Centro de Investigación en Matemáticas, A.C.

CIMAT

Optimización del Diseño Muestral para Indicadores Económicos

T E S I S

Que para obtener el grado de

Maestro en Ciencias en Estadística Oficial

P r e s e n t a :

Diana Gabriela Cedeño Robles

Director de Tesis:

Dr. Rodrigo Macías

Co-Director de Tesis:

Mtro. En E. Sergio Nava

Guanajuato, Gto. Diciembre de 2009

Índice

INTRODUCCIÓN

La situación económica nacional presenta cambios cada vez más dinámicos y frecuentes, lo que implica que la información estadística deba de ser generada acorde con la velocidad de los mismos, y así apoyar adecuadamente las decisiones que deben tomar los diversos usuarios de la información.

Bajo esta premisa se elabora la Encuesta Mensual de Opinión Empresarial **(EMOE)**, la cual generará mensualmente indicadores que permitan establecer la situación y expectativas del sector manufacturero con base a la opinión de los dirigentes empresariales de las unidades económicas seleccionadas. Esto implica la búsqueda de nuevas alternativas y mejoras para la generación de información.

Las Encuestas de Opinión son una técnica de investigación que ha alcanzado una sólida posición en las sociedades modernas, ya que se les reconoce una gran capacidad predictiva, constituyéndose como un recurso indispensable para orientar las decisiones de los actores públicos y privados.

La característica principal de estas encuestas es la agilidad en las preguntas, que facilita a su vez la rapidez de las respuestas, permitiendo obtener resultados confiables en plazos breves. Su versatilidad permite la realización de preguntas puntuales en momentos específicos de la coyuntura económica para tener una perspectiva más cercana y objetiva de la inmediata evolución de la economía.

El cuestionario combina la opinión y tendencia a través de preguntas que se formulan en forma de tres alternativas excluyentes para cada variable (aumentó, permaneció igual y disminuyó) con lo que se construye una serie de indicadores que permita el análisis coyuntural del sector manufacturero.

El esquema de levantamiento de estas encuestas es mensual, consultándose a los directivos de las empresas seleccionadas a partir de la primera semana de cada mes y los resultados son presentados en los primeros días del mes siguiente.

Cabe hacer mención que en los países europeos tienen gran relevancia y difusión este tipo de encuestas y en América Latina se está dando un impulso importante para su realización y armonización por parte de la Comisión Económica para América Latina y el Caribe **(CEPAL)**.

Ante la necesidad de contar con un instrumento estadístico más acorde con la dinámica que presenta el sector, se inició en el año 2000 los trabajos para una reformulación del proyecto para ubicarlos dentro del marco de trabajo de las encuestas de opinión a nivel internacional, las cuales permiten en el corto plazo captar la situación de las empresas y el sector en el que se desarrollan y la percepción del entorno económico a través de la opinión de los directivos empresariales. Estos instrumentos estadísticos tienen gran importancia en los países desarrollados y se han convertido en una herramienta analítica fundamental entre los profesionales e instituciones especializadas en el análisis económico, que por su naturaleza mantienen un seguimiento constante de la coyuntura económica.

Por tal motivo, el presente trabajo tiene como objetivo el análisis exhaustivo de esquema de diseño a fin de proponer mejoras en la asignación de los tamaños de muestra y/o disminución de la misma, ampliándolo para un diseño estratificado considerando una encuesta de opinión multiobjetivo. Es por ello que comenzaremos por definir el muestreo estratificado, considerando las propiedades de los estimadores y la distribución multivariada, ésta última debido a que en la mayoría de las encuestas de opinión se presentan más de dos posibles respuestas, aplicando así los estadísticos referentes a la varianza y covarianza requeridos por la forma de construcción de los indicadores económicos.

Es así como en el segundo capítulo se presenta el diseño actual de la Encuesta Mensual de Opinión Empresarial la cual tiene como objetivo la estimación de los siguientes indicadores: Índice de Confianza del Productor, Indicador de Pedidos Manufactureros e Índice Agregado de Tendencia, siendo ésta encuesta la base para la aplicación de las técnicas de optimización aplicadas en la asignación del tamaño de la muestra por estrato, se eligió ésta encuesta debido a que es una de las pocas encuestas de opinión nacionales con una periodicidad mensual y que por las funciones desarrolladas dentro del Instituto Nacional de Estadística y Geografía (INEGI) se cuanta con el material necesario para la generación de simulaciones y aplicación de los distintos métodos. Además se conoce la forma exacta del cálculo de los indicadores lo cual es imprescindible para la estimación de su varianza misma que es el fundamento para los ejercicios y la toma de decisiones.

En el tercer capítulo se estudian los métodos de asignación óptima para un muestreo estratificado univariado, debido a que se toma cada uno de los objetivos de la encuesta por separado, comparándolos entre ellos, pues a lo largo de las aplicaciones se observa que presentan una alta correlación. Dentro de ésta optimización se aplican la asignación de Neyman y el método BFGS (Broyden, Fletcher, Goldfarb y Shanno) perteneciente a los métodos quasi-Newton, éstos métodos se eligieron después del análisis de varios, debido a que las características de la encuesta y de la información poblacional son factibles para su aplicación.

Finalmente se realizaron algunos ejercicios cambiando los tamaños de muestra y aplicando los procesos nuevamente, cuyo objetivo fue tener los elementos necesarios y suficientes para presentar una propuesta confiable en la mejora del diseño muestral.

CAPITULO 1.- CONCEPTOS BASICOS

I. Muestreo Estratificado

Comenzaremos por definir el muestreo estratificado, en el cual resulta conveniente fraccionar la población original de N unidades en L subpoblaciones de tamaño N_h tales que $N_1 + N_2 + \ldots + N_L = N$; de tal forma que cada unidad de la población pertenece solo a una subdivisión y la unión de todas ellas conforma la población original.
Si se extrae una muestra independiente de cada subdivisión y posteriormente se reúne la información para obtener las estimaciones globales de la población, tal esquema de muestreo se le conoce como ***muestreo estratificado*** y a cada subdivisión considerada de manera independiente se denomina ***estrato***.

Para ayudarnos a saber concretamente cuando se debe realizar el muestreo estratificado, se pueden considerar las siguientes razones:

1. Se pretende proteger contra la posibilidad de obtener una mala muestra, es decir, hablando nuevamente de la naturaleza de la población, que esta sea demasiado heterogénea y proporcione información que sea poco representativa.
2. Si los datos deseados deben tener una precisión conocida en algunas subdivisiones de la población.
3. Una muestra estratificada podrá administrase, de manera más conveniente, a un menor costo. Esto puede verse claramente en la forma de recolección de datos, en un estrato podrá aplicarse una encuesta telefónica y en otro estrato una encuesta personal.
4. Si se hace correctamente, dará estimaciones más precisas (con menor varianza).

Si la naturaleza de la investigación presenta por lo menos una de estas necesidades es conveniente utilizar el muestreo estratificado, Cochran (1977).

Notación

El subíndice $h = 1,2,\ldots,H$ denota el *estrato* e $i = 1,2,\ldots,N_h$ la unidad dentro del estrato h. Además

N	Número total de unidades en el marco
N_h	Número total de unidades en el estrato h
n_h	Número de unidades en la muestra en el estrato h
y_{hi}	Valor obtenido para la i-ésima unidad en el estrato h
$W_h = \frac{N_h}{N}$	Tamaño relativo del estrato h
$f_h = \frac{n_h}{N_h}$	Fracción de muestreo en el estrato h
$\bar{Y}_h = \frac{\sum_{j=1}^{N_h} y_{hi}}{N_h}$	Media poblacional del estrato h

$$\bar{y}_h = \frac{\sum_{i=1}^{n_h} y_{hi}}{n_h}$$ Media muestral en el estrato h

$$S_h^2 = \frac{\sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}{N_h - 1}$$ Varianza poblacional en el estrato h

II. Estimadores y sus propiedades

Un estimador es una función de la muestra que se utiliza para estimar un parámetro desconocido de la población. El estimador de la media poblacional usado en un muestreo aleatorio estratificado es $\bar{y}_{st}$ ("st" significa estratificado), donde:

$$\bar{y}_{st} = \frac{\sum_{h=1}^{H} N_h \bar{y}_h}{N} = \sum_{h=1}^{H} W_h \bar{y}_h \qquad (1)$$

donde $N=N_1+N_2+\ldots+N_L$.
El estimador de $\bar{y}_{st}$ en general, no es el mismo que el estimador de la media muestral $\bar{y}$ que es,

$$\bar{y} = \frac{\sum_{h=1}^{H} n_h \bar{y}_h}{n} \qquad (2)$$

La diferencia radica en que las estimaciones a partir de los estratos reciben sus ponderaciones correctas N_h/N.

Las propiedades principales de la estimación $\bar{y}_{st}$ se bosquejan en los siguientes teoremas[1].

Insesgabilidad

Ésta es una de las propiedades deseables para un estimador. El término insesgado se refiere al hecho de que la esperanza del estimador de la media poblacional usada en muestreo estratificado, es decir, $\bar{y}_{st}$, coincide con la media de la población $\bar{Y}$. El siguiente teorema demuestra esta propiedad.

Teorema 1. *Si en cada estrato la estimación muestral* $\bar{y}_h$ es insesgada, entonces $\bar{y}_{st}$ es una estimación insesgada de la media poblacional $\bar{Y}$.

$$E(\bar{y}_{st}) = E\left(\sum_{h=1}^{H} W_h \bar{y}_h\right) \qquad (3)$$

[1] Cochran William G. *Técnicas de Muestreo*, Compañía Editorial Continental,1990

$$= \sum_{h=1}^{H} W_h \bar{Y}_h$$
$$= \bar{Y}.$$

Varianza de los estimadores

Esta propiedad es importante ya que un estimador con menor varianza es más probable que dé lugar a estimaciones que estén más próximas al verdadero valor del parámetro.

Teorema 2. *Si las muestras se extraen independientemente en los diferentes estratos, entonces*

$$V(\bar{y}_{st}) = \sum_{h=1}^{H} W_h^2 V(\bar{y}_h), \qquad (4)$$

Donde ($V(\bar{y}_h)$ es la varianza de $\bar{y}_h$ sobre muestras repetidas del estrato de prueba h.

Teorema 3. *Para el muestreo aleatorio estratificado, la varianza de la estimación* $\bar{y}_{st}$ es

$$V(\bar{y}_{st}) = \frac{1}{N^2} \sum_{h=1}^{H} N_h (N_h - n_h) \frac{S_h^2}{n_h} = \sum_{h=1}^{H} W_h^2 \frac{S_h^2}{n_h} (1 - f_h) \qquad (5)$$

Estimaciones de la varianza para muestras estratificadas

Si se toma una muestra aleatoria simple dentro de cada estrato, una estimación insesgada de S_h^2 es

$$s_h^2 = \frac{1}{n_h - 1} \sum_{i=1}^{n_h} \left(y_{hi} - \bar{y}_h \right)^2 , \qquad (6)$$

entonces se obtiene el siguiente resultado.

Teorema 4. *Con muestreo aleatorio estratificado, una estimación insesgada de la varianza de* $\bar{y}_{st}$ *es*

$$\hat{V}(\bar{y}_{st}) = \sum_{h=1}^{H} \frac{W_h^2 s_h^2}{n_h} - \sum_{h=1}^{H} \frac{W_h s_h^2}{N} \qquad (7)$$

Recordando que para calcular éste estimador debe haber cuando menos dos unidades provenientes cada uno de los estratos, a continuación se presenta la distribución multinomial debido a que el presente estudio se basa en una encuesta de opinión con más de dos respuestas.

III. Distribución Multinomial

La distribución multinomial se puede ver como una generalización del binomial en el que, en lugar de tener dos posibles resultados, tenemos *r* resultados posibles.

Supongamos que el resultado de una determinada experiencia puede ser r valores distintos: A_1, A_2, ..., A_r cada uno de ellos con probabilidad p_1, p_2, ..., p_r, respectivamente.

$$P(A_1) = p_1; \quad P(A_2) = p_2; \quad ... \quad P(A_r) = p_r; \quad con \sum_{i=1}^{r} P(A_i) = 1;$$

Si repetimos la experiencia n veces en condiciones independientes, podemos preguntarnos la probabilidad de que el suceso A_1 aparezca k_1 veces, el suceso A_2, k_2 veces y así sucesivamente:

$$P\left[(A_1 = k_1) \cap (A_2 = k_2) \cap ... \cap (A_r = k_r)\right]$$

Al modelo estadístico que nos da dicha probabilidad se le denomina *Multinomial*, y su función de densidad viene dada por:

$$f(k_1, k_2, ..., k_r) = P\left[(A_1 = k_1) \cap (A_2 = k_2) \cap ... \cap (A_r = k_r)\right] = \frac{n!}{k_1! k_2! ... k_r!} p_1^{K_1} p_2^{K_2} ... p_r^{K_r}$$

$$con \quad \sum_{i=1}^{r} P(A_i) = 1 \quad y \quad \sum_{i=1}^{r} k_i = n;$$

como se ve, el modelo Multinomial queda definido por los parámetros (n, p_1, p_2, ..., p_r). La fórmula anterior puede deducirse de forma análoga al caso Binomial. En realidad, si tomamos $r = 2$ tenemos exactamente el modelo Binomial.

Se debe destacar que este modelo es un ejemplo de *distribución multivariante*, es decir, de distribución conjunta de varias (r) variables aleatorias. En efecto, si definimos la variable aleatoria X_1 como *número de veces que se produce el suceso A_1 de un total de* n *experiencias,* y así sucesivamente, tenemos un conjunto de r variables aleatorias discretas cuya función de densidad conjunta (valorada a la vez) viene definida por la anterior fórmula. Nótese que si consideramos cada una de estas variables X_i ($i = 1, 2, ..., r$) por separado, su distribución es la Binomial de parámetros n y p_i.

