

03061

Instituto de Investigación en Matemáticas Aplicadas y Sistemas

UNIDAD ACADÉMICA DE LOS CICLOS PROFESIONAL
Y DE POSTGRADO DEL C.C.H.

U . N . A . M .

**DISEÑO ESTADÍSTICO DE UNA ENCUESTA
INDUSTRIAL MENSUAL**

TESIS CON
FALLA DE ORIGEN

T E S I S
QUE PARA OBTENER EL GRADO DE:
**MAESTRA EN ESTADÍSTICA E
INVESTIGACION DE OPERACIONES**
PRESENTA LA ACTUARIO
ESTRELLA DE LOURDES LUEVANO MOYA

MEXICO, D. F.



2002



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

Agradezco al Dr. Ignacio Méndez Ramírez su valiosa ayuda en el desarrollo de este trabajo y al personal y autoridades de la Secretaría de Programación y Presupuesto, por el apoyo y las facilidades que me proporcionaron.

A mi hija

A mi esposo

A mi madre

A mis hermanos

DISEÑO ESTADÍSTICO DE UNA ENCUESTA INDUSTRIAL MENSUAL

INDICE

Pag.

INTRODUCCION.

I. CONSIDERACIONES GENERALES.	1
II. EL USO DE COMPONENTES PRINCIPALES PARA EL DISEÑO DE LA ENCUESTA.	5
III. CLASIFICACION DE LOS ESTABLECIMIENTOS EN MAYORES, MEDIANOS Y MENORES.	14
III.1 Establecimientos Menores.	15
III.2 Establecimientos Medianos y Mayores	17
IV. DISEÑO DE LA MUESTRA.	26
IV.1 Subestratificación y Procedimiento de Selección	26
IV.2 Tamaño de Muestra y su Afijación.	29
IV.3 Estimadores.	32
V. SIMULACION DE ESTIMADORES.	55
V.1 Estimadores de Totales.	58
V.2 Estimadores de razón.	63
V.3 Estimadores de Promedios en Dominios de Estudio.	70

CONCLUSIONES

BIBLIOGRAFIA

CUADROS

ANEXOS

APENDICES

INTRODUCCION

En este trabajo se propone una metodología para realizar el diseño estadístico de una Encuesta Industrial Mensual (EIM). Esta metodología está basada en las ideas desarrolladas y en las experiencias adquiridas durante mi participación en el grupo de trabajo que realizó el diseño estadístico de la Encuesta Mensual de la Industria Manufacturera (EMIM) que se está levantando actualmente en la Dirección General de Estadística (D.G.E.) de la Coordinación General de los Servicios Nacionales de Estadística, Geografía e Informática - (C.G.S.N.E.G.I.)

Cabe señalar que el diseño que se presenta no es exactamente el de la EMIM sino que se introducen algunos cambios tendientes a desarrollar los aspectos técnicos de interés de esta tesis, los cuales se señalan en las secciones correspondientes.

El diseño estadístico de la encuesta comprende la elaboración del diseño muestral y la construcción del marco de establecimientos industriales.

Debido a las diferencias existentes en el comportamiento económico entre clases de actividad y entre un establecimiento y otro se consideró conveniente definir un criterio que clasificara a los establecimientos por clase de actividad en mayores, medianos y menores. Tal criterio se estableció a través del análisis de componentes principales (Morrison, 1967) aplicado a cuatro variables que carac

terizan a los establecimientos industriales desde el punto de vista económico, las cuales se tienen registradas en el marco muestral que se obtuvo mediante el censo de establecimientos industriales de 1976. (Secretaría de Programación y Presupuesto, Dirección General de Estadística, 1979).

Por medio del análisis de componentes principales, se construyó un índice I_{e_1} para establecimientos por clase de actividad y dependiendo del valor del índice en cada establecimiento, se les asignó un orden a éstos dentro de la clase a la que pertenecen. Asimismo se construyó un índice general para las clases de actividad I_c , el cual ordena a las mismas de acuerdo a su contribución al total de la actividad económica del sector industrial.

Una vez construido el índice para establecimientos, se hizo un análisis económico y estadístico, como resultado del cual se clasificó a los establecimientos por clase de actividad, en menores y no menores.

Debido a que la EIM sólo cubre a los establecimientos no menores, se calculó un nuevo índice considerado solamente a esos establecimientos, el cual se denota por I_{e_2}

Al conjunto de establecimientos no menores se le aplicó un criterio estadístico que a su vez permitió clasificar a los establecimientos en mayores y medianos (Glasser, 1962).

Con los establecimientos mayores y medianos se formó un directorio que contiene los datos de identificación y ubicación de cada establecimiento. Para la construcción de ese directorio se utilizaron los registros del Censo Industrial de 1976, los cuales fueron actualizados con la información de altas, bajas y cambios que se reciben en el Departamento de Directorios de esta Institución.

Con el objeto de reducir la heterogeneidad en las características a observar, los establecimientos medianos fueron subestratificados. La muestra de la encuesta está constituida por todos los establecimientos mayores de las 37 clases de actividad que cubre la encuesta y por una muestra de los establecimientos medianos de las mismas clases, seleccionada con muestreo aleatorio irrestricto dentro de cada subestrato, con afijación óptima del tamaño de muestra y utilizando como variable de estratificación al índice I_{e_2} .

Se presentan las fórmulas para obtener estimaciones a nivel nacional y por clase de actividad, a partir de la información de la muestra.

Finalmente se realiza una simulación para observar algunas estadísticas importantes de los estimadores y verificar si se satisface la precisión establecida.

I. CONSIDERACIONES GENERALES.

Para el desarrollo integral de la encuesta se formaron cuatro grupos de trabajo, cuya denominación y funciones se mencionan a continuación:

DENOMINACION	FUNCIONES
Diseño Conceptual	<p>Determinar la cobertura conceptual de la encuesta, esto es, la información a captarse, especificando las definiciones y clasificaciones necesarias.</p> <p>Definir la unidad de observación, la cobertura geográfica, el nivel de publicación de los resultados y la información auxiliar en que se basa el diseño estadístico de la encuesta.</p> <p>Determinar las clases de actividad que abarca la encuesta.</p>
Diseño Estadístico.	<p>Realizar el diseño de la muestra, el cual incluye la especificación del proceso de selección de la muestra y el proceso para obtener los estimadores muestrales de los valores poblacionales.</p> <p>Diseñar y construir el marco muestral.</p>

DENOMINACION

FUNCIONES

<i>Diseño Logístico.</i>	<i>Elaborar la red de actividades y el presupuesto de la encuesta, incluyendo la determinación de los recursos necesarios y las fechas en que tienen que estar disponibles.</i> <i>Realizar la prueba piloto.</i> <i>Elaborar el manual de procedimientos de la encuesta.</i> <i>Ejecutar y controlar el trabajo de campo.</i>
<i>Diseño del Procesamiento Electrónico.</i>	<i>Prestar apoyo para el diseño estadístico de la encuesta y para el control del trabajo de campo.</i> <i>Realizar el procesamiento de la información desde que se reciben los cuestionarios, hasta que se obtienen los resultados.</i>

Las funciones señaladas permiten tener idea de la delimitación de responsabilidades aunque, desde luego, no pretenden ser exhaustivas, pues se tendrían que proporcionar mayores detalles, que salen del contexto del presente trabajo.

Por tal motivo, simplemente se mencionan los aspectos que afectan directamente el trabajo del grupo estadístico y los que permiten un conocimiento general de las características de la encuesta, tales como: la información disponible, la unidad de observación, el tipo de cuestionario, la forma de captación, el tipo de procesamiento de datos, etc.

La información disponible para el diseño estadístico de la encuesta era la del Censo Industrial realizado en 1976, en el cual se captó in formación de los establecimientos que realizaron actividades durante todo o parte del año de 1975.

La información censal se tenía disponible en archivos magnéticos, a nivel de cada establecimiento de las 240 clases de actividad que integran la industria manufacturera y de la transformación. La clasifica ción de los establecimientos censados en 1976, se realizó de acuerdo al Catálogo Mexicano de Actividades Económicas (C.M.A.E.) de 1976 (Se cretaría de Programación y Presupuesto. Dirección General de Estadística, 1979), en el cual se tienen cuatro niveles de agregación, que son: división, grupo, subgrupo y clase de actividad. Esta última se identifica por medio de cuatro dígitos en los que el primero indica la división, el segundo el grupo, el tercero el subgrupo y el cuarto el orden de la clase de actividad dentro del correspondiente subgrupo.

La encuesta cubre las 37 clases de actividad que se presentan en el Anexo N° 1.

Con objeto de que los resultados de la encuesta fueran comparables con los del censo, se decidió adoptar la misma definición de establecimiento industrial, según la cual, un establecimiento industrial es "toda unidad económica que en una sola ubicación física, realiza principalmente actividades de extracción, maquila, ensamble, procesamiento o transformación total o parcial de uno o varios productos. La información del establecimiento debe incluir además, las actividades que se realicen en

la misma ubicación física, vinculadas al proceso productivo, tales como: mantenimiento, investigación, almacenaje, etc." (Secretaría de Programación y Presupuesto. Dirección General de Estadística, 1979).

La decisión de utilizar como unidad última de muestreo de la encuesta al establecimiento industrial, definido arriba, permitió utilizar como marco de muestreo a los registros censales y tener disponible, a nivel de cada establecimiento, las 64 variables que se captaron en el censo.

A partir de esta información, el grupo de Diseño Conceptual seleccionó cuatro variables, que son: promedio de personal ocupado total (P.O.), producción bruta total (P), activos fijos (A.F.) y valor agregado censal (V.A.), las cuales se definen en el Anexo N° 2.

Estas cuatro variables fueron utilizadas como información auxiliar para el diseño estadístico de la encuesta, las cuales miden la importancia económica de los establecimientos y, además, están relacionadas con los aspectos más importantes que se pueden captar de los establecimientos industriales.

II. EL USO DE COMPONENTES PRINCIPALES PARA EL DISEÑO DE LA ENCUESTA.

En el diseño de la encuesta se utilizaron tres índices, que se denotan por I_{e_1} , I_{e_2} e I_c . Los dos primeros se denominan índices para establecimientos y el último, índice para clases de actividad.

Los índices para establecimientos, como su nombre lo indica, se calculan para cada establecimiento. La diferencia entre estos dos índices es en cuanto a la población para la que se calculan. Para I_{e_1} se utilizan todos los establecimientos y para I_{e_2} solo los denominados no menores. Los criterios para clasificar a los establecimientos se describen en el siguiente capítulo. El cálculo de estos dos índices se realiza en forma independiente dentro de cada clase de actividad.

El método para calcular los índices fue el análisis de componentes principales, aplicado a las variables: P.O., P., A.F. y V.A.

El insumo para calcular los índices para establecimientos fueron los valores, a nivel establecimiento de las variables mencionadas. En tanto que, para el índice I_c se usaron los totales, a nivel clase de actividad, de las cuatro variables.

El método de componentes principales fue desarrollado por Pearson (1901) y por Hotelling (1933) para analizar la estructura de la matriz de covarianza o de correlación de un conjunto de datos. Mediante este método, las cuatro variables mencionadas, que se denotan

por X_1, X_2, X_3 y X_4 se transforman lineal y ortogonalmente en cuatro nuevas variables Y_1, Y_2, Y_3 y Y_4 que se denominan componentes principales. Estas se obtienen de manera que la primera explica la mayoría de la variabilidad de los datos originales; la siguiente componente explica la mayor parte de la variabilidad, sujeta a no estar correlacionada con la primera componente, y así sucesivamente.

Debido a que las unidades de medida de las variables originales son muy diferentes (P., A.F. y V.A. en pesos y P.O. en número de personas ocupadas), las componentes lineales de las variables originales tendrían poco significado. Por tal motivo se consideró conveniente usar las variables en forma estandarizada y utilizar la matriz de correlación para el análisis de los datos; a fin de obtener una visión más exacta del patrón de dependencia.

La primera componente principal se expresa como:

$$Y_1 = a_{11} X_1 + a_{21} X_2 + \dots + a_{k1} X_k$$

y en forma matricial como

$$Y_1 = X a_1 \quad \dots (1)$$

La varianza de Y_1 es

$$\sigma_{Y_1}^2 = \sum_{i=1}^K \sum_{j=1}^K a_{i1} a_{j1} \rho_{ij} \quad \dots (2)$$

$$= a_1' R a_1$$

donde R es la matriz de correlación de las X_i 's.

Se desea seleccionar a_j de manera que se maximice la varianza de Y_j . Sin embargo, es claro que deben imponerse algunas restricciones sobre a_j pues, de lo contrario, esta varianza podría hacerse infinitamente grande, así que se normaliza estableciendo que:

$$a_j' a_j = 1 \quad \dots (3)$$

El problema es, entonces, maximizar la ecuación (2), sujeta a la restricción (3).

Se define:

$$Q = a_j' R a_j - \lambda_j (a_j' a_j - 1)$$

donde λ_j es un multiplicador de Lagrange.

Entonces, derivando Q con respecto a a_j se obtiene:

$$\frac{\partial Q}{\partial a_j} = 2 R a_j - 2 \lambda_j a_j$$

Igualando a cero esta derivada

$$R a_j = \lambda_j a_j \Leftrightarrow (R - \lambda_j I) a_j = 0 \quad \dots (4)$$

Se observa que a_1 es el vector característico correspondiente a la raíz característica λ_1 .

Ahora, premultiplicando la ecuación (4) por a_1' y de acuerdo a la ecuación (2) se tiene que

$$v_{y_1}^2 = \lambda_1 a_1' a_1 = \lambda_1$$

Por lo cual, se debe escoger λ_1 como la raíz característica más grande de R .

La segunda componente principal se define como

$$y_2 = X a_2$$

donde a_2 se selecciona de tal forma que se maximice

$$v_{y_2}^2 = a_2' R a_2$$

sujeta a las restricciones

$$a_2' a_2 = 1$$

$$a_1' a_2 = 0$$

en donde la segunda restricción expresa que y_2 no esté correlacionada con y_1 .

se define:

$$Q = a_2' R a_2 - \lambda_2 (a_2' a_2 - 1) - \mu (a_1' a_2)$$

donde λ_2 y μ son multiplicadores de Lagrange.

Derivando Q con respecto a a_2 e igualando esta derivada a cero, se obtiene que

$$\frac{\partial Q}{\partial a_2} = 2 R a_2 - 2 \lambda_2 a_2 - \mu a_1 = 0 \quad \dots (5)$$

Premultiplicando esta ecuación por a_1' se observa que

$$2 a_1' R a_2 - \mu = 0 \quad \dots (6)$$

Para obtener el valor de μ se premultiplica la ecuación (4) por a_2' , quedando:

$$a_2' R a_1 = \lambda_1 a_2' a_1 = 0 \quad \dots (7)$$

Entonces, para que se satisfagan simultáneamente las ecuaciones (6) y (7), se debe tener $\mu = 0$. Substituyendo el valor de μ en la ecuación (5) se observa que

$$R a_2 = \lambda_2 a_2 \Leftrightarrow (R - \lambda_2 I) a_2$$

de lo cual resulta obvio que λ_2 debería seleccionarse como la segunda raíz característica más grande de R .

Las restantes componentes principales se van encontrando sucesivamente de los otros vectores característicos.

La j -ésima componente principal se expresa como:

$$V_j = a_{1j} X_1 + a_{2j} X_2 + \dots + a_{kj} X_j$$

donde los coeficientes son los elementos del vector característico de la matriz de correlación R , correspondientes a la j -ésima raíz característica más grande.

El signo algebraico y la magnitud de las a_{ij} indica la dirección e importancia de la contribución de la i -ésima respuesta a la j -ésima componente.

La traza de la matriz de correlación ($\text{tr } R$) es igual al número de variables, en este caso cuatro, y la proporción de la varianza total atribuible a la j -ésima componente es $\lambda_j / \text{tr } R$.

Las correlaciones de las respuestas con la j -ésima componente principal están dadas por el vector $\sqrt{\lambda_j} a_{ij}$, $j = 1, 4$; y la varianza de la componente j -ésima, que es λ_j , se obtiene sumando estas correlaciones elevadas al cuadrado.

La obtención de los índices, de acuerdo al método expuesto, se realizó utilizando el programa Factor Analysis del paquete estadístico SPSS (Nie, 1975). En el apéndice No. 1 se presenta un listado de salida de este programa, que contiene la media, la desviación estándar y la matriz de correlación de las variables originales; así como el número de establecimientos con los que se realizó el análisis.

El método de factorización utilizado es el PAI (factorización principal sin iteración) con unos en la diagonal principal de la matriz de correlación, mediante el cual se obtienen las componentes principales.

El programa calcula la proporción de la varianza total atribuible a la componente j -ésima, usando la relación $\lambda_j / \text{tr}R$, $j = \overline{1, 4}$. Estos valores se imprimen en orden decreciente de magnitud.

En el cuadro No. 1 se presentan estos resultados para el índice I_{e_1} y en el cuadro No. 2 los correspondientes al índice I_{e_2} . Para el índice I_{e_1} , las proporciones de la varianza atribuibles a la primera, segunda, tercera y cuarta componente principal son 82.4%, 10.8%, 5.0% y 1.8%, respectivamente.

En el listado de salida aparece otra columna que contiene el porcentaje de varianza acumulada de la primera componente, de las dos primeras componentes, de las tres primeras y, por último, de las cuatro componentes, que es 100, puesto que con las cuatro componentes se explica la totalidad de la varianza de los datos originales.

Las correlaciones de las variables originales con la j -ésima componente se obtienen como $\sqrt{\lambda_j} a_{ij}$ y se imprimen bajo el encabezado "Factor Matrix", para las componentes con valor característico mayor de uno.

Los Cuadros N° 3 y 4 contienen estas correlaciones para los índices Ie_1 e Ie_2 , respectivamente.

La correlación del índice de clases con las variables: P.O., P., A.F. y V.A. es de 0.825, 0.950, 0.887 y 0.962. respectivamente.

Por último, se obtiene la traspuesta de la matriz "Factor Matrix" y se multiplica por la inversa de la matriz de correlación. De este producto se obtiene la matriz llamada "factor score coefficients" que se usa para calcular los índices.

El índice Ie_1 para el establecimiento i -ésimo se calcula como sigue, usando los "factor score coefficients" del cuadro N° 5 para la clase de actividad de que se trate.

$$Ie_{1i} = FSC_1 X_{1i} + FSC_2 X_{2i} + FSC_3 X_{3i} + FSC_4 X_{4i}$$

donde X_{1i} , X_{2i} , X_{3i} y X_{4i} son los valores estandarizados de las variables originales, para el establecimiento i -ésimo.

FSC_1 es el "factor score coefficient" del factor 1 (o primera componente principal) para la variable X_1 . FSC_2 , FSC_3 y FSC_4 se definen de manera análoga para las variables X_2 , X_3 y X_4 .

El índice Ie_2 para cada establecimiento se obtiene en la misma forma que el Ie_1 , sólo que utilizando los "factor score coefficients" - que aparecen en el Cuadro N° 6.

Por último para calcular el índice I_c para la j -ésima clase de actividad del sector industrial se usa la relación.

$$I_{c_j} = 0.250 X_{1j} + 0.288 X_{2j} + 0.269 X_{3j} + 0.292 X_{4j}; \quad j = \overline{1, 240}$$

En donde X_{1j} , X_{2j} , X_{3j} y X_{4j} son los totales estandarizados de las cuatro variables, para la clase j -ésima.

III. CLASIFICACION DE LOS ESTABLECIMIENTOS EN MAYORES, MEDIANOS Y MENORES.

Con base en los índices antes calculados, los establecimientos industriales se clasificaron en menores y no menores, y estos últimos en mayores y medianos.

Los establecimientos menores tienen las siguientes características:

- Aportan porcentajes muy bajos del total de las variables de interés.
- La mayoría de los establecimientos quedan en esta categoría, en casi todas las clases de actividad, lo cual puede apreciarse en el cuadro N° 7, elaborado con información censal de 1975.
- Presentan una gran dinámica de cambio, puesto que son los que más frecuentemente cambian de domicilio, de propietario, se dan de baja o de alta, cambian de actividad, etc.
- Se encuentran más dispersos geográficamente que los demás establecimientos.

La baja contribución de los establecimientos menores a la actividad industrial, aunada a su dispersión geográfica y al elevado costo y dificultad de construir y mantener actualizado un directorio de esos establecimientos, fueron las consideraciones básicas que llevaron a excluir a los establecimientos menores de esta encuesta, quedando su captación para una etapa posterior, a través de un marco muestral de áreas.

El procedimiento para clasificar a los establecimientos en menores y

no menores se presenta en la Sección III.1.

Para los establecimientos no menores se construyó un directorio a partir de los registros del Censo Industrial de 1976, que contiene la siguiente información para cada establecimiento: número de folio censal, clase de actividad, entidad federativa, municipio, categoría, nombre del propietario o razón social, domicilio y teléfono.

Este directorio se actualizó con la información de altas, bajas y cambios, que se reciben en el Departamento de Directorios y Marcos Estadísticos de esta institución.

