

SEGUNDAS JORNADAS DE MATEMÁTICA Y ESTADÍSTICA
APLICADAS A LA ECONOMÍA, CONTABILIDAD Y GESTIÓN DE EMPRESA
18 al 20 de setiembre de 1969

DISEÑO DE UNA MUESTRA DE
ESTABLECIMIENTOS INDUSTRIALES

TRABAJO N° 21

PRESENTADO POR: JORGE D. FISCHMAN
Instituto Nacional de Estadística
y Censos

DISEÑO DE MUESTRA DE ESTABLECIMIENTOS INDUSTRIALESRESUMEN

- 1º) La muestra ha sido seleccionada con el objetivo de mantener una encuesta periódica sobre variaciones de variables económicas tales como volumen físico de la producción, salarios, horas obrero trabajadas.
- 2º) Dada la falta de un padrón actualizado de establecimientos industriales se ha debido recurrir a dos tipos de selección
- a) Una selección de establecimientos correspondientes al Censo Económico 1964.
 - b) Una selección de establecimientos no captados por el Censo ya sea por omisión o por haber comenzado a funcionar con posterioridad a 1964.
- 3º) La selección de establecimientos de la parte a) se efectúa en forma aleatoria estratificada para cada subgrupo de la Clasificación Internacional C.I.I.U., por separado. El criterio utilizado para la estratificación es el valor de producción por establecimiento.
- 4º) La construcción de estratos y la determinación del número los mismos se hace para cada subgrupo, por el método del profesor Tore Dalenius.
- 5º) La fijación del número de establecimientos a seleccionar para cada estrato se hace en forma proporcional al producto del número de establecimientos por el desvío standard del valor de producción.
- 6º) Cada subgrupo contiene un estrato cuyos establecimientos son de inclusión forzosa en la muestra.
- 7º) La selección correspondiente a la parte b) se hace a partir de una muestra de áreas geográficas que comprende tres etapas.
- Interior del país: 1a. etapa: departamento provinciales.
2a. etapa: fracciones departamentales.
3a. etapa: radios censales.
- Capital Federal: 1a. etapa: circunscripciones electorales.
2a. etapa: subdivisiones de las circunscripciones o zonas.
3a. etapa: cuadradas manzanas.
- 8º) Para el interior del país las unidades de primera y segunda etapa, se eligen con probabilidades proporcionales a la raíz cuadrada del valor de producción censal. Las unidades de tercera etapa tienen igual probabilidad de selección.
- 9º) Para la Capital Federal las unidades se eligen con probabilidades proporcionales a la raíz cuadrada del valor de producción para las tres etapas.
- 10º) Las estimaciones de las variaciones periódicas respecto al período base se hacen por el método del cociente.

los establecimientos, valor agregado, número de horas trabajadas, cantidad de obreros ocupados.

La adopción de un criterio en particular estaría dada por la disponibilidad de información y en caso de tener una información completa, en el grado de confiabilidad que merezca cada información.

Número de los estratos y sus límites

A continuación supondremos que el criterio utilizado para la selección de establecimientos en cada estrato es el valor de producción.

Antes de seleccionar los establecimientos debemos proceder a determinar el número de estratos a construir y los límites de cada estrato.

En términos generales un mayor grado de estratificación resulta en una mayor eficiencia del muestreo; es decir, mayor precisión para igual cantidad de establecimientos.

Pero el aumento de la eficiencia no es proporcional al número de estratos, de modo tal que existe un punto desde el cual un mayor número de estratos no contribuye fuertemente a un aumento de la eficiencia y en cambio resulta engorroso y anti-económico. Habrá entonces que fijar el número de estratos, mediante un criterio práctico el cual variará según el número de establecimientos comprendidos en el grupo y su distribución por valor de producción.

En cuanto a los límites de los estratos existe un método de Dalenius, para establecerlos en forma eficiente.

Este método supone que originalmente se tiene una tabla de frecuencias para el valor de producción de los establecimientos que integran un grupo de actividades.

