



Distr.
RESTRINGIDA

LC/MVD/R.102

Febrero de 1993

ORIGINAL: ESPAÑOL

CEPAL
Comisión Económica para América Latina y el Caribe
Oficina de Montevideo

ARCHIVO
único ejemplar

**DISEÑO MUESTRAL PARA LOS SECTORES
AGROALIMENTICIO Y METALMECANICO:
DOCUMENTO METODOLOGICO**

INDICE

	<u>Página</u>
1. Objetivos y universo a investigar	3
2. Unidad de observación	4
3. Niveles de desagregación de los resultados	5
4. Diseño muestral	6
5. Afijación de la muestra	9
6. Plan de estimadores y sus varianzas	10
7. Consideraciones finales	14



1. OBJETIVOS Y UNIVERSO A INVESTIGAR

La Encuesta sobre Políticas de Recursos Humanos se propone investigar, mediante el muestreo estadístico, diversas variables relacionadas con el empleo en las empresas medianas y grandes de ciertos sectores de la Industria Manufacturera en todo el país. Las variables a investigar se refieren a la dotación de personal de las empresas, sexo y edad de los empleados, ausentismo, formas de selección y reclutamiento, evaluación del desempeño, incentivos y capacitación.

El universo objeto de la investigación está constituido por las empresas que en el territorio nacional realizan principalmente actividades manufactureras del tipo agroindustrial, fabricación de productos de la madera y metalmecánica, sectores hacia los cuales se pretende extender la línea de investigación que la CEPAL iniciara con relación a la Industria Exportadora de nuestro país. Los recursos disponibles para la Encuesta permitirán entrevistar en el orden de las 100 empresas, lo que obliga acotar el universo incluyendo solo a las empresas medianas y grandes que, a pesar de su reducido número, representan una porción importante del empleo en dichos sectores de actividad.

AGRUPACION DE RAMAS	PESO DE LAS EMPRESAS MEDIANAS Y GRANDES (50 o más pers. ocupadas)
Agroindustrias	60,4%
Madera y Papel	25,0%
Metalmecánica	39,6%

Fuente: Dirección General de Estadística y Censos
(DGEyC) III Censo Económico Nacional - 1988

Si se clasifica a las empresas del universo por actividad y por su tamaño, se obtiene el siguiente cuadro.

UNIVERSO CORREGIDO

Agrupación de Ramas	Total	Tramo de personal ocupado				
		50-99	100-199	200-499	500 y +	
Total	273	137	77	37	22	
Agroindustria	172	80	48	26	18	
Madera y Papel	17	10	3	1	3	
Metalmecánica	84	47	26	10	1	

Fuente: DGEyC - Registro Permanente de Actividades Económicas

2. UNIDAD DE OBSERVACION

Las unidades estadísticas comunmente utilizadas en la investigación de las actividades comprendidas son, dependiendo de la temática a estudiar, la empresa, el local físico asiento de la actividad, el establecimiento tipo CIUU o la "unidad por clase de actividad" (conjunto de establecimientos de una empresa que en una misma región se dedican a la misma actividad principal).

Dadas las características de la investigación a llevarse a cabo en esta Encuesta, se ha entendido como mas apropiado elegir a la empresa como unidad de observación. Las políticas de personal y la capacitación, son variables cuyo comportamiento responde a decisiones que se toman al nivel de la empresa (mas que al nivel del local físico o del establecimiento).

Si se selecciona la empresa como unidad de observación deben tenerse en cuenta las siguientes limitaciones.

1. A la hora de presentar resultados por rama de actividad, debe establecerse un claro criterio para clasificar en una única categoría a las empresas con actividad en mas de una rama. Téngase presente que por la propia dinámica de la actividad, económica las empresas pueden cambiar su actividad principal aún en el corto plazo (de un trimestre a otro), pero para alterar el código de la actividad principal es necesario que transcurra al menos un año (para evitar la influencia de la estacionalidad).

2. En segundo lugar, aunque el diseño no lo prevé a priori, es posible presentar resultados desagregados por regiones del país, aunque con un nivel de error no predecible. Para obtener tabulaciones a nivel regional es necesario clasificar las unidades de observación por regiones. Cualquiera sea el criterio de regionalización, siempre existirán empresas con actividad en dos o mas regiones, y para estos casos será necesario o bien partir de la unidad de observación o bien asignarla a la región donde desarrolla la mayor parte de su actividad.