Posteriormente se verificó telefónicamente la vigencia de los establecimientos del marco.

Los establecimientos no menores fueron a su vez clasificados como mayores y medianos, con el fin de disminuir la varianza de los estimadores, utilizando el procedimiento que se describe en la sección III.2.

III.1 Establecimientos Menores.

Con el objeto de obtener un criterio para clasificar a los establecimientos en menores y no menores, se elaboraron y analizaron cuadros por clase de actividad, como el que aparece en el apéndice N° 2, en los que se observan: la distribución del índice para establecimientos Ie_1 , la distribución del número de establecimientos y la de las cuatro variables bajo estudio.

Para obtener estos cuadros, se le restó al valor del índice I_{e_j} de cada establecimiento el mínimo de los índices I_{e_j} de esa clase y se le sumó la cantidad 0.5 a fin de hacerlos positivos, lo cual es sólo una traslación que no afecta su significado y permite obtener los porcentajes acumulados para cada una de las variables. El rango del índice se dividió en 50 niveles de igual longitud. Para cada nivel aparece el número de establecimientos incluidos, la suma de los índices I_{e_j} de esos establecimientos y el porcentaje acumulado de cada variable.

Es importante señalar que para llevar a cabo el análisis de los cuadros, se utilizaron simultáneamente los índices I_{e_j} e I_c . - El primero, para ordenar a los establecimientos dentro de su clase y el segundo, como apoyo para decidir a partir de qué nivel se clasificarían los establecimientos como menores, dependiendo de la importancia de la clase de actividad, la cual es determinada por el índice I_c .

La clasificación final de los establecimientos en menores y no menores, se hizo en base a los siguientes puntos:

- La importancia de cada clase de actividad, en cuanto a su contribución a la actividad económica industrial del país, la cual se mide mediante el índice I_c .

- El porcentaje representado en cada nivel, respecto a la cobertura de las cuatro variables.
- El número de establecimientos acumulado en los diferentes niveles y los puntos donde existían incrementos muy grandes en tal número y cuya respectiva aportación al total de cada una de las cuatro variables antes mencionadas era muy baja.

De un total de 22,925 establecimientos de las 37 clases que cubre la encuesta, resultó un número de no menores igual a 3,131 (13.7% del total de establecimientos). El total de establecimientos menores y no menores por clase de actividad se presenta en el Cuadro N° 8 y los porcentajes de las variables que cubren los establecimientos no menores de cada clase aparecen en el Cuadro N° 9.

III.2 Establecimientos Medianos y Mayores.

Los establecimientos no menores presentan aún una gran variabilidad y concentración de las variables de interés, en unos cuantos establecimientos. Para este tipo de poblaciones asimétricas se ha recomendado (Fretes, 1971) incluir en la muestra con certeza a las unidades más grandes y seleccionar del resto una muestra aleatoria, debido a que las primeras son las que más contribuyen a la varianza de los estimadores.

Con base en lo anterior, se decidió formar dos estratos con los

establecimientos no menores. Los del primer estrato se denominaron "mayores" y los del segundo "medianos".

El índice Ie_2 se usó como variable auxiliar para definir los estratos de establecimientos mayores y medianos. Esto resulta más conveniente que usar una sola de las cuatro variables consideradas en la construcción del índice.

Para clasificar a los establecimientos no menores en mayores y medianos, se utilizó un procedimiento propuesto por Glasser - - (1962), cuyo objetivo es encontrar el número óptimo de unidades a incluir con certeza en una muestra de tamaño n , de manera que se minimice la varianza del estimador del promedio de la variable que se esté considerando, en este caso, del índice Ie_2 . El procedimiento se describe a continuación.

Los N establecimientos no menores que forman una clase de actividad, se disponen en orden ascendente, de acuerdo al valor del índice Ie_2 , quedando un arreglo de la siguiente forma:

$$y_{(1)}, y_{(2)}, \dots, y_{(N)}$$

Puede considerarse que los N establecimientos quedan clasificados en dos estratos. El primero formado por los i establecimientos más grandes, que entran en la muestra con certeza; y el segundo por los $N-i$ establecimientos restantes, de los cuales se selecciona una muestra aleatoria de tamaño $n-i$. Entonces, el promedio poblacional puede expresarse como:

$$\mu = \frac{N-l}{N} \mu_{(N-l)} + \frac{l}{N} \mu_{(l)}$$

donde

$$\mu = \frac{1}{N} \sum_{j=1}^N y_{(j)}$$

$$\mu_{(l)} = \frac{1}{l} \sum_{j=N-l+1}^N y_{(j)}$$

$$\mu_{(N-l)} = \frac{1}{N-l} \sum_{j=1}^{N-l} y_{(j)}$$

Es fácil ver que el promedio μ puede estimarse mediante

$$\hat{\mu} = \frac{N-l}{N} \bar{y}_{(n-l)} + \frac{l}{N} \mu_{(l)}$$

donde

$\bar{y}_{(n-l)}$ = Media de las $n-l$ unidades seleccionadas en la muestra del segundo estrato.

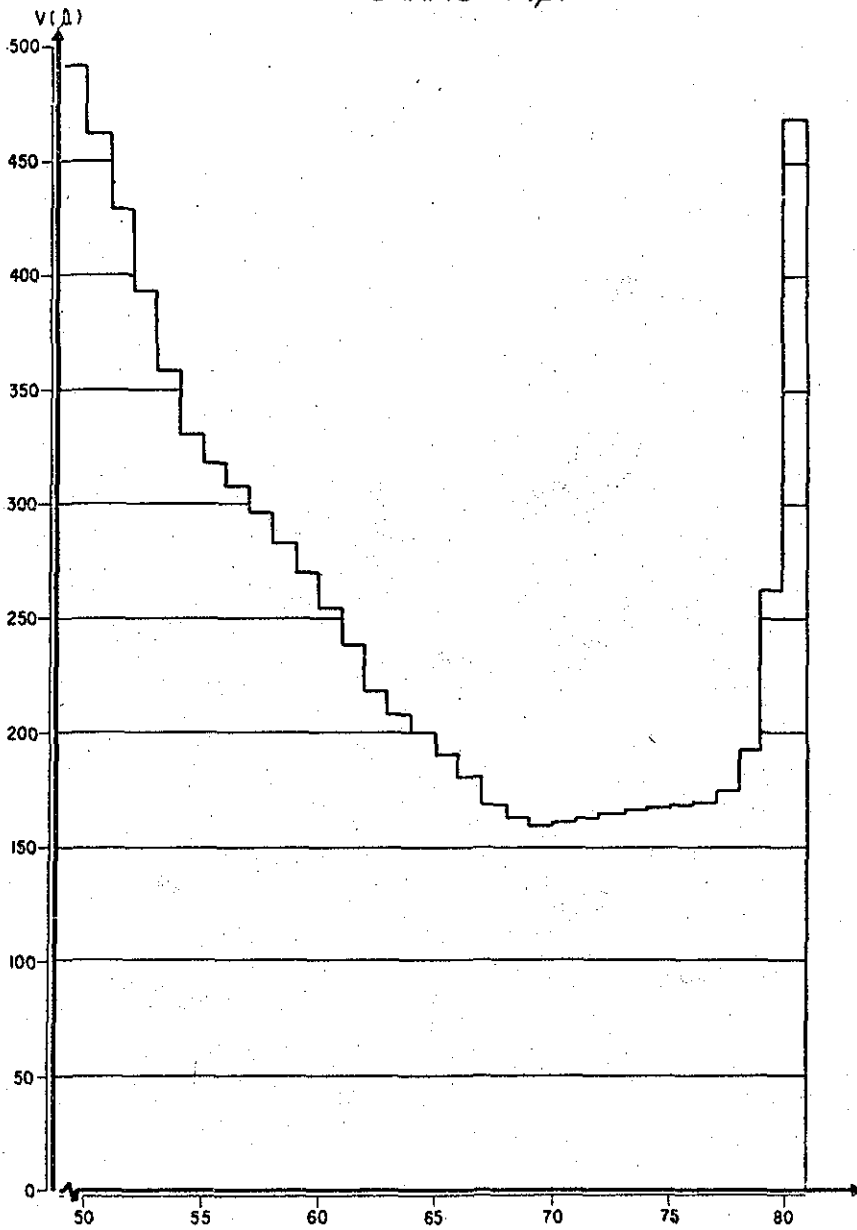
Puesto que $\bar{y}_{(n-l)}$ es un estimador insesgado de $\mu_{(N-l)}$, entonces la varianza de $\hat{\mu}$ es:

$$\begin{aligned}
V(\hat{\mu}) &= V\left(\frac{N-i}{N} \bar{y}_{(n-i)} + \frac{i}{N} \mu_{(i)}\right) \\
&= \left(\frac{N-i}{N}\right)^2 V(\bar{y}_{(n-i)}) \\
&= \left(\frac{N-i}{N}\right)^2 \left(1 - \frac{n-i}{N-i}\right) \frac{S^2_{(N-i)}}{n-i} \\
&= \left(\frac{N-i}{N}\right)^2 \left(1 - \frac{n-i}{N-i}\right) \frac{N-i}{n-i} \frac{\sigma^2_{(N-i)}}{(N-i-1)} \\
&= \left(\frac{N-i}{N}\right)^2 \frac{N-n}{N-i-1} \frac{\sigma^2_{(N-i)}}{n-i}
\end{aligned}$$

El procedimiento consiste en calcular la varianza de $\hat{\mu}$ en forma iterativa empezando con $i=0$ hasta $i=n-1$. De estos valores de $V(\hat{\mu})$ se selecciona el menor y se observa el valor de i para el que se obtuvo ese mínimo. Ese número indica el total de establecimientos que entran al estrato de mayores.

En el apéndice N° 3 se presenta como ejemplo el listado de salida de la clase de actividad 3222. Los valores de $V(\hat{\mu})$ para $i = 50, 51, \dots, 81$ se presentan en la gráfica que aparece en la siguiente página, en donde los valores de $V(\hat{\mu})$ han sido multiplicados por 10^7 para apreciar mejor los resultados. Los valores de $V(\hat{\mu})$ para $i = 0, 1, \dots, 49$ fueron omitidos de la gráfica, pero sus valores corresponderían a: 23912, 17989, 13488, 11044, 9600, 8483, 7561, 6808, 6192, 5620, 5248, 4941, 4696, 4451, 4211, 3986, 3756, 3556, 3362, 3193, 3019, 2879, 2735, 2621, 2502, 2391, 2275, 2166, 2049, 1947, 1838, 1737, 1633, 1532, 1429, 1318, 1213, 1108, 995, 947, 900, 857, 815, 769, 720, 673, 637, 597, 559 y 522 respectivamente, en la escala de medida utilizada.

GRAFICA V ($\hat{\mu}$)



Puede observarse en la gráfica que el valor mínimo de $V(\hat{\mu}_i)$ se localiza en el punto $i = 70$, lo cual indica que el número de establecimientos mayores en esa clase debe ser 70.

Un problema que se presentó en la aplicación del método fue que el tamaño de muestra n no era conocido, por lo que tuvo que ser estimado. Este valor de n se denominó "tamaño de muestra inicial", con objeto de no confundirlo con el tamaño de muestra de la encuesta.

El tamaño de muestra inicial n , de cada clase de actividad se calculó a partir del valor del índice Ie_2 para cada establecimiento, usando la fórmula que corresponde a un muestreo aleatorio estratificado con igual probabilidad de selección y con asignación óptima del tamaño de muestra, dado que se observó una gran variabilidad dentro de los establecimientos no menores. Cabe señalar que esto no implica que en esta etapa se haya seleccionado una muestra de n elementos de la población, simplemente se requirió obtener un tamaño de muestra inicial que fue utilizado en la fórmula de la $V(\hat{\mu}_i)$. La confiabilidad que se usó para esta muestra inicial fue del 95%, con un error máximo permitido del 5% del promedio del índice $Ie_2^{1/}$.

1/ En el diseño de la EMIM se utilizó el 5% del rango del índice Ie_2 , como máximo error de muestreo permitido, para calcular el tamaño de muestra inicial.

Para calcular el tamaño de muestra inicial se formaron cuatro estratos por clase de actividad. Estos se definieron dividiendo el rango del índice Ie_2 en partes iguales^{1/}

Se consideraron costos iguales en cada estrato debido a que no se contaba con información para determinar los costos reales.

Las fórmulas fueron las siguientes (Méndez, 1976):

$$\bar{n} = \frac{\left[\sum_{h=1}^4 N_h S_h \right]^2}{N^2 V_0 + \sum_{h=1}^4 N_h S_h^2}$$

donde

$$V_0 = \left[\frac{.05 (\text{Promedio del Índice } Ie_2)}{1.96} \right]^2$$

el punto 1.96 cumple con

$$P \left[\hat{\bar{y}} - 1.96 \sqrt{V(\hat{\bar{y}})} \leq \bar{y} \leq \hat{\bar{y}} + 1.96 \sqrt{V(\hat{\bar{y}})} \right] = 0.95$$

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2$$

y_{hi} = Valor del índice Ie_2 en el establecimiento i -ésimo del estrato h .

^{1/} En cambio, para calcular el tamaño de muestra de la encuesta, se consideró un número variable de estratos por clase de actividad y el procedimiento de estratificación que se describe en el siguiente capítulo.

$$\bar{y}_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y_{hi}$$

N_h = Total de establecimientos en el estrato h .

Cuando N_h era igual a 1 para algún estrato, se restaba una unidad al valor de N y no se consideraba ese estrato para el cálculo del tamaño de muestra inicial. Al final se incrementaba en una unidad el valor obtenido de n .

En el Cuadro N° 10 se presenta el tamaño de muestra inicial y el porcentaje que representa con respecto al número de establecimientos no menores por clase de actividad.

Al aplicar el método de Glasser (1962), de un total de 3,131 establecimientos no menores de las 37 clases que cubre la encuesta, resultaron 1,811 mayores y 1,320 medianos (7.9% y 5.7% del total de establecimientos, respectivamente).

En el cuadro N° 11 se muestra el número de establecimientos mayores y los porcentajes de las cuatro variables que cubren esos establecimientos, con respecto a los no menores, por clase de actividad y en el cuadro N° 12 se presenta esa misma información para los establecimientos medianos.

Como puede observarse en esos cuadros, los establecimientos mayores, que representan poco más de la mitad de los establecimientos no menores, aportan casi la totalidad de las cuatro variables, en la mayoría de las clases de actividad. Esto confirma que el método efectivamente separa a los establecimientos más grandes de cada

clase de actividad.

En el apéndice N° 4 se presenta un listado de salida del programa que calcula las coberturas de los establecimientos mayores y medianos.

**TESIS CON
FALLA DE CÉGEN**

IV. DISEÑO DE LA MUESTRA.

El diseño de la muestra se realizó en forma independiente para cada clase de actividad, puesto que se requerían estimaciones por separado a ese nivel.

Como se mencionó en el capítulo anterior, los establecimientos fueron clasificados en tres estratos: mayores, medianos y menores. De esta manera se redujo la heterogeneidad de los mismos dentro de cada estrato, en cuanto a características económicas se refiere. Sin embargo, debido a que aún existen marcadas diferencias entre los establecimientos medianos, se decidió subestratificarlos, a fin de lograr una mayor precisión en las estimaciones^{1/}.

A continuación se describen los métodos que se utilizaron para subestratificar y seleccionar la muestra, para calcular el tamaño de ésta y su afijación; así como las fórmulas para calcular los estimadores y sus varianzas o errores cuadrático medios.

IV.1 Subestratificación y Procedimiento de Selección.

La variable que se utilizó para la subestratificación fue el índice I_{e_2} .

En lo referente al número de subestratos se decidió que fuera variable, en función del número de establecimientos medianos (N_2) en cada clase de actividad, esto es, se formaron más estratos en las clases con mayor cantidad de establecimientos. Para las clases con menos de 30 establecimientos no se consideró

^{1/} En el diseño muestral de la EMIM los establecimientos medianos no se subestratificaron.

necesario realizar la subestratificación. En las demás clases se formaron de 2 a 6 estratos de acuerdo al siguiente criterio:

Si $30 \leq N_2 \leq 44$ se formaron dos subestratos.

$45 \leq N_2 \leq 59$ se formaron tres subestratos.

$60 \leq N_2 \leq 74$ se formaron cuatro subestratos.

$75 \leq N_2 \leq 89$ se formaron cinco subestratos.

$90 \leq N_2$ se formaron seis subestratos.

Una vez decidido el número de subestratos a formarse, estaba el problema de determinar los límites de los mismos, de manera que se minimizara la varianza del estimador del promedio, bajo un esquema de muestreo aleatorio estratificado con afijación óptima del tamaño de muestra. Los resultados teóricos en este sentido suponen que la variable de estratificación, denotada por y , es precisamente la variable en estudio y que se conoce la distribución de y , la cual es continua y tiene función de densidad $f(y)$, $a \leq y \leq b$. Para formar L estratos el rango de y es cortado en puntos $y_0 < y_1 < \dots < y_L$. Sean μ_h y σ_h la media y la desviación estándar del estrato h y $w_h = N_h/N$ la razón del número de unidades del estrato h al total de unidades en la población.

Por métodos de cálculo estándar, Dalenius (1950) mostró que el óptimo v_h satisface las ecuaciones

$$\sigma_h + \frac{(y_h - \mu_h)^2}{\sigma_h} = \sigma_{h+1} + \frac{(y_h - \mu_{h+1})^2}{\sigma_{h+1}}$$

Estas ecuaciones son difíciles de resolver, puesto que las cantidades μ_h y σ_h dependen de y_h .

Sin embargo, se han desarrollado algunas reglas sencillas para establecer los límites de los estratos, que permiten obtener una aproximación de la solución óptima.

Probablemente la mejor solución aproximada ha sido dada por Dalenius y Hodges (1957), la cual se conoce como la regla del cum $\sqrt{f(y)}$, donde $f(y)$ denota la función de densidad de la variable de estratificación. Esta regla es la que se utilizó para determinar los límites de los subestratos de los establecimientos medianos y consiste en calcular los valores de $\sqrt{f(y)}$, acumular estos valores y dividir el total acumulado, T , en L partes aproximadamente iguales. Los límites de estrato aproximados son

$$y_1 = \frac{T}{L}; y_2 = \frac{2T}{L}; \dots; y_h = \frac{hT}{L}$$

En la práctica, la variable de estratificación viene en intervalos de clase y hay que hacer aproximaciones gruesas. La regla supone intervalos de clase de la misma longitud d , y debe modificarse cuando los intervalos de clase tienen longitud d_y . En este caso, se acumulan los valores de $\sqrt{d_y f(y)}$ y se divide la aproximación en partes aproximadamente iguales (Kish, 1965).

En el cuadro N° 13 se muestra el tamaño de los subestratos obtenidos aplicando la regla del cum $\sqrt{f(y)}$.

Una justificación no rigurosa de esta regla fue dada por Dalenius y Hodges (1959).

En un estudio realizado por Cochran (1961) la regla del $\sqrt{f(y)}$ es comparada con otros tres métodos sencillos para determinar límites de estratos.

Para la comparación se utilizaron 8 poblaciones reales aproximadamente del tipo de las que se encuentran en el muestreo de instituciones, es decir, con distribución sesgada en la cual existen muchas instituciones pequeñas y unas pocas grandes. La comparación se realizó para subdivisiones en 2, 3 y 4 estratos. De los resultados obtenidos se observa que, en términos generales, la regla del $\sqrt{f(y)}$ funciona mejor que las demás.

En lo referente al procedimiento para seleccionar la muestra de la encuesta, como se ha mencionado anteriormente los establecimientos mayores entran con certeza. Los establecimientos medianos, una vez subestratificados (en su caso) son seleccionados con igual probabilidad y sin reemplazo dentro de cada subestrato, siendo independiente la selección de un subestrato a otro.

IV.2 Tamaño de Muestra y su Afijación.

El criterio para calcular el tamaño de muestra fue minimizar los costos para una varianza fija V_0 y utilizar afijación óptima del tamaño de muestra en los subestratos.

Para determinar el tamaño de muestra cuando se utiliza la afijación óptima, es necesario considerar los costos de recolección; para los establecimientos medianos se supuso costos iguales en los distintos substratos, debido a que el método de recolección en todos los casos es la entrevista directa o diferida y se considera que no hay diferencias significativas en la distribución geográfica de los establecimientos para los diferentes substratos.

El tamaño de muestra se calculó para cada una de las 4 variables consideradas, utilizando la siguiente fórmula (Méndez, 1970):

$$n = \frac{\left[\sum_{h=1}^L N_h S_h \right]^2}{V_0 + \sum_{h=1}^L N_h S_h^2}$$

donde:

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{y}_h)^2$$

y_{hi} = Valor de la variable Y del establecimiento i -ésimo del substrato h .

$$\bar{y}_h = \frac{y_h}{N_h} = \frac{\sum_{i=1}^{N_h} y_{hi}}{N_h}$$

y_h = Total de la variable Y , en el subestrato h .

\bar{y}_h = Promedio de la variable Y , en el subestrato h .