Si para cada intervalo de clase de la tabla tenemos f_k = frecuencia de los establecimientos en el k -ésimo intervalo, u_k = relación entre las longitudes del k -ésimo y del primer intervalo, el método de Dalenius consiste en obtener los valores acumulados de $\sqrt{f_k u_k}$, dividir $\sum_{k=1}^n \sqrt{f_k u_k}$ por el número de estratos que se desean y fijar el punto de corte para cada estrato en el extremo de cada intervalo cuyo valor acumulado de $\sqrt{f_k u_k}$, corresponde a un múltiplo de ese cociente.

Ejemplo: Supongamos tener para un grupo de establecimientos la siguiente variable de frecuencias del valor de producción.
distribución

INTERVALO		FR	VR	FAVR	\sqrt{FAVR}	$\Sigma \sqrt{FAVR}$
Menos de \$	100.000	10	1	10	3,16	
\$ 100.001 -	300.000	6	2	12	3,46	6,62
300.001 -	500.000	9	2	18	4,24	10,86
500.001 -	1.000.000	8	5	40	6,22	17,08
1.000.001 -	1.500.000	15	5	75	8,66	25,74
1.500.001 -	2.000.000	7	5	35	5,91	31,65
2.000.001 -	3.000.000	4	10	40	6,32	37,97
3.000.001 -	5.000.000	3	20	60	7,74	45,71

Supongamos que mediante un criterio práctico se haya fijado el número de estratos en 4. Habrá que fijar los límites de cada estrato, en los extremos de los intervalos para los que el valor acumulado de \sqrt{FAVR} resulta un múltiplo de $\frac{45,71}{4} = 11,44$ es decir 11,44; 22,88; 34,32; 45,77.

Los estratos resultantes son:

0 -	500.000
500.001 -	1.500.000
1.500.001 -	2.000.000
2.000.001 -	5.000.001

La eficiencia de este método ha sido probada y además constituye un método operativo, sin intervención de factores subjetivos.

Determinación del número de establecimientos por estrato

Para determinar la cantidad de establecimientos a seleccionar en cada estrato para el cálculo del índice deben fijarse dos criterios: la distribución de establecimientos por estratos y el número total de establecimientos.

Distribución de establecimientos por estrato

En lo que sigue supondremos que el costo de encuestar a un establecimiento es independiente de las características del mismo:

su valor de producción, ubicación geográfica, etc.....de modo tal que el costo mínimo para una precisión fija, se alcance cuanto menor sea el número de establecimientos necesarios para alcanzar esa precisión.

En este caso, resulta óptimo establecer para cada estrato un tamaño de muestra proporcional a $\frac{V_A S_A}{S^2}$ donde V_A es el número total de establecimientos en el estrato y S_A el desvío standard del valor de producción en el estrato.

Tamaño total de la muestra por grupo

De disponerse de presupuesto ilimitado, el tamaño de muestra para el total del grupo, se fijará en base al error máximo tolerable de muestreo.

La fórmula para calcularla en caso de haber efectuado adjudicaciones óptimas por estrato, será:

donde el error máximo de muestreo tolerable está dado por V = varianza máxima del valor de producción. En caso que el presupuesto sea limitado, se fijará un tamaño total para cada grupo en base al presupuesto disponible y se podrá calcular el error de muestreo en que se incurre con respecto al valor de producción.

Este procedimiento clásico ha sido ligeramente modificado. Esta modificación consiste en incluir para cada sub-grupo un primer estrato de inclusión forzosa.

El estado actual de esta parte del trabajo es el siguiente:

se ha procedido a la formación de estratos para cada sub-grupo y se ha realizado el programa correspondiente para permitir que el equipo de cómputo IBM perteneciente al Instituto Nacional de Estadística y Censos proceda al cálculo de las varianzas de cada estrato para cada sub-grupo.

Diseño de muestra de áreas geográficas

Este diseño de muestra comprende tres etapas: cada una de las cuales toma como unidades de muestreo áreas geográficas cada vez menores.