Por otra parte, para el desarrollo de la varianza de los indicadores es necesario conocer los estadísticos de ésta distribución, presentados a continuación:

ESPERANZA, VARIANZA Y COVARIANZA

Suponiendo que, la distribución de N_i es Binom(n, p_i). Por lo tanto, se tiene que

$$E(N_i) = np_i$$

$$Var(N_i) = np_i(1 - p_i)$$

Recordemos que, la distribución multinomial es una generalización del binomial, entonces, $N_i + N_j \sim Binom(n, p_i + p_j)$, por lo que

$$Var(N_i + N_j) = Var(N_i) + Var(N_j) + 2Cov(N_i,N_j)$$

$$n(p_i + p_j)(1 - p_i - p_j) = np_i(1 - p_i) + np_j(1 - p_j) + 2Cov(N_i,N_j)$$

$$Cov(N_i,N_j) = -np_ip_j$$

Una vez que se cuenta con los elementos necesarios para el desarrollo del cálculo de la varianza de los indicadores, se presenta el esquema de diseño utilizado y se describe la forma detallada de la construcción de los índices, con lo que podremos aplicar las definiciones anteriores.

CAPITULO 2.- Encuesta Mensual de Opinión Empresarial (EMOE)[2]

En el contexto del muestreo estratificado y las encuestas multiobjetivo se presenta ésta encuesta a fin de tener el insumo necesario para ejemplificar las técnicas de optimización estudiadas, motivo por el cual se muestra a continuación tanto el esquema de muestreo utilizado actualmente, como el procedimiento de cálculo de los índices mencionados en el objetivo de la encuesta.

Aquí es importante puntualizar que se tiene 3 esquemas diferentes, cuyas diferencias radican en la desagregación de las estimaciones que se requieren:

- El primero llamado "tradicional" se centra en generar estimaciones nacionales con un tamaño de muestra de 511 empresas incluyendo 397 con certeza.
- El segundo es una muestra "ampliada" la cual se produce como resultado de la necesidad de estimar los indicadores para los Subsectores de Actividad Económica que representan aproximadamente el 70% de los ingresos totales del sector manufacturero, con lo cual se obtiene un incremento de 424 empresas dando una muestra total de 935 empresas.
- En cuanto al tercer diseño denominada "completa" contempla la parte restante de los subsectores agrupados en un solo dominio, generando un incremento de 297 empresas con respecto a la ampliada, dando un total de 1232 empresas.

Para el presente trabajo se utiliza la muestra tradicional como parámetro de comparación ya que las estimaciones que se presentan son nacionales, sin tomar en cuenta la actividad económica, sin embargo nos servimos de los datos de la muestra completa para realizar las simulaciones con incremento de la muestra por estrato.

I. Diseño Muestral 2006-2009

1. Objetivo de la encuesta

Generar mensualmente Indicadores cualitativos tales como el Índice de Confianza del Productor, Índice de Pedidos Manufactureros e Indicador Agregado de Tendencia sobre la actividad económica que permita conocer anticipadamente su comportamiento y sirva de apoyo en la toma de decisiones del Sector público y privado.

2. Población objetivo

Son las empresas manufactureras del país con más de 100 personas ocupadas.

3. Cobertura geográfica

Los resultados de la encuesta permiten obtener estimaciones nacionales.

4. Diseño de la muestra

Se caracteriza por ser probabilístico, por lo cual los resultados obtenidos de la encuesta pueden generalizarse a toda la población objeto de estudio y también es posible medir los errores de las estimaciones, obtenidas de la encuesta.

[2] Datos publicados por el INEGI mediante su página www.inegi.org.mx

4.1 Marco de la encuesta

Se utilizan como referencia dos fuentes: el marco para el diseño de la muestra anterior, y la Encuesta Mensual de la Industria Manufacturera, actualizado con la problemática detectada en las mismas encuestas, definiendo un marco actual de 4 427 empresas, la base del directorio está formado con los resultados definitivos de los Censos Económicos 2004 referente al sector industrias manufactureras.

4.2 Unidad de muestreo

La unidad de muestreo para la encuesta son las empresas manufactureras.

4.3 Estratificación

Las empresas del marco muestral se clasificaron en cuatro grupos con base en el personal ocupado[3] de los Censos Económicos 2004 para Manufacturas presentado a continuación:

Estrato	Rango en Personal Ocupado	Número de Empresas
Total		**4 427**
1	1 001 y más	397
2	501 a 1 000	599
3	251 a 500	1 008
4	101 a 250	2 423

5. Esquema de muestreo

El esquema de muestreo es probabilístico y estratificado:

a) Probabilístico
Las unidades de selección tienen una probabilidad conocida y distinta de cero de ser seleccionadas.

b) Estratificado
Las unidades primarias de muestreo con características similares se agruparon para formar estratos.

6. Tamaño de la muestra

Para calcular el tamaño de la muestra no se consideran las empresas con más de 1,000 personas ocupadas, pertenecientes al estrato 1, las cuales se censarán. Para el resto de los estratos se utiliza el índice de confianza del productor como variable de referencia utilizando la siguiente expresión:

$$n_o = \frac{z^2 s^2 Deff}{d^2(1-TNR)} \qquad \text{donde:} \quad d = rI$$

[3] Personal ocupado. Remunerado, no remunerado y suministrado por otra razón social.

Aplicando factor de finitud:

$$n = \frac{n_o}{1 + \frac{n_o}{N}}$$

Donde:

n_o	=	tamaño de muestra sin corrección por finitud.
n	=	tamaño de muestra con corrección por finitud.
N	=	número de empresas en el marco.
I	=	índice de confianza del productor[4].
r	=	nivel de error relativo.
TNR	=	tasa de no Respuesta.
Deff	=	efecto de diseño[5].
s^2	=	varianza estimada del índice.
z	=	valor en tablas para una distribución normal estándar.

Considerando una confianza del 95%; error relativo del 4%, tasa de no Respuesta del 20%; efecto de diseño de 0.96 e índice de confianza del productor del 49.64% se obtuvo un tamaño de muestra de 114 empresas.[6]

7. Afijación de la muestra

La muestra se distribuyó mediante la expresión Neyman, calculando la varianza respecto al personal ocupado, excepto para el estrato 1 que fue incluido con certeza, según la siguiente expresión:

$$n_h = n\left(\frac{N_h S_h}{\sum_{h=2}^{H} N_h S_h}\right)$$

Donde:

n_h	=	muestra en el h-ésimo estrato.
n	=	muestra total.
N_h	=	total de empresas en el h-ésimo estrato.
S_h	=	desviación estándar del h-ésimo estrato, de la variable de personal ocupado.
H	=	número de estratos.

Presentando la siguiente distribución:

[4] Referido a marzo 2004.

[5] Definido como el cociente de la varianza en la estimación del diseño utilizado, entre la varianza obtenida considerando un muestreo aleatorio simple para un mismo tamaño de muestra.

[6] Sin incluir las empresas del estrato 1.

Estrato	Número de Empresas
Total	**511**
1	397
2	41
3	35
4	38

8. Selección de la muestra

La selección de la muestra fue aleatoria e independiente en cada estrato. Incluyendo con certeza en la muestra las empresas con más de 1 000 personas ocupadas.

9. Cálculo de los factores de expansión

Los factores de expansión son elaborados mediante el siguiente procedimiento:

$$F_h = \frac{N_h}{n_h}$$

Donde:

F_h = factor de expansión del h-ésimo estrato.

N_h = total de empresas en el h-ésimo estrato.

n_h = muestra en el h-ésimo estrato.

9.1 Ajuste por no Respuesta

El ajuste de factores de expansión por no Respuesta se realiza para cada uno de los dominios a nivel estrato.

$$F_h^{'} = F_h \frac{ns_h}{nr_h}$$

Donde:

F'_h = factor de expansión ajustado del h-ésimo estrato.

ns_h = número de empresas seleccionadas, en el h-ésimo estrato.

nr_h = número de empresas con respuesta, en el h-ésimo estrato.

F_h = factor de expansión del h-ésimo estrato.

10. Estimaciones

El estimador total de la característica X es:

$$\hat{X} = \sum_{h=1}^{H} \sum_{i=1}^{n_h} F_{hi} X_{hi}$$

Donde:

$\hat{X}$ = estimador total en la característica X.

F_{hi} = factor de expansión del i-ésimo elemento, de el h-ésimo estrato.

X_{hi} = característica a estimar reportada en el i-ésimo elemento, en el h-ésimo estrato.

H = número de estratos.

n_h = número de unidades en la muestra en el h-ésimo estrato.

Para la estimación de proporciones, tasas y promedios se utiliza el estimador de razón:

$$\hat{R} = \frac{\hat{X}}{\hat{Y}}$$

Donde, $\hat{Y}$ se define en forma análoga a $\hat{X}$.

11. Estimación de las precisiones

A diferencia del cálculo presentado en ésta investigación, actualmente para la evaluación de los errores de muestreo de las estimaciones requeridas se usa el método de Jacknife obteniéndose la siguiente fórmula para estimar la precisión de $\hat{\theta}$:

$$\hat{V}_{JK}(\hat{\theta}) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} \left[\hat{\theta}_{(hj)} - \hat{\theta}\right]^2$$

Donde:

$\hat{V}_{JK}(\hat{\theta})$ = varianza Jacknife del estimador $\hat{\theta}$.

$\hat{\theta}_{(hj)}$ = estimador en la misma forma que $\hat{\theta}$ pero sin utilizar la j-ésima observación, en el h-ésimo estrato.

n_h = número de elementos en el h-ésimo estrato.

H = número de estratos.

$\hat{\theta}$ = parámetro a estimar.

Las estimaciones de la desviación estándar (D.E.), coeficiente de variación (C.V.) y efecto de diseño (DEFF) se calculan mediante las siguientes expresiones:

$$\text{D.E.} = \sqrt{\hat{V}(\hat{\theta})} \qquad \text{C.V.} = \frac{\sqrt{\hat{V}(\hat{\theta})}}{\hat{\theta}} \qquad \text{DEFF} = \frac{\hat{V}_{JK}(\hat{\theta})}{\hat{V}(\hat{\theta})_{MAS}}$$

II. Cálculo de Índices y sus Varianzas

Dado que el objetivo de la encuesta es generar indicadores cualitativos de tendencia sobre la actividad de las empresas de la industria manufacturera, que permitan conocer anticipadamente el comportamiento del sector y sirvan de apoyo a la toma de decisiones en materia política industrial, es necesario conocer el procedimiento exacto para su cálculo así como su varianza.

ÍNDICE DE CONFIANZA DEL PRODUCTOR (ICP)

Para utilizar el "índice de confianza del productor" como variable de referencia en el cálculo del tamaño de muestra es necesario encontrar una expresión que nos permita asociar n (el tamaño de muestra) con el estimador $\hat{\theta}$ para un error ℓ con una probabilidad α , es decir, se pretende encontrar un valor de n tal que:

$$P(/\hat{\theta} - \theta / \geq \ell) = \alpha$$

Donde

$\ell = z_{\alpha} \sqrt{V(\hat{\theta})}$ …. (1)

z = *Valor de la abscisa bajo la distribución de normal estándar para un nivel de confianza de* $(1-\alpha)$

Ahora bien, el cálculo del índice de confianza del productor se realiza de la siguiente manera:

1. Se conforma de 6 preguntas, de las cuales las dos primeras se promedian, por lo cual se parte de hecho que ésta conformado por 5 preguntas:

    a. *Comparando la situación actual del país con la de hace un año ¿Cree que este momento es el adecuado para que se realicen inversiones? (preg.5.1)*
    b. *Comparando la situación actual de su empresa con la de hace un año ¿Cree que este momento es el adecuado para que se realicen inversiones? (preg. 5.2)*
    c. *¿Cómo considera usted la situación económica del país hoy en día comparada con la de hace 12 meses? (preg. 8.1)*
    d. *¿Cómo considera usted que será la situación económica del país dentro de 12 meses, respecto a la actual? (preg. 8.2)*
    e. *¿Cómo considera usted la situación económica de su empresa hoy en día comparada con la de hace 12 meses? (preg. 8.3)*
    f. *¿Cómo considera usted la situación económica de su empresa hoy en día comparada con la de hace 12 meses? (preg.8.4)*

2. Para cada una de éstas preguntas se saca una tabla de frecuencias expandidas por estrato- respuesta,

$$y_{khj} = \sum_{i=1}^{n_{khj}} F_{khji}$$

Donde

$$F_{khji} = \frac{N_h}{n'_h},$$

N_h = Total de empresas en el marco del h-ésimo estrato

n'_h = total de empresas entrevistados del h-ésimo estrato

3. Se obtienen los totales poblacionales y se calcula la fracción de muestreo:

Estrato	Población N_h	$w_h = \frac{N_h}{N}$
1	397	0.0897
2	599	0.1353
3	1008	0.2277
4	2423	0.5473
	4427	

Donde:

N_h = Total de empresas en la población para el h-ésimo estrato

N = Total de empresas en la población

4. Entonces tenemos que:

$\hat{P}_{khj} = \frac{y_{khj}}{n_h}$ con k = pregunta, h = estrato y j = respuesta.