N_h = Número de establecimientos en el subestrato h .

$$V_0 = \left[\frac{.05 \{ \text{Total de la Variable } Y \}}{1.96} \right]^2$$

Donde V_0 se determina de manera que el máximo error de muestreo permitido sea del 5% del total de la variable Y ^{1/} para los establecimientos no menores, con un nivel de confianza del 95%

La afijación óptima de la muestra a los subestratos es como sigue (Raj, 1968):

$$n_h = \frac{n N_h S_h}{\sum_{h=1}^L N_h S_h}$$

^{1/} En el diseño de la EMIM se utilizó el 5% del rango de la variable Y , como máximo error de muestreo permitido.

De los 4 valores de n_h obtenidos para el subestrato h , se eligió el mayor de ellos como tamaño de muestra. En esta forma se satisfacen simultáneamente los requerimientos de error máximo permitido y confiabilidad para las cuatro variables.

El tamaño de muestra para la encuesta es de 2,013 establecimientos. En el cuadro N° 14 se presentan los tamaños de muestra por clase de actividad y en el apéndice N° 5 se muestra un ejemplo de los listados de salida del programa que calcula los tamaños de muestra.

IV.3 Estimadores.

Los estimadores se presentan por clase de actividad, para la población objetivo de la encuesta, es decir, para los establecimientos no menores. Estos estimadores se refieren a: totales, promedios, razón entre dos variables y estimación para dominios de estudio.

La presentación es por medio de cuadros en los que aparecen los estimadores de la característica que se está considerando (total, promedio, etc.) para establecimientos mayores, para medianos y para no menores; la varianza o el error cuadrático medio (ECM) real y sus estimadores; y el intervalo de confianza aproximado al 5% para las estimaciones, sin considerar los errores ajenos al muestreo. En esos cuadros N_1 es el total de establecimientos mayores y N_2 el total de medianos.

La población de establecimientos no menores al dividirse en dos estratos: mayores y medianos, dentro de cada clase de actividad hace necesario calcular estimadores separados para estos dos estratos y posteriormente calcular el estimador combinado.

En el caso de establecimientos mayores (estrato 1), los estimadores son iguales a los valores poblacionales, ya que la selección de las unidades es con certeza.

En el caso de establecimientos medianos (estrato 2) el esquema de muestreo es estratificado; siendo la selección con igual probabilidad y sin reemplazo dentro de cada subestrata

Todos los estimadores para establecimientos medianos y no menores, requieren de estimaciones separadas para cada subestrata del estrato 2.

Las estimaciones separadas necesarias son las siguientes:

a) Total del Subestrato.

Para cada uno de los L subestratos en que se dividió el estrato de establecimientos medianos, un estimador insesgado del total de la variable Y es: (Raj, 1968).

$$\hat{Y}_{2h} = N_{2h} \bar{y}_{2h}$$

donde:

N_{2h} = Total de establecimientos en el subestrato h del estrato 2.

$\bar{y}_{2h} = \frac{\sum y_{2hi}}{n_{2h}}$ = Medía muestral de la variable Y del subestrato h del estrato 2.

$$= \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi}$$

n_{2h} = Número de establecimientos en muestra en el subestrato h del estrato 2.

y_{2hi} = Valor de la variable Y del establecimiento i -ésimo de la muestra, en el subestrato h , del estrato 2.

b) Varianza real y estimada por subestrato (Raj, 1968).

$$V(\hat{y}_{2h}) = \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] S_{2hy}^2$$

donde:

$$S_{2hy}^2 = \frac{1}{N_{2h} - 1} = \sum_{i=1}^{N_{2h}} (y_{2hi} - \bar{y}_{2h})^2$$

Un estimador de la varianza de \hat{y}_{2h} es (Raj, 1968).

$$V(\hat{y}_{2h}) = \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] s_{2hy}^2$$

donde

$$s_{2hy}^2 = \frac{1}{n_{2h} - 1} \sum_{i=1}^{n_{2h}} (y_{2hi} - \bar{y}_{2h})^2$$

IV.3.1 Estimadores de totales.

Los estimadores de totales se presentan en el cuadro N° 15 para los establecimientos mayores, para los medianos y para los no menores.

a) Para los establecimientos mayores el total estimado de la variable es igual al total poblacional, el cual se denota por y_1 .

$$y_1 = \sum_{i=1}^{N_1} y_{1i}$$

donde

y_{1i} = Valor de la variable y del i -ésimo establecimiento mayor.

N_1 = Número de establecimientos mayores.

b) El estimador del total de la variable Y para el estrato de establecimientos medianos se obtiene a partir de los estimadores de totales de los subestratos que lo forman. Para una población de N_2 elementos dividida en L subestratos, de la cual se selecciona una muestra probabilística independiente de cada subestrato, se sabe que (Raj, 1968).

$$\hat{y}_2 = \sum_{h=1}^L \hat{y}_{2h}$$

donde

\hat{y}_2 = Estimador del total de la variable Y en el estrato de establecimientos medianos.

\hat{y}_{2h} = Estimador del total de la variable Y en el h -ésimo subestrato del estrato de establecimientos medianos.

L = Número de subestratos.

La varianza de \hat{y}_2 es (Raj, 1968).

$$V(\hat{y}_2) = \sum_{h=1}^L V(\hat{y}_{2h})$$

Un estimador de la varianza de \hat{y}_2 es (Raj, 1968).

$$\hat{V}(\hat{y}_2) = \sum_{h=1}^L \hat{V}(\hat{y}_{2h})$$

c) Para obtener el estimador del total de la variable Y para establecimientos no menores; denotado por \hat{Y} se utiliza el total poblacional Y_1 , de los establecimientos mayores y el estimador \hat{Y}_2 de los establecimientos medianos, aplicando el principio usado en el inciso anterior, siendo

$$\hat{Y} = Y_1 + \hat{Y}_2$$

$$V(\hat{Y}) = V(Y_1 + \hat{Y}_2) = V(\hat{Y}_2)$$

$$\hat{V}(\hat{Y}) = \hat{V}(Y_1 + \hat{Y}_2) = \hat{V}(\hat{Y}_2)$$

IV.3.2 Estimadores de Promedios.

En ocasiones es importante estimar el promedio de alguna variable dentro de una clase de actividad, por ejemplo: el promedio de obreros por establecimiento o la producción promedio, las ventas promedio, etc. Esto permite realizar comparaciones entre clases de actividad, además de otros usos importantes.

En el cuadro N° 16 se presentan los estimadores de promedios para establecimientos mayores, medianos y no menores.

a) Para los establecimientos mayores el promedio poblacional se denota por \bar{Y}_1 y se obtiene como

$$\bar{Y}_1 = \frac{Y_1}{N_1}$$

- b) La media poblacional de la variable Y de los establecimientos medianos se denota por \bar{Y}_2 y se obtiene como

$$\bar{Y}_2 = \frac{Y_2}{N_2}$$

donde

Y_2 = total poblacional de la variable Y para los establecimientos medianos.

N_2 = Número de establecimientos medianos.

\bar{Y}_2 puede estimarse mediante [Raj, 1968].

$$\hat{\bar{Y}}_2 = \frac{\hat{Y}_2}{N_2}$$

Siendo la varianza de $\hat{\bar{Y}}_2$ la siguiente

$$V(\hat{\bar{Y}}_2) = V\left(\frac{\hat{Y}_2}{N_2}\right) = \frac{1}{N_2^2} V(\hat{Y}_2)$$

Un estimador de la varianza de $\hat{\bar{Y}}_2$ es

$$\hat{V}(\hat{\bar{Y}}_2) = \hat{V}\left(\frac{\hat{Y}_2}{N_2}\right) = \frac{1}{N_2^2} \hat{V}(\hat{Y}_2)$$

- c) Para los establecimientos no menores el estimador del promedio de la variable Y se obtiene como

$$\hat{V} = \frac{\hat{y}}{N} = \frac{y_1 + \hat{y}_2}{N_1 + N_2}$$

Siendo la varianza de y calculada como

$$V(\hat{y}) = V\left(\frac{y_1 + \hat{y}_2}{N_1 + N_2}\right) = \frac{V(\hat{y}_2)}{(N_1 + N_2)^2}$$

Un estimador de la varianza de \hat{V} es

$$\hat{V}(\hat{V}) = \hat{V}\left(\frac{y_1 + \hat{y}_2}{N_1 + N_2}\right) = \frac{\hat{V}(\hat{y}_2)}{(N_1 + N_2)^2}$$

IV.3.3 Estimadores de la razón de dos variables.

En ocasiones resulta importante estimar la razón de dos variables, a partir de los datos de una muestra. Por ejemplo, la razón de la producción y del personal ocupado, para una clase de actividad particular. Estas estimaciones resultan muy útiles para analizar las características de las clases de actividad y realizar comparaciones entre ellas.

A diferencia de los estimadores de totales y promedios mencionados anteriormente, el estimador de razón es sesgado. Por tal motivo, es preferible usar el error cuadrático medio (ECM) en lugar de la varianza, como una medida de dispersión del estimador (Cochran, 1977).

Como en los dos cuadros anteriores, en el cuadro N° 17 se presentan los estimadores para establecimientos mayores, medianos y no menores.

a) Para el estrato de establecimientos mayores, la razón de las variables Y y X está dada por el cociente de los totales poblacionales de los establecimientos mayores.

$$R_1 = \frac{y_1}{x_1} = \frac{\sum_{i=1}^{N_1} y_{1i}}{\sum_{i=1}^{N_1} x_{1i}}$$

b) Para el estrato de establecimientos medianos se utilizó el estimador de razón combinado que está dado por el cociente de los estimadores \hat{y}_2 y \hat{x}_2 . (Raj, 1968).

$$\hat{R}_2 = \frac{\hat{y}_2}{\hat{x}_2} = \frac{\sum_{h=1}^L N_{2h} \bar{y}_{2h}}{\sum_{h=1}^L N_{2h} \bar{x}_{2h}}$$

donde

$$\bar{y}_{2h} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi}$$

$$\bar{x}_{2h} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} x_{2hi}$$

El error cuadrático medio de \hat{R}_2 es, aproximadamente (Raj, 1968).

$$ECM(\hat{R}_2) \doteq \frac{1}{X_2^2} \left[V(\hat{y}_2) + R_2^2 V(\hat{x}_2) - 2 R_2 \text{Cov}(\hat{y}_2, \hat{x}_2) \right]$$

donde

$$R_2 = \frac{y_2}{x_2}$$

$$V(\hat{y}_2) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] S_{2hy}^2$$

$$V(\hat{x}_2) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] S_{2hx}^2$$

$$\text{Cov}(\hat{y}_2, \hat{x}_2) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] \sum_{i=1}^{n_{2h}} \frac{(y_{2hi} - \bar{y}_{2h})(x_{2hi} - \bar{x}_{2h})}{N_{2h} - 1}$$

$$S_{2hy}^2 = \frac{1}{N_{2h}-1} \sum_{i=1}^{N_{2h}} (y_{2hi} - \bar{y}_{2h})^2; \quad \bar{y}_{2h} = \frac{1}{N_{2h}} \sum_{i=1}^{N_{2h}} y_{2hi}$$

$$S_{2hx}^2 = \frac{1}{N_{2h}-1} \sum_{i=1}^{N_{2h}} (x_{2hi} - \bar{x}_{2h})^2; \quad \bar{x}_{2h} = \frac{1}{N_{2h}} \sum_{i=1}^{N_{2h}} x_{2hi}$$

Un estimador del error cuadrático medio aproximado de \hat{R}_2 es:

$$\widehat{ECM}(\hat{R}_2) \approx \frac{1}{\hat{x}_2^2} \left[\hat{V}(\hat{y}_2) + \hat{R}_2^2 \hat{V}(\hat{x}_2) - 2 \hat{R}_2 \widehat{Cov}(\hat{y}_2, \hat{x}_2) \right]$$

donde

$$\hat{V}(\hat{y}_2) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] \delta_{2hy}^2$$

$$\hat{V}(\hat{x}_2) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] \delta_{2hx}^2$$

$$\widehat{Cov}(\hat{y}_2, \hat{x}_2) = \sum_{h=1}^L N_{2h}^2 \left[1 - \frac{n_{2h}}{N_{2h}} \right] \sum_{i=1}^{n_{2h}} \frac{(y_{2hi} - \bar{y}_{2h})(x_{2hi} - \bar{x}_{2h})}{n_{2h}(n_{2h}-1)}$$

$$\delta_{2hy}^2 = \frac{1}{n_{2h}-1} \sum_{i=1}^{n_{2h}} (y_{2hi} - \bar{y}_{2h})^2; \quad \bar{y}_{2h} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi}$$

$$\delta_{2hx}^2 = \frac{1}{n_{2h}-1} \sum_{i=1}^{n_{2h}} (x_{2hi} - \bar{x}_{2h})^2; \quad \bar{x}_{2h} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} x_{2hi}$$

El intervalo de confianza al 95% aproximado es:

$$\left[\hat{R}_2 - 1.96 \sqrt{\widehat{ECM}(\hat{R}_2)} \leq R_2 \leq \hat{R}_2 + 1.96 \sqrt{\widehat{ECM}(\hat{R}_2)} \right]$$

- c) El estimador de razón combinado para los establecimientos no menores es el siguiente:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}} = \frac{y_1 + \hat{y}_2}{x_1 + \hat{x}_2}$$

El error cuadrático medio aproximado del estimador de razón \hat{R} está dado por

$$ECM(\hat{R}) \approx \frac{1}{\hat{X}^2} [V(\hat{Y}) + R^2 V(\hat{X}) - 2R \text{Cov}(\hat{Y}, \hat{X})]$$

donde

$$R = \frac{Y}{X} = \frac{y_1 + y_2}{x_1 + x_2}$$

$$V(\hat{Y}) = V(\hat{y}_2)$$

$$V(\hat{X}) = V(\hat{x}_2)$$

$$\text{Cov}(\hat{Y}, \hat{X}) = \text{Cov}(y_1 + \hat{y}_2, x_1 + \hat{x}_2) = \text{Cov}(\hat{y}_2, \hat{x}_2)$$

$$= \text{Cov} \left(\sum_{h=1}^L \hat{y}_{2h}, \sum_{h=1}^L \hat{x}_{2h} \right) = \sum_{h=1}^L \text{Cov}(\hat{y}_{2h}, \hat{x}_{2h})$$

$$= \sum_{h=1}^L \text{Cov}(N_{2h} \bar{y}_{2h}, N_{2h} \bar{x}_{2h}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) S_{2hy} S_{2hx}$$

$$= \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) \frac{1}{N_{2h} - 1} \sum_{i=1}^{N_{2h}} (y_{2hi} - \bar{y}_{2h})(x_{2hi} - \bar{x}_{2h})$$

Un estimador aproximado del error cuadrático medio de \hat{R} es:

$$\widehat{ECM}(\hat{R}) \approx \frac{1}{\hat{X}^2} [\hat{V}(\hat{Y}) + \hat{R}^2 \hat{V}(\hat{X}) - 2\hat{R} \widehat{\text{Cov}}(\hat{Y}, \hat{X})]$$

donde

$$\hat{V}(\hat{Y}) = \hat{V}(\hat{Y}_2)$$

$$\hat{V}(\hat{X}) = \hat{V}(\hat{X}_2)$$

$$\hat{Cov}(\hat{Y}, \hat{X}) = \hat{Cov}(Y_1 + \hat{Y}_2, X_1 + \hat{X}_2)$$

$$= \hat{Cov}(\hat{Y}_2, \hat{X}_2)$$

El intervalo de confianza de \hat{R} aproximado al 95% es:

$$\left[\hat{R} - 1.96 \sqrt{\widehat{ECM}(\hat{R})} \leq R \leq \hat{R} + 1.96 \sqrt{\widehat{ECM}(\hat{R})} \right]$$

IV.3.4 Estimación en dominios de estudio.

Las estimaciones de una encuesta pueden obtenerse para algunas porciones o subdivisiones de la población, a las que la Subcomisión de Muestreo de la Organización de las Naciones Unidas (1950) ha denominado "dominios de estudio", (Cochran, 1977). En este caso, los dominios de estudio están constituidos por establecimientos que poseen cierta característica α , por ejemplo, aquellos que fabrican un producto determinado o los que realizan actividades de maquila.

Para dominios de estudio se presentan estimadores de totales y de promedios, en los cuadros 18 y 19, respectivamente.

IV.3.4.1 Estimador de total, varianza e intervalo de confianza aproximado.

a) Para el estrato de establecimientos mayores, el total

de la variable Y de los establecimientos que poseen la característica α , es el valor poblacional.

$$y_1^\alpha = \sum_{i=1}^{N_1} y_{1i} s_i$$

donde

$$s_i = \begin{cases} 1 & \text{si el establecimiento } i\text{-ésimo posee la} \\ & \text{característica } \alpha. \\ 0 & \text{en caso contrario.} \end{cases}$$

b) En el estrato de establecimientos medianos, un estimador insesgado del total de la variable Y , de los establecimientos que poseen la característica α es (Cochran, 1977).

$$\hat{y}_2^\alpha = \sum_{h=1}^L \hat{y}_{2h}^\alpha = \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi} s_i$$

Es fácil observar que \hat{y}_2^α es un estimador insesgado de y_2^α , puesto que

$$\begin{aligned} E[\hat{y}_2^\alpha] &= E\left[\sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi} s_i \right] \\ &= \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} E(y_{2hi} s_i) \\ &= \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} \sum_{i=1}^{N_{2h}} \frac{y_{2hi} s_i}{N_{2h}} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{h=1}^L \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} \sum_{i=1}^{N_{2h}} y_{2hi} s_i \\
 &= \sum_{h=1}^L \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2h}^{\alpha} = \sum_{h=1}^L y_{2h}^{\alpha} = y_2^{\alpha}
 \end{aligned}$$

La varianza del estimador del total de la variable y , de establecimientos que poseen la característica α , en el estrato de establecimientos medianos es:

$$\begin{aligned}
 &V(\hat{y}_2^{\alpha}) \\
 &= V\left(\sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi} s_i\right) \\
 &= \sum_{h=1}^L V\left(\frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi} s_i\right) \\
 &= \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) \frac{1}{n_{2h}-1} \left[\sum_{i=1}^{n_{2h}} (y_{2hi} s_i - \overline{y_{2hi} s_i})^2\right]
 \end{aligned}$$

y el estimador de la varianza es (Cochran, 1977)

$$\begin{aligned}
 &\hat{V}(\hat{y}_2^{\alpha}) \\
 &= \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) \frac{1}{n_{2h}-1} \left[\sum_{i=1}^{n_{2h}} (y_{2hi} s_i - \overline{y_{2hi} s_i})^2\right]
 \end{aligned}$$

El intervalo de confianza aproximado al 95% es:

$$\left[\hat{y}_2^{\alpha} - 1.96 \sqrt{\hat{V}(\hat{y}_2^{\alpha})} \leq y_2^{\alpha} \leq \hat{y}_2^{\alpha} + 1.96 \sqrt{\hat{V}(\hat{y}_2^{\alpha})} \right]$$

- c) El estimador del total de la variable Y para esta establecimientos no menores, que poseen la característica α , es

$$\hat{y}^{\alpha} = y_1^{\alpha} + \hat{y}_2^{\alpha}$$

La varianza de \hat{y}^{α} es:

$$V(\hat{y}^{\alpha}) = V(\hat{y}_2^{\alpha})$$

Un estimador de la varianza de \hat{y}^{α} es:

$$\hat{V}(\hat{y}^{\alpha}) = \hat{V}(\hat{y}_2^{\alpha})$$

El intervalo de confianza aproximado al 95% es:

$$\left[\hat{y}^{\alpha} - 1.96 \sqrt{\hat{V}(\hat{y}^{\alpha})} \leq y^{\alpha} \leq \hat{y}^{\alpha} + 1.96 \sqrt{\hat{V}(\hat{y}^{\alpha})} \right]$$

IV.3.4.2 Estimador de promedio, error cuadrático medio e intervalo de confianza aproximado.

- a) Para el estrato de establecimientos mayores, el promedio poblacional de la variable Y , de los esta establecimientos con la característica α , es

$$\bar{y}_1^{\alpha} = \frac{y_1^{\alpha}}{N_1^{\alpha}}$$

N_1 = total de establecimientos con la característica α en el estrato de establecimientos mayores.

- b) Para el estrato de establecimientos medianos el estimador del promedio de la variable Y , en los establecimientos que poseen la característica α es:

$$\hat{y}_2^\alpha = \frac{\hat{Y}_2^\alpha}{\hat{N}_2^\alpha}$$

\hat{N}_2^α = estimador del número de establecimientos con la característica α en el estrato de medianos.

Es importante hacer notar que en este estimador interviene la estimación del número de establecimientos del estrato de medianos que poseen la característica α , denotado por \hat{N}_2^α .

Esto se debe a que se desconoce el verdadero número de establecimientos con la característica α en el estrato de medianos (N_2^α).

El hecho de que tanto el numerador como el denominador del estimador \hat{y}_2^α estén en términos de una estimación, hace que \hat{y}_2^α tome la forma de una estimación de razón, motivo por el cual se le dará a éste el mismo tratamiento que se dió al estimador de razón.