3. NIVELES DE DESAGREGACION DE LOS RESULTADOS

La forma en que podrán presentarse los resultados de la investigación está determinada por el diseño muestral. Este responde, a su vez, a las necesidades de información y su compatibilización con los recursos disponibles para la investigación.

El diseño muestral contempla una doble estratificación: por actividades y por tamaño de las unidades de observación.

El primer criterio permite agrupar en una misma clase y analizar conjuntamente empresas relativamente homogéneas desde el punto de vista de la producción.

La estratificación por tamaño de las empresas se justifica por las posibles diferencias que la caracterizan, no solo por la porción del empleo que detentan. De acuerdo con los resultados a obtenerse, se podrá investigar si las empresas de diferente tamaño presentan diferencias importantes en cuanto a las diferentes variables investigadas. Sin embargo, es necesario observar que de acuerdo con el diseño muestral propuesto, la precisión de las estimaciones por tamaño de las empresas no se fija a priori, y por tanto, en virtud del reducido tamaño de la muestra, sería conveniente colapsar los dos estratos mas pequeños para realizar inferencias confiables.

Sin perjuicio de lo anterior, y dependiendo de los resultados en el sorteo de la muestra, podrán presentarse otras desagregaciones de los datos, vinculados con posibles regionalizaciones y otras variables de interés contenidas en los cuestionarios.

Es importante señalar que el diseño muestral y el tamaño de la muestra no permiten realizar inferencia confiable para cada cruce de rama y tamaño, sino solamente para cada agrupación de ramas incluidas en el universo. Para presentar datos por regiones, deberá analizarse previamente el número de casos que resultan en el sorteo en cada región, de forma de garantizarse una información mínima. Probablemente será necesario establecer una regionalización en pocas categorías, de forma de obtener una mínimo de confiabilidad de las estimaciones.

La flexibilidad del diseño (lo cual se comenta en el capítulo siguiente) permite agregar o suprimir actividades en el universo a investigar. Incluso es posible incorporar otros tramos de personal ocupado (vg.: 20 a 49 personas ocupadas) a condición de realizar algunas pequeñas modificaciones en el diseño muestral, pero con un incremento importante en el tamaño de la muestra y, por tanto, en el costo de la investigación.

4. DISEÑO MUESTRAL

El diseño muestral que se propone es el muestreo aleatorio estratificado en una sola etapa de selección con:

- inclusión forzosa de las empresas "grandes" (con 200 y mas personas ocupadas).
- afijación óptima o de Neyman para los estratos de las empresas con 50 a 199 personas ocupadas.

La estratificación propuesta es doble: por actividades y por tamaño de las empresas. Las actividades se clasifican en 3 grupos, tal como se presentaron en la sección 1. El tamaño se define a partir de los siguientes tramos:

<u>Tramo</u>	<u>Personal Ocupado</u>
1	50 - 99
2	100-199
3	200-499
4	500 y más

Como se observa, quedan fuera de este diseño las empresas con menos de 50 personas ocupadas. La exclusión responde a razones de costo: se trata de un número muy grande de unidades que por su tamaño aportan individualmente muy poco al empleo total.

En virtud del nivel de desagregación deseado de los resultados, cada grupo de actividades se trata como un subuniverso independiente de los demás, el cual se estratifica en función de su personal ocupado.

El supuesto básico es que, al interior de cada grupo, las variables investigadas están alta y positivamente correlacionadas con la variable personal ocupado. Entonces es ventajosa la estratificación a través de esta última para realizar estimaciones sobre las demás variables, en la medida que prefijada una precisión y seguridad, la aplicación del muestreo aleatorio estratificado óptimo garantiza el tamaño mas pequeño posible para la muestra.

El tamaño de la muestra en cada grupo se calcula a partir de una precisión del 5% del total del personal ocupado en cada agrupación de ramas, y una seguridad del 95% en la estimación del total del personal ocupado.

$$P\left(\left| \sum_{h=1}^4 N_h \bar{X}_h - \sum_{h=1}^4 N_h \mu_h \right| < 0.05N\mu \right) = 0.95 \quad (1)$$

donde: $0.05.N\mu$ = precisión (5% del total poblacional)

0.95 = seguridad

N = total de empresas en la población (en la agrupación de ramas)

N_h = total de empresas en la población del estrato h

μ_h = media del personal ocupado en el estrato h

μ = media del personal ocupado en la población

\bar{X}_h = media muestral en el estrato h (estimador de μ_h)

n_h = tamaño de la muestra en el estrato h.