5. Por lo tanto, el balance para la k-ésima pregunta por estrato resulta:

$B_{kh} = \sum_{j=1}^{5} a_j p_{khj}$ dado $a_j = \{0, 0.25, 0.50, 0.75, 1\}$

6. Con lo que decimos que para la k-ésima pregunta se tiene:

$$D_k = \frac{N_1}{N} B_{k1} + \frac{N_2}{N} B_{k2} + \frac{N_3}{N} B_{k3} + \frac{N_4}{N} B_{k4}$$

$$= w_1 B_{k1} + w_2 B_{k2} + w_3 B_{k3} + w_4 B_{k4} = \sum_{h=1}^{4} w_h B_{kh}$$

7. Finalmente, se tiene:

$I = \frac{1}{5} \sum_{k=1}^{5} D_k$ donde I = Índice de Confianza del Productor (ICP)

Ahora bien, ya que se conoce la forma precisa del cálculo del ICP también es necesario conocer su varianza, para ello se comenzará con el cálculo de $V(B_{kh})$ que será de utilidad, entonces se tiene que $B_{kh} = \sum_{j=1}^{5} a_j \frac{y_{khj}}{n_h}$ y se sabe que

y_{kij} es multinomial así $Vy_{khj}) = n_h p_{khj}(1 - p_{khj})$ y $Cov(y_{khj}, y_{khl}) = -n_h p_{khj} p_{khl}$ por lo tanto[7]:

$$V(B_{kh}) = \frac{1}{n_h^2}\sum_{j=1}^{5} a_j^2 V(y_{khj}) + \frac{2}{n_h^2}\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l Cov(y_{khj}, y_{khl})$$

$$= \frac{1}{n_h^2}\sum_{j=1}^{5} a_j^2 n_h p_{khj}(1 - p_{khj}) - \frac{2}{n_h^2}\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l n_h p_{khj} p_{khl}$$

$$= \frac{1}{n_h}\sum_{j=1}^{5} a_j^2 p_{khj}(1 - p_{khj}) - \frac{2}{n_h}\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl}$$

$$V(B_{kh}) = \frac{1}{n_h}\left(\sum_{j=1}^{5} a_j^2 p_{khj}(1 - p_{khj}) - 2\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl}\right)$$

Ahora bien,

$D_k = \sum_{h=1}^{4} w_h B_{kh}$ entonces suponiendo independencia

$$V(D_k) = \sum_{h=1}^{4} w_h^2 V(B_{kh}) + \cancel{2\sum_{h=1}^{4}\sum_{j=1}^{h-1} w_h^2 w_j^2 Cov(B_{kh}, B_{kj})}$$

Para comprobar ésta suposición se generaron 10,000 muestras calculando las B_{kh} para cada índice y se generaron las matrices de correlación. Concluyendo que existen elementos suficientes para suponer que las covarianzas son cero, finalmente:

$V(I) = \frac{1}{25}\sum_{k=1}^{5} V(D_k) + \cancel{2\sum_{k=2}^{5}\sum_{m=1}^{k-1} Cov(D_k, D_m)}$ sustituyendo

$$V(I) = \frac{1}{25}\sum_{k=1}^{5}\sum_{h=1}^{4} w_h^2 V(B_{kh}) = \frac{1}{25}\sum_{k=1}^{5}\sum_{h=1}^{4} w_h^2\left(\frac{1}{n_h}\left(\sum_{j=1}^{5} a_j^2 p_{khj}(1 - p_{khj}) - 2\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl}\right)\right)$$

Aplicando éste procedimiento se obtuvieron los resultados de la serie 2005 a 2008 presentados en el anexo A.

ÍNDICE AGREGADO DE TENDENCIA (IAT)

Debido a que el cálculo de los índices es muy similar, se sigue el mismo procedimiento que en el anterior partiendo de que el índice de tendencia se forma de la siguiente manera:

1. Se conforma de 5 preguntas:

[7] Utilizando el teorema del valor esperado y la varianza de una función lineal. Mendenhal, "Estadística con Aplicaciones". Pág. 210

a. *Para los períodos indicados respecto al inmediato anterior señale como se comportó el volumen físico de producción de su empresa (preg 1.1)*
b. *¿Cómo considera usted el porcentaje de utilización de planta y equipo que registró y registrará su empresa para los períodos indicados? (preg. 1.2)*
c. *¿Cómo considera usted la variación en la demanda nacional de sus productos, medida por el volumen de unidades vendidas para los periodos indicados respecto al inmediato anterior? (preg. 2.2)*
d. *¿Cómo evolucionó el volumen físico de las exportaciones de su empresa para los períodos siguientes respecto al inmediato anterior? (preg. 2.4)*
e. *¿Cómo evolucionó y estima que lo hará el número de trabajadores en su empresa, en los períodos indicados respecto al inmediato anterior? (preg. 6.1)*

2. Para el caso de la pregunta “b” , antes de julio 2007 primero se clasificaban las respuestas de la siguiente manera:

$$Si \begin{cases} Histórico > Estimado \Rightarrow \Pr eg.1.2 = 1 \\ Histórico = Estimado \Rightarrow \Pr eg.1.2 = 3 \\ Histórico < Estimado \Rightarrow \Pr eg.1.2 = 5 \end{cases}$$

Una vez clasificada se continuaba el procedimiento global, descrito en los siguientes pasos.

3. Cada una de estas preguntas se pondera utilizando una variable auxiliar, conforme a la siguiente tabla:

Pregunta	VARIABLE	PONDERADOR
a	Producción	Ventas
b	Producción	Activo Fijo
c	Demanda Interna	Ventas
d	Demanda Externa	Ventas
e	Personal ocupado	Personal ocupado

Aplicando la siguiente fórmula para la k-ésima pregunta:

$$y_{khj} = \sum_{i=1}^{n_{hj}} Pond_{khji} * F_{khji}$$

Donde:

F_{khji}	=	Factor de expansión en el i-ésimo elemento en el h-ésimo estrato.
y_{khj}	=	Respuesta del i-ésimo elemento en el h-ésimo estrato para la k-ésima pregunta.
$Pond_{khji}$	=	Ponderador del i-ésimo elemento en el h-ésimo estrato para la k-ésima pregunta.
H	=	Número de estratos.
n_{khj}	=	Número de unidades en la muestra en el h-ésimo estrato para la k-ésima pregunta.

4. Se saca una tabla de frecuencias ponderadas (sumando la variable ventas) por estrato-respuesta, tomando como base las cifras de la sección "Estimado".

5. Después tenemos que:

$\hat{P}_{khj} = \frac{y_{khj}}{n_h}$ con k = pregunta, h = estrato y j = respuesta. Por lo tanto, el balance para la k-ésima pregunta por estrato resulta:

$$B_{kh} = \sum_{j=1}^{5} a_j p_{khj} \quad \text{dado } a_j = \{0, 0.25, 0.50, 0.75, 1\}$$

6. Así podemos decir que para cada k-ésima pregunta se tiene:

$$D_k = \frac{N_1}{N} B_{k1} + \frac{N_2}{N} B_{k2} + \frac{N_3}{N} B_{k3} + \frac{N_4}{N} B_{k4}$$

$$= w_1 B_{k1} + w_2 B_{k2} + w_3 B_{k3} + w_4 B_{k4} = \sum_{h=1}^{4} w_h B_{kh}$$

donde k= pregunta a

7. Finalmente, se tiene:

$I = \frac{1}{5}\sum_{k=1}^{5} D_k$ donde I = Índice Agregado de Tendencia (IAT)

De igual manera que para el ICP, se tiene la varianza de B_{kh} cómo se presenta a continuación:

$$V(B_{kh}) = \frac{1}{n_h}\left(\sum_{j=1}^{5} a_j^2 p_{khj}(1 - p_{khj}) - 2\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl}\right)$$

Luego entonces la varianza del índice es:

$$V(I)=\frac{1}{25}\sum_{k=1}^{5}\sum_{h=1}^{4}w_h^2V(B_{kh})=\frac{1}{25}\sum_{k=1}^{5}\sum_{h=1}^{4}w_h^2\left(\frac{1}{n_h}\left(\sum_{j=1}^{5}a_j^2\,p_{khj}(1-p_{khj})-2\sum_{j=2}^{5}\sum_{l=1}^{j-1}a_j a_l\,p_{khj}\,p_{khl}\right)\right)$$

Se presentan las varianzas estimadas para la serie mensual 2005-2009 en el anexo A

INDICADOR DE PEDIDOS MANUFACTUREROS (IPM)

Para la conformación del IPM, se ponderan cinco indicadores de manera decreciente como se muestra a continuación:

a) Demanda (30 por ciento).

2.1¿Cómo considera usted el comportamiento en el volumen total de pedidos a su empresa provenientes del mercado interno y externo, para los periodos indicados respecto al inmediato anterior?

b) Producción (25 por ciento).

1.1Para los periodos indicados respecto al inmediato anterior señale ¿Cómo se comportó el volumen físico de producción de su empresa?

c) Personal Ocupado (20 por ciento)

6.1¿Cómo evolucionó y estima que lo hará el número de obreros y empleados dependientes de su empresa, o de otra razón social que laboran en la misma, en los periodos indicados respecto al inmediato anterior?

d) Entrega de Insumos (15 por ciento).

3.2(A) ¿Para los periodos indicados con respecto al mes anterior, cómo califica la oportunidad de la entrega de insumos por parte de los proveedores?

e) Inventarios (10 por ciento)

3.1 ¿Cómo evolucionó el volumen físico de sus inventarios de insumos y bienes intermedios para los periodos indicados respecto al inmediato anterior?

A fin de conocer el proceso del cálculo del indicador se describen las diferencias con respecto al cálculo anterior del IAT,

1. Cada una de estas preguntas se pondera utilizando una variable auxiliar, conforme a la siguiente tabla:

Pregunta	VARIABLE	PONDERADOR
a	Demanda	Ventas
b	Producción	Ventas
c	Personal Ocupado	Personal ocupado
d	Entrega de Insumos	Ventas
e	Inventarios	Ventas

2. Se pondera conforme a los pesos indicados en la descripción de las preguntas, es decir:

    a) Demanda (30 por ciento).
    b) Producción (25 por ciento).
    c) Personal Ocupado (20 por ciento).
    d) Entrega de Insumos (15 por ciento).
    e) Inventarios (10 por ciento).

Siguiendo el mismo procedimiento del índice anterior tenemos que el indicador de pedidos manufactureros es:

$$IPM = 0.30 * D_a + 0.25 * D_b + 0.20 * D_c + 0.15 * D_d + 0.10 * D_e$$

En cuanto a su varianza, tenemos que

$B_{kh} = \sum_{j=1}^{5} a_{kj} \frac{y_{khj}}{n_h}$ donde a_j cambia por a_{kj} debido a que en éste indicador para la pregunta 4 las ponderaciones son diferentes

$a_{4j}=\{1.0,0.75,0.5,0.25,0\}$, ahora bien,

se sabe que

$$V(B_{kh}) = \frac{1}{n_h}\left(\sum_{j=1}^{5} a_{kj}^2 p_{khj}(1-p_{khj}) - 2\sum_{j=2}^{5}\sum_{l=1}^{j-1} a_{kj} a_{kl} p_{khj} p_{khl}\right)$$

Entonces,

$D_k = \sum_{h=1}^{4} w_h B_{hi}$ y suponiendo independencia $V(D_k) = \sum_{h=1}^{4} w_h^2 V(B_{kh})$

Obteniendo:

$$V(IPM) = V(0.30 * D_a + 0.25 * D_b + 0.20 * D_c + 0.15 * D_d + 0.10 * D_e)$$

Finalmente, sea

$$v = (0.30, 0.25, 0.20, 0.15, 0.10)$$

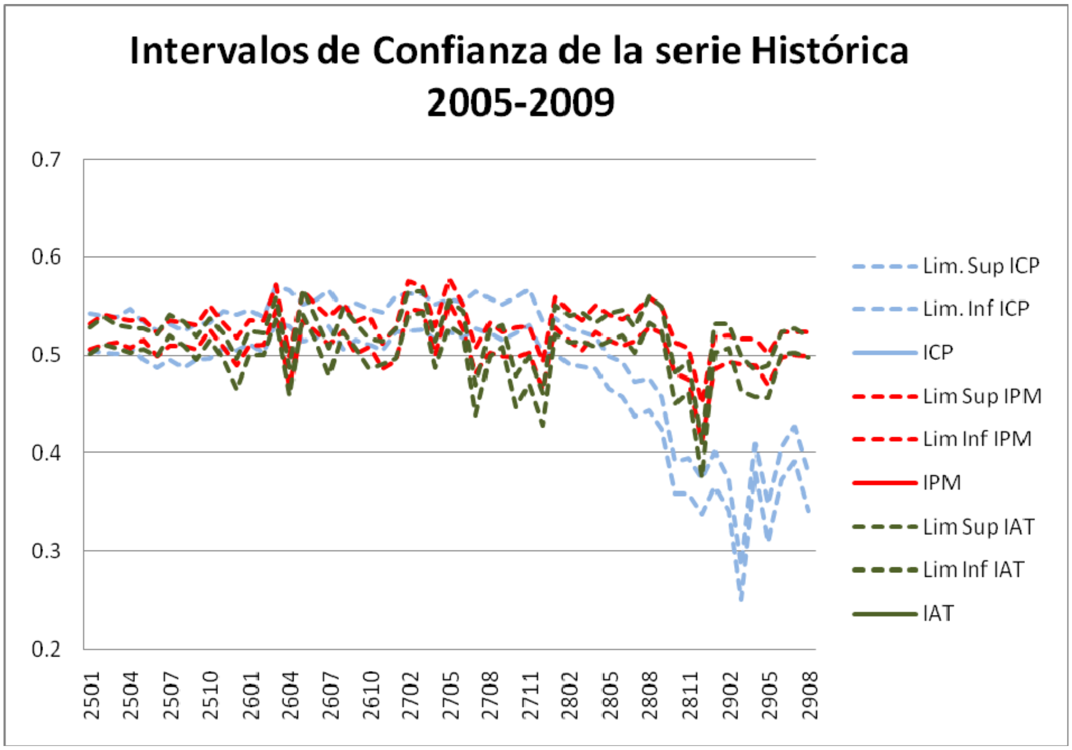
entonces

$$V(IPM) = V\left(\sum_{k=1}^{5} v_k D_k\right) = \sum_{k=1}^{5} v_k^2 V(D_k) = \sum_{k=1}^{5} v_k^2 \sum_{h=1}^{4} w_h^2 V(B_{kh})$$

$$V(IPM) = \sum_{h=1}^{4} w_h^2 \sum_{k=1}^{5} v_k^2 V(B_{kh})$$

Ver las estimaciones de la serie mensual 2005-2009 en el Anexo A.