Ahora bien, un estimador insesgado del total de establecimientos con la característica α es:

$$\hat{N}_2^\alpha = \sum_{h=1}^L \hat{N}_{2h}^\alpha = \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{n_{2h}} \delta_i$$

Para calcular la varianza de \hat{N}_2^α se observa que \hat{N}_{2h}^α puede expresarse como

$$N_{2h}^\alpha = N_{2h} \frac{N_{2h}^\alpha}{N_{2h}} = N_{2h} P_{2h}$$

donde

P_{2h} = Proporción de establecimientos en la población, que poseen la característica α .

El estimador \hat{N}_{2h}^α puede expresarse como

$$\hat{N}_{2h}^\alpha = N_{2h} \hat{P}_{2h}$$

donde

$$\hat{P}_{2h} = \frac{n_{2h}^\alpha}{n_{2h}} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} \delta_i$$

Entonces (Cochran, 1977).

$$V(\hat{N}_2^\alpha) = \sum_{h=1}^L V(\hat{N}_{2h}^\alpha) = \sum_{h=1}^L N_{2h}^2 V(\hat{P}_{2h})$$

$$= \sum_{h=1}^L N_{2h}^2 \frac{N_{2h}^\alpha}{N_{2h}} \left[1 - \frac{N_{2h}^\alpha}{N_{2h}} \right] \frac{(N_{2h} - n_{2h})}{n_{2h}(N_{2h} - 1)}$$

Como estimador de la varianza se tiene (Cochran, 1977).

$$\hat{V}(\hat{N}_2^\alpha) = \sum_{h=1}^L N_{2h}^2 \hat{V}(\hat{P}_{2h})$$

$$= \sum_{h=1}^L N_{2h}^2 \frac{(N_{2h} - n_{2h})}{(n_{2h} - 1)N_{2h}} \frac{n_{2h}^\alpha}{n_{2h}} \left[1 - \frac{n_{2h}^\alpha}{n_{2h}} \right]$$

La covarianza de \hat{Y}_2^α y \hat{N}_2^α está dada por

$$\text{Cov}(\hat{Y}_2^\alpha, \hat{N}_2^\alpha)$$

$$= \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] \sum_{i=1}^{N_{2h}} \frac{(y_{2hi} s_i - \frac{y_{2hi} s_i}{N_{2h}}) (s_i - \frac{N_{2h}^\alpha}{N_{2h}})}{N_{2h} - 1}$$

$$= \sum_{h=1}^L \frac{(N_{2h} - n_{2h})(N_{2h} - N_{2h}^\alpha)}{n_{2h}(N_{2h} - 1)} \sum_{i=1}^{N_{2h}} y_{2hi} s_i$$

donde

$$\overline{y_{2hi} s_i} = \sum_{i=1}^{N_{2h}} \frac{y_{2hi} s_i}{N_{2h}}$$

El estimador de la covarianza $\widehat{Cov}(\widehat{Y}_2^\alpha, \widehat{N}_2^\alpha)$ es:

$$\begin{aligned} & \widehat{Cov}(\widehat{Y}_2^\alpha, \widehat{N}_2^\alpha) \\ &= \sum_{h=1}^L N_{2h}^2 \left(1 - \frac{n_{2h}}{N_{2h}}\right) \sum_{i=1}^{n_{2h}} \frac{(y_{2hi} s_i - \overline{y_{2hi} s_i}) (s_i - \frac{n_{2h}}{N_{2h}})}{N_{2h} (n_{2h} - 1)} \end{aligned}$$

donde

$$\overline{y_{2hi} s_i} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y_{2hi} s_i$$

El error cuadrático medio para el estimador \widehat{Y}_2^α es:

$$\begin{aligned} & ECM(\widehat{Y}_2^\alpha) \\ &= \frac{1}{(N_2^\alpha)^2} \left[V(\widehat{Y}_2^\alpha) + (\widehat{Y}_2^\alpha)^2 V(\widehat{N}_2^\alpha) - 2 \widehat{Y}_2^\alpha \widehat{Cov}(\widehat{Y}_2^\alpha, \widehat{N}_2^\alpha) \right] \end{aligned}$$

Un estimador del error cuadrático medio es:

$$\widehat{ECM}(\hat{y}_2^\alpha)$$

$$= \frac{1}{(\hat{N}_2^\alpha)^2} \left[\hat{V}(\hat{y}_2^\alpha) + (\hat{y}_2^\alpha)^2 \hat{V}(\hat{N}_2^\alpha) - 2 \hat{y}_2^\alpha \widehat{Cov}(\hat{y}_2^\alpha, \hat{N}_2^\alpha) \right]$$

El intervalo de confianza para el estimador es:

$$\left[\hat{y}_2^\alpha - 1.96 \sqrt{\widehat{ECM}(\hat{y}_2^\alpha)} \leq y_2^\alpha \leq \hat{y}_2^\alpha + 1.96 \sqrt{\widehat{ECM}(\hat{y}_2^\alpha)} \right]$$

- c) El estimador \hat{y}^α de la media de la variable Y para los establecimientos no menores que poseen la característica α está dado por

$$\hat{y}^\alpha = \frac{y_1^\alpha + \hat{y}_2^\alpha}{N_1^\alpha + \hat{N}_2^\alpha}$$

$$= \frac{1}{N_1^\alpha + \hat{N}_2^\alpha} \left[\sum_{i=1}^{N_1} y_{1i} s_i + \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{i=1}^{N_{2h}} y_{2hi} s_i \right]$$

Para calcular el error cuadrático medio del estimador \hat{y}^α se tienen las expresiones:

La varianza del numerador \hat{y}^α es:

$$V(y_1^\alpha + \hat{y}_2^\alpha) = V(\hat{y}_2^\alpha)$$

y un estimador es:

$$\widehat{V}(y_1^\alpha + \hat{y}_2^\alpha) = \widehat{V}(\hat{y}_2^\alpha)$$

TESIS CON FALLA DE COLGEN

La varianza en el denominador de \hat{V} es igual a la varianza para el estrato de medianos, es decir:

$$V(N_1^x + \hat{N}_2^x) = V(\hat{N}_2^x)$$

y donde un estimador de la varianza es:

$$\hat{V}(N_1^x + \hat{N}_2^x) = \hat{V}(\hat{N}_2^x)$$

La covarianza entre el numerador y el denominador de \hat{V} es

$$\begin{aligned} \text{Cov}(y_1^x + \hat{y}_2^x, N_1^x + \hat{N}_2^x) \\ &= E[(y_1^x + \hat{y}_2^x - E(y_1^x + \hat{y}_2^x))(N_1^x + \hat{N}_2^x - E(N_1^x + \hat{N}_2^x))] \\ &= E[(y_1^x + \hat{y}_2^x - y_1^x - y_2^x)(N_1^x + \hat{N}_2^x - N_1^x - N_2^x)] \\ &= E[(\hat{y}_2^x - E(\hat{y}_2^x))(\hat{N}_2^x - E(\hat{N}_2^x))] = \text{Cov}(\hat{y}_2^x, \hat{N}_2^x) \end{aligned}$$

Un estimador de la covarianza es:

$$\begin{aligned} \hat{\text{Cov}}(y_1^x + \hat{y}_2^x, N_1^x + \hat{N}_2^x) \\ &= \hat{\text{Cov}}(\hat{y}_2^x, \hat{N}_2^x) \end{aligned}$$

El error cuadrático medio es entonces:

$$ECM(\hat{V}^{\alpha}) = \frac{V(\hat{V}_2^{\alpha}) + (\bar{V}^{\alpha})^2 V(\hat{N}_2^{\alpha}) - 2\bar{V}^{\alpha} \text{Cov}(\hat{V}_2^{\alpha}, \hat{N}_2^{\alpha})}{(N_1^{\alpha} + N_2^{\alpha})^2}$$

El error cuadrático medio estimado está dado por

$$\widehat{ECM}(\hat{V}_2^{\alpha}) = \frac{\hat{V}(\hat{V}_2^{\alpha}) + (\hat{V}^{\alpha})^2 \hat{V}(\hat{N}_2^{\alpha}) - 2\hat{V}^{\alpha} \widehat{\text{Cov}}(\hat{V}_2^{\alpha}, \hat{N}_2^{\alpha})}{(N_1^{\alpha} + N_2^{\alpha})^2}$$

El intervalo de confianza aproximado al 95% para el estimador \hat{V}^{α} está dado por

$$\left[\hat{V}^{\alpha} - 1.96 \sqrt{\widehat{ECM}(\hat{V}^{\alpha})} \leq V^{\alpha} \leq \hat{V}^{\alpha} + 1.96 \sqrt{\widehat{ECM}(\hat{V}^{\alpha})} \right]$$

Los estimadores de promedios en dominio de estudio mencionados arriba y sus errores cuadrático medios se presentan en el cuadro N° 19.

VI. SIMULACION DE ESTIMADORES.

En la teoría del muestreo se establece que a partir de una muestra probabilística es posible obtener estimaciones de los parámetros poblacionales y construir intervalos de confianza para esas estimaciones, los cuales contendrían al valor real en aproximadamente el 95%^{1/} de los casos si se seleccionaran muchas muestras del mismo tamaño y con el mismo procedimiento de selección.

A fin de comprobar empíricamente en qué medida se cumple lo anterior y analizar algunas características importantes de los estimadores, - tales como: la varianza, el sesgo, el coeficiente de variación, el error de muestreo, etc; se decidió realizar un estudio de simulación para lo cual se seleccionaron 25 muestras aleatorias distintas en cada clase de actividad.

Los estimadores que se analizaron son: de totales, de razón y de promedios en dominios de estudio. El estudio se realizó separadamente para los establecimientos medianos y para los no menores de cada clase de actividad. Sin embargo, los resultados se presentan solamente para los establecimientos no menores, visto que ese es el nivel al que se va a publicar la información.

Las clases de actividad: 2391, 2392, 2821, 3022, 3031, 3072, 3429, 3593, 3632, 3692, 3817 y 3891 fueron excluidas del estudio por no po

^{1/} En esta encuesta se usó el 95%, aunque este porcentaje podría haber sido diferente.

debe seleccionarse en ellas 25 muestras aleatorias distintas del tamaño requerido.

Para calcular el número de muestras distintas de tamaño n que pueden seleccionarse de una población de N elementos, se utiliza la fórmula

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{(N-n+1)(N-n+2)\dots(N)}{(1)(2)\dots(n)}$$

Cuando se tienen L substratos, el número de muestras distintas es

$$\binom{N_1}{n_1} \binom{N_2}{n_2} \dots \binom{N_L}{n_L}$$

donde N_i y n_i es el tamaño poblacional y muestral del estrato i , respectivamente.

Para seleccionar aleatoriamente a los establecimientos que quedaron incluidos en cada muestra se utilizó la subrutina RANDU (INIVAC, 1972) que produce una serie de números pseudo-aleatorios X_1, X_2, \dots, X_n , distribuidos uniformemente en el intervalo $[0, 1)$, utilizando el método congruencial mixto (Naylor, 1966). Estos números aleatorios se transforman a números aleatorios enteros Y_1, Y_2, \dots, Y_n , distribuidos uniformemente en el intervalo $[0, N]$ mediante

$$Y_i = [X_i * N + 1] \quad i = 1, 2, \dots, n$$

Los paréntesis [] indican la parte entera de número que aparece dentro de ellos.

Los valores y_i corresponden a los elementos de la población que entran en la muestra; estableciéndose controles para que no haya elementos repetidos en la misma.

En las clases en que se subestratificó, la selección se realizó en cada subestrato en forma separada.

Para cada muestra seleccionada se calcularon los valores que se mencionan más adelante, usando la información del marco muestral. La notación utilizada es semejante a la de capítulos anteriores, sólo que se agregó el subíndice k , indicando que se trata de la k -ésima simulación.

V.I Estimadores de totales.

La simulación del estimador de totales se realizó para las variables: P.O., P., A.F. y V.A. Las estadísticas obtenidas se describen a continuación.

1. Total poblacional de la variable Y.

$$Y_0 = Y_1 + Y_2$$

$Y_1 =$ Total de la variable Y de los establecimientos mayores.

$$= \sum_{i=1}^{N_1} Y_{1i}$$

$Y_{1i} =$ Valor de la variable Y en el i -ésimo establecimiento mayor.

$N_1 =$ Número de establecimientos mayores.

$Y_2 =$ Total de la variable Y de los establecimientos media
nos.

$$= \sum_{i=1}^{N_2} Y_{2i}$$

$Y_{2i} =$ Valor de la variable Y en el i -ésimo establecimiento mediano.

$N_2 =$ Número de establecimientos medianos.

2. Media poblacional de la variable Y .

$$\bar{Y}_0 = \frac{Y_0}{N}$$

donde

$$N = N_1 + N_2$$

3. Máximo de los 25 estimadores.

$$\text{Max}(\hat{Y}_{0,k}) = \text{Max} \{ \hat{Y}_{0,1}; \hat{Y}_{0,2}; \dots; \hat{Y}_{0,25} \}$$

donde

$$\hat{Y}_{0,k} = Y_1 + \hat{Y}_{2,k}$$

$$\hat{Y}_{2,k} = \sum_{h=1}^L N_{2h} \bar{Y}_{2h,k}$$

$L =$ Número de subestratos.

$\bar{Y}_{2h,k} =$ Media muestral de la variable Y en el subestrato h ,
en la k -ésima simulación.

$$= \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} Y_{2hi,k}$$

$y_{2hi,k}$ = Valor de la variable y del i -ésimo establecimiento en la muestra del sustrato h , en la k -ésima simulación.

n_{2h} = Tamaño de muestra en el sustrato h .

N_{2h} = Número de establecimientos en el sustrato h .

4. Mínimo de los 25 estimadores.

$$\text{Min}(\hat{y}_{0,k}) = \text{Min} \{ \hat{y}_{0,1}; \hat{y}_{0,2}; \dots; \hat{y}_{0,25} \}$$

5. Rango de los 25 estimadores.

$$\text{Rango}(\hat{y}_{0,k}) = \text{Max}(\hat{y}_{0,k}) - \text{Min}(\hat{y}_{0,k})$$

6. Media de los totales estimados.

$$\bar{\hat{y}}_{0,k} = \frac{1}{25} \sum_{k=1}^{25} \hat{y}_{0,k}$$

7. Error absoluto máximo de $\hat{y}_{0,k}$.

$$\text{EAMax}(\hat{y}_{0,k}) = \text{Max} \{ |\hat{y}_{0,1} - y_0|, |\hat{y}_{0,2} - y_0|, \dots, |\hat{y}_{0,25} - y_0| \}$$

8. Error absoluto mínimo de $\hat{y}_{0,k}$.

$$\text{EAMin}(\hat{y}_{0,k}) = \text{Min} \{ |\hat{y}_{0,1} - y_0|, |\hat{y}_{0,2} - y_0|, \dots, |\hat{y}_{0,25} - y_0| \}$$

9. Promedio del error absoluto de $\hat{y}_{0,k}$.

$$\overline{EA}(\hat{y}_{0,k}) = \frac{1}{25} \sum_{k=1}^{25} |\hat{y}_{0,k} - y_{0,k}|$$

10. Varianza de $\hat{y}_{0,k}$.

$$V(\hat{y}_{0,k}) = V(\hat{y}_{2,k})$$

$$V(\hat{y}_{2,k}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) S_{2hy}^2$$

donde

$$S_{2hy}^2 = \frac{1}{N_{2h} - 1} \sum_{i=1}^{N_{2h}} (y_{2hi} - \bar{y}_{2h})^2$$

11. Desviación estandar de $\hat{y}_{0,k}$.

$$DE(\hat{y}_{0,k}) = \sqrt{V(\hat{y}_{0,k})}$$

12. Estimación de la desviación estandar de $\hat{y}_{0,k}$ en base a las simulaciones

$$\widehat{DE}(\hat{y}_{0,k}) = \sqrt{\widehat{V}(\hat{y}_{0,k})}$$

donde

$$\widehat{V}(\hat{y}_{0,k}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{y}_{0,k} - \bar{\hat{y}}_{0,k})^2$$

13. Coeficiente de variación de $\hat{V}_{0,k}$, ignorando el sesgo

$$CV(\hat{V}_{0,k}) = \frac{\widehat{DE}(\hat{V}_{0,k})}{V_0} \times 100$$

14. Estimación del error cuadrático medio de $\hat{V}_{0,k}$ en base a las simulaciones

$$\widehat{ECM}(\hat{V}_{0,k}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{V}_{0,k} - V_0)^2$$

15. Coeficiente de variación total de $\hat{V}_{0,k}$ en base a las simulaciones.

$$CVT(\hat{V}_{0,k}) = \frac{\sqrt{\widehat{ECM}(\hat{V}_{0,k})}}{V_0}$$

16. Estimación del sesgo de $\hat{V}_{0,k}$, en base a las simulaciones.

$$\widehat{\text{sesgo}}(\hat{V}_{0,k}) = \sqrt{|\widehat{ECM}(\hat{V}_{0,k}) - \widehat{V}(\hat{V}_{0,k})|}$$

17. Proporción del sesgo ($\hat{V}_{0,k}$) relativo a la $\widehat{DE}(\hat{V}_{0,k})$

$$\text{Prop. Sesgo}(\hat{V}_{0,k}) = \frac{\widehat{\text{sesgo}}(\hat{V}_{0,k})}{\widehat{DE}(\hat{V}_{0,k})}$$

18. Número de veces que el valor verdadero V_0 quedó dentro del intervalo de confianza aproximado del 95%.

$$\left[\hat{V}_{0,k} - 1.96 \sqrt{\widehat{V}(\hat{V}_{0,k})}, \hat{V}_{0,k} + 1.96 \sqrt{\widehat{V}(\hat{V}_{0,k})} \right]$$

En el cuadro N° 20 aparecen algunos resultados obtenidos y en el apéndice N° 6 se presenta un listado de salida del estimador de totales para establecimientos no menores.

V.2 Estimadores de razón.

La simulación del estimador de razón se realizó para el cociente P/P.O. Las estadísticas obtenidas se describen a continuación:

1. Total poblacional de la variable Y.

$$Y_0 = Y_1 + Y_2$$

2. Total poblacional de la variable X.

$$X_0 = X_1 + X_2$$

donde

X_1 = Total de la variable X de los establecimientos mayores.

$$= \sum_{i=1}^{N_1} X_{1i}$$

X_{1i} = Valor de la variable X para el i-ésimo establecimiento mayor.

X_2 = Total de la variable X de los establecimientos medianos.

$$= \sum_{i=1}^{N_2} X_{2i}$$

X_{2i} = Valor de la variable X del i-ésimo establecimiento mediano.

3. Coeficiente de correlación entre Y y X

$$\rho = \frac{\text{Cov}(Y_i, X_i)}{\sqrt{\text{Var}(Y_i)} \sqrt{\text{Var}(X_i)}}$$

donde

$$\text{Cov}(Y_i, X_i) = \frac{1}{N} \sum_{i=1}^N (y_{0i} - \bar{y}_0)(x_{0i} - \bar{x}_0)$$

$$\text{Var}(Y_i) = \frac{1}{N} \sum_{i=1}^N (y_{0i} - \bar{y}_0)^2$$

$$\text{Var}(X_i) = \frac{1}{N} \sum_{i=1}^N (x_{0i} - \bar{x}_0)^2$$

4. Coeficiente de variación de $\hat{y}_{0,k}$.

$$CV(\hat{y}_{0,k}) = \frac{\sqrt{V(\hat{y}_{0,k})}}{y_0}$$

5. Coeficiente de variación de $\hat{X}_{0,k}$.

$$CV(\hat{X}_{0,k}) = \frac{\sqrt{V(\hat{X}_{0,k})}}{\hat{X}_{0,k}}$$

Donde

$$\hat{X}_{0,k} = X_1 + \hat{X}_{2,k} = X_1 + \sum_{h=1}^L N_{2h} \bar{X}_{2h,k}$$

$\bar{X}_{2h,k}$ = Media muestral del substrato h en la k-ésima simulación.

$$\text{Var}(\hat{X}_{0,k}) = V(\hat{X}_{2,k})$$

$$\text{Var}(\hat{X}_{2,k}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) S_{2hx}^2$$

$$S_{2hx}^2 = \frac{1}{N_{2h}-1} \sum_{i=1}^{N_{2h}} (X_{2hi} - \bar{X}_{2h})^2$$

$$\bar{X}_{2h} = \frac{1}{N_{2h}} \sum_{i=1}^{N_{2h}} X_{2hi}$$

X_{2hi} = Valor de la variable X para el i-ésimo establecimiento del substrato h.