De acuerdo con el diseño muestral propuesto, se tiene:

$$\begin{aligned} n_3 &= N_3 \\ n_4 &= N_4 \end{aligned} \quad (2)$$

Para la determinación de los restantes tamaños muestrales (n_1 y n_2) se parte de la ecuación (1) y mediante transformaciones convenientes se obtiene:

$$n = n_1 + n_2 = \left(\sum_1^2 N_h \sigma_h \right)^2 / \left(\frac{0.05N\mu}{1.96} \right)^2 + \sum_1^2 N_h \sigma_h^2 \quad (3)$$

Donde: σ_h^2 = varianza del personal ocupado al interior del estrato

Finalmente, para determinar cada n_h ($h=1,2$) se aplica la formula de Neyman:

$$n_h = n \cdot \frac{N_h \sigma_h}{\sum_1^2 N_h \sigma_h} \quad (4)$$

Una sola corrección debe realizarse a los resultados obtenidos, la cual consiste en imponer en los estratos aleatorios la condición:

$$n_h \geq 2 \quad (5)$$

Esta condición es necesaria para posibilitar la estimación de la varianza de los estimadores, tal como se explica en la sección 6. Esta corrección, de tener que realizarse, garantiza la precisión y seguridad deseadas, y hasta puede incrementarlas.

En resumen, el diseño muestral para cada grupo de actividades es el muestreo aleatorio estratificado óptimo con las restricciones (2) y (5).

Una característica de este diseño es su flexibilidad: es posible realizar algunos cambios sin necesidad de modificar lo esencial del diseño. Así es posible suprimir alguna de las agrupaciones de actividades o agregar nuevos grupos con la única condición de sortear la muestra de estos grupos, adicionarla a la ya seleccionada y redefinir el universo investigado.

También es posible adicionar al universo el tramo de 20 a 49 personas ocupadas (para un grupo, algunos o todos los grupos) y aún el tramo de 10 a 19, con la condición de volver

a determinar el tamaño de la muestra en cada estrato de cada agrupación de ramas. En tal caso, el tamaño de la muestra se incrementará notablemente (en virtud del gran número de unidades en los tramos más pequeños) pero se reducirán mínimamente los tamaños n_1 y n_2 . Para evitar importantes modificaciones en el diseño, se sugiere que cuando estas reducciones sean mínimas, se mantengan los tamaños (n_1 , n_2) y simplemente se adicionen a la muestra las unidades sorteadas para el nuevo tramo.

Otro problema que podría plantearse es el de obtener resultados al nivel de una rama de actividad específica, que en este diseño ha sido agrupada con otras ramas. Si se desea mantener la precisión y seguridad planeadas, entonces será necesario redefinir los subuniversos y volver a sortear las muestras respectivas (correspondientes a los subuniversos redefinidos) manteniéndose inalterado el resto de la muestra original, con un tamaño total final mayor en virtud de la desagregación realizada.

5. AFIJACION DE LA MUESTRA

En el cuadro siguiente se presentan los tamaños de muestra por agrupación de ramas y su afijación por tramos de personal ocupado. Para la determinación de los tamaños de muestra se tuvieron en cuenta las relaciones y restricciones de (1) a (5) enunciadas en la sección anterior.

La experiencia del muestreo enseña que allí donde el tamaño de la muestra en un estrato es $n_h = 2$, se corre el riesgo de una "caída" de la muestra (por problemas de no contacto o de rechazo), razón por la cual, cuando el tamaño en la población lo permitía, el tamaño mínimo de la muestra en cada estrato se aumentó a 3 unidades. De esta forma, aún en casos de caída de la muestra en algún estrato, será igualmente posible el cálculo de las varianzas de los estimadores (si el tamaño efectivo de la muestra resulta por lo menos de tamaño 2 en cada estrato).

Por último, si en algunos de los estratos superiores a muestrearse resultaba que algún $N_h \leq 3$, entonces se tomó la decisión de bajar el límite de la inclusión forzosa de manera de incluir dichos estratos.

