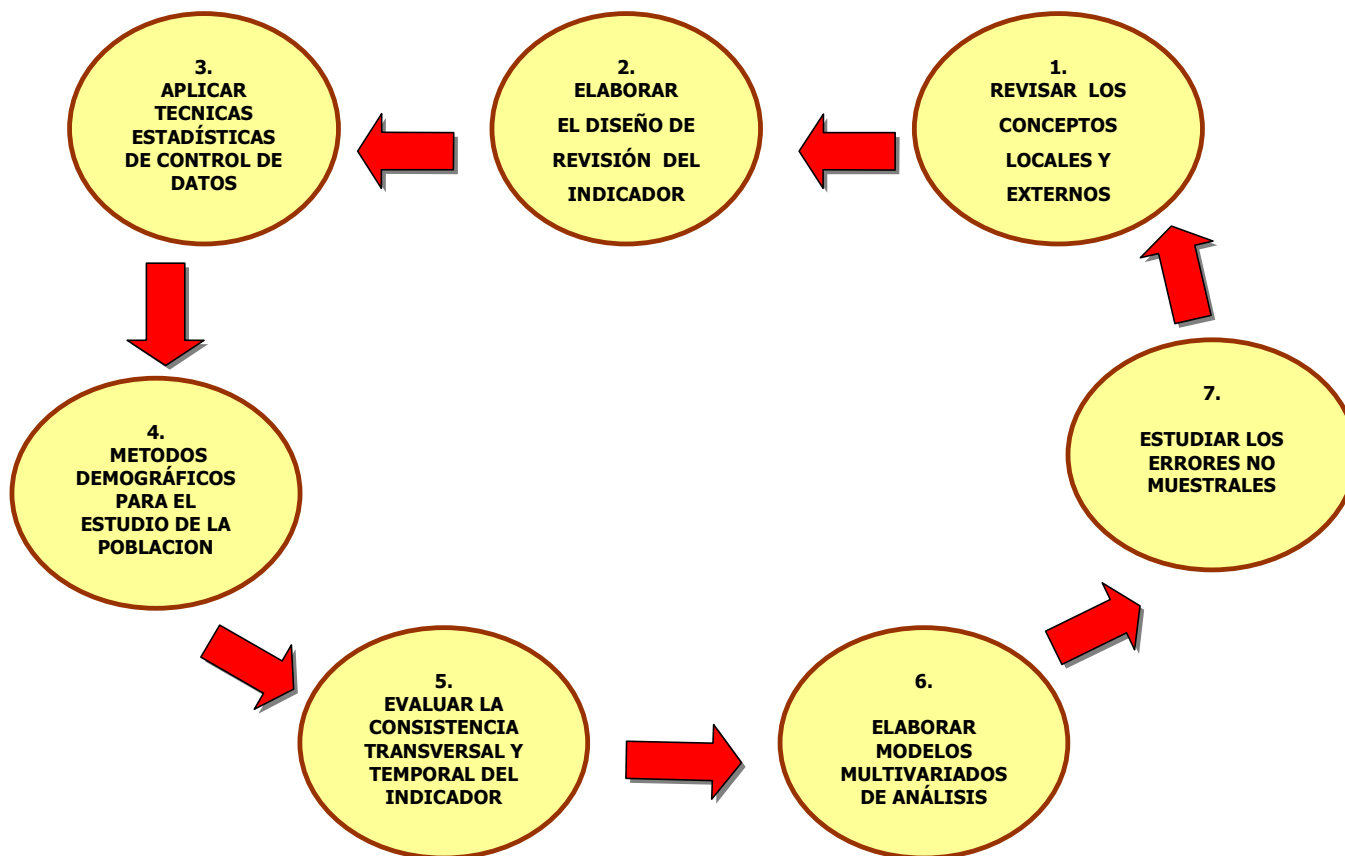
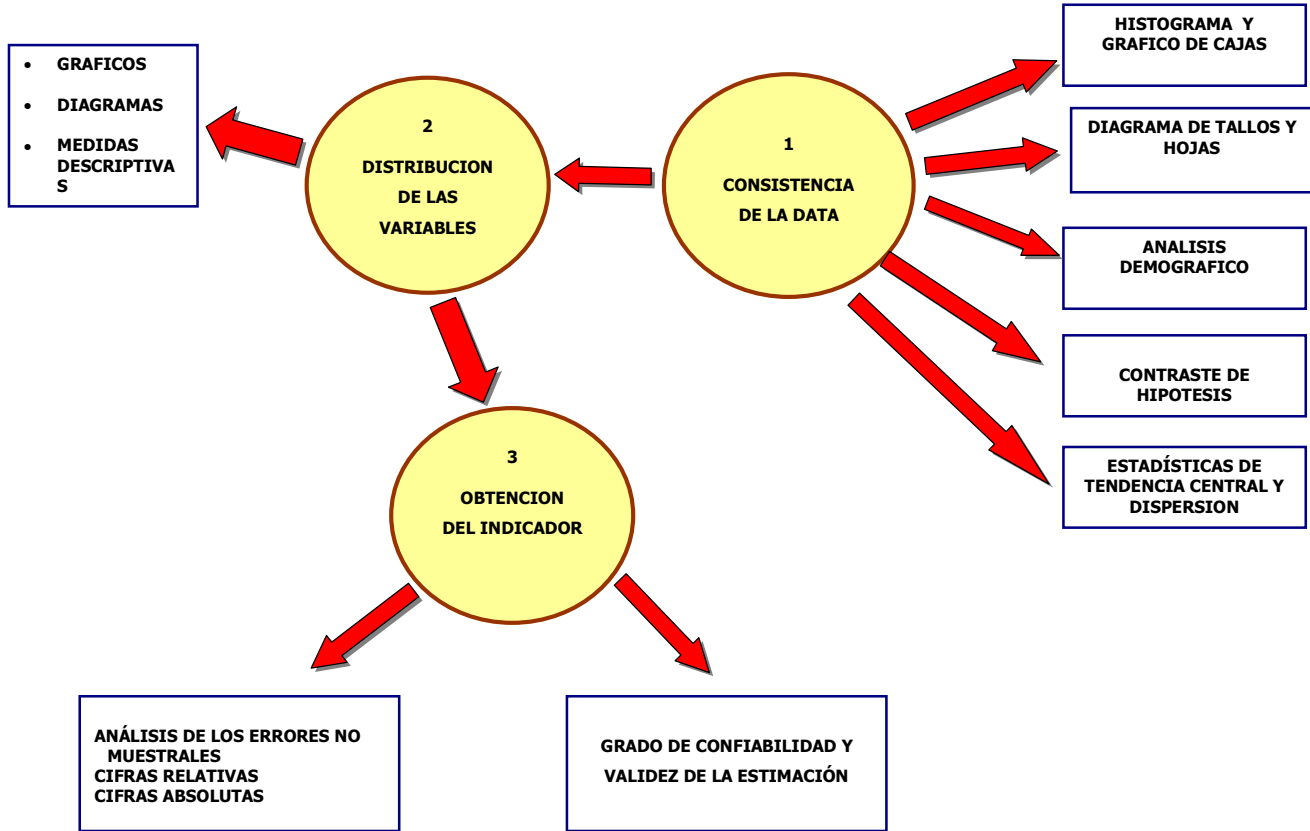




DIAGRAMA DE PROCESOS



## 2. ANALISIS DISCRIMINANTE