6. Razón verdadera de Y a X .

$$R = \frac{y_0}{x_0} = \frac{y_1 + y_2}{x_1 + x_2}$$

7. Covarianza entre $\hat{y}_{0,k}$ y $\hat{x}_{0,k}$.

$$\text{Cov}(\hat{y}_{0,k}, \hat{x}_{0,k}) = \text{Cov}(\hat{y}_{2,k}, \hat{x}_{2,k})$$

donde

$$\text{Cov}(\hat{y}_{2,k}, \hat{x}_{2,k}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) S_{2hx} S_{2hy}$$

8. Error cuadrático medio aproximado de $\hat{R}_{0,k}$.

$$\text{ECM}(\hat{R}_{0,k}) = \frac{1}{x_0^2} \left[V(\hat{y}_{0,k}) + R^2 V(\hat{x}_{0,k}) - 2R \text{Cov}(\hat{y}_{0,k}, \hat{x}_{0,k}) \right]$$

9. Raíz cuadrada del error cuadrático medio aproximado de $\hat{R}_{0,k}$.

10. Sesgo de $\hat{R}_{0,k}$.

$$\text{Sesgo}(\hat{R}_{0,k}) = R \left[\text{CV}(\hat{x}_{0,k}) \right]^2 - \frac{\text{Cov}(\hat{y}_{0,k}, \hat{x}_{0,k})}{x_0^2}$$

11. Sesgo de $\hat{R}_{0,k}$ relativo a la razón R .

$$\text{Sesgo rel.}(\hat{R}_{0,k}) = \frac{|\text{Sesgo}(\hat{R}_{0,k})|}{R}$$

12. Sesgo de $\hat{R}_{0,k}$ relativo a la raíz cuadrada del ECM aproximado de $\hat{R}_{0,k}$.

$$\text{Sesgo rel.}(\hat{R}_{0,k}) = \frac{|\text{Sesgo}(\hat{R}_{0,k})|}{\sqrt{\text{ECM}(\hat{R}_{0,k})}}$$

13. Coeficiente de variación de $\hat{R}_{0,k}$.

$$CV(\hat{R}_{0,k}) = \frac{\sqrt{\text{ECM}(\hat{R}_{0,k})}}{R}$$

14. Máximo de los 25 estimadores.

$$\text{Max}(\hat{R}_{0,k}) = \text{Max}\{\hat{R}_{0,1}; \hat{R}_{0,2}; \dots, \hat{R}_{0,25}\}$$

donde

$$\hat{R}_{0,k} = \frac{\hat{y}_{0,k}}{\hat{x}_{0,k}} = \frac{y_1 + \hat{y}_{2,k}}{x_1 + \hat{x}_{2,k}}$$

15. Mínimo de los 25 estimadores.

$$\text{Mín}(\hat{R}_{0,k}) = \text{Mín}\{\hat{R}_{0,1}; \hat{R}_{0,2}; \dots, \hat{R}_{0,25}\}$$

16. Rango de los 25 estimadores.

$$\text{Rango}(\hat{R}_{0,k}) = \text{Max}(R_{0,x}) - \text{Mín}(R_{0,k})$$

17. Error absoluto máximo de $\hat{R}_{0,k}$.

$$EAMax(\hat{R}_{0,k}) = \text{Max } |\hat{R}_{0,1} - R|, |\hat{R}_{0,2} - R|, \dots, |\hat{R}_{0,25} - R|$$

18. Error absoluto mínimo de $\hat{R}_{0,k}$.

$$EAMin(\hat{R}_{0,k}) = \text{Min } |\hat{R}_{0,1} - R|, |\hat{R}_{0,2} - R|, \dots, |\hat{R}_{0,25} - R|$$

19. Promedio del error absoluto de $\hat{R}_{0,k}$.

$$\overline{EA}(\hat{R}_{0,k}) = \frac{1}{25} \sum_{k=1}^{25} |\hat{R}_{0,k} - R|$$

20. Promedio de los 25 estimadores.

$$\overline{\hat{R}}_{0,k} = \frac{1}{25} \sum_{k=1}^{25} \hat{R}_{0,k}$$

21. Estimación del ECM de $\hat{R}_{0,k}$, en base a las simulaciones.

$$\widehat{ECM}(\hat{R}_{0,k}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{R}_{0,k} - R)^2$$

22. Raíz cuadrada de $\widehat{ECM}(\hat{R}_{0,k})$

23. Estimación de la varianza de $\hat{R}_{0,k}$, en base a las simulaciones.

$$\widehat{V}(\hat{R}_{0,k}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{R}_{0,k} - \overline{\hat{R}}_{0,k})^2$$

24. Estimación de la desviación estándar de $\hat{R}_{0,k}$, en base a las simulaciones.

$$\widehat{DE}(\hat{R}_{0,k}) = \sqrt{\widehat{V}(\hat{R}_{0,k})}$$

25. Estimación del sesgo de $\hat{R}_{0,k}$.

$$\widehat{\text{Sesgo}}(\hat{R}_{0,k}) = \sqrt{|\widehat{ECM}(\hat{R}_{0,k}) - \widehat{V}(\hat{R}_{0,k})|}$$

26. Estimación del sesgo de $\hat{R}_{0,k}$ relativo a $\bar{\hat{R}}_{0,k}$.

$$\widehat{\text{Sesgo rel.}}_1(\hat{R}_{0,k}) = \frac{\widehat{\text{Sesgo}}(\hat{R}_{0,k})}{\bar{\hat{R}}_{0,k}}$$

27. Estimación del sesgo de $\hat{R}_{0,k}$ relativo a la raíz cuadrada del $\widehat{ECM}(\hat{R}_{0,k})$

$$\widehat{\text{Sesgo rel.}}_2(\hat{R}_{0,k}) = \frac{\widehat{\text{Sesgo}}(\hat{R}_{0,k})}{\sqrt{\widehat{ECM}(\hat{R}_{0,k})}}$$

28. Estimación del sesgo de $\hat{R}_{0,k}$ relativo a la desviación estándar estimada de $\hat{R}_{0,k}$.

$$\widehat{\text{Sesgo rel.}}(\hat{R}_{0,k}) = \frac{\widehat{\text{Sesgo}}(\hat{R}_{0,k})}{\widehat{DE}(\hat{R}_{0,k})}$$

29. Estimación del coeficiente de variación de $\hat{R}_{0,k}$.

$$\hat{CV}(\hat{R}_{0,k}) = \frac{\sqrt{\widehat{ECM}(\hat{R}_{0,k})}}{\hat{R}_{0,k}}$$

30. Número de veces que el valor verdadero R cayó en el intervalo de confianza aproximado del 95%.

$$\left[\hat{R}_{0,k} - 1.96 \sqrt{\widehat{ECM}(\hat{R}_{0,k})}, \hat{R}_{0,k} + 1.96 \sqrt{\widehat{ECM}(\hat{R}_{0,k})} \right]$$

En el cuadro N° 21 aparecen algunos de los resultados obtenidos y en el apéndice N° 7 se presenta un ejemplo de listado de salida de la simulación del estimador de razón.

V.3 Estimadores de promedios en dominios de estudio.

La simulación del estimador de promedios en dominios de estudio, esto es, el estimador de medias para los establecimientos que poseen cierta característica α , se realizó para la variable P.O. La característica α que se usó fue el tener más de 20 personas ocupadas. Las estadísticas obtenidas se describen a continuación.

1. Total poblacional de la variable Y para los establecimientos que poseen la característica α .

$$Y^\alpha = Y_1^\alpha + Y_2^\alpha$$

donde

$Y_1^\alpha =$ Total de la variable Y para los establecimientos ma
yores que poseen la característica α

$$= \sum_{i=1}^{N_1} Y_{1i} S_i$$

$Y_2 =$ Total de la variable Y para los establecimientos me
dianos que poseen la característica α

$$= \sum_{i=1}^{N_2} Y_{2i} S_i$$

$$S_i = \begin{cases} 1 & \text{Si el establecimiento } i\text{-ésimo posee la caracterís-} \\ & \text{tica } \alpha \\ 0 & \text{en caso contrario.} \end{cases}$$

2. Número de establecimientos que poseen la característica

$$N^\alpha = N_1^\alpha + N_2^\alpha$$

$$= \sum_{i=1}^N S_i$$

donde

$N_1^\alpha =$ Número de establecimientos mayores con la caracterís
tica α

$N_2^{\times} =$ Número de establecimientos medianos con la característica \times

3. Coeficiente de correlación entre y_i , s_i y s_i

$$\rho = \frac{\text{Cov}(y_i s_i, s_i)}{\sqrt{V(y_i s_i)} \sqrt{V(s_i)}}$$

donde

$$\text{Cov}(y_i s_i, s_i) = \frac{1}{N} \sum_{i=1}^N (y_{0i} s_i - \overline{y_{0i} s_i}) (s_i - \bar{s}_i)$$

$$= \left[\frac{1}{N} - \frac{N^{\times}}{N^2} \right] \sum_{i=1}^N y_{0i} s_i$$

$$V(y_i s_i) = \frac{1}{N} \sum_{i=1}^N (y_{0i} s_i - \overline{y_{0i} s_i})^2$$

$$= \frac{1}{N} \sum_{i=1}^N y_{0i}^2 s_i - \left[\frac{1}{N} \sum_{i=1}^N y_{0i} s_i \right]^2$$

$$V(s_i) = \frac{1}{N} \sum_{i=1}^N (s_i - \bar{s}_i)^2$$

$$= \frac{N^{\times}}{N} \left[1 - \frac{N^{\times}}{N} \right]$$

4. Coeficiente de variación de $\hat{y}_{0,k}^{\times}$

$$CV(\hat{y}_{0,k}^{\times}) = \frac{\sqrt{V(\hat{y}_{0,k}^{\times})}}{y_0^{\times}}$$

donde

$$V(\hat{y}_{0,k}^{\alpha}) = V(\hat{y}_{2,k}^{\alpha})$$

$$V(\hat{y}_{2,k}^{\alpha}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left[1 - \frac{n_{2h}}{N_{2h}} \right] \frac{1}{N_{2h}^{-1}} \sum_{i=1}^{N_{2h}} (y_{2hi} s_i - \overline{y_{2hi} s_i})^2$$

5. Coeficiente de variación de $\hat{N}_{0,k}^{\alpha}$

$$CV(\hat{N}_{0,k}^{\alpha}) = \frac{\sqrt{V(\hat{N}_{0,k}^{\alpha})}}{N^{\alpha}}$$

donde

$$\hat{N}_{0,k}^{\alpha} = N_1^{\alpha} + \hat{N}_{2,k}^{\alpha}$$

$$V(\hat{N}_{0,k}^{\alpha}) = V(\hat{N}_{2,k}^{\alpha})$$

$$V(\hat{N}_{2,k}^{\alpha}) = \sum_{h=1}^L N_{2h}^2 \left[\frac{N_{2h}^{\alpha}}{N_{2h}} \right] \left[1 - \frac{N_{2h}^{\alpha}}{N_{2h}} \right] \left[\frac{N_{2h} - n_{2h}}{n_{2h} (N_{2h}^{-1})} \right]$$

N_{2h}^{α} = Número de establecimientos medianos en el estrato h
que poseen la característica α

L = Número de subestratos.

6. Media poblacional de la variable Y para los establecimientos que poseen la característica α

$$\bar{y}^{\alpha} = \frac{y^{\alpha}}{N^{\alpha}}$$

TESIS CON FALLA DE ORIGEN

7. Covarianza entre $\hat{Y}_{0,k}^{\sim}$ y $\hat{N}_{0,k}^{\sim}$.

$$\text{Cov}(\hat{Y}_{0,k}^{\sim}, \hat{N}_{0,k}^{\sim}) = \text{Cov}(\hat{Y}_{2,k}^{\sim}, \hat{N}_{2,k}^{\sim})$$

$$\text{Cov}(\hat{Y}_{2,k}^{\sim}, \hat{N}_{2,k}^{\sim}) = \sum_{h=1}^L \frac{(N_{2h} - n_{2h}) (N_{2h} - N_{2h}^{\sim})}{n_{2h} (N_{2h} - 1)} \sqrt{\frac{N_{2h}}{L-1}} y_{2hi} s_i$$

8. Error cuadrático medio aproximado de $\bar{Y}_{0,k}$.

$$\text{ECM}(\hat{Y}_{0,k}^{\sim}) = \frac{1}{N^{\sim 2}} [V(\hat{Y}_{0,k}^{\sim}) + (\bar{Y}^{\sim})^2 V(\hat{N}_{0,k}^{\sim}) - 2\bar{Y}^{\sim} \text{Cov}(\hat{Y}_{0,k}^{\sim}, \hat{N}_{0,k}^{\sim})]$$

9. Raíz cuadrada del $\text{ECM}(\hat{Y}_{0,k}^{\sim})$.

10. Sesgo de $\hat{Y}_{0,k}^{\sim}$.

$$\text{Sesgo}(\hat{Y}_{0,k}^{\sim}) = \bar{Y}^{\sim} [CV(\hat{N}_{0,k}^{\sim})]^2 - \frac{\text{Cov}(\hat{Y}_{0,k}^{\sim}, \hat{N}_{0,k}^{\sim})}{(\hat{N}_{0,k}^{\sim})^2}$$

11. Sesgo de $\hat{Y}_{0,k}^{\sim}$ relativo a la media $\bar{Y}_{0,k}^{\sim}$.

$$\text{Sesgo rel.}(\hat{Y}_{0,k}^{\sim}) = \frac{|\text{Sesgo}(\hat{Y}_{0,k}^{\sim})|}{\bar{Y}_{0,k}^{\sim}}$$

12. Sesgo de $\hat{Y}_{0,k}^{\sim}$ relativo a la raíz cuadrada del ECM aproximado de $Y_{0,k}$.

$$\text{Sesgo rel.}(\hat{Y}_{0,k}^{\sim}) = \frac{|\text{Sesgo}(\hat{Y}_{0,k}^{\sim})|}{\sqrt{\text{ECM}(\hat{Y}_{0,k}^{\sim})}}$$

13. Coeficiente de variación de $\hat{y}_{0,k}^\alpha$.

$$CV(\hat{y}_{0,k}^\alpha) = \frac{\sqrt{ECM(\hat{y}_{0,k}^\alpha)}}{\bar{y}_0^\alpha}$$

14. Máximo de los 25 estimadores.

$$\text{Max}(\hat{y}_{0,k}^\alpha) = \text{Max}\{\hat{y}_{0,1}^\alpha; \hat{y}_{0,2}^\alpha; \dots; \hat{y}_{0,25}^\alpha\}$$

donde

$$\hat{y}_{0,k}^\alpha = \frac{\hat{y}_{0,k}^\alpha}{\hat{N}_{0,k}^\alpha} = \frac{y_1^\alpha + \hat{y}_{2k}^\alpha}{\hat{N}_{0,k}^\alpha}$$

15. Mínimo de los 25 estimadores.

$$\text{Min}(\hat{y}_{0,k}^\alpha) = \text{Min}\{\hat{y}_{0,1}^\alpha; \hat{y}_{0,2}^\alpha; \dots; \hat{y}_{0,25}^\alpha\}$$

16. Rango de los 25 estimadores.

$$\text{Rango}(\hat{y}_{0,k}^\alpha) = \text{Max}(\hat{y}_{0,k}^\alpha) - \text{Min}(\hat{y}_{0,k}^\alpha)$$

17. Error absoluto máximo de $\hat{y}_{0,k}^\alpha$.

$$EAMax(\hat{y}_{0,k}^\alpha) = \text{Max}\left\{\left|\hat{y}_{0,1}^\alpha - \bar{y}_0^\alpha\right|, \left|\hat{y}_{0,2}^\alpha - \bar{y}_0^\alpha\right|, \dots, \left|\hat{y}_{0,25}^\alpha - \bar{y}_0^\alpha\right|\right\}$$

18. Error absoluto mínimo de $\hat{y}_{0,k}^{\infty}$.

$$EAMin(\hat{y}_{0,k}^{\infty}) = \text{Min}\{|\hat{y}_{0,1}^{\infty} - \bar{y}_0^{\infty}|, |\hat{y}_{0,2}^{\infty} - \bar{y}_0^{\infty}|, \dots, |\hat{y}_{0,25}^{\infty} - \bar{y}_0^{\infty}|\}$$

19. Promedio del error absoluto de $\hat{y}_{0,k}^{\infty}$.

$$\bar{EA}(\hat{y}_{0,k}^{\infty}) = \frac{1}{25} \sum_{k=1}^{25} |\hat{y}_{0,k}^{\infty} - \bar{y}_0^{\infty}|$$

20. Promedio de los 25 estimadores.

$$\bar{\hat{y}}_{0,k}^{\infty} = \frac{1}{25} \sum_{k=1}^{25} \hat{y}_{0,k}^{\infty}$$

21. Estimación del ECM de $\hat{y}_{0,k}^{\infty}$, en base a las simulaciones.

$$\widehat{ECM}(\hat{y}_{0,k}^{\infty}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{y}_{0,k}^{\infty} - \bar{\hat{y}}_{0,k}^{\infty})^2$$

22. Raíz cuadrada del $\widehat{ECM}(\hat{y}_{0,k}^{\infty})$

23. Estimación de $V(\hat{y}_{0,k}^{\infty})$ en base a las simulaciones.

$$\widehat{V}(\hat{y}_{0,k}^{\infty}) = \frac{1}{24} \sum_{k=1}^{25} (\hat{y}_{0,k}^{\infty} - \bar{\hat{y}}_{0,k}^{\infty})^2$$

24. Raíz cuadrada de $\widehat{V}(\hat{y}_{0,k}^{\infty})$

25. Estimación del sesgo de $\hat{y}_{0,k}^{\infty}$, en base a las simulaciones.

$$\widehat{\text{Sesgo}}(\hat{y}_{0,k}^{\infty}) = \sqrt{|\widehat{ECM}(\hat{y}_{0,k}^{\infty}) - \widehat{V}(\hat{y}_{0,k}^{\infty})|}$$

26. Estimación del sesgo de $\hat{y}_{0,k}^{\sim}$ relativo a $\bar{y}_{0,k}^{\sim}$.

$$\widehat{\text{Sesgo rel.}}_1(\hat{y}_{0,k}^{\sim}) = \frac{\widehat{\text{Sesgo}}(\hat{y}_{0,k}^{\sim})}{\bar{y}_{0,k}^{\sim}}$$

27. Estimación del sesgo de $\hat{y}_{0,k}^{\sim}$ relativo a la raíz cuadrada del $\widehat{\text{ECM}}(\hat{y}_{0,k}^{\sim})$

$$\widehat{\text{Sesgo rel.}}_2(\hat{y}_{0,k}^{\sim}) = \frac{\widehat{\text{Sesgo}}(\hat{y}_{0,k}^{\sim})}{\sqrt{\widehat{\text{ECM}}(\hat{y}_{0,k}^{\sim})}}$$

28. Estimación del sesgo de $\hat{y}_{0,k}^{\sim}$ relativo a la raíz cuadrada de $\widehat{V}(\hat{y}_{0,k}^{\sim})$.

$$\widehat{\text{Sesgo rel.}}(\hat{y}_{0,k}^{\sim}) = \frac{\widehat{\text{Sesgo}}(\hat{y}_{0,k}^{\sim})}{\sqrt{\widehat{V}(\hat{y}_{0,k}^{\sim})}}$$

29. Estimación del coeficiente de variación de $\hat{y}_{0,k}^{\sim}$.

$$\widehat{\text{CV}}(\hat{y}_{0,k}^{\sim}) = \frac{\sqrt{\widehat{\text{ECM}}(\hat{y}_{0,k}^{\sim})}}{\bar{y}_{0,k}^{\sim}}$$

30. Número de veces que el valor real $\bar{y}_{0,k}^{\sim}$ cayó en el intervalo de confianza aproximado del 95%.

$$\left[\bar{y}_{0,k}^{\sim} - 1.96 \sqrt{\widehat{\text{ECM}}(\hat{y}_{0,k}^{\sim})}, \bar{y}_{0,k}^{\sim} + 1.96 \sqrt{\widehat{\text{ECM}}(\hat{y}_{0,k}^{\sim})} \right]$$

En el cuadro N° 22 aparecen algunos de los resultados obtenidos y en el apéndice N° 8 se presenta un ejemplo de listado de salida de la simulación del estimador de razón.

CONCLUSIONES.

En este estudio se han aplicado varias técnicas de muestreo para producir un esquema adecuado para la realización de una Encuesta Industrial Mensual. Se entiende por adecuado que sea viable económicamente y factible su trabajo de campo (estos aspectos no se discuten en este trabajo por concentrarse en el diseño estadístico, sin embargo la prueba piloto demostró esta viabilidad); además que produzca estimadores con poco error.

Los resultados de la simulación señalan que el procedimiento es adecuado, ya que los coeficientes de variación para varios estimadores estuvieron siempre por debajo del 3.7%, lo cual es muy aceptable. Por otro lado los intervalos de confianza funcionan efectivamente, puesto que alrededor de un 95% de las veces cubren los parámetros. Además, los porcentajes de error de estimación absoluto máximo son menores del 5% para totales y razones. Para promedios en dominios como es lógico este error aumenta, pero únicamente en 12% de las clases se tuvieron errores de estimación absolutos máximos entre 5 y 10%.