Debido a que algunas de las series utilizadas son preliminares y para no crear confusión entre los datos presentados en éste estudio y los datos publicados por INEGI, se evitó mostrar los datos estimados de los indicadores (ICP, IAT e IPM), únicamente se muestran estadísticos y la siguiente gráfica en la que se presentan los intervalos de confianza, a fin de tener una idea más clara del impacto económico reflejado en la percepción de la opinión empresarial y su tendencia en el tiempo.



Una vez conocidos los antecedentes de la encuesta y la forma de cálculo de los índices, nos enfocaremos a la aplicación de la optimización general la cual será dirigida hacia la asignación de la muestra, presentada en el siguiente capítulo.

CAPITULO 3.- ASIGNACIÓN ÓPTIMA

Si se ha determinado el tamaño de muestra para la población, es necesario establecer el porcentaje que corresponderá a cada uno de los estratos, por lo cual se da un vistazo general al muestreo estratificado univariado ya que el diseño de la encuesta utiliza únicamente la variable de personal ocupado para la estratificación de las empresas. Continuando con la asignación de Neyman que es la aplicada en el diseño actual, finalmente se presenta una nueva propuesta con el fin de optimizar ésta distribución.

I. Muestreo Estratificado Univariado

En el muestreo estatificado, los valores de los tamaños de muestras n_h en los respectivos estratos, los elige quien hace el muestreo. Se pueden seleccionar para minimizar $V(\bar{y}_{st})$.

La función de costo más simple es de la forma

$$\cos to = C = c_o + \sum c_h n_h \tag{8}$$

Dentro de cualquier estrato, el costo es proporcional al tamaño de la muestra, pero el costo por unidad c_h puede variar entre estratos. El término c_0 representa un costo fijo.

Teorema 5. *Con una función de costo lineal para un muestreo estratificado, la varianza de la media estimada* $\bar{y}_{st}$ *es un mínimo para un costo específico C y el costo es un mínimo para una varianza específica* $V(\bar{y}_{st})$ donde n_h es proporcional a $W_h S_h / \sqrt{c_h}$. (9)

Con lo cual se presentan los siguientes problemas

1) Elegir los n_h para minimizar V con C especificada, o bien,
2) Elegir los n_h para minimizar C con V especificada.

En términos del tamaño total de muestra n_h en un estrato, tenemos como resultado del teorema,

$$\frac{n_h}{n} = \frac{W_h S_h / \sqrt{c_h}}{\sum W_h S_h / \sqrt{c_h}} \tag{10}$$

Esto da lugar a las siguientes reglas de conducta. En un estrato dado, se toma una muestra más grande si

1) El estrato es más grande
2) El estrato es más variable internamente
3) El muestreo es más barato en el estrato.

Para completar la asignación se toma n_h en términos de n y la solución depende de si la muestra se escoge para satisfacer un costo total especificado C o para dar una varianza especificada V para y_{st}. Si el costo es fijo y se resuelve la ecuación para n. Esto da

$$n = \frac{(C - c_0)\sum(N_h S_h / \sqrt{c_h})}{\sum(N_h S_h / \sqrt{c_h})} \tag{11}$$

Si V es fija, se sustituye la n_h óptima en la fórmula para $V(\bar{y}_{st})$. Encontramos

$$n = \frac{(\sum N_h S_h / \sqrt{c_h}) \sum (N_h S_h / \sqrt{c_h})}{V + (1/N) \sum W_h S_h^2} \tag{12}$$

donde $W_h = N_h / N$.

Un caso importante se da cuando $c_h = c$, esto es, si el costo por unidad es el mismo en todos los estratos. Es costo se convierte en $C = c_0 + cn$ y la asignación óptima para un costo fijo se reduce a la asignación óptima para un tamaño de muestra fijo. El resultado de este caso especial es como sigue.

II. Asignación Neyman

Teorema 6. *En un muestreo estratificado la varianza $V(\bar{y}_{st})$ se minimiza para un tamaño de muestra fijo n si*

$$n_h = n \frac{W_h S_h}{\sum W_h S_h} = n \frac{N_h S_h}{\sum N_h S_h} \tag{13}$$

ésta asignación se llama asignación de Neyman.

Por el momento sólo hemos abarcado el muestreo estratificado univariado, sin embargo la mayoría de las encuestas contemplan el problema de estimar varias características de la población por lo que a continuación se revisarán algunos conceptos.

Ejemplo. Utilizando el promedio histórico de las varianzas y aplicando la afijación de Neyman para cada uno de los índices se obtuvo la siguiente distribución de la muestra (n=114)[8]

Promedio de Desviación

	Estrato			
	1	2	3	4
N	397	599	1008	2423
n- Actual	397	41	35	38
ICP	0.04356859	0.19812815	0.39297583	0.83004573
Proporción		0.04698	0.15682	0.79620
Muestra 1	**397**	**5**	**18**	**91**
IPM	0.0398256	0.15780242	0.29050928	0.63788273
Proporción		0.04890	0.15150	0.79960
Muestra 2	**397**	**6**	**17**	**91**
IAT	0.0389521	0.15427327	0.29114386	0.59884159
Proporción		0.05031	0.15977	0.78992
Muestra 3	**397**	**6**	**18**	**90**

[8] Tamaño demuestra obtenido en el diseño nacional de la muestra tradicional.

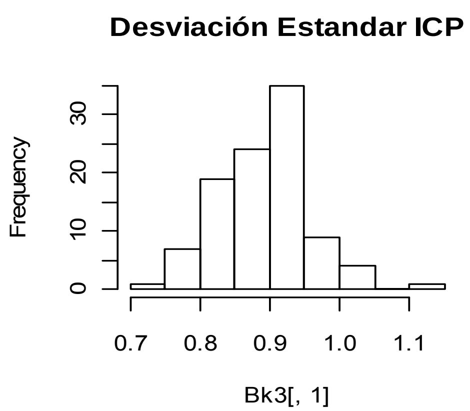
Utilizando Var Mayor por estrato

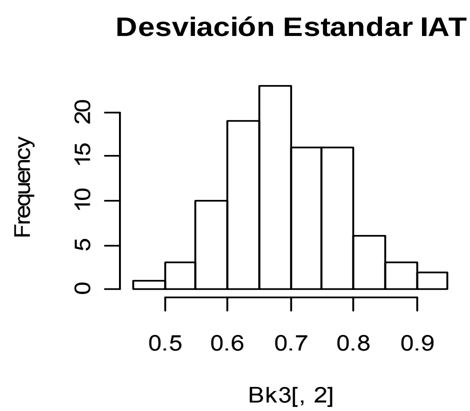
	0.04356859	0.19812815	0.39297583	0.83004573
Proporción		0.04698	0.15682	0.79620
Muestra 4	**397**	**5**	**18**	**91**

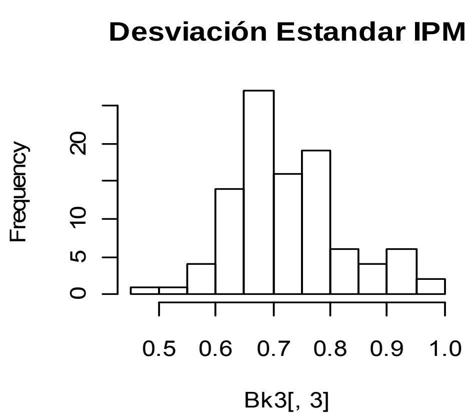
Se presenta además un cuarto ejercicio tomando el máximo de las varianzas por estrato. Concluyendo que el incrementar la muestra en el estrato 4 podría contribuir en la disminución de la varianza global, además de que se podría disminuir la muestra en el estrato 2, ésta hipótesis de llevó a cabo seleccionando 1000 muestras y replicando los cálculos obteniendo los siguientes resultados

Desviación Estándar Promedio de 1000 muestras simuladas		
Índice	Asignación Actual	Asignación Propuesta
ICP	**0.928723**	**0.8912**
IPM	**0.727575**	**0.7232**
IAT	**0.699956**	**0.6935**

A fin de ampliar el panorama de resultados se presentan los histogramas de la desviación estándar calculada para las muestras seleccionadas de los 3 indicadores:







También se obtuvieron los estadísticos básicos:

	Asignación Neyman		
Estadístico	ICP	IAT	IPM
Min. :	0.7485	0.494	0.4925
1st Qu.:	0.8445	0.6222	0.6594
Median :	0.899	0.6852	0.7058
Mean :	0.8912	0.6935	0.7232
3rd Qu.:	0.9276	0.7553	0.7776
Max. :	1.1381	0.9324	0.9634
NA's :		1	

Con lo que se tiene suficientes elementos para concluir que con el mismo tamaño de muestra al asignar mayor número de empresas en el estrato 4 y disminuir en el estrato 2 se disminuir la varianza de los tres indicadores, sin embargo se no se puede dejar de lado que al disminuir la muestra sustancialmente como en el estrato 2 el problema de la no Respuesta puede ocasionar conflictos en los cálculos.

III. Optimización, Método Cuasi-Newton

Estos métodos son similares a los métodos de gradiente conjugado en el sentido de que se basan principalmente en propiedades de las funciones cuadráticas. Sin embargo, en el método del gradiente conjugado, la principal fortaleza de la búsqueda se deriva del uso de las direcciones conjugadas de búsqueda, mientras que los métodos de Quasi-Newton están diseñados para imitar más directamente las características positivas del método de Newton pero usando sólo información de primer orden.
Todos los métodos de este grupo generan las direcciones a usarse para generar el siguiente punto con:

$$s(X^{(k)}) = -A^{(k)}\nabla f(X^{(k)}) \qquad (3.3.1)$$

donde $A^{(k)}$ es una matriz de $N \times N$ a la que se denomina la métrica.
A los métodos que emplean direcciones de esta forma, se les suele llamar métodos de métrica variable, porque A cambia a cada iteración. Para ser precisos, un método de métrica variable es un método quasi-Newton si está diseñado de tal forma que satisfaga la siguiente propiedad cuadrática:

$$\Delta x = C^{-1}\Delta g \qquad (3.3.2)$$

Supongamos una recursión para estimar la inversa de la matriz Hessiana:

$$A^{(k+1)} = A^{(k)} + A_c^{(k)}$$

donde $A_c^{(k)}$ es una corrección a la métrica actual.

La idea es formar $A_c^{(k)}$ tal que la secuencia $A^{(0)}, A^{(1)}, A^{(2)},..., A^{(k+1)}$ aproxime $H^{-1} = \nabla^2 f(x^*)^{-1}$, porque al lograr esto, una búsqueda lineal adicional producirá x* si f(x) es cuadrática.
Supongamos que una aproximación de C^{-1} es de la forma:

$$C^{-1} = \beta A^{(k)} \quad (3.3.4)$$

donde β es un escalar.
Sin embargo, se quiere que la aproximación satisfaga:

$$\Delta x^{(k)} = A^{(k)} \Delta g^{(k)}$$

Pero esto es imposible, puesto que necesitamos $A^{(k)}$ para encontrar $\Delta g^{(k)}$, donde:

$$\Delta x^{(k)} = x^{(k+1)} - x^{(k)}$$
$$\Delta g^{(k)} = g\left(x^{(k+1)}\right) - g(x^{(k)})$$

Se sugiere entonces, que la nueva aproximación satisfaga:

$$\Delta x^{(k)} = \beta A^{(k+1)} - \Delta g^{(k)}$$

combinando las ecuaciones, tenemos:

$$A_c^{(k)} \Delta g^{(k)} = \left(\frac{1}{\beta}\right) \Delta x^{(k)} - A^{(k)} \Delta g^{(k)}$$

y al verificarse por sustitución directa, se tiene que:

$$A_c^{(k)} = \frac{1}{\beta}\left(\frac{\Delta x^{(k)} y^T}{y^T \Delta g^{(k)}}\right) - \frac{A^{(k)} \Delta g^{(k)} z^T}{z^T \Delta g^{(k)}}$$

es una solución. Como "*y*" y "*z*" son vectores arbitrarios, esta es realmente una familia de soluciones. Por esta razón, se propone un método iterativo de la forma:

$$x^{k+1} = x^k - t^k S^k g^k \qquad g^k = \nabla f(x^k)$$

Donde S^k es una matriz que aproxima a $\left[\nabla^2 f(x^k)\right]^{-1}$ y $t^k \geq 0$ minimiza f sobre $x^k - \Delta g^k$ para $\Delta \geq 0$ (paso exacto o aproximado).
Se presenta uno de los métodos que permiten construir iterativamente la matriz S^k de manera que se verifiquen las siguientes condiciones:

i. Definida Positiva: Si S^k es definida positiva $\Rightarrow S^{k+1}$ también lo es
ii. Aproximación de la matriz hessiana inversa de $f : x^k \to \bar{x}, S^k \to \left[\nabla^2 f\left(\bar{x}\right)\right]^{-1}$ para $k \to \infty$.

La forma de construir estos métodos asegura:

- Convergencia Global
- Rapidez de convergencia mayor que lineal
- Convergencia a un mínimo local

Método de BFGS

Este método fue propuesto en 1970 por Broyden y refinado posteriormente por Fletcher, Goldfarb y Shanno. Pertenece a la misma familia del método de Davidon-Fletcher-Powell, pero se le considera más poderoso. A continuación se presenta la descripción general del método para su aplicación:

METODO BFGS (Broyden-Fletcher-Goldfarb-Shanno)

Etapa 0: Seleccionar un punto inicial $x^0 \in \mathbb{R}^n$
Inicializar $S_0 = I_{n \times n}$
$k = 0$

Etapa 1: Calcular $g^k = \nabla f(x^k)$
Si $\|g^k\| \approx 0 \Longrightarrow STOP$
Si no, calcular $x^{k+1} = x^k - t^k S^k g^k$
donde $t_k \geq 0$ se escoge según regla de Goldstein
y seguir a Etapa 2.