La forma en que se han tabulado los datos correspondientes al Censo Económico ha hecho que deban considerarse en cada etapa distintos tipos de unidades para la Capital Federal y para el interior del país.

Unidades para el interior del país

a) Unidades de primera etapa

Las unidades de primera etapa para el interior del país son los departamentos provinciales.

Los mismos han sido elegidos aleatoriamente con probabilidades proporcionales a la raíz cuadrada de su valor de producción.

La hipótesis de trabajo utilizada para fijar estas probabilidades consiste en que los departamentos provinciales con mayor valor de producción en el año 1963, son los que han tenido mayor posibilidad de absorber nuevos establecimientos.

Pero esta probabilidad de absorber nuevos establecimientos no puede considerarse como directamente proporcional al valor de producción correspondiente al Censo.

En efecto, puede llegar a presentarse, en los departamentos con mayor valor de producción un fenómeno de saturación en la posibilidad de creación de nuevos establecimientos.

En algunos casos ha debido procederse al agrupamiento de varios departamentos que integran una sola unidad de primera etapa.

Hay departamentos provinciales que en el último Censo Económico tenían un valor de producción industrial nula.

Si bien esto indica que los mencionados departamentos deben poseer seguramente una vida económica basada totalmente en la agricultura y ganadería, no pueden ser excluidos de la muestra otorgándoseles una probabilidad nula de aparición.

Otro criterio que podría haberse utilizado optativamente para no excluir estos departamentos es el de otorgárseles una probabilidad de selección arbitrariamente pequeña, pero este criterio no resulta conveniente por razones que se explican más adelante.

b) Unidades de segunda etapa

Las unidades de segunda etapa para el interior del país son las fracciones departamentales.

También se han elegido con probabilidades proporcionales a la raíz cuadrada de su valor de producción censal basándose en el mismo criterio de fijación de probabilidades que en la primera etapa.

También en la segunda etapa se ha debido proceder en algunos casos a la agrupación de fracciones para integrar unidades de segunda etapa.

c) Unidades de tercera etapa

Las unidades de tercera etapa son los radios censales.

Como para ellos no se disponía de información auxiliar sobre su valor de producción se han tomado con probabilidades de selección igual para cada unidad secundaria.

Razón final de muestreo

Las unidades de primera etapa se han estratificado según su valor de producción y para cada uno de estos estratos se han fijado razones finales de muestreo, resultante del producto de las probabilidades de las tres etapas según lo muestra el siguiente cuadro:

Valor de producción (miles de \$ 1963)	Razón final de muestreo
50 - 1.000	1/200
1.000 - 3.000	1/150
3.000 - 5.000	1/120
5.000 - 10.000	1/75
+ de 10.000	1/50

Es aquí donde se hace presente la dificultad para proceder por el método optativo de no exclusión de los departamentos con valor de producción censal nulo. En efecto, para llegar con ellos a una razón final de 1/200 correspondiente al primer estrato, habrá que fijar probabilidades muy altas en la 2a. o 3a. etapa, dada la probabilidad reducida de la primera etapa.

Unidades de muestreo para la Capital Federal

La muestra de áreas geográficas para la Capital Federal también comprende tres etapas:

a) Unidades de primera etapa

Son las circunscripciones electorales. Han sido elegidas con probabilidades proporcionales a la raíz cuadrada de su valor de producción.

b) Unidades de segunda etapa

Se han formado dividiendo a las circunscripciones seleccionadas en áreas lo más homogéneas posibles. Las probabilidades de selección para las unidades de segunda etapa se han fijado en forma proporcional a la raíz cuadrada del valor de producción.

c) Unidades de tercera etapa

Son las manzanas que integran las unidades de segunda etapa. Han sido seleccionadas con probabilidades proporcionales a la raíz cuadrada de su valor de producción.

✓ Índice del volumen físico de producción

Una de las principales informaciones que desean obtenerse mediante la encuesta para la que ha sido diseñada esta muestra es un índice de volumen físico de producción.