Se considera que se ha efectuado un desarrollo de tecnología en estadística aplicada que es de utilidad dentro de las encuestas de establecimientos industriales y quizá aplicable a otras poblaciones muy asimétricas, tales como otros establecimientos de tipo comercial o agrícola. Además, este desarrollo indudablemente que tiene utilidad desde el punto de vista didáctico, para la maestría y la especialización en estadística de la UNAM.

BIBLIOGRAFIA CONSULTADA

- Anton, H. (1973), "Elementary Linear Algebra", John Wiley and Sons, Inc., New York.
- Cochran, W. G. (1961), "Comparison of Methods for Determining Stratum Boundaries", *Bull. Int. Stat. Inst.*, 38.
- Cochran, W. G. (1977), "Sampling Techniques", John Wiley and Sons, 3rd. Ed., New York.
- Dalenius, T. and Hodges, J. L. (1959), "Minimum Variance Stratification". *Journal of the American Statistical Association.* 54, 88-101.
- Fretes, G.R. (1971), "La Planificación de la Encuesta Industrial", ILPES, Santiago de Chile.
- Glasser, G. J. (1962), "On the Complete Coverage of Large Units in a Statistical Study", *Review of the International Statistical Institute*, 30, 28-32.
- Hotelling, H. (1933), "Analysis of a Complex of Statistical Variables into Principal Components", *J. Educ. Psychol.*, 24, 417-441, 498-520.
- Johnston, J. (1979), "Métodos de Econometría", Vicens-Vives, S. A., 3a. Ed., España.
- Kish, L. (1965), "Survey Sampling", John Wiley and Sons, Inc., New York.
- Kish, L. and Anderson, D. W. (1978), "Multivariate and Multipurpose Stratification", *J.A.S.A.*, 73, 367, 24-34.

- Méndez, R. I. (1976), "Conceptos muy Elementales del Muestreo con Enfoque en la Determinación Práctica del Tamaño de Muestra", UNAM, IIMAS, Comunicaciones Técnicas, Serie Azul, Vol. 3, N° 25.
- Morrison, F. D. (1967), "Multivariate Statistical Methods", McGraw-Hill, New York.
- Naylor, T. H., et. al. (1966), "Computer Simulation Techniques", John Wiley and Sons, Inc., New York.
- Nie, N. H., et. al. (1975), "Statistical Package for the Social Sciences", McGraw-Hill, 2nd. Ed., New York.
- Pearson, K. (1901). "On Lines and Planes of Closest Fit to a System of Points in Space". *Phil. Mag.*, 2, 6th Series, 557-572.
- Raj, D. (1968), "Sampling Theory", McGraw-Hill, New York.
- Secretaría de Programación y Presupuesto, Dirección General de Estadística (1979), "X Censo Industrial 1976, Datos de 1975, Resumen General, Tomo 1", México.
- UNIVAC (1972), "Large Scale Systems, Math Pack, Programmers Reference", UP-7542, Rev. 1-B, U.S.A.

CUADROS

CUADRO N.º 1

VARIANZA EXPLICADA POR LAS COMPONENTES PRINCIPALES (POBLACION: TODOS LOS ESTABLECIMIENTOS)

Clase de Actividad	Varianza Explicada por la Componente Principal			
	Primera	Segunda	Tercera	Cuarta
2051	88.3	6.8	3.5	1.4
2318	97.6	1.2	1.0	0.2
2391	91.0	6.0	2.6	0.4
2392	96.4	2.4	1.2	0.0
2412	80.5	11.3	7.4	0.8
2413	97.9	1.4	0.6	0.1
2414	86.8	8.1	4.6	0.5
2415	87.6	7.8	3.4	1.2
2519	99.5	0.3	0.1	0.1
2711	87.3	8.6	2.9	1.2
2821	89.9	6.0	2.7	1.4
2921	92.6	4.0	2.6	0.8
3022	84.6	10.2	4.3	0.9
3031	89.3	6.1	3.6	1.0
3072	90.4	7.1	1.6	0.9
3091	79.1	14.6	5.4	0.9
3219	93.8	3.8	1.4	1.0
3222	90.9	5.6	2.9	0.6
3223	92.5	4.9	1.7	0.9
3229	85.2	6.5	6.4	1.9
3354	86.8	7.9	4.5	0.8
3429	89.5	8.0	2.1	0.4
3531	91.0	4.4	3.9	0.7
3593	97.0	2.1	0.6	0.3
3596	98.5	0.7	0.5	0.3
3632	81.5	11.6	5.9	1.0
3639	92.7	5.4	1.5	0.4
3692	96.4	2.7	0.7	0.2
3694	93.8	4.8	1.0	0.4
3695	93.6	4.0	2.0	0.4
3697	87.6	10.0	1.4	1.0
3729	83.2	11.9	4.1	0.8
3813	94.4	3.9	1.0	0.7
3815	95.3	3.9	0.7	0.1
3817	96.0	2.7	1.1	0.2
3891	98.9	1.0	0.1	0.0
3992	93.4	5.1	1.4	.1

82

CUADRO N° 2

VARIANZA EXPLICADA POR LAS COMPONENTES PRINCIPALES (POBLACION:
ESTABLECIMIENTOS NO MENORES)

Clase de Actividad.	Varianza Explicada por la Componente Principal			
	Primera	Segunda	Tercera	Cuarta
2051	85.3	8.6	4.4	1.7
2318	97.3	1.4	1.0	0.3
2391	88.8	7.4	3.3	0.5
2392	94.9	3.3	1.7	0.1
2412	76.2	14.5	8.3	1.0
2413	98.1	1.5	0.4	0.0
2414	83.1	10.6	5.7	0.6
2415	86.3	8.5	3.7	1.5
2519	99.7	0.1	0.1	0.1
2711	83.6	10.6	4.2	1.6
2821	84.6	9.2	4.1	2.1
2921	90.9	4.7	3.4	1.0
3022	82.5	11.3	5.2	1.0
3031	87.3	7.3	4.3	1.1
3072	86.0	10.5	2.3	1.2
3091	77.7	14.9	6.4	1.0
3219	91.6	5.2	1.9	1.3
3222	86.7	7.3	4.4	1.6
3223	91.3	5.8	1.8	1.1
3229	84.5	7.3	6.2	2.0
3354	83.7	9.7	5.4	1.2
3429	84.8	12.1	2.6	0.5
3531	88.9	5.6	4.7	0.8
3593	96.6	2.8	0.5	0.1
3596	98.5	0.7	0.5	0.3
3632	80.4	12.0	6.4	1.2
3639	90.9	6.8	1.9	0.4
3692	95.6	3.4	0.9	0.1
3694	94.2	4.1	1.2	0.5
3695	92.3	4.9	2.3	0.5
3697	85.4	11.6	1.7	1.3
3729	79.5	15.1	4.0	1.4
3813	94.4	3.9	1.0	0.7
3815	95.3	3.9	0.7	0.1
3817	96.0	2.7	1.1	0.2
3891	98.9	1.0	0.1	0.0
3992	93.4	5.1	1.4	0.1

83

CUADRO N.º 3

CORRELACION DEL INDICE I_{e_7} CON LAS CUATRO VARIABLES

Clase de Actividad	Correlación de I_{e_7} con			
	P.O.	P.	A.F.	V.A.
2051	.955	.919	.975	.909
2318	.985	.993	.982	.991
2391	.918	.981	.962	.953
2392	.969	.987	.974	.996
2412	.877	.838	.880	.987
2413	.982	.994	.985	.998
2414	.923	.962	.886	.953
2415	.892	.966	.907	.976
2519	.997	.999	.996	.998
2711	.954	.943	.860	.976
2821	.958	.954	.903	.977
2921	.948	.980	.944	.976
3022	.892	.939	.905	.941
3031	.901	.957	.942	.977
3072	.970	.982	.944	.905
3091	.911	.953	.716	.954
3219	.959	.978	.951	.985
3222	.947	.989	.918	.959
3223	.976	.982	.923	.966
3229	.887	.964	.912	.914
3354	.896	.935	.925	.971
3429	.909	.943	.944	.986
3531	.931	.973	.937	.974
3593	.988	.985	.984	.983
3596	.992	.993	.990	.995
3632	.853	.966	.822	.961
3639	.975	.979	.913	.983
3692	.990	.988	.959	.990
3694	.985	.978	.930	.980
3695	.966	.988	.937	.979
3697	.979	.975	.826	.956
3729	.870	.941	.849	.982
3813	.950	.984	.963	.988
3815	.975	.995	.942	.992
3817	.984	.991	.966	.978
3891	.991	.999	.988	.999
3992	.924	.987	.960	.992

CUADRO N° 4

CORRELACION DEL INDICE I_{e_2} CON LAS CUATRO VARIABLES

Clase de Actividad	Correlación de I_{e_2} con			
	P.O	P.	A.F.	V.A.
2051	.942	.891	.971	.887
2318	.982	.991	.984	.989
2391	.895	.976	.953	.944
2392	.956	.980	.965	.995
2412	.840	.789	.867	.983
2413	.983	.995	.985	.999
2414	.897	.951	.857	.938
2415	.868	.960	.912	.971
2519	.998	.999	.998	.999
2711	.937	.921	.823	.969
2821	.934	.931	.845	.965
2921	.937	.974	.933	.969
3022	.875	.929	.890	.939
3031	.878	.955	.930	.971
3072	.955	.973	.931	.844
3091	.897	.946	.711	.949
3219	.945	.971	.933	.979
3222	.897	.970	.906	.949
3223	.974	.978	.907	.962
3229	.887	.964	.912	.914
3354	.880	.913	.901	.964
3429	.849	.908	.936	.985
3531	.910	.966	.926	.967
3593	.988	.981	.982	.981
3596	.992	.993	.991	.995
3632	.873	.965	.790	.947
3639	.968	.975	.888	.979
3692	.988	.985	.947	.990
3694	.981	.976	.944	.981
3695	.960	.985	.921	.974
3697	.975	.970	.793	.946
3729	.774	.956	.849	.973
3813	.950	.984	.963	.988
3815	.975	.995	.942	.992
3817	.984	.991	.966	.978
3891	.991	.999	.988	.999
3992	.924	.987	.960	.992

CUADRO N° 5

MATRIZ "FACTOR SCORE COEFFICIENTS" PARA CALCULAR I_e

Clase de Actividad	P.O.	P.	A.F.	V.A.
2051	.270	.260	.276	.257
2318	.252	.254	.252	.254
2391	.252	.270	.264	.262
2392	.251	.256	.253	.259
2412	.273	.260	.273	.306
2413	.251	.254	.251	.255
2414	.265	.277	.255	.275
2415	.254	.276	.259	.279
2519	.250	.251	.250	.251
2711	.273	.270	.246	.279
2821	.266	.265	.251	.272
2921	.256	.265	.255	.264
3022	.264	.277	.268	.278
3031	.252	.268	.264	.274
3072	.268	.272	.261	.250
3091	.288	.301	.226	.302
3219	.256	.261	.254	.263
3222	.260	.272	.252	.264
3223	.264	.265	.249	.261
3229	.262	.285	.270	.270
3354	.258	.269	.266	.279
3429	.254	.263	.264	.276
3531	.256	.267	.257	.268
3593	.255	.254	.254	.253
3596	.252	.252	.251	.252
3632	.262	.296	.252	.295
3639	.263	.264	.246	.265
3692	.257	.256	.249	.257
3694	.262	.261	.248	.261
3695	.258	.264	.250	.261
3697	.279	.278	.236	.273
3729	.261	.283	.255	.295
3813	.252	.261	.255	.262
3815	.256	.261	.247	.260
3817	.256	.258	.252	.255
3891	.251	.253	.250	.253
3992	.247	.264	.257	.266

CUADRO N° 6

MATRIZ "FACTOR SCORE COEFICIENTS" PARA CALCULAR I_{e_2}

Clase de Actividad	P.O.	P.	A.F.	V.A.
2051	.276	.261	.285	.260
2318	.252	.255	.253	.254
2391	.252	.275	.268	.266
2392	.252	.258	.254	.262
2412	.276	.259	.285	.323
2413	.250	.254	.251	.254
2414	.270	.286	.258	.282
2415	.252	.278	.264	.281
2519	.250	.251	.250	.250
2711	.280	.275	.246	.290
2821	.276	.275	.250	.285
2921	.258	.268	.257	.266
3022	.265	.282	.270	.284
3031	.252	.274	.266	.278
3072	.278	.283	.271	.246
3091	.289	.304	.229	.306
3219	.258	.265	.255	.267
3222	.259	.280	.261	.274
3223	.267	.268	.248	.263
3229	.262	.285	.270	.270
3354	.263	.273	.269	.288
3429	.250	.268	.276	.290
3531	.256	.272	.261	.272
3593	.256	.254	.254	.254
3596	.252	.252	.251	.252
3632	.272	.300	.246	.295
3639	.266	.268	.244	.269
3692	.258	.258	.248	.259
3694	.260	.259	.251	.260
3695	.260	.267	.250	.264
3697	.286	.284	.232	.277
3729	.243	.301	.267	.306
3813	.252	.261	.255	.262
3815	.256	.261	.247	.260
3817	.256	.258	.258	.255
3891	.251	.253	.250	.253
3992	.248	.264	.257	.266

CUADRO N° 7
 PORCENTAJES ACUMULADOS DE LAS CUATRO VARIABLES PARA DISTINTOS NIVELES DE PERSONAL
 OCUPADO

Nivel de P.O.	Porcentaje Acumulado de					
	Número de Establecimientos	P.O.	P.	A.F.	V.A.	
Más de 750	.22	22.65	31.72	35.67	31.76	
750 - 501	.41	30.66	39.88	46.54	41.37	
500 - 351	.69	38.63	49.86	58.25	51.63	
350 - 251	1.02	45.46	57.37	65.95	59.46	
250 - 176	1.51	52.57	65.92	73.32	67.54	
175 - 101	2.59	62.40	76.44	82.42	77.43	
100 - 76	3.33	66.89	80.70	85.64	81.27	
75 - 51	4.66	72.56	85.97	89.56	86.27	
50 - 26	7.48	79.50	91.47	93.82	91.57	
25 - 16	10.35	83.35	94.16	95.85	94.13	
15 - 6	19.36	88.79	97.14	98.05	97.09	
Menos de 6	100.00	100.00	100.00	100.00	100.00	

88

CUADRO N° 8

NUMERO DE ESTABLECIMIENTOS MENORES, NO MENORES Y TOTAL POR CLASE DE ACTIVIDAD

Clase de Actividad	Número de Establecimientos		
	Menores	No Menores	Total
2051	426	28	454
2318	160	35	195
2391	14	16	30
2392	39	11	50
2412	5282	105	5387
2413	182	48	230
2414	282	51	333
2415	225	98	323
2519	1516	28	1544
2711	2882	291	3173
2821	29	32	61
2921	4149	253	4402
3022	54	40	94
3031	36	33	69
3072	6	8	14
3091	75	78	153
3219	159	56	215
3222	170	102	272
3223	137	50	187
3229	0	670	670
3354	1811	233	2044
3429	39	7	46
3531	348	134	482
3593	238	21	259
3596	495	63	558
3632	74	44	118
3639	337	83	420
3692	83	28	111
3694	204	51	255
3695	43	55	98
3697	119	69	188
3729	180	73	253
3813	0	44	44
3815	0	56	56
3817	0	36	36
3891	0	18	18
3992	0	83	83
TOTAL	19,794	3,131	22,925

CUADRO N° 9

PORCENTAJES DE LAS CUATRO VARIABLES QUE CUBREN LOS ESTABLECIMIENTOS
NO MENORES POR CLASE DE ACTIVIDAD

Clase de Actividad	Porcentaje cubierto de la variable			
	P.O	P.	A.F.	V.A.
2051	76.68	81.61	76.05	82.09
2318	78.52	85.02	85.84	83.69
2391	95.58	97.75	98.87	97.52
2392	90.21	97.08	98.27	96.61
2412	57.92	81.48	80.95	75.00
2413	68.45	72.78	74.69	76.21
2414	68.63	82.42	82.67	80.46
2415	84.22	90.93	88.92	90.56
2519	38.00	51.98	57.52	48.45
2711	63.11	81.44	75.96	78.32
2821	90.03	94.07	94.88	93.62
2921	53.46	74.32	72.06	71.81
3022	95.69	97.88	96.78	97.92
3031	96.24	97.69	98.10	97.87
3072	88.74	98.65	97.80	98.10
3091	92.09	94.27	96.39	95.09
3219	80.40	87.36	85.49	87.01
3222	84.35	90.21	87.23	89.87
3223	92.55	96.90	96.65	96.60
3229	100.00	100.00	100.00	100.00
3354	59.69	83.18	83.41	78.52
3429	90.17	97.18	96.74	94.37
3531	81.87	88.49	88.63	88.88
3593	69.83	83.63	85.81	86.37
3596	74.11	83.31	85.77	85.02
3632	92.86	95.94	94.82	95.74
3639	84.37	90.17	87.93	91.24
3692	90.19	93.38	92.10	94.13
3694	84.88	93.03	90.03	92.46
3695	95.65	97.71	96.47	97.91
3697	90.96	94.48	93.86	94.09
3729	95.35	94.97	95.13	95.60
3813	100.00	100.00	100.00	100.00
3815	100.00	100.00	100.00	100.00
3817	100.00	100.00	100.00	100.00
3891	100.00	100.00	100.00	100.00
3992	100.00	100.00	100.00	100.00

CUADRO N° 10

TAMAÑO DE MUESTRA INICIAL Y PORCENTAJE QUE ESTE REPRESENTA CON RESPECTO AL NUMERO DE ESTABLECIMIENTOS NO MENORES POR CLASE DE ACTIVIDAD.

Clase de Actividad	Tamaño de Muestra Inicial	Porcentaje
2051	25	89
2318	27	77
2391	15	94
2392	10	91
2412	30	76
2413	23	48
2414	38	75
2415	73	74
2519	11	39
2711	183	63
2821	30	94
2921	148	58
3022	34	85
3031	28	85
3072	7	88
3091	62	79
3219	47	84
3222	82	80
3223	41	82
3229	329	49
3354	142	61
3429	6	86
3531	102	76
3593	20	95
3596	42	67
3632	39	87
3639	64	77
3692	26	93
3694	42	82
3695	47	85
3697	58	84
3729	60	82
3813	38	86
3815	42	75
3817	33	92
3891	17	94
3992	72	87

91

CUADRO N° 11

NUMERO DE ESTABLECIMIENTOS MAYORES Y PORCENTAJES QUE CUBREN LAS CUATRO VARIABLES CON RESPECTO A LOS NO MENORES, POR CLASE DE ACTIVIDAD.

Clase de Actividad	N° Estab. Mayores	Porcentaje de la Variable			
		P.O.	P.	A.F.	V.A.
2051	21	92	95	91	92
2318	19	80	81	88	79
2391	10	95	94	96	95
2392	9	97	98	99	98
2412	67	91	89	91	89
2413	18	71	72	71	77
2414	33	84	91	93	91
2415	65	89	91	95	93
2519	9	73	75	81	74
2711	150	80	85	83	84
2821	27	95	96	95	97
2921	117	76	81	80	81
3022	33	96	98	98	97
3031	26	95	97	97	97
3072	6	87	98	99	96
3091	52	91	92	95	93
3219	41	91	92	94	93
3222	70	88	90	91	90
3223	35	94	96	92	96
3229	267	87	93	89	92
3354	117	79	83	84	82
3429	4	97	94	97	86
3531	88	89	92	93	91
3593	18	96	97	99	99
3596	33	84	87	92	89
3632	37	97	99	98	98
3639	51	91	92	86	91
3692	24	97	97	97	98
3694	36	91	94	93	95
3695	44	96	98	95	97
3697	48	91	93	94	93
3729	48	91	88	85	90
3813	36	99	99	99	99
3815	37	98	99	99	99
3817	32	99	99	99	99
3891	15	99	99	99	99
3992	68	99	99	99	99

CUADRO N° 12

NUMERO DE ESTABLECIMIENTOS MEDIANOS Y PORCENTAJES QUE CUBREN LAS CUATRO VARIABLES CON RESPECTO A LOS NO MENORES POR CLASE DE ACTIVIDAD.