Etapa 2: Calcular:
$p^k = x^{k+1} - x^k \quad q^k = g^{k+1} - g^k$

$$S^{k+1} = S^k + \left[1 + \frac{(q^k)^t S^k q^k}{(p^k)^t q^k}\right] \frac{p^k (p^k)^t}{(p^k)^t q^k} - \frac{\left[p^k (q^k)^t S^t + S^k q^k (p^k)^t\right]}{(p^k)^t q^k}$$

$k = k + 1$ y volver a Etapa 1.

Con lo cual se quiere minimizar la varianza en función del tamaño de muestra para cada uno de los índices, por lo que las funciones a minimizar son las siguientes:

$$\min_n V(ICP) = \frac{1}{25} \sum_{k=1}^{5} \sum_{h=1}^{4} w_h^2 \left(\frac{1}{n_h} \left(\sum_{j=1}^{5} a_j^2 p_{khj} (1 - p_{khj}) - 2 \sum_{j=2}^{5} \sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl} \right) \right)$$

$$\min_n V(IAT) = \frac{1}{25} \sum_{k=1}^{5} \sum_{h=1}^{4} w_h^2 \left(\frac{1}{n_h} \left(\sum_{j=1}^{5} a_j^2 p_{khj} (1 - p_{khj}) - 2 \sum_{j=2}^{5} \sum_{l=1}^{j-1} a_j a_l p_{khj} p_{khl} \right) \right)$$

$$\min_n V(IPM) = \sum_{h=1}^{4} w_h^2 \sum_{k=1}^{5} v_k^2 V(B_{kh})$$

Donde $\quad \sum_{h=1}^{4} n_h = n$

Ahora bien, éste es un método de optimización general que se aplicó para optimizar la asignación del tamaño de muestra actual (114 empresas), obteniendo la siguiente distribución:

Distribución por Estrato Optimizando con BFGS				
	Estrato			
Parámetro	**1**	**2**	**3**	**4**
N	397	599	1008	2423
Actual	397	41	35	38
ICP (Asignación I)	397	18	30	66
IPM (Asignación II)	397	16	32	66
IAT (Asignación III)	397	16	36	62

Con el fin de poder comparar la distribución se replicaron nuevamente 1,000 muestras de las cuales se obtuvieron los siguientes resultados y los histogramas presentados en el anexo B:

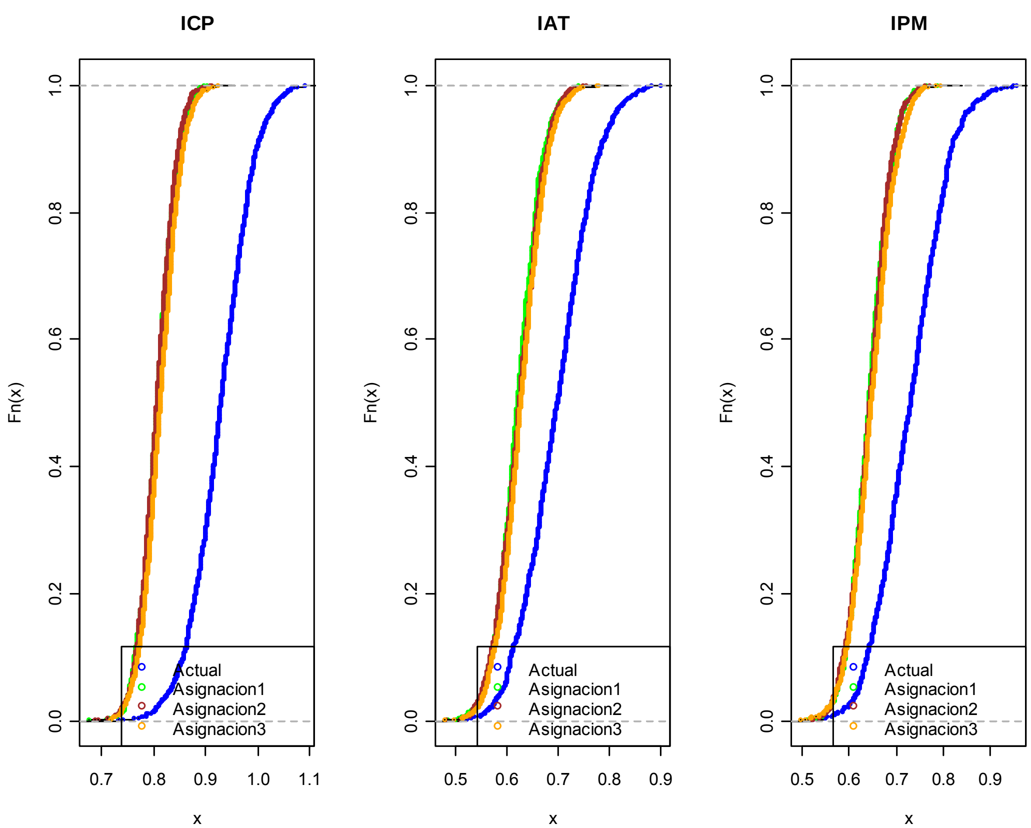
	Asignación Actual		
Estadístico	ICP	IAT	IPM
Min. :	0.7155	0.4520	0.5192
1st Qu.:	0.8950	0.6558	0.6833
Median :	0.9303	0.6969	0.7251
Mean :	0.9301	0.6998	0.7291
3rd Qu.:	0.9664	0.7462	0.7726
Max. :	1.1398	0.9416	0.9941

	Asignación II		
Estadístico	ICP	IAT	IPM
Min. :	0.7132	0.4837	0.4964
1st Qu.:	0.7873	0.5907	0.6131
Median :	0.8075	0.6217	0.6430
Mean :	0.8078	0.6218	0.6425
3rd Qu.:	0.8297	0.6508	0.6728
Max. :	0.9360	0.7834	0.7869

	Asignación I		
Estadístico	ICP	IAT	IPM
Min. :	0.7030	0.4938	0.4958
1st Qu.:	0.7836	0.5907	0.6120
Median :	0.8053	0.6196	0.6399
Mean :	0.8060	0.6191	0.6407
3rd Qu.:	0.8284	0.6469	0.6694
Max. :	0.9083	0.7527	0.7863

	Asignación III		
Estadístico	ICP	IAT	IPM
Min. :	0.7044	0.4837	0.4996
1st Qu.:	0.7915	0.5928	0.6144
Median :	0.8161	0.6260	0.6456
Mean :	0.8144	0.6242	0.6457
3rd Qu.:	0.8375	0.6552	0.6751
Max. :	0.9129	0.7662	0.8107

Gráficamente se tiene:



Así podemos observar que aunque las asignaciones son muy parecidas existen pequeñas diferencias en los resultados presentado menor varianza la asignación I, si realizamos la comparación con respecto a la actual, la distribución en cualquiera de las asignaciones presenta menor variación concluyendo que si incrementamos la muestra en el estrato 4 y la disminuimos en el estrato 2 contribuimos a la minimizar la varianza.

En consecuencia con los resultados y como parte de las aportaciones para las mejoras del diseño, se sugiere la aplicación de éste método y la incorporación de las variables de interés tomando la de mayor varianza ya que con ella se cubriría el resto.

Sin dejar de lado el tamaño de muestra, en el siguiente capítulo se presenta una serie de ejercicios diferenciados por la variable de interés involucrada en dicho cálculo.

CAPITULO 4.- TAMAÑO DE MUESTRA

Como se mostró en el capítulo III los parámetros utilizados para el cálculo del tamaño de muestra son: confianza del 95%; error relativo del 4%, tasa de no Respuesta del 20% y un efecto de diseño de 0.96 obteniendo un tamaño de muestra de 114 empresas sin incluir las empresas con certeza. Es por ello que para verificar los parámetros se llevó a cabo los siguientes ejercicios, pudiendo con ello tener los elementos para sugerir la disminución de la muestra.

I. Error Relativo

Para verificar que el error relativo de las estimaciones es a los más el 4% como se especificó en el diseño, se simulan 1,000 muestras y se calculan las estimaciones de los índices y sus diferencias con respecto a la publicación realizada por INEGI para el mes de Junio 2009, con lo cual se obtuvo un valor máximo mayor al 4%, pero observamos que el tercer cuartil para los tres índices es menor al 3%, como se presenta en la siguiente tabla,

	Error		
Estadístico	ICP	IAT	IPM
Min. :	-3.79126	-4.7164	-2.595
1st Qu.:	-1.02432	-0.2087	0.898
Median :	0.054152	0.922	1.916
Mean :	-0.00131	0.8977	1.861
3rd Qu.:	0.970675	1.9756	2.831
Max. :	4.087255	5.127	5.798

Analizando la información se tiene que el porcentaje de casos que reflejan un error mayor al 4% es:

Porcentaje de casos cuyo error es mayor al 4% estimado		
ICP	IAT	IPM
0.1	2.5	6.2

En general, se puede confiar que en el 94% de los casos el porcentaje de error estimado se cumple, sin embargo no se tienen elementos suficientes para poder disminuir éste porcentaje de error en el cálculo del tamaño de muestra.

II. Cálculo de Tamaño de Muestra

Para el cálculo de tamaño de muestra ejemplificado en ésta sección se aplica la misma fórmula del diseño (Ver capítulo II), utilizando los siguientes parámetros:

Parámetros para el cálculo de Tamaño de Muestra			
Índice	Total	Varianza	Deff
ICP*	0.49644	0.01	0.96
Min(iat)	0.392512	0.010421	0.893381
Min(ipm)	0.430651	0.009253	0.917421

*Con los parámetros del diseño actual

Obteniendo los siguientes tamaños de muestra:

n	Error					
	0.01	0.02	0.03	0.04	0.05	0.06
ICP*	1277	419	198	114	73	51
IAT	1687	615	299	174	113	79
IPM	1422	484	230	133	86	60

*Con los parámetros del diseño actual

Entonces para un error del 4% tenemos una muestra total de:

Indicador	Error	Muestra
	0.04	Total
ICP	114	511
IAT	174	571
IPM	133	530

Ahora bien, bajo el mismo error se aplica el método de optimización BFSG que proporciona una mejor distribución de la muestra, distribuyéndose de la siguiente forma:

Distribución por Estrato Optimizando con BFGS					
	Estrato				
Parámetro	1	2	3	4	Total
N	397	599	1008	2423	4427
ICP	397	41	35	38	511
IAT	397	25	55	94	571
IPM	397	18	38	77	530

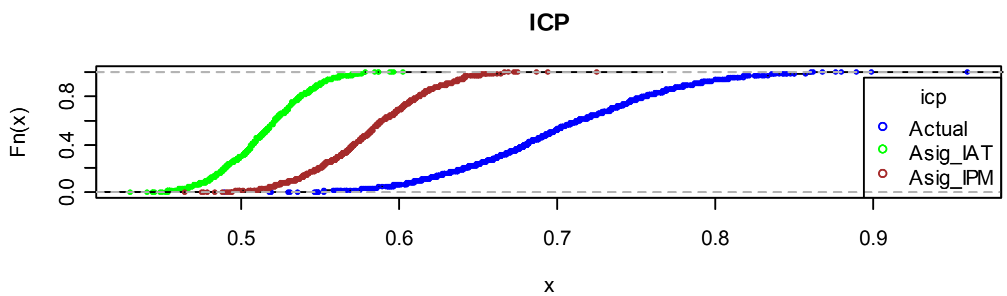
Simulando 1,000 muestras se obtuvieron los siguientes estadísticos de las varianzas observadas y los histogramas presentados en el anexo C:

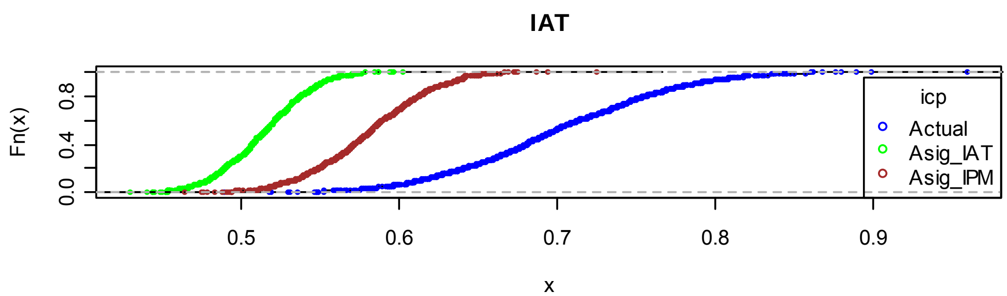
	Asignación Actual		
Estadístico	ICP	IAT	IPM
Min. :	0.7155	0.4520	0.5192
1st Qu.:	0.8950	0.6558	0.6833
Median :	0.9303	0.6969	0.7251
Mean :	0.9301	0.6998	0.7291
3rd Qu.:	0.9664	0.7462	0.7726
Max. :	1.1398	0.9416	0.9941

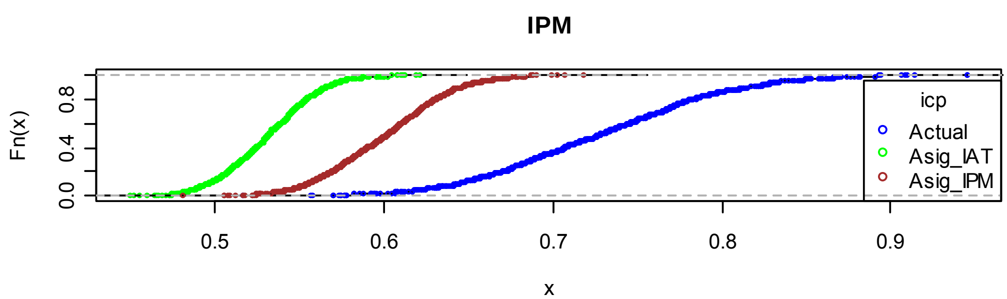
	Asignación para IAT		
Estadístico	ICP	IAT	IPM
Min. :	0.5994	0.4362	0.4503
1st Qu.:	0.6499	0.4970	0.5132
Median :	0.6628	0.5144	0.5319
Mean :	0.6634	0.5152	0.5314
3rd Qu.:	0.6766	0.5344	0.5499
Max. :	0.7445	0.6033	0.6083

	Asignación para IPM		
Estadístico	**ICP**	**IAT**	**IPM**
Min. :	0.6559	0.4591	0.4809
1st Qu.:	0.7299	0.5566	0.5759
Median :	0.7487	0.5809	0.5998
Mean :	0.7489	0.5811	0.5995
3rd Qu.:	0.7703	0.6058	0.6246
Max. :	0.8394	0.6905	0.6972

Con lo cual observamos que la media más alta se percibe en la muestra actual, también la distancia entre el máximo y mínimo dentro de cada asignación es mayor para éste mismo caso, sin embargo como ya lo hemos analizado antes esto se debe a la distribución de la muestra, para un análisis general se presentan las siguientes gráficas:







En resumen, al incrementar el tamaño de muestra no es significativa la diferencia de la varianza, con lo que podemos concluir que el tamaño de muestra actual es suficiente para la estimación a nivel nacional, sin embargo su distribución en los diferentes estratos sí influye en el incremento del error para los diferentes indicadores, por lo que la modificación de ésta puede contribuir a una mejor precisión.