A los efectos de la aplicación de las fórmulas (3) y (4) se calcularon las medias (μ) de personal ocupado para cada agrupación de ramas y se aproximaron las varianzas de los estratos

$$(\sigma_h^2)$$

suponiendo uniformidad (distribución uniforme discreta) del personal ocupado de las empresas en cada estrato poblacional.

MUESTRA CORREGIDA

Agrupación de Ramas	TOTAL	Tramo de personal ocupado			500 y +
		50-99	100-199	200-499	
Total	90	14	17		3722
Agroindustria	52	3	5		2618
Madera y Papel	10	3	3		13
Metalmecánica	28	8	9		101

Nota: Estos resultados son preliminares, sujetos a eventuales modificaciones en el marco.

6. PLAN DE ESTIMADORES Y SUS VARIANZAS

Los estimadores que aquí se proponen se corresponden con el diseño muestral de la sección 4. Por tanto, valen para el universo de las empresas con 50 o mas personas ocupadas.

Sea X una variable cualquiera de las investigadas (personal ocupado, desempeño, etc.). El subíndice "i" se refiere a la empresa i-ésima de la muestra, el subíndice "h" denota el tramo de una agrupación y el subíndice "g" se refiere a la agrupación de ramas.

Supongamos que interesa estimar el parámetro "total en la población de la variable X":

$$T = \sum_g \sum_{h=1}^4 \sum_{i=1}^{N_{hg}} x_{ihg} \quad (6)$$

donde x_{ihg} es el valor de la variable X en la empresa i del estrato h de la agrupación g, y N_{hg} es el número de empresas en la población en el estrato h de la agrupación g.

El estimador que se propone es:

$$\tilde{T} = \sum_g \sum_{h=1}^4 \frac{N_{hg}}{n_{hg}} \sum_{i=1}^{n_{hg}} x_{ihg} \quad (7)$$

donde n_{hg} es el número de empresas en el estrato h de la agrupación g en la muestra.

La última sumatoria corresponde al valor de la X para todas las empresas de la muestra en el estrato h de la agrupación g. El coeficiente N/n expande los resultados de la muestra a toda la población del estrato h en la agrupación g. Si lo que se quiere calcular es un estimador del total de X en la agrupación g. (v.g.: Agroindustria), entonces:

$$\tilde{T}_g = \sum_{h=1}^4 \frac{N_{hg}}{n_{hg}} \sum_{i=1}^{n_{hg}} x_{ihg} \quad (8)$$

El total de X para la población total se estima sumando los \tilde{T}_g para todas las actividades.

Si lo que quiere estimarse es la proporción de empresas que en la población poseen un cierto atributo (empresas que invierten, que demandan capacitación, que exportan mas del 50% de sus ventas, etc.), entonces el problema puede resolverse a través de una variable Y dicotómica que toma los valores 0 ó 1.

Cuando una empresa posee el atributo, entonces $y = 1$, y cuando no lo posee, $y = 0$. El parámetro $p =$ proporción de empresas que en la población poseen el atributo, puede definirse como sigue:

$$p = \frac{\sum_g \sum_{h=1}^4 \sum_{i=1}^{N_{hg}} y_{ihg}}{\sum_g \sum_{h=1}^4 N_{hg}} \quad (9)$$

Sea:

$$N = \sum_g \sum_{h=1}^4 N_{hg}$$

El estimador que se propone es:

$$\tilde{p} = \frac{1}{N} \sum_g \sum_{h=1}^4 \frac{N_{hg}}{n_{hg}} \sum_{i=1}^{n_{hg}} y_{ihg} \quad (10)$$

Si solo interesa estimar la proporción de los que poseen el atributo en la agrupación g , entonces:

$$P_g = \frac{\sum_{h=1}^4 \sum_{i=1}^{N_{hg}} y_{ihg}}{\sum_{h=1}^4 N_{hg}}$$

Sea

$$N_g = \sum_{h=1}^4 N_{hg}$$

El estimador que se propone para P_g es:

$$\hat{P}_g = \frac{1}{N_g} \sum_{h=1}^4 \frac{N_{hg}}{n_{hg}} \sum_{i=1}^{n_{hg}} y_{ihg} \quad (11)$$

Todos los estimadores planteados tienen la propiedad de ser insesgados de sus respectivos parámetros.