Clase de Actividad	N° Estab. Medianos	Porcentaje de la Variable			
		P.O.	P.	A.F.	V.A.
2051	7	8	5	9	8
2318	16	20	19	12	21
2391	6	5	6	4	5
2392	2	3	2	1	2
2412	38	9	11	9	11
2413	30	29	28	29	23
2414	18	16	9	7	9
2415	33	11	9	5	7
2519	19	27	25	19	26
2711	141	20	15	17	16
2821	5	5	4	5	3
2921	136	24	19	20	19
3022	7	4	2	2	3
3031	7	5	3	3	3
3072	2	13	2	1	4
3091	26	9	8	5	7
3219	15	9	8	6	7
3222	32	12	10	9	10
3223	15	6	4	8	4
3229	403	13	7	11	8
3354	116	21	17	16	18
3429	3	3	6	3	14
3531	46	11	8	7	9
3593	3	4	3	1	1
3596	30	16	13	8	11
3632	7	3	1	2	2
3639	32	9	8	14	9
3692	4	3	3	3	2
3694	15	9	6	7	5
3695	11	4	2	5	3
3697	21	9	7	6	7
3729	25	9	12	15	10
3813	8	1	1	1	1
3815	19	2	1	1	1
3817	4	1	1	1	1
3891	3	1	1	1	1
3992	15	1	1	1	1

CUADRO N° 13

NUMERO DE ESTABLECIMIENTOS EN CADA SUBSTRATO POR CLASE DE ACTIVIDAD

Clase de Actividad	Substrato						Total
	1	2	3	4	5	6	
2051	7						7
2318	16						16
2391	6						6
2392	2						2
2412	24	14					38
2413	13	17					30
2414	18						18
2415	18	15					33
2519	19						19
2711	19	17	17	18	27	43	141
2821	5						5
2921	17	21	22	21	30	25	136
3022	7						7
3031	7						7
3072	2						2
3091	26						26
3219	15						15
3222	15	17					32
3223	15						15
3229	43	50	52	89	69	100	403
3354	16	14	10	38	13	25	116
3429	3						3
3531	14	10	22				46
3593	3						3
3596	11	19					30
3632	7						7
3639	21	11					32
3692	4						4
3694	15						15
3695	11						11
3697	21						21
3729	25						25
3813	8						8
3815	19						19
3817	4						4
3891	3						3
3992	15						15
TOTAL							1320

CUADRO N.º 14

TAMAÑO DE MUESTRA POR CLASE DE ACTIVIDAD

Clase de Actividad	Tamaño de Muestra		
	Mayores	Medianos	No Menores
2051	21	3	24
2318	19	6	25
2391	10	2	12
2392	9	2	11
2412	67	6	73
2413	18	20	38
2414	33	4	33
2415	65	4	69
2519	9	10	19
2711	150	19	169
2821	27	2	29
2921	117	20	137
3022	33	2	35
3051	26	2	28
3072	6	2	8
3091	52	2	54
3219	41	2	43
3222	70	4	74
3223	35	2	37
3229	267	12	279
3354	117	21	138
3429	4	3	7
3531	88	6	94
3593	18	2	20
3596	33	4	37
3632	37	2	39
3639	51	8	59
3692	24	2	26
3694	36	3	39
3695	44	2	46
3697	48	3	51
3729	48	10	58
3813	36	2	38
3815	37	2	39
3817	32	2	34
3891	15	2	17
3992	68	2	70
TOTAL	1,811	202	2,013

95

ESTIMADORES DE TOTALES, VARIANZAS E INTERVALOS DE CONFIANZA

	MAYORES	MEDIANOS	NO MENORES
TOTAL	$N_1 \sum_{k=1}^L y_{1k}$	$\hat{y}_2 = \sum_{k=1}^L \hat{y}_{2k}$	$\hat{y} = y_1 + \hat{y}_2$
VARIANZA	0	$v(\hat{y}_2) = \sum_{k=1}^L v(\hat{y}_{2k})$	$v(\hat{y}) = v(\hat{y}_2)$
VARIANZA ESTIMADA	0	$\hat{v}(\hat{y}_2) = \sum_{k=1}^L \hat{v}(\hat{y}_{2k})$	$\hat{v}(\hat{y}) = \hat{v}(\hat{y}_2)$
INTERVALO DE CONFIANZA	—	$ y_2 - \hat{y}_2 \leq 1.96 \sqrt{\hat{v}(\hat{y}_2)}$	$ y - \hat{y} \leq 1.96 \sqrt{\hat{v}(\hat{y})}$

ESTIMADORES DE PROMEDIOS, VARIANZAS E INTERVALOS DE CONFIANZA

	MAYORES	MEDIANOS	NO MENORES
MEDIA	$\bar{Y}_1 = \frac{Y_1}{N_1}$	$\hat{Y}_2 = \frac{\hat{Y}_2}{N_2}$	$\hat{Y} = \frac{Y_1 + \hat{Y}_2}{N_1 + N_2}$
VARIANZA	0	$V(\hat{Y}_2) = \frac{V(\hat{Y}_2)}{N_2^2}$	$V(\hat{Y}) = \frac{V(\hat{Y}_2)}{(N_1 + N_2)^2}$
VARIANZA ESTIMADA	0	$\hat{V}(\hat{Y}_2) = \frac{\hat{V}(\hat{Y}_2)}{N_2^2}$	$\hat{V}(\hat{Y}) = \frac{\hat{V}(\hat{Y}_2)}{(N_1 + N_2)^2}$
INTERVALO DE CONFIANZA	—	$ \bar{Y}_2 - \hat{Y}_2 \leq 1.96 \sqrt{\hat{V}(\hat{Y}_2)}$	$ \bar{Y} - \hat{Y} \leq 1.96 \sqrt{\hat{V}(\hat{Y})}$

CUADRO N° 17

ESTIMADORES DE RAZON DE DOS CARACTERISTICAS, Y y X, ERRORES CUADRATICO MEDIOS (E.C.M.) E INTERVALOS DE CONFIANZA

	MAYORES	MEDIANOS	NO MENORES
RAZON	$R_1 = \frac{Y_1}{X_1}$	$\hat{R}_2 = \frac{\hat{Y}_2}{\hat{X}_2}$	$\hat{R} = \frac{\hat{Y}_1 + \hat{Y}_2}{X_1 + \hat{X}_2}$
ECH	0	$ECH(\hat{R}_2) = \frac{1}{X_2^2} [V(\hat{Y}_2) + R_2^2 V(\hat{X}_2) - 2R_2 Cov(\hat{Y}_2, \hat{X}_2)]$	$ECH(\hat{R}) = \frac{1}{(X_1 + X_2)^2} [V(\hat{Y}_2) + R^2 V(\hat{X}_2) - 2RCov(\hat{Y}_2, \hat{X}_2)]$
$\hat{E}CH$	0	$\hat{E}CH(\hat{R}_2) = \frac{1}{\hat{X}_2^2} [V(\hat{Y}_2) + R_2^2 V(\hat{X}_2) - 2R_2 Cov(\hat{Y}_2, \hat{X}_2)]$	$\hat{E}CH(\hat{R}) = \frac{1}{(\hat{X}_1 + \hat{X}_2)^2} [V(\hat{Y}_2) + R^2 V(\hat{X}_2) - 2RCov(\hat{Y}_2, \hat{X}_2)]$
INTERVALO DE CONFIANZA	—	$ R_2 - \hat{R}_2 \leq 1.96 \sqrt{ECH(\hat{R}_2)}$	$ R - \hat{R} \leq 1.96 \sqrt{ECH(\hat{R})}$

CUADRO N° 18

ESTIMADORES DE TOTALES EN DOMINIOS DE ESTUDIO, VARIANZAS E INTERVALOS DE CONFIANZA

	MAYORES	MEDIANOS	NO MENORES
TOTAL	y_1^{∞}	$\hat{y}_2^{\infty} = \sum_{h=1}^L \frac{N_{2h}}{n_{2h}} \sum_{\lambda=1}^{n_{2h}} y_{2hc\lambda}$	$\hat{y}^{\infty} = \hat{y}_1^{\infty} + \hat{y}_2^{\infty}$
VARIANZA	0	$V(\hat{y}_2^{\infty}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) \frac{1}{N_{2h}^{-1}} \sum_{\lambda=1}^{n_{2h}} (y_{2hc\lambda} - \bar{y}_{2hc})^2$	$V(\hat{y}^{\infty}) = V(\hat{y}_2^{\infty})$
VARIANZA ESTIMADA	0	$\hat{V}(\hat{y}_2^{\infty}) = \sum_{h=1}^L \frac{N_{2h}^2}{n_{2h}} \left(1 - \frac{n_{2h}}{N_{2h}}\right) \frac{1}{n_{2h}^{-1}} \sum_{\lambda=1}^{n_{2h}} (y_{2hc\lambda} - \bar{y}_{2hc})^2$	$\hat{V}(\hat{y}^{\infty}) = \hat{V}(\hat{y}_2^{\infty})$
INTERVALO DE CONFIANZA	—	$ y_2^{\infty} - \hat{y}_2^{\infty} \leq 1.96 \sqrt{\hat{V}(\hat{y}_2^{\infty})}$	$ y^{\infty} - \hat{y}^{\infty} \leq 1.96 \sqrt{\hat{V}(\hat{y}^{\infty})}$

99

CUADRO N.º 19

ESTIMADORES DE PROMEDIOS EN DOMINIOS DE ESTUDIO, ERRORES CUADRÁTICO MEDIOS (ECM) E INTERVALOS DE CONFIANZA

	MAYORES	MEDIANOS	NO MENORES
PROMEDIO	$\bar{y}_1 = \frac{y_1}{N_1}$	$\hat{\bar{y}}_2 = \frac{\hat{y}_2}{N_2}$	$\hat{\bar{y}}_2 = \frac{y_1 + \hat{y}_2}{N_1 + \hat{N}_2}$
ECM	0	$ECM(\hat{\bar{y}}_2) = \frac{1}{(N_2)^2} \left[V(\hat{y}_2) + (V_2)^2 V(\hat{N}_2) - 2V_2 Cov(\hat{y}_2, \hat{N}_2) \right]$	$ECM(\hat{\bar{y}}_2) = \frac{1}{(N_1 + \hat{N}_2)^2} \left[V(\hat{y}_2) + (V_2)^2 V(\hat{N}_2) - 2V_2 Cov(\hat{y}_2, \hat{N}_2) \right]$
\hat{ECM}	0	$\hat{ECM}(\hat{\bar{y}}_2) = \frac{1}{(\hat{N}_2)^2} \left[\hat{V}(\hat{y}_2) + (\hat{V}_2)^2 \hat{V}(\hat{N}_2) - 2\hat{V}_2 \hat{Cov}(\hat{y}_2, \hat{N}_2) \right]$	$\hat{ECM}(\hat{\bar{y}}_2) = \frac{1}{(N_1 + \hat{N}_2)^2} \left[\hat{V}(\hat{y}_2) + (\hat{V}_2)^2 \hat{V}(\hat{N}_2) - 2\hat{V}_2 \hat{Cov}(\hat{y}_2, \hat{N}_2) \right]$
INTERVALO DE CONFIANZA	---	$\left \bar{y}_2 - \hat{\bar{y}}_2 \right \leq 1.96 \sqrt{ECM(\hat{\bar{y}}_2)}$	$\left \bar{y}_2 - \hat{\bar{y}}_2 \right \leq 1.96 \sqrt{ECM(\hat{\bar{y}}_2)}$

CUADRO N° 20

RESULTADOS DE LA SIMULACION DEL ESTIMADOR DE TOTALES (ESTABLECIMIENTOS NO MENORES)

Estadística	Nivel	Porcentaje de Clases de Actividad Incluidas en ese nivel para la Variable		
		P.O.	A.F.	V.A.
$CVT(\hat{Y}_{0,k}) \times 100$	0 - .25	12	16	12
	.25 - .50	16	28	16
	.50 - .75	28	32	24
	.75 - 1.00	28	20	32
	1.00 - 2.10	16	4	16
Prop. Sesgo $(\hat{Y}_{0,k}) \times 100$	0 - 10	12	4	12
	10 - 20	56	76	72
	20 - 30	8	4	16
	30 - 50	24	16	-
	100 - 96	72	68	60
Porcentaje de caídas en el intervalo.	92 - 88	28	32	40
	0 - 0.5	12	16	12
Porcentaje de error absoluto máximo respecto al total real.	0.5 - 1.0	20	20	8
	1.0 - 1.5	24	32	28
	1.5 - 2.0	20	16	28
	2.0 - 5.0	24	16	24
				12

101

CUADRO N.º 21

RESULTADOS DE LA SIMULACION DEL ESTIMADOR DE RAZON (ESTABLECIMIENTOS NO MENORES)

Estadística	Nivel	Porcentaje de Clases de Actividad para la razón P/P.O.
$\hat{CV}(\hat{R}_{0,k}) \times 100$	0 - .25	20
	.25 - .50	24
	.50 - .75	36
	.75 - 1.00	16
	1.00 - 2.90	4
$\hat{Sesgo} \text{ rel}(\hat{R}_{0,k}) \times 100$	0 - 15	24
	15 - 20	24
	20 - 30	36
	30 - 42	16
	100 - 96	60
Porcentaje de caídas en el intervalo	92 - 84	40
	0 - 1	28
Porcentaje del error absoluto máximo con respecto a la razón real	1 - 2	20
	2 - 3	32
	3 - 5	20

102

CUADRO N° 22

RESULTADOS DE LA SIMULACION DEL ESTIMADOR DE PROMEDIOS EN DOMINIOS DE ESTUDIO
(ESTABLECIMIENTOS NO MENORES)

Estadística	Nivel	Porcentaje de Clases de Actividad, para la variable P.O.
$\hat{CV}(\hat{Y}_{0,k}^{\infty}) \times 100$	0 - 1	24
	1 - 1.5	16
	1.5 - 2	24
	2 - 2.5	28
	2.5 - 3.7	8
sesgo $rel_2(\hat{Y}_{0,k}^{\infty}) \times 100$	0 - 5	20
	5 - 15	24
	15 - 20	44
	20 - 46	12
Porcentaje de caídas en el intervalo	100 - 96	84
	92	16
Porcentaje del error absoluto máximo con respecto al promedio real	0 - 2	24
	2 - 3	20
	3 - 4	32
	4 - 5	12
	5 - 10	12

TESIS CON
FALLA DE ORIGEN

ANEXOS

164

ANEXO N° 1

105

CLASES DE ACTIVIDAD QUE CUBRE LA ENCUESTA

- 2051 *Pasteurización, rehidratación, homogeneización y envasado de leche.*
- 2318 *Fabricación de encajes, cintas y otros tejidos de poca anchura; incluso pasamanería.*
- 2391 *Fabricación de telas impermeabilizadas e impregnadas con diversos materiales.*
- 2392 *Fabricación de alfombras, tapetes y tapices de fibras blandas.*
- 2412 *Confeccción de trajes, sacos, pantalones y otra ropa exterior para hombre; excepto camisas y uniformes.*
- 2413 *Confeccción de uniformes militares, deportivos, escolares, industriales y para otros usos.*
- 2414 *Confeccción de camisas.*
- 2415 *Confeccción de ropa exterior para niños y niñas.*
- 2519 *Fabricación de calzado y pantuflas; excepto los moldeados de hule o plástico.*
- 2711 *Fabricación de muebles; excepto los de metal y los de plástico moldeado.*
- 2821 *Fabricación de envases de papel.*
- 2921 *Impresión y encuadernación.*
- 3022 *Fabricación y mezcla de insecticidas y otros plaguicidas.*
- 3031 *Fabricación de hule y resinas sintéticas.*
- 3072 *Fabricación y refinación de sebo, grasas y aceites animales para usos industriales.*
- 3091 *Fabricación de pegamentos, adhesivos, impermeabilizantes, aprestos y productos similares.*
- 3219 *Fabricación de linóleos, calzado y otros productos de hule.*
- 3222 *Fabricación de envases, envolturas y rollos de plástico.*
- 3223 *Fabricación por moldeo o extrusión de calzado y juguetes de plástico.*
- 3229 *Fabricación por moldeo o extrusión de otros artículos de plástico.*

- 3354 Fabricación de mosaicos, tubos, bloques, postes y similares a base de cemento.
- 3429 Fundición, refinación, laminación, extrusión y estiraje de metales no ferrosos; excepto cobre y aluminio.
- 3531 Fabricación de estructuras metálicas para la construcción y tanques metálicos en general.
- 3593 Fabricación de alambres, telas metálicas y otros productos de alambre.
- 3596 Fabricación de piezas metálicas por fundición y moldeo; excepto para maquinaria, equipo y material de transporte.
- 3632 Fabricación, ensamble y reparación de maquinaria, equipo y tractores para las industrias extractivas y de la construcción.
- 3639 Fabricación y ensamble de maquinaria, equipo y sus partes para otras industrias específicas.
- 3692 Fabricación, ensamble y reparación de grúas, montacargas y otras máquinas para transportar o levantar.
- 3694 Fabricación, ensamble y reparación de bombas, rociadores y extinguidores.
- 3695 Fabricación de válvulas metálicas.
- 3697 Fabricación e instalación de equipos y aparatos de aire acondicionado, calefacción y refrigeración.
- 3729 Fabricación de partes, dispositivos y accesorios para equipo y aparatos de radio, televisión y comunicaciones.
- 3813 Fabricación de motores y sus partes para vehículos automóviles.
- 3815 Fabricación de partes para el sistema de suspensión de vehículos automóviles.
- 3817 Fabricación de partes y accesorios para el sistema eléctrico de vehículos automóviles.
- 3891 Fabricación y ensamble de motocicletas, bicicletas y otros vehículos de pedal.
- 3992 Fabricación de lápices, gomas, plumas y otros artículos y útiles para oficina, dibujo y pintura artística.

ANEXO N.º 2

108

VARIABLES UTILIZADAS EN EL DISEÑO ESTADÍSTICO DE LA ENCUESTA.

El diseño estadístico de la Encuesta se basa en cuatro variables, que son: promedio de personal ocupado total (P.O.), producción bruta total (P), activos fijos brutos (A.F.) y valor agregado censal (V.A.), las cuales se definen como sigue (Secretaría de Programación y Presupuesto, Dirección General de Estadística, 1979):

P.O.: Es el promedio correspondiente al personal ocupado total en tres fechas (31 de diciembre de 1974, 30 de junio de 1975 y 31 de diciembre de 1975). El personal ocupado total es aquel de planta y eventual que estaba empleado en el establecimiento en las fechas señaladas. Excluye a las personas con licencia ilimitada o pensionadas y a las que prestaron servicios a base exclusivamente de honorarios, igualas y comisiones, cuando no formaban parte del personal del establecimiento.

P.: Es el valor a precio de venta de los productos que el establecimiento elaboró en 1975, incluida la variación (+) de inventarios de productos en proceso; lo cobrado por servicios de maquila, reparación y mantenimiento; el valor del activo fijo producido por el establecimiento para su propio uso; el margen bruto por compra venta de mercancías y otros ingresos brutos.

A.F.: Es la suma de los valores a costo de adquisición de: maquinaria y equipo de producción, edificios, construcciones y obras, mobiliario, equipo de transporte y otros activos fijos tangibles, incluso terrenos.

V.A.: Es el valor que resulta de restar a la producción bruta total, los insumos totales; siendo estos, el valor de las materias primas y auxiliares, combustibles y lubricantes, envases y empaques, así como otros bienes y servicios consumidos por el establecimiento durante el año de 1975.

APENDICES

111

APENDICE N° 1

SALIDA DEL PROGRAMA FACTOR ANALYSIS DEL SPSS

Este apéndice presenta la salida del Programa Factor Analysis del SPSS, para la clase 3222. Consta de cinco hojas. La primera hoja contiene la siguiente información. En la primera columna aparecen las cuatro variables para las que se realizó el análisis, la segunda contiene la media, la tercera la desviación estándar y la cuarta el número de establecimientos de la clase. La segunda hoja muestra la matriz de correlación de las cuatro variables. En la tercera hoja se observa lo siguiente: La primera columna especifica las cuatro variables, la segunda, los estimadores de la comunalidad; la tercera, el factor o componente principal de que se trate; la cuarta, el valor característico de cada factor; la quinta el porcentaje de variación explicado por cada factor y la sexta, el porcentaje de variación acumulado de la columna anterior. En la cuarta hoja aparecen la matriz "factor" obtenida utilizando factorización principal sin iteraciones (opción PAI), y la matriz de comunalidades.

Por último, en la quinta hoja, aparece la matriz "Factor Score Coeficients" que sirve para calcular el índice I_e , según lo indicado en el Capítulo III.

TEMIS CON
FALLA DE ORIGEN

SALIDA DEL PROGRAMA FACTOR ANALYSIS DEL SPSS

CLASE DE ACTIVIDAD 5272 29 MAR 81 PAGE 368

FACTOR ANALYSIS FOR CENSUS: PART 2

FILE NAME (CREATION DATE = 20 MAR 81)
SUBFILE 572

(1)	(2)	(3)	(4)
PARAMETER	MEAN	STANDARD DEV	CASES
PER	37.8619	77.9724	272
PAO	9525517500	19956482500	272
VAL	38888432	8758182500	272
VAL	3809163750	2072801250	272

113

**TESIS CON
FALLA DE ORIGEN**

20 MAR 81 PAGE 370

FACTOR ANALYSIS FOR CENSOYM. PART 2

FILE NAME (CREATION DATE = 20 MAR 81)

SUBFILE = 573

(1)	(2)	(3)	(4)	(5)	(6)
VARIABLE	EST COMMUNALITY	FACTOR	EIGENVALUE	PCT OF VAR	CUM PCT
PCA	1.40810	1	3.63750	90.9	90.9
PROG	1.00000	2	.22122	5.6	96.5
AC1	1.00000	3	-.11559	2.9	99.4
TAL	1.00000	4	.02269	.6	100.0



115

00200
PF

FACTOR ANALYSIS FOR CENSOEN, PART 2

20 MAR 81 PAGE 372

FILE NAME (CREATION DATE = 20 MAR 81)
SUEFILL 573

FACTOR SCORE COEFFICIENTS

FACTOR 1

BCR 426032
PRDP 45592
AC1 454233
VAL 28378

117

APENDICE N° 2

PORCENTAJES ACUMULADOS PARA LAS VARIABLES POR CLASE DE ACTIVIDAD

Este apéndice contiene la salida del programa que divide el rango del índice en 50 niveles y muestra los porcentajes acumulados de las variables en cada nivel, de la clase 3222. La primera columna indica el número del nivel. La segunda y tercera columnas muestran el intervalo del Índice I_{e_j} . La cuarta columna contiene la suma de los índices I_{e_j} de los establecimientos de cada nivel y la quinta el número de establecimientos por nivel. La sexta columna es el acumulado de la columna anterior. Finalmente, las columnas 7 a la 12 muestran los porcentajes acumulados del valor del índice I_{e_j} , del número de establecimientos, del P.O., de la P., de los A. F. y del V.A., respectivamente para cada nivel.