CONCLUSION

1. Es claro que una de las preocupaciones principales en la planeación de una encuesta por muestreo probabilístico se orienta a determinar con certeza el número de observaciones necesarias para poder garantizar la precisión y confianza requerida en los estimadores calculados. En este sentido, es importante destacar que las organizaciones ejecutoras están conscientes que una mala decisión puede limitar la validez de los resultados y en ocasiones llegar a invalidar la investigación completa.

2. Las encuestas que se realizan en las oficinas estadísticas internacionales se pueden considerar de propósitos múltiples. Sin embargo, es práctica común que la determinación del tamaño de muestra se efectúe como si sólo se desearan obtener estimaciones para una sola variable. Posteriormente, se forman dominios de estudios y se analiza la información pudiéndola evaluar mediante simulación a fin de analizar si las restricciones impuestas por el tamaño de muestra conducen a la pérdida de confiabilidad en las relaciones de causalidad observadas en subpoblaciones de interés y de ésta manera formular recomendaciones de política con validez estadística.

3. Para el caso de una sola variable la literatura documenta cómo determinar el tamaño de muestra cuando se desea estimar una proporción binomial y se dispone de un valor que aproxime su varianza. Sin embargo, en la práctica sucede frecuentemente que las variables que se desean estimar tienen una distribución multinomial, y los procedimientos que habitualmente se aplican son métodos de remuestreo, sin embargo como una aportación se afina el método directo de la varianza para cada uno de los indicadores siendo ésta el insumo fundamental tanto para el tamaño de muestra como para su asignación.

4. Una vez que se tiene la expresión de la varianza se han desarrollado procedimientos matemáticos que permiten explorar alternativas y generan soluciones al problema de determinar el tamaño de óptimo de muestra y su asignación en los diferentes estratos en que se divide la población. Concluyendo que en ocasiones el problema puede encontrarse en la distribución de la muestra dentro de los estratos como se observó al momento de distribuir el mismo tamaño de muestra diferenciando el esquema de asignación, provocando una disminución de la varianza para los tres indicadores.

5. Por lo tanto, la única manera de saber si se cumplieron con las expectativas de confiabilidad de los estimadores definidas en el diseño de la muestra se logra a partir del cálculo de los errores de muestreo, tanto para la variable principal como para todos aquellos indicadores que se analizarán a partir de la información generada.

6. Para lograr la elaboración de diseños de muestra óptimos es necesario disponer de información sobre la varianza de los estimadores.

7. Finalmente, como resultado de la evaluación del diseño muestral de la Encuesta Mensual de Opinión Empresarial relacionado con la muestra tradicional se puede concluir que,

    a. El error determinado al momento del cálculo de la muestra se encuentra dentro del intervalo generado con la información estimada.

b. Con éste error se tiene que el tamaño de muestra de 114 empresas es suficiente para realizar las estimaciones nacionales[9].
c. Sin embargo, la asignación de la muestra para los estratos se puede mejorar aplicando el método quasi-Newton (BFGS) cuya distribución disminuye la varianza de los tres indicadores estimados.

8. Acabaremos con decir, que fruto de esta tesis, ahora se cuenta con el CODIGO para realizar los cálculos de los indicados como los que son publicados mensualmente por el INEGI producto de los primeros 3 programas presentados en el anexo D.

[9] Sin contar las 397 empresas del estrato 1 que entran con certeza en la muestra.

Anexo A

Cálculo de la varianza, Desviación Estándar e Intervalo de Confianza
<u>Índice de Confianza del Productor</u>
Serie Mensual 2005-2009

Mes	Var_Ind	SD_Ind
2501 Ene-05	9.85E-05	0.00992265
2502 Feb-05	9.91E-05	0.00995698
2503 Mar-05	9.72E-05	0.00985674
2504 Abr-05	9.97E-05	0.00998265
2505 May-05	1.02E-04	0.01009133
2506 Jun-05	9.74E-05	0.00986998
2507 Jul-05	9.13E-05	0.00955581
2508 Ago-05	9.69E-05	0.00984141
2509 Sep-05	9.44E-05	0.00971782
2510 Oct-05	9.14E-05	0.00955802
2511 Nov-05	9.82E-05	0.00991199
2512 Dic-05	9.06E-05	0.00952083
2601 Ene-06	8.41E-05	0.00917068
2602 Feb-06	9.25E-05	0.00961639
2603 Mar-06	9.80E-05	0.00990076
2604 Abr-06	9.69E-05	0.00984346
2605 May-06	9.96E-05	0.00997956
2606 Jun-06	1.02E-04	0.01007912
2607 Jul-06	8.96E-05	0.00946639
2608 Ago-06	1.09E-04	0.01041922
2609 Sep-06	9.73E-05	0.0098643
2610 Oct-06	9.35E-05	0.0096671
2611 Nov-06	1.03E-04	0.01014072
2701 Ene-07	8.98E-05	0.00947439
2702 Feb-07	9.51E-05	0.00975076
2703 Mar-07	9.22E-05	0.00960231
2704 Abr-07	8.87E-05	0.00941965
2705 May-07	7.71E-05	0.00877952
2706 Jun-07	8.83E-05	0.00939813
2707 Jul-07	9.43E-05	0.00971067
2708 Ago-07	8.29E-05	0.0091036
2709 Sep-07	8.99E-05	0.00948049
2710 Oct-07	9.46E-05	0.00972615
2711 Nov-07	9.18E-05	0.00958141
2712 Dic-07	7.88E-05	0.00887623
2801 Ene-08	8.66E-05	0.0093035
2802 Feb-08	8.89E-05	0.00942804
2803 Mar-08	8.46E-05	0.00919669

Mes	Var_Ind	SD_Ind
2804 Abr-08	7.60E-05	0.00871941
2805 May-08	7.81E-05	0.00883786
2806 Jun-08	8.29E-05	0.00910711
2807 Jul-08	8.32E-05	0.00911954
2808 Ago-08	7.57E-05	0.00870316
2809 Sep-08	7.42E-05	0.0086146
2810 Oct-08	6.94E-05	0.00833136
2811 Nov-08	8.40E-05	0.0091626
2812 Dic-08	8.75E-05	0.00935209
2901 Ene-09	7.84E-05	0.00885535
2902 Feb-09	6.83E-05	0.00826441
2903 Mar-09	7.43E-05	0.00861787
2904 Abr-09	5.87E-05	0.00766349
2905 May-09	9.47E-05	0.00973207
2906 Jun-09	7.48E-05	0.00864981
2907 Jul-09	8.58E-05	0.0092649
2908 Ago-09	9.54E-05	0.00976696

Cálculo de la varianza, Desviación Estándar e Intervalo de Confianza
<u>Índice de Pedidos Manufactureros</u>
Serie Mensual 2005-2009

Mes	Var_Ind	SD_Ind
2501 Ene-05	4.64E-05	0.00680988
2502 Feb-05	6.15E-05	0.00784216
2503 Mar-05	4.10E-05	0.00640696
2504 Abr-05	5.32E-05	0.0072911
2505 May-05	3.29E-05	0.00573202
2506 Jun-05	3.59E-05	0.00599123
2507 Jul-05	4.36E-05	0.00660492
2508 Ago-05	3.62E-05	0.00601411
2509 Sep-05	3.92E-05	0.00626175
2510 Oct-05	3.99E-05	0.00631516
2511 Nov-05	4.20E-05	0.00647779
2512 Dic-05	5.25E-05	0.00724627
2601 Ene-06	4.79E-05	0.00692106
2602 Feb-06	4.28E-05	0.00654023
2603 Mar-06	4.63E-05	0.00680642
2604 Abr-06	6.42E-05	0.0080117
2605 May-06	6.42E-05	0.00801431
2606 Jun-06	5.08E-05	0.00712866
2607 Jul-06	5.26E-05	0.00725208
2608 Ago-06	5.35E-05	0.00731626

Mes	Var_Ind	SD_Ind
2610 Sep-06	6.99E-05	0.00835847
2610 Oct-06	6.61E-05	0.00813264
2611 Nov-06	4.62E-05	0.00679663
2701 Ene-07	7.84E-05	0.00885513
2702 Feb-07	5.59E-05	0.00747537
2703 Mar-07	3.83E-05	0.00618583
2704 Abr-07	5.82E-05	0.00762976
2705 May-07	4.02E-05	0.00634251
2706 Jun-07	3.97E-05	0.00629957
2707 Jul-07	4.61E-05	0.00678618
2708 Ago-07	6.29E-05	0.00793314
2709 Sep-07	3.96E-05	0.00629378
2710 Oct-07	6.43E-05	0.00801641
2711 Nov-07	4.62E-05	0.00679813
2712 Dic-07	5.21E-05	0.00722144
2801 Ene-08	6.03E-05	0.00776315
2802 Feb-08	6.29E-05	0.00793173
2803 Mar-08	6.45E-05	0.00803053
2804 Abr-08	4.59E-05	0.00677461
2805 May-08	4.43E-05	0.00665676
2806 Jun-08	5.08E-05	0.00712415
2807 Jul-08	5.63E-05	0.00750296
2808 Ago-08	6.11E-05	0.00781786
2809 Sep-08	4.96E-05	0.00704388
2810 Oct-08	6.72E-05	0.00819964
2811 Nov-08	6.85E-05	0.00827796
2812 Dic-08	1.09E-04	0.01042111
2901 Ene-09	7.09E-05	0.00841804
2902 Feb-09	5.09E-05	0.00713108
2903 Mar-09	4.82E-05	0.00694153
2904 Abr-09	5.08E-05	0.00712596
2905 May-09	6.75E-05	0.00821766
2906 Jun-09	4.64E-05	0.00680844
2907 Jul-09	3.90E-05	0.00624488
2908 Ago-09	4.24E-05	0.00651381

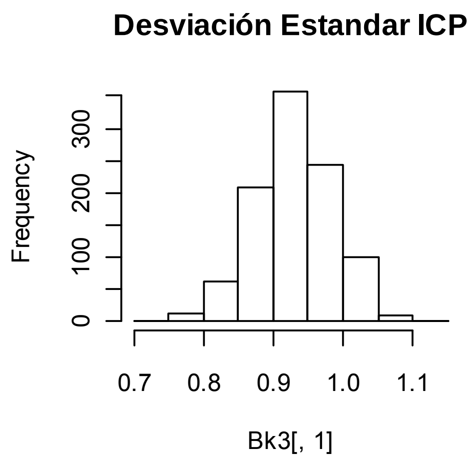
Cálculo de la varianza, Desviación Estándar e Intervalo de Confianza
<u>Índice de Agregado de Tendencia</u>
Serie Mensual 2005-2009

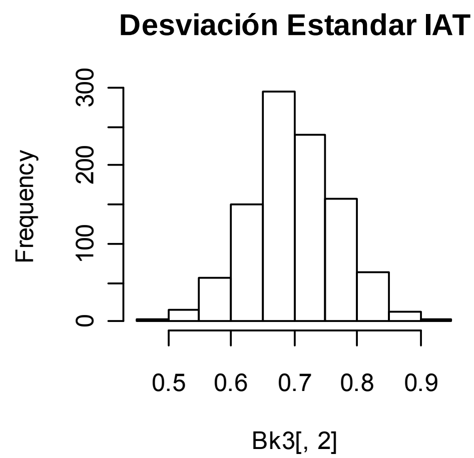
Mes	Var_Ind	SD_Ind
2501 Ene-05	4.95E-05	0.00703755
2502 Feb-05	5.57E-05	0.00746028
2503 Mar-05	3.52E-05	0.00592921
2504 Abr-05	4.52E-05	0.00672407
2505 May-05	3.12E-05	0.0055826
2506 Jun-05	3.43E-05	0.00586019
2507 Jul-05	3.17E-05	0.00562632
2508 Ago-05	4.24E-05	0.00651445
2509 Sep-05	3.74E-05	0.00611747
2510 Oct-05	3.29E-05	0.00573569
2511 Nov-05	4.07E-05	0.00638128
2512 Dic-05	6.54E-05	0.00808414
2601 Ene-06	4.19E-05	0.00647162
2602 Feb-06	3.55E-05	0.00595626
2603 Mar-06	3.42E-05	0.00584979
2604 Abr-06	5.66E-05	0.00752093
2605 May-06	4.47E-05	0.00668686
2606 Jun-06	3.64E-05	0.00602949
2607 Jul-06	5.65E-05	0.00751707
2608 Ago-06	3.79E-05	0.00615637
2609 Sep-06	4.75E-05	0.0068894
2610 Oct-06	6.00E-05	0.00774809
2611 Nov-06	3.70E-05	0.00608436
2701 Ene-07	5.98E-05	0.0077339
2702 Feb-07	3.85E-05	0.00620663
2703 Mar-07	3.12E-05	0.00558745
2704 Abr-07	4.99E-05	0.00706656
2705 May-07	4.15E-05	0.00644117
2706 Jun-07	4.09E-05	0.00639182
2707 Jul-07	5.87E-05	0.00766216
2708 Ago-07	4.53E-05	0.00672845
2709 Sep-07	3.22E-05	0.00567609
2710 Oct-07	6.53E-05	0.00808139
2711 Nov-07	5.38E-05	0.00733338
2712 Dic-07	6.98E-05	0.00835246
2801 Ene-08	5.46E-05	0.00738585
2802 Feb-08	5.10E-05	0.00714434
2803 Mar-08	5.49E-05	0.00741047
2804 Abr-08	4.11E-05	0.00641283