Los datos de la muestra permitirán calcular estimaciones puntuales para cada parámetro investigado, utilizando las fórmulas (1) a (4). Pero si además se pretende construir intervalos de confianza para los parámetros, entonces será necesario conocer la varianza de los estimadores planteados. Las fórmulas siguientes permiten estimar dichas varianzas. Para simplificar la notación, se anotará:

$$\bar{x}_{hg} = \frac{1}{n_{hg}} \sum_{i=1}^{n_{hg}} x_{ihg}, \quad \bar{y}_{hg} = \frac{1}{n_{hg}} \sum_{i=1}^{n_{hg}} y_{ihg}, \quad \frac{n_{hg}}{N_{hg}} = f_{hg}$$

Entonces:

$$\hat{V}(\hat{T}_g) = \sum_{h=1}^2 N_{hg}^2 \cdot \frac{1-f_{hg}}{n_{hg}} \cdot \frac{\sum_{i=1}^{n_{hg}} (x_{ihg} - \bar{x}_{hg})^2}{n_{hg} - 1} \quad (2')$$

$$\hat{V}(\hat{T}) = \sum_g \hat{V}(\hat{T}_g) \quad (1')$$

$$\hat{V}(\hat{P}_g) = \sum_{h=1}^2 \left(\frac{N_{hg}}{N_g}\right)^2 \cdot \frac{1-f_{hg}}{n_{hg}} \cdot \frac{\sum_{i=1}^{n_{hg}} (y_{ihg} - \bar{y}_{hg})^2}{n_{hg} - 1} \quad (4')$$

$$\hat{V}(\hat{P}) = \sum_g \left(\frac{N_g}{N}\right)^2 \hat{V}(\hat{P}_g) \quad (3')$$

Como podrá observarse, en la fórmula de la varianza no se toma en cuenta la variabilidad de los tramos 3 y 4, pues en todos los casos estos son censados. Por ejemplo, para el tramo 4 resulta:

$$1 - f_{4g} = 1 - \frac{n_{4g}}{N_{4g}} = 1 - \frac{N_{4g}}{N_{4g}} = 0$$

En aquellos grupos en que otros estratos son censados puede mantenerse la misma fórmula pues por el motivo recién expuesto también contribuirán con un cero en la fórmula de la varianza.

Sea θ uno cualquiera de los parámetros recién definidos y sea $\hat{\theta}$ su respectivo estimador. Entonces puede construirse un intervalo de confianza para θ utilizando la siguiente expresión:

$$\hat{\theta} \pm 2\sqrt{\hat{V}(\hat{\theta})}$$

En virtud del reducido tamaño de la muestra, en muchos casos no será posible aplicar la aproximación del Teorema Central del Limite a la distribución de $\hat{\theta}$, pero aún en estos casos el intervalo propuesto contendrá el verdadero valor del parámetro con una probabilidad bastante alta (del orden del 95% cuando la distribución de $\hat{\theta}$ es aproximadamente normal).

7. CONSIDERACIONES FINALES

En primer lugar, tal como se diseñó la muestra, para la variable de estratificación la estimación de la media del personal ocupado se realiza con un 95% de seguridad y un 5% de precisión. Se demuestra que, en el caso de los estimadores propuestos, para una seguridad del 95%, la precisión es del 5% o aun mejor.

En segundo lugar, se reitera que la muestra no es autoponderada, esto es, los resultados de la muestra no pueden agregarse simplemente (por ejemplo para estimar totales, promedios o proporciones), sino que previamente deben clasificarse por tramo y grupo y luego ponderarse adecuadamente para obtener los estimadores deseados.

En tercer lugar, a pesar del reducido tamaño de la muestra se logra, a través del diseño muestral propuesto, minimizar el error muestral. Entonces, para obtener resultados precisos y confiables, es necesario controlar especialmente los errores ajenos al muestreo. Esto significa que será preciso poner especial atención en factores tales como:

- El intento reiterado de ubicar al informante originalmente sorteado y recién en último caso apelar a la lista de unidades suplentes.
- Insistir especialmente por la respuesta en los tramos 3 y 4 pues en ellos la muestra es un censo y no hay posibilidad de apelar a eventuales suplentes.
- La anotación cuidadosa de las genuinas respuestas de los informantes en los cuestionarios.
- La validación de los cuestionarios y la codificación de las preguntas abiertas.
- El control de la digitación.
- El control del marco y de la muestra antes de proceder a la expansión.