TESIS CON FALLA DE ORIGEN

CLASE DE ACTIVIDAD 3222

DIRECCION GENERAL DE DISEÑO
DEPARTAMENTO DE ESTADISTICA

PORCENTAJES ACUMULADOS PARA LAS VARIABLES POR CLASE DE ACTIVIDAD

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
NIVEL	NIVELES DEL INDOCE	TOTAL POR NIVEL INDOCE	INDICE	NIVEL ESTABE	ACUM. EST. POR NIVEL	INDICE ESTABE	PER-OCCUP	ACTIVOS	AGREGADO		
1	10-9171	A	10-7089	10-9171020	1	1	04080	00368	08226	04856	09451
2	10-7088	A	10-7035	0000000	0	1	04080	00368	08226	04856	09451
3	10-5006	A	10-2921	0000000	0	1	04080	00368	08226	04856	09451
4	10-2921	A	10-0837	0000000	0	1	04080	00368	08226	04856	09451
5	10-0837	A	9-8754	0000700	0	1	04080	00368	08226	04856	09451
6	9-8754	A	9-6671	0000000	0	1	04080	00368	08226	04856	09451
7	9-6671	A	9-4587	0000000	0	1	04080	00368	08226	04856	09451
8	9-4587	A	9-2504	0000000	0	1	04080	00368	08226	04856	09451
9	9-2504	A	9-0420	0000000	0	1	04080	00368	08226	04856	09451
10	9-0420	A	8-8337	0003000	0	1	04080	00368	08226	04856	09451
11	8-8337	A	8-6253	0000000	0	1	04080	00368	08226	04856	09451
12	8-6253	A	8-4170	0000000	0	1	04080	00368	08226	04856	09451
13	8-4170	A	8-2087	0000000	0	1	04080	00368	08226	04856	09451
14	8-2087	A	7-9920	0000000	0	1	04080	00368	08226	04856	09451
15	8-0003	A	7-7820	0000000	0	1	04080	00368	08226	04856	09451
16	7-7820	A	7-5836	0000000	0	1	04080	00368	08226	04856	09451
17	7-5836	A	7-3753	0000300	0	1	04080	00368	08226	04856	09451
18	7-3753	A	7-1669	0000000	0	1	04080	00368	08226	04856	09451
19	7-1669	A	6-9586	0000000	0	1	04080	00368	08226	04856	09451
20	6-9586	A	6-7503	0000000	0	1	04080	00368	08226	04856	09451
21	6-7503	A	6-5419	6-6268315	1	2	06557	00735	12601	13200	08602
22	6-5419	A	6-3336	0000000	0	2	06557	00735	12601	13200	08602
23	6-3336	A	6-1252	6-1757227	1	3	08865	01103	14291	17721	15443
24	6-1252	A	5-9169	0000000	0	3	08865	01103	14291	17721	15443
25	5-9169	A	5-7086	0000000	0	3	08865	01103	14291	17721	15443
26	5-7086	A	5-5002	0000000	0	3	08865	01103	14291	17721	15443
27	5-5002	A	5-2919	0000000	0	3	08865	01103	14291	17721	15443
28	5-2919	A	5-0835	0000000	0	3	08865	01103	14291	17721	15443
29	5-0835	A	4-8752	0000000	0	3	08865	01103	14291	17721	15443
30	4-8752	A	4-6668	0000000	0	3	08865	01103	14291	17721	15443
31	4-6668	A	4-4585	0000000	0	3	08865	01103	14291	17721	15443
32	4-4585	A	4-2502	0000000	0	3	08865	01103	14291	17721	15443
33	4-2502	A	4-0418	4-6156091	1	4	10590	01471	17837	21089	17697
34	4-0418	A	3-8335	4-0781297	1	5	12121	01638	21855	23968	23797
35	3-8335	A	3-6251	3-5400000	0	5	13609	02206	24481	25602	25539
36	3-6251	A	3-4168	3-4000000	0	7	15099	02806	24818	25602	25539
37	3-4168	A	3-2084	3-2600000	0	7	16593	03394	25495	27674	28132
38	3-2084	A	3-0001	3-1200000	0	9	18083	04000	25992	27674	28132
39	3-0001	A	2-7918	3-1180657	1	9	19577	04614	26495	27674	28132
40	2-7918	A	2-5834	2-6290728	2	11	21071	05228	26998	27674	28132
41	2-5834	A	2-3751	2-4675789	2	12	22565	05842	27501	27674	28132
42	2-3751	A	2-1667	11-3977351	1	13	24059	06456	28004	27674	28132
43	2-1667	A	1-9584	8-1974527	1	18	25741	06668	28216	27674	28132
44	1-9584	A	1-7501	12-9173302	4	22	28005	06882	28216	27674	28132
45	1-7501	A	1-5417	14-6294145	9	38	31900	07197	28216	27674	28132

46	1.5817	A	1.3378	4.7886032	5	41	41877	-15499	-62878	-70782	-69392	-68988
47	1.3334	A	1.1250	18.6691012	12	52	71877	-30291	42078	-10252	-75377	-73456
48	1.1250	A	0.9187	19.1187775	19	74	85641	-27266	57182	-89452	-93377	-93352
49	0.9187	A	-7083	25.1383848	28	102	63358	-42687	88591	-87473	-93008	-93008
50	-7083	A	-5000	97.2412862	170	212	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

APENDICE N° 3

DETERMINACION DE ESTABLECIMIENTOS MAYORES Y MEDIANOS

Este apéndice contiene la salida del programa que calcula la $V(\hat{\mu})$ de la clase 3222 y localiza el valor mínimo. En la parte superior del listado aparece la clase de actividad, el total de establecimientos no menores y el tamaño de muestra inicial. La primera columna contiene el valor de i , desde $i = 0$ hasta el número de establecimientos no menores que tiene esa clase de actividad. En la segunda columna aparecen los números de folio de los establecimientos no menores y en la tercera el valor del índice Ie_2 de cada establecimiento. Finalmente, en la cuarta columna se muestran los valores de $V(\hat{\mu})$ obtenidos para el correspondiente valor de i .

En la parte inferior se imprime el mínimo valor de la varianza y el total de establecimientos mayores y medianos.

21

DOZCO
7
1950

DETERMINACION DE ESTABLECIMIENTOS MAYORES Y MEDIANOS

CLASE DE ACTIVIDAD	3222	TOTAL DE ESTABLECIMIENTOS	102	TAMANO INICIAL DE MUESTRA	82
(1)	(2)	(3)	(4)	(5)	(6)
FOLIO	INDICE DEL ESTABLECIMIENTO	VARIANZA DEL ESTIMADOR (N-I)			
0			.0023912		
1	763	6.3158662	.0013889		
2	25001	5.6076507	.0013988		
3	846	9.4169129	.0011099		
4	834	3.6509358	.0009600		
5	862	3.3228412	.0038463		
6	833	3.0721048	.0007561		
7	25005	2.8260871	.0006688		
8	818	2.6898945	.0006192		
9	2	2.5981751	.0005620		
10	890	2.3012398	.0005298		
11	863	2.1772562	.0004991		
12	836	2.0522992	.0004696		
13	791	2.5272458	.0034951		
14	804	1.9958199	.0009211		
15	25006	1.9863770	.0039986		
16	756	1.2310869	.0001758		
17	722	1.8565956	.0005556		
18	837	1.8232998	.0033362		
19	830	1.7532231	.0003193		
20	771	1.7997660	.0003019		
21	716	1.6553770	.0002879		
22	761	1.6566918	.0002235		

3229319

221

TESIS CON
FALLA DE ORIGEN

00200

23	832	1.5460809	.0302621
24	793	1.55576350	.0002502
25	33012	1.5226620	.3323381
26	742	1.5153139	.0002275
27	738	1.482491	.0002166
28	619	1.4882491	.3322289
29	770	1.4328667	.0001947
30	720	1.4318724	.3304836
31	740	1.3923181	.0301237
32	796	1.3659301	.0001633
33	813	1.2610069	.3321532
34	764	1.2363642	.0001429
35	837	1.3479890	.0301318
36	779	1.3159925	.0301213
37	812	1.2268859	.0001108
38	794	1.2971897	.3309995
39	737	1.1259367	.0300947
40	835	1.112326	.3300900
41	817	1.0679249	.0000657
42	835	1.0700436	.0000615
43	841	1.0735421	.3300769
44	703	1.0676100	.0000720
45	796	1.0484242	.3300673
46	751	1.0080598	.0000637
47	715	1.0000015	.0000597
48	853	.9864790	.3303559
49	803	.9526740	.0000522
50	805	.9405458	.0000492
51	838	.9266239	.0000462

3229316

121

DD200
17
14

52	843	.5726882	.3333429
53	724	.9250559	.0300392
54	846	.9027660	.0000359
55	855	.8729882	.3333431
56	704	.8189728	.0000318
57	804	.8014556	.0000308
58	806	.8001358	.0300296
59	708	.7969430	.0200283
60	602	.7876361	.3333270
61	735	.7858512	.0300255
62	740	.7831240	.0000239
63	788	.7809306	.0330219
64	783	.7807666	.0000209
65	757	.7826614	.0300200
66	732	.7801270	.0000191
67	787	.7208918	.0000181
68	765	.7189726	.3330169
69	811	.6986912	.0000162
70	863	.6771335	.8000160
71	776	.6602137	.0000161
72	859	.6591247	.0000163
73	30033	.6556198	.3330165
74	792	.6555377	.0000167
75	739	.6553926	.0000168
76	747	.6533231	.3330169
77	854	.6487536	.0000170
78	816	.6411107	.3330175
79	707	.6236052	.8000193

3229317

129

**TESIS CON
FALLA DE ORIGEN**

80	760	+602123	+330263
81	754	+598723	+000070
82	715	+592081	
83	702	+590521	
84	725	+509054	
85	789	+505382	
86	745	+5761217	
87	716	+575587	
88	75003	+5150524	
89	762	+5697824	
90	858	+5625656	
91	706	+532371	
92	752	+5500037	
93	760	+5338075	
94	784	+5180667	
95	776	+5170257	
96	849	+5165217	
97	702	+5131124	
98	768	+5130161	
99	746	+5086854	
100	829	+5043104	
101	801	+5008614	
102	786	+5000000	
MIRINO VALOR DE LA VARIANZA			+0000160
TOTAL DE ESTABLECIMIENTOS MAYORES			70
TOTAL DE ESTABLECIMIENTOS MEDIANOS			32

02020
 12
 02020

TALLERES DE LA NACION
 3229316

APENDICE N° 4

COBERTURAS DE MAYORES Y MEDIANOS

Este apéndice contiene la salida del programa que calcula las coberturas en total y en porcentaje de los establecimientos mayores y medianos, para las variables P.O., P., A.F., V.A. y número de establecimientos, de la clase 3222. Contiene la clase de actividad, el total de establecimientos mayores en esa clase y el total de medianos.

La primera columna indica la variable a la que se refiere la información. La segunda contiene el total de cada variable de los establecimientos no menores. La tercera columna contiene el total de la variable para los establecimientos mayores y la cuarta columna el porcentaje de cada variable que representa el total de los establecimientos mayores con respecto al total de los no menores. Las columnas quinta y sexta son análogas a la tercera y cuarta, sólo que se refieren a los establecimientos medianos.

TESIS CON FALLA DE ORIGEN

000000

COBERTURAS DE MAYORES Y MEDIANOS

CLASE DE ACTIVIDAD 3222

TOTAL DE MAYORES TOTAL MEDIANOS

TOTAL 32

COBERTURAS Y PORCENTAJES DE MAYORES Y MEDIANOS

(1) (2) (3) (4) (5) (6)

TOTAL MAYORES MEDIANOS

VARIABLE	CLASE	COBERTURA	COBERTURA	COBERTURA	%
P	778000-000	684900-000	682315-000	913000-000	117685-000
PA	778000-000	188181-010	836315-000	215998-000	103688-000
A	778245-000	667063-029	818220-000	8627883-000	864737-001
VA	778245-000	667063-029	818220-000	8627883-000	864737-001
EST.	102000-001	711000-022	88478-000	320000-000	11328-000

APENDICE N° 5

TAMAÑO DE MUESTRA

Este apéndice contiene la salida del programa que calcula el tamaño de muestra, de la clase 3222. Las abreviaturas que aparecen en la primera columna significan lo siguiente:

MPO = Tamaño de muestra para la variable P.O.

MP = Tamaño de muestra para la variable P.

MAF = Tamaño de muestra para la variable A.F.

MVA = Tamaño de muestra para la variable V.A.

NH = Número de establecimientos medianos.

MAXIM = Máximo de los tamaños de muestra de las cuatro variables.

Las columnas [2] a la [7] contienen la información correspondiente a cada estrato. La columna [8] es la suma de las columnas [2] a la [7]. La columna [9] presenta los tamaños de muestra sin estratificar.

CLASE DE ACTIVIDAD NO. 3222

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
EST. 1	EST. 2	EST. 3	EST. 4	EST. 5	EST. 6	TOTAL	MSEST	
MPO	2	0	0	0	0	4	4	4
MP	2	0	0	0	0	4	4	3
MAF	2	0	0	0	0	4	4	3
MVA	2	0	0	0	0	4	4	3
NH	15	0	0	0	0	32	32	0
HRXIM	2	0	0	0	0	4	4	4

621

APENDICE N° 6

SIMULACION DEL ESTIMADOR DE TOTALES

Este apéndice muestra la salida del programa que realiza la simulación del estimador de totales, de la clase 3222. La primera columna contiene las estadísticas que se obtuvieron, las cuales están descritas en la sección V.1 en el mismo orden en que aparecen en el listado.

Las columnas (2) a la (4) contienen la información obtenida para las variables P.O., P., A.F. y V.A., respectivamente.

TESIS CON FALLA DE ORIGEN

SIMULACION DEL ESTIMADOR DE TOTALES

(1)	(2)	(3)	(4)	(5)	(6)
ESTADÍSTICAS	PERSONAL OCUPADO	PRODUCCION BRUTA	ACTIVO FIJO	VALOR AGREGADO	VALOR AGREGADO
TCTA REAL Y	.773800+004	.208317+010	.725899+009		.770266+009
MEDIA REAL Y	.760588+002	.204232+008	.711449+007		.755163+007
ESTIMADOR P31MC	.786988+004	.211057+010	.753251+009		.777854+009
ESTIMADOR P31PC	.770773+004	.205260+010	.715432+009		.768235+009
RANG ESTIMADOR	.142753+003	.359933+008	.178184+008		.146188+008
PESEA DE LOS TOTALES ESTIMADOS	.777859+004	.207979+010	.726537+009		.770122+009
EROR ABSOLUTIC P31MC ESTIM.	.918831+002	.274039+008	.108665+008		.750738+007
EROR ABSOLUTIC P31PC ESTIM.	.203896+001	.822126+003	.153568+006		.239975+005
PROMPIO DEL EROR RESOLUTIC	.418332+002	.203511+007	.368705+007		.233766+007
VARIANZA DEL ESTIMADOR	.304068+004	.143923+013	.206908+014		.154657+014
DESVIACION STANDARD	.557429+002	.119968+008	.454879+007		.395263+007
ESTIMACION DE LA DEEV. STAN.	.428132+002	.120803+008	.456727+007		.410834+007
COEFICIENTE DE VARIACION	.531658+000	.374911+000	.629362+000		.533566+000
EROR CUADRATIC PESJE	.222583+004	.151547+015	.213702+014		.162243+014
COEFICIENTE DE VARIACION T.	.608179+002	.590948+003	.637013+002		.522928+002
SEGC	.198207+002	.234809+007	.714295+006		.808802+006
PROPORCION DEL SESGO	.462958+000	.196035+000	.156394+000		.194868+000
NUMERO DE CAIDAS EN ELINTER.	24	23	23		25

APENDICE N° 7

SIMULACION DEL ESTIMADOR DE RAZON

Este apéndice muestra la salida del programa que realiza la simulación del estimador de razón, de la clase 3222.

Especifica el número de establecimientos mayores, medianos y medianos en muestra.

La primera columna contiene las estadísticas que se obtuvieron, las cuales están descritas en la sección V.2 en el mismo orden en que aparecen en el listado.

Las columnas (2) y (3) contienen la información obtenida para los establecimientos medianos y no menores, respectivamente, para la razón P/P.O.

**LEJIS CON
FALLA DE ORIGEN**

STIMULACION DEL

ESTIMADOR DE RAZON

CLASE DE ACTIVIDAD 3222

MEDIANOS 32 MEDIANOS EN MUESTRA 4

PATCHES 70

RAZON = P.B/P.O

(5)

(7)

MEDIANOS NO MENORES

TOTY	.215991+009	.208317+010
TGTY	.913000+003	.775800+004
CCRENY	.342822+000	.817264+000
CHARTY	.554831+001	.575891+002
CHARTY	.603969+001	.710781+002
RAZON	.236573+006	.268519+006
CCVITY	.347206+009	.513204+007
ECM	.573892+009	.502172+004
SEELP	.355600+005	.193547+002
SEJEC	.327930+004	.450482+004
SEVAGE	.540207+002	.412511+002
SEVAGE	.701743+000	.772737+004
PAJEST	.272460+004	.262438+006
PJNEST	.197316+004	.103999+005
RANCO	.810847+003	.608077+004
EMAXE	.452548+005	.134793+003
EMLINE	.101606+004	.235159+004
PEARET	.180303+005	.267350+006
PEBEST	.28186+006	.750134+007
SECP	.447291+009	.273886+004
SEELCP	.811493+005	.639020+007
EWAF	.392666+009	.252768+004
SEENAF	.198788+005	.109411+004
ESJOC	.359040+004	.594209+002
ESJOC	.523476+001	.384872+000
ESJPEST	.344464+000	.448742+000
ESJAFR	.574900+000	.148742+000
INTERV	.828990+001	.148742+000

133

APENDICE N.º 8

SIMULACION DEL ESTIMADOR DE PROMEDIOS EN DOMINIOS DE ESTUDIO

Este apéndice muestra la salida del programa que realiza la simulación del estimador de promedios en dominios de estudio, de la clase 3222.

Especifica el número de establecimientos mayores, medianos y medianos en muestra.

La primera columna contiene las estadísticas que se obtuvieron, las cuales están descritas en la sección V.3 en el mismo orden en que aparecen en el listado.

Las columnas (2) y (3) contienen la información obtenida para los establecimientos no menores y medianos, respectivamente, para la variable P.O.

DP
DOCO

SIMULACION DEL

* ESTIMADOR DE PROMEDIOS EN DOMINIOS *

CLASE DE ACTIVIDAD 3222

VARIABLE P.O.

MAYORES 70 (1) MEDIANOS 32 (2) MEDIANOS EN MUESTRA 2 (3)

	(1)	(2)	(3)
	MAYORES	MEDIANOS	MEDIANOS EN MUESTRA
	94	24	24
TOTY	.762000+004		.775000+003
TOTN			.784091+000
CORRYA	.325635+000		.110476+000
CVARTY	.112361+001		.101968+000
CVARTN	.280396+001		.322917+002
MEDIA	.818638+002		.176722+003
CVWYTA	.176722+003		.376833+001
ECH	.200281+001		.193888+001
SSECC	.142927+001		.507770+001
SESGO	.349659+001		.189982+003
SEVCOE	.241358+003		.189982+003
ECHARM	.176316+001		.600350+001
MAREST	.836137+002		.379621+002
MINEST	.780833+002		.284850+002
RANSG	.540581+001		.841717+001
EAPAXE	.281557+001		.497047+001
EAMINE	.534333+001		.869741+001
PEAEST	.116495+001		.156028+001
MEDEST	.809934+002		.329950+002
ECCM	.201848+001		.389622+001
S0SECC	.149207+001		.197338+001
EVAR	.209741+001		.356121+001
S0EVAR	.146425+001		.184181+001
ESSEGO	.280094+000		.594195+000
ESPHED	.346878+002		.180071+001
ESVSEC	.197746+000		.301088+001
ESVEVA	.193990+000		.318730+000
ECVVAR	.178610+001		.570352+001
ANTIERV		24	