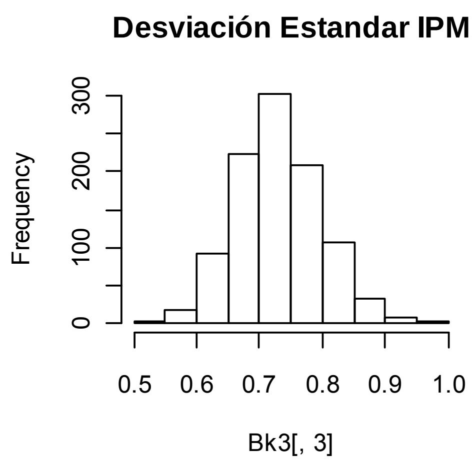
Mes	Var_Ind	SD_Ind
2805 May-08	5.40E-05	0.00734779
2806 Jun-08	4.62E-05	0.00679815
2807 Jul-08	4.28E-05	0.00654252
2808 Ago-08	4.90E-05	0.0070018
2809 Sep-08	4.28E-05	0.00654335
2810 Oct-08	6.50E-05	0.00806041
2811 Nov-08	6.89E-05	0.00830187
2812 Dic-08	8.35E-05	0.00914007
2901 Ene-09	5.87E-05	0.0076625
2902 Feb-09	4.01E-05	0.00633276
2903 Mar-09	8.56E-05	0.00925284
2904 Abr-09	4.86E-05	0.0069739
2905 May-09	7.18E-05	0.00847526
2906 Jun-09	3.49E-05	0.00590427
2907 Jul-09	4.30E-05	0.00655763
2908 Ago-09	2.95E-05	0.00543065

Anexo B

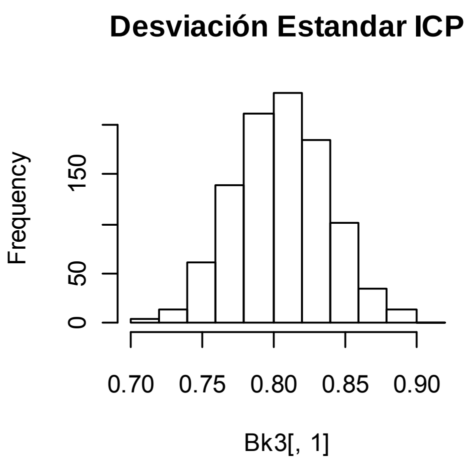
<u>Histograma de las 1,000 muestras simuladas utilizando como parámetro el ICP (Asignación Actual)</u>

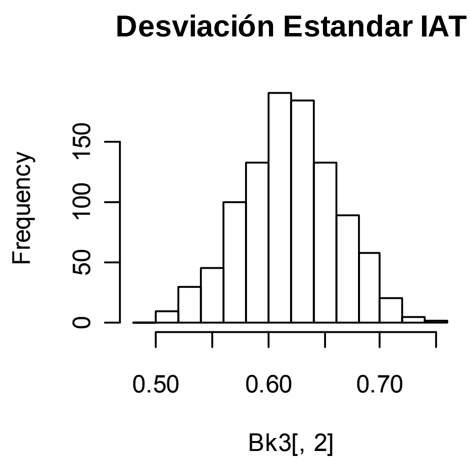


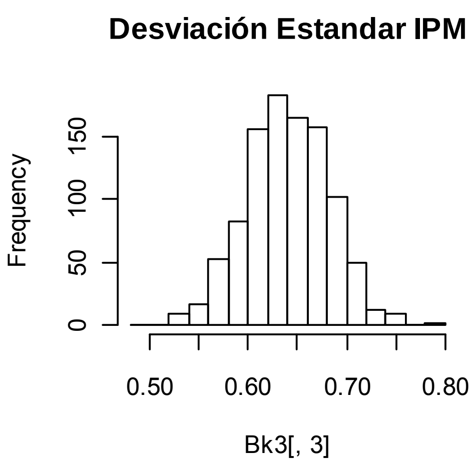




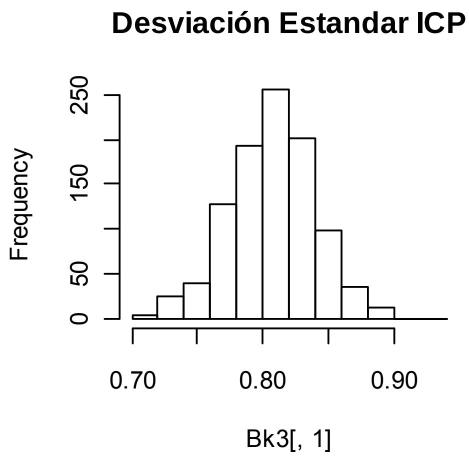
Histograma de las 1,000 muestras simuladas utilizando como parámetro el ICP (Asignación I) y la optimización de la asignación de la muestra mediante el método BFGS

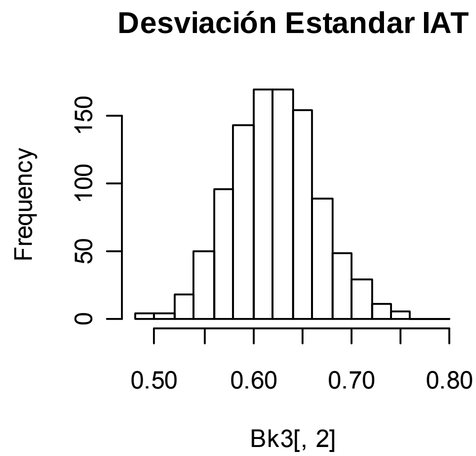


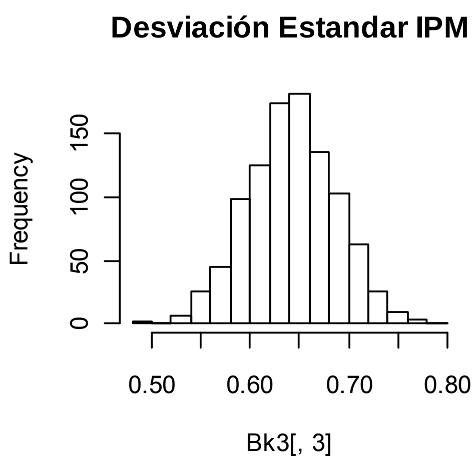




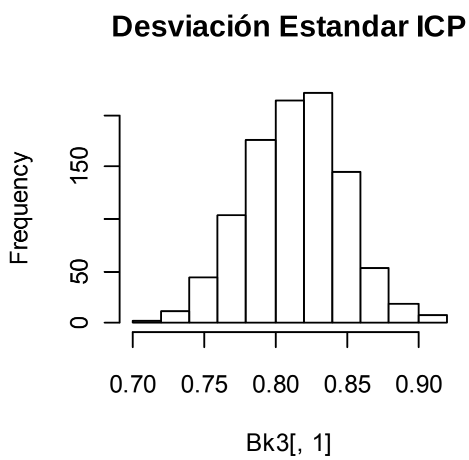
Histograma de las 1,000 muestras simuladas utilizando como parámetro el IPM (Asignación II) y la optimización de la asignación de la muestra mediante el método BFGS

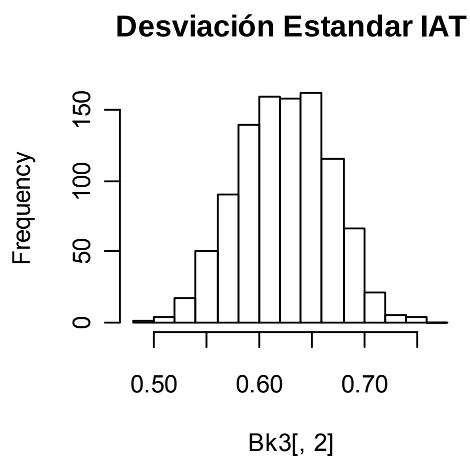


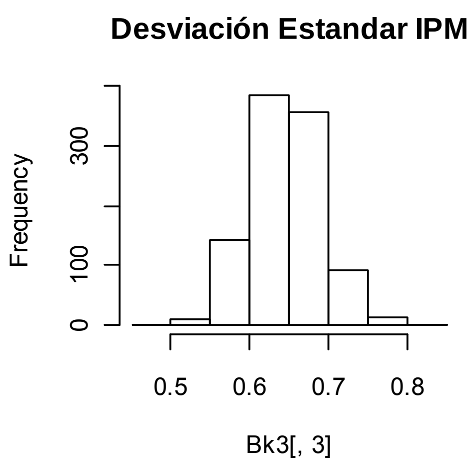




Histograma de las 1,000 muestras simuladas utilizando como parámetro el IAT (Asignación III) y la optimización de la asignación de la muestra mediante el método BFGS

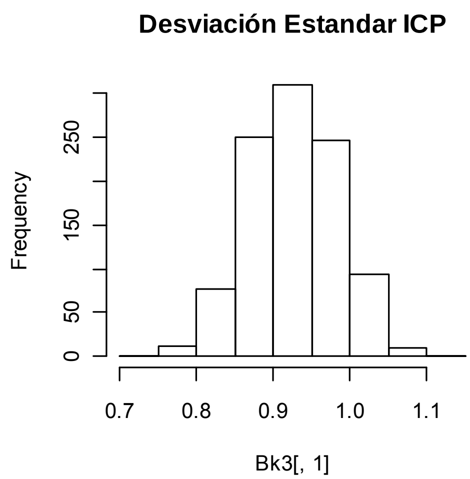


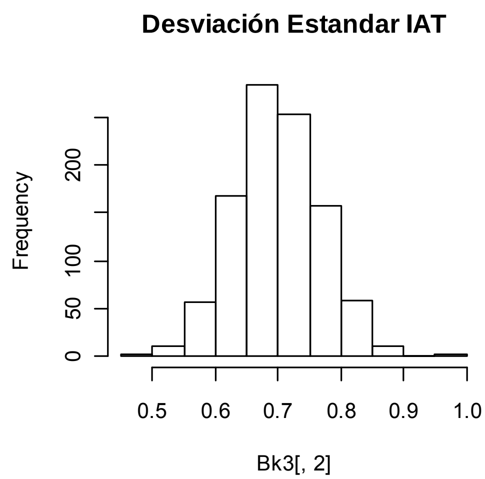


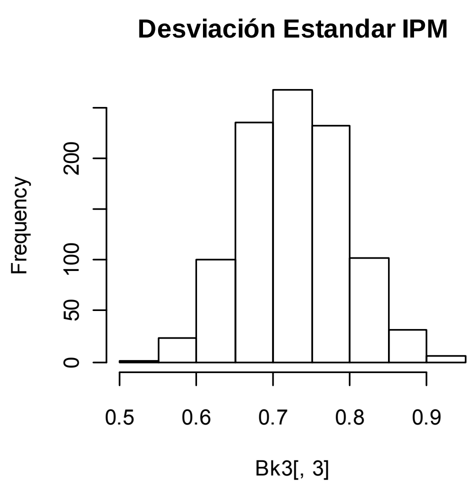


Anexo C

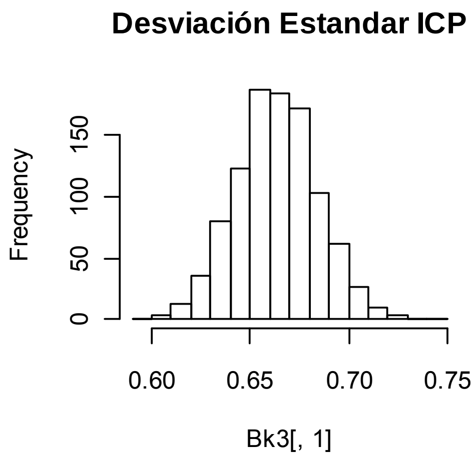
Histograma de las 1,000 muestras simuladas utilizando como parámetro el ICP para muestras de 511 empresas y la optimización de la asignación de la muestra mediante el método BFGS