Esto hace interesante agregar a estas páginas las fórmulas más clásicas de cálculo de índice de volumen físico de producción y la forma en que los mismos pueden ser estimados a través de una muestra de establecimientos encuestados.

I - Nomenclatura utilizada

o En lo que sigue se utilizará la siguiente nomenclatura
hi : Precio unitario del producto *h*-ésimo en el *i*-ésimo establecimiento del grupo *j* de la clasificación internacional

C.I.I.U. para el período base.

q_{ijk}^0 : cantidad producida del producto k , en el establecimiento i del j -ésimo grupo para el período base.

p_{ijk}^t : precio unitario para el período t .

q_{ijk}^t : cantidad producida en el período t .

I.A - Formas de cálculo de un índice de volumen físico

Se conocen dos tipos clásicos de índice de volumen físico.

a) Índice de volumen físico de Laspeyres con coeficientes de ponderación fijos.

$$\hat{Q}_{Ljk} = \frac{\sum_{i=1}^m p_{ijk}^0 q_{ijk}^t}{\sum_{i=1}^m p_{ijk}^0 q_{ijk}^0}$$

b) Índice de cantidades con coeficientes de ponderación variable de Paasche.

A su vez, cada uno de estos índices puede calcularse como cociente entre un índice de valor de la producción:

$$V_{Rj} = \frac{\sum_{i=1}^m p_{ijk}^t q_{ijk}^t}{\sum_{i=1}^m p_{ijk}^0 q_{ijk}^0}$$

y un índice de precios.

Así

$$Q_{Ljk} = \hat{V}_{Rj} / \hat{P}_{Phj}$$

donde

$$\hat{P}_{Phj} = \frac{\sum_{i=1}^m p_{ijk}^t q_{ijk}^t}{\sum_{i=1}^m p_{ijk}^0 q_{ijk}^0}$$

es un índice de precios con coeficientes de ponderación variable de Paasche.

(n = número de establecimientos del grupo, seleccionados para el cálculo del índice).

y

donde

$$\hat{Q}_{pkj} = \hat{V}_{kj} / P_{Lhj}$$

$$\hat{P}_{Lkj} = \frac{\sum_{l=1}^m p_{lyk}^t q_{lyk}^o}{\sum_{l=1}^m p_{lyk}^o q_{lyk}^t}$$

es un índice de precios de Laspeyres con coeficientes de ponderación fijos.

Ya calculado el índice para cada producto del grupo se procede a construir un índice general para el grupo, como promedio ponderado de los índices para cada producto.

Para cada producto se elige como coeficiente de ponderación el valor de producción del producto en el año base o para el período actual, según el tipo de índice.

Así por ejemplo:

$$Q_{Lj} = \frac{1}{V_j^o} \sum_{k=1}^L w_{kj}^o \hat{Q}_{Lkj}$$

donde

$$V_j^o = \sum_{l=1}^m \sum_{k=1}^L p_{lyk}^o q_{lyk}^o$$

$$w_{kj}^o = \frac{\sum_{l=1}^m p_{lyk}^o q_{lyk}^o}{V_j^o}$$

(L = número de productos comprendidos en el grupo).

Cuando el índice de cantidades utilizado es un índice de Paasche de coeficientes de ponderación variable, también conviene utilizar coeficientes de ponderación variable en la construcción del índice global:

$$\hat{Q}_{pj} = \frac{1}{V_j^t} \sum_{k=1}^L w_{kj}^t \hat{Q}_{pkj}$$

donde

$$V_j^t = \sum_{l=1}^m \sum_{k=1}^L p_{lyk}^t q_{lyk}^t$$

y

$$w_{kj}^t = \frac{\sum_{l=1}^m p_{lyk}^t q_{lyk}^t}{V_j^t}$$

representa el valor de producción total para los establecimientos seleccionados en el período base.

V_{R1} representa el valor de producción del artículo k para los establecimientos seleccionados en el año base.

Análogamente V_{Rt} para el período t .

Cuando el índice de cantidades se calcula como cociente entre un índice de valores y un índice de precios, deben utilizarse los mismos coeficientes de ponderación, para el cálculo del índice global, que los empleados en el cálculo directo.

Además, si se hubiese dispuesto de un padrón actualizado de establecimientos, hubiese bastado con realizar una muestra de los mismos según la metodología de la parte a.

En ese caso se hubiese podido efectuar conjuntamente con las estimaciones de los índices, una estimación de los errores de muestreo en los mismos, por las fórmulas correspondientes a una estimación de cocientes.

La introducción de la muestra de áreas geográficas para captar nuevos establecimientos incide sobre el cálculo de los errores en los índices y las fórmulas finales de cálculo de error dependen de la forma en que se extraiga la muestra de nuevos establecimientos, que aún no ha sido terminada de determinar.

Cálculo del error de muestreo en el índice de cada producto

Si la muestra se diseña en forma aleatoria estratificada, con determinación de los límites y números de estratos, distribución de establecimientos por estratos y tamaño total de muestra en la forma explicada anteriormente, podrá determinarse el error de muestra para cada índice de cada producto y en cada período.

Daremos la fórmula correspondiente a la varianza de cada índice, cuando se utilizan coeficientes de ponderación fijos. Cuando se calcula un índice de Paasche, el error de muestreo se obtiene en forma análoga.

Utilizamos para la fórmula la siguiente nomenclatura:

$$\begin{aligned} x_{yjk} &= k y_k q_{yjk} \\ y_{jk} &= k y_k q_{yjk} \end{aligned}$$

La varianza del índice \hat{Q}_{Lkj} estará dada por:

$$V(\hat{Q}_{Lkj}) = \frac{(N-m)}{N(m-1)\bar{x}^2} \left(\sum y_{jk}^2 + \hat{Q}^2 \sum x_{yjk}^2 - 2\hat{Q} \sum x_{yjk} y_{jk} \right)$$

donde N: total de establecimientos en el grupo
n: tamaño total de la muestra
X: promedio de los valores x por establecimientos

Para un índice de Paasche, la fórmula sigue siendo válida pero con la siguiente nomenclatura:

$$x_{yk} = \frac{p_{yk}^0 q_{yk}}{p_{yk}^1 q_{yk}}$$
$$y_{yk} = \frac{p_{yk}^1 q_{yk}}{p_{yk}^0 q_{yk}}$$

Conviene señalar los siguientes hechos:

- 1) La muestra de establecimientos se selecciona en el período base, momentos en el cual se efectúa la estratificación.
- 2) El tamaño de muestra se fija en base al error tolerable en la producción total (en valor) y no en las cantidades físicas de cada producto.

Este diseño, para que resulte válido, está apoyado en 2 supuestos, que no son de cumplimiento general.

El primer supuesto consiste en que la producción de cada artículo está correlacionada con el valor total de producción del establecimiento.

Puede verse fácilmente que existen serias posibilidades de que este supuesto no se cumpla. Por ejemplo, tomemos el caso de un producto cuya elaboración no resulta conveniente para los establecimientos grandes por la dificultad de colocarlos en mercado en gran escala.

O el caso de un establecimiento muy grande que se dedique a la elaboración de un conjunto parcial de los productos elaborados en el grupo.

El segundo supuesto consiste en suponer que la estructura del valor de producción por establecimiento se mantiene a través del tiempo de tal modo que la estratificación efectuada para el período inicial siga siendo útil en los sucesivos.

Basta para que no se cumplan estos supuestos casos de crecimiento de establecimientos con valor de producción inicial menor o de caída de establecimientos con valor de producción inicial mejor.

Se acompañan los siguientes apéndices.

- 1) Programa del cálculo de varianzas para la muestra de establecimientos censados.
- 2) Listado de los sub-grupos de integran la investigación, su magnitud y estratos que los componen.