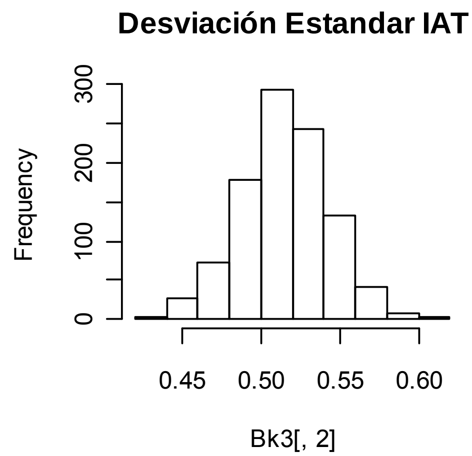


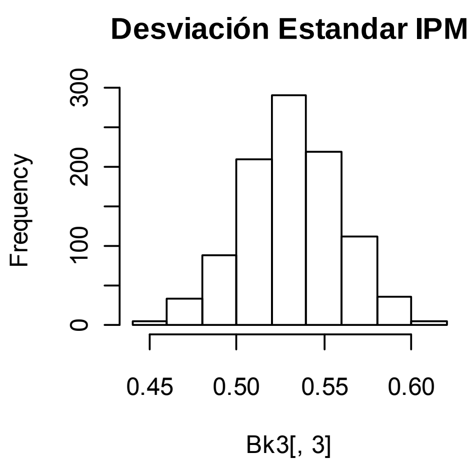




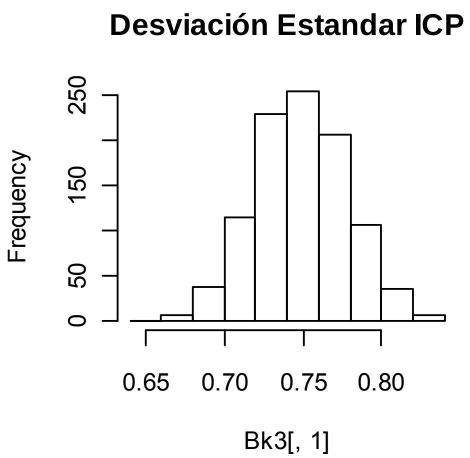
Histograma de las 1,000 muestras simuladas utilizando como parámetro el IAT para muestras de 571 empresas y la optimización de la asignación de la muestra mediante el método BFGS

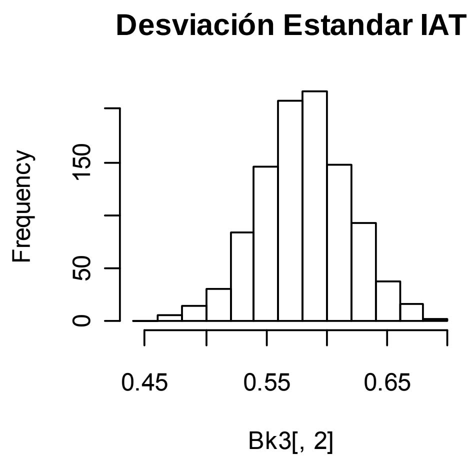


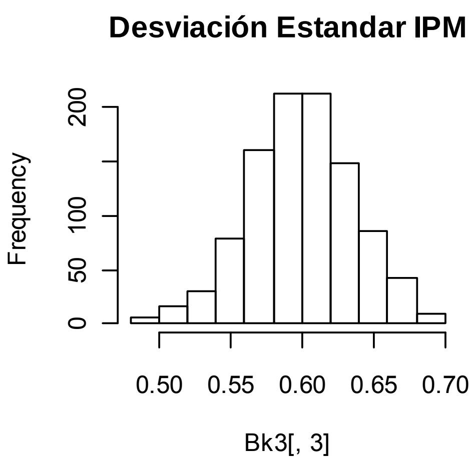




Histograma de las 1,000 muestras simuladas utilizando como parámetro el IPM para muestras de 530 empresas y la optimización de la asignación de la muestra mediante el método BFGS







Anexo D
Código de Programas en R-project

¡)Cálculo de IPC y su varianza para la serie 2005-2009

```
nICP<-with(bases,{
preg1<-xtabs(~ESTRATO + I5_1 + MES, data = bases)[,as.character(c(2:3)),]
preg2<-xtabs(~ESTRATO + I5_2 + MES , data = bases)[,as.character(c(2:3)),]
preg3<-xtabs(~ESTRATO + S8_1 + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(~ESTRATO + S8_2 + MES , data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(~ESTRATO + S8_3 + MES , data = bases)[,as.character(c(1:5)),]
preg6<-xtabs(~ESTRATO + S8_4 + MES, data = bases)[,as.character(c(1:5)),]
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

pICP<-with(bases,{
preg1<-xtabs(FACTOR~ESTRATO + I5_1 + MES, data = bases)[,as.character(c(2:3)),]
preg2<-xtabs(FACTOR~ESTRATO + I5_2 + MES , data = bases)[,as.character(c(2:3)),]
preg3<-xtabs(FACTOR~ESTRATO + S8_1 + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(FACTOR~ESTRATO + S8_2 + MES , data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(FACTOR~ESTRATO + S8_3 + MES , data = bases)[,as.character(c(1:5)),]
preg6<-xtabs(FACTOR~ESTRATO + S8_4 + MES, data = bases)[,as.character(c(1:5)),]
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

prop.pICP<-pICP
for (i in 1:length(pICP))
   prop.pICP[[i]]<-prop.table(pICP[[i]],c(1,3))

PESOS<- list(c(0,1),
                      c(0,1),
                      c(0,0.25,0.5,0.75,1.0),
                      c(0,0.25,0.5,0.75,1.0),
                      c(0,0.25,0.5,0.75,1.0),
                      c(0,0.25,0.5,0.75,1.0))


ICP<-numeric()
for (iii in  levels(as.factor(bases$MES)))
{
cat(iii,"   ")

PREG<-list(pICP[[1]][,,iii],pICP[[2]][,,iii],pICP[[3]][,,iii],pICP[[4]][,,iii],pICP[[5]][,,iii],pICP[[6]][,,iii])

TNORESP<- PORC <- vector("list",6)

#Datos Poblacionales

        Ni <- c(397,600,1008,2422)
        ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
        Wi <-prop.table(Ni)

#Ponderacion
IND <- VAR <- rep(0,6)

for(ii in 1:6)
{
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(nICP[[ii]][,,iii] )
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi

        fi       <- Ni/nki
```

```
		porcentaje <- prop.pICP[[ii]][,,iii]  #prop.table(pregunta,1)
		Bk        <- porcentaje %*%a #Suma Bki
		Dk        <- t(Wi)%*%Bk   #sum(Wi*Bk)

		IND[ii] <- Dk

		PORC[[ii]] <- porcentaje

#Varianza

		Var_Bki<-(porcentaje*(1-porcentaje)) %*%  (a^2)  ##rowSums((aj^2)*(porcentaje)*(1-porcentaje))

		nr <- nrow(porcentaje)
		nc <- ncol(porcentaje)

		aa <- rep(0,nr)
		for(i in 1:nr)
		for(j in 2:nc)
		for(h in 1:(j-1))
			aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
		Cov_Bki <- aa*2

		Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

		VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100
SD_IND<-sqrt(VAR)


pesos <- c(1/10,1/10,rep(1/5,4))

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
SD_Ind<-sqrt(Var_Ind)
sqrt(Var_Ind)*100
print(Indice)

ICS<-Indice+1.96*SD_Ind
ICI<-Indice-1.96*SD_Ind
unido<-cbind(Indice,Var_Ind,SD_Ind,ICS,ICI)
dimnames(unido)[[1]]<-iii

ICP<-rbind(ICP,unido)
}

##
# Minimizando la muestra en cada estrato, teniendo como fijo el tamaño total de la muestra
##


FunMinVar <- function(nnki)
{

		nki     <- c(nnki,NNN-sum(nnki))

VAR <- rep(0,6)
for(ii in 1:6)
{
# ii <- 1

		if(ii <= 2)
			a       <- c(0,1)
		else
			a       <- c(0,0.25,0.5,0.75,1.0)

		aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

		porcentaje <- PORC[[ii]]
		Bk        <- rowSums(porcentaje*aj) #Suma Bki
		Dk        <- sum(Wi*Bk)
```

```
#Varianza

        Var_Bki<- rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        VAR[ii] <- Var_Bk
}

        pesos   <- c(1/10,1/10,rep(1/5,4))
        VarInd <- sum(pesos*pesos*VAR)
        return(VarInd*1000000)
}
```

ii) Cálculo de IAT y su varianza para la serie 2005-2009

```
nIAT<-with(bases,{
preg1<-xtabs(~ESTRATO + P1_1EST + MES , data = bases)[,as.character(c(1:5)),]
preg2<-xtabs(~ESTRATO + P1_2EST + MES, data = bases)[,as.character(c(1:5)),]
preg3<-xtabs(~ESTRATO + D2_2EST + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(~ESTRATO + D2_4EST + MES, data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(~ESTRATO + P6_1EST + MES, data = bases)[,as.character(c(1:5)),]
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

pIAT<-with(bases,{
preg1<-xtabs(FACTOR*VENTAS~ESTRATO + P1_1EST + MES , data = bases)[,as.character(c(1:5)),]
preg2<-xtabs(FACTOR*ACTIVO~ESTRATO + P1_2EST + MES, data = bases)[,as.character(c(1:5)),]
preg3<-xtabs(FACTOR*VENTAS~ESTRATO + D2_2EST + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(FACTOR*VENTAS~ESTRATO + D2_4EST + MES, data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(FACTOR*PERSONAL~ESTRATO + P6_1EST + MES, data = bases)[,as.character(c(1:5)),]
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

prop.pIAT<-pIAT
for (i in 1:length(pIAT))
  prop.pIAT[[i]]<-prop.table(pIAT[[i]],c(1,3))

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0))


IAT<-numeric()
for (iii in  levels(as.factor(bases$MES)))
{
PREG<-list(pIAT[[1]][,,iii],pIAT[[2]][,,iii],pIAT[[3]][,,iii],pIAT[[4]][,,iii],pIAT[[5]][,,iii])

TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales

        Ni <- c(397,600,1008,2422)
        ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
        Wi <-prop.table(Ni)

#Ponderacion
IND <- VAR <- rep(0,5)
```

```
for(ii in 1:5)
{
	pregunta <- PREG[[ii]]
	a	<- PESOS[[ii]]

	aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

	nki	<- rowSums(nIAT[[ii]][,,iii] )
  nri	<- ndi-nki
  TNORESP[[ii]] <- nri/ndi

	fi	<- Ni/nki
	porcentaje <- prop.pIAT[[ii]][,,iii]  #prop.table(pregunta,1)
	Bk	<- porcentaje %*%a #Suma Bki
	Dk	<- t(Wi)%*%Bk   #sum(Wi*Bk)

	IND[ii] <- Dk

	PORC[[ii]] <- porcentaje

#Varianza

	Var_Bki<-(porcentaje*(1-porcentaje)) %*%  (a^2)  ##rowSums((aj^2)*(porcentaje)*(1-porcentaje))

	nr <- nrow(porcentaje)
	nc <- ncol(porcentaje)

	aa <- rep(0,nr)
	for(i in 1:nr)
	for(j in 2:nc)
	for(h in 1:(j-1))
		aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
	Cov_Bki <- aa*2

	Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

	VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100
SD_IND<-sqrt(VAR)


pesos <- rep(1/5,5)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
SD_Ind<-sqrt(Var_Ind)
sqrt(Var_Ind)*100
print(Indice)

ICS<-Indice+1.96*SD_Ind
ICI<-Indice-1.96*SD_Ind
unido<-cbind(Indice,Var_Ind,SD_Ind,ICS,ICI)
dimnames(unido)[[1]]<-iii
dimnames(unido)[[1]]<-iii

IAT<-rbind(IAT,unido)
}
```

iii) Cálculo de IPM y su varianza para la serie 2005-2009

```
nIPM<-with(bases,{
preg1<-xtabs(~ESTRATO + D2_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg2<-xtabs(~ESTRATO + P1_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg3<-xtabs(~ESTRATO + P6_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(~ESTRATO + I3_2AEST + MES, data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(~ESTRATO + I3_1EST + MES, data = bases)[,as.character(c(1:5)),]
```

```
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

pIPM<-with(bases,{
preg1<-xtabs(FACTOR*VENTAS~ESTRATO + D2_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg2<-xtabs(FACTOR*VENTAS~ESTRATO + P1_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg3<-xtabs(FACTOR*PERSONAL~ESTRATO + P6_1EST + MES, data = bases)[,as.character(c(1:5)),]
preg4<-xtabs(FACTOR*VENTAS~ESTRATO + I3_2AEST + MES, data = bases)[,as.character(c(1:5)),]
preg5<-xtabs(FACTOR*VENTAS~ESTRATO + I3_1EST + MES, data = bases)[,as.character(c(1:5)),]
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

  prop.pIPM<-pIPM
for (i in 1:length(pIPM))
  prop.pIPM[[i]]<-prop.table(pIPM[[i]],c(1,3))

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(1.0,0.75,0.5,0.25,0),
                        c(0,0.25,0.5,0.75,1.0))


IPM<-numeric()
for (iii in  levels(as.factor(bases$MES)))
{
cat(iii,"   ")
PREG<-list(pIPM[[1]][,,iii],pIPM[[2]][,,iii],pIPM[[3]][,,iii],pIPM[[4]][,,iii],pIPM[[5]][,,iii])

TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales

        Ni <- c(397,600,1008,2422)
        ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
        Wi <-prop.table(Ni)

#Ponderacion
IND <- VAR <- rep(0,5)

for(ii in 1:5)
{
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(nIPM[[ii]][,,iii] )
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi

        fi        <- Ni/nki
        porcentaje <- prop.pIPM[[ii]][,,iii]  #prop.table(pregunta,1)
        Bk        <- porcentaje %*%a #Suma Bki
        Dk        <- t(Wi)%*%Bk   #sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-(porcentaje*(1-porcentaje)) %*%  (a^2)  ##rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2
```

```
        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100
SD_IND<-sqrt(VAR)

pesos <- c(.3,.25,.2,.15,.1)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
SD_Ind<-sqrt(Var_Ind)
sqrt(Var_Ind)*100
print(Indice)

ICS<-Indice+1.96*SD_Ind
ICI<-Indice-1.96*SD_Ind
unido<-cbind(Indice,Var_Ind,SD_Ind,ICS,ICI)
dimnames(unido)[[1]]<-iii

IPM<-rbind(IPM,unido)
}
```

iv) Método BFGS para asignación de tamaño de muestra por estrato utilizando la varianza del ICP

```
PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-xtabs(FACTOR~ESTRATO + I5_1 , data = subconjunto)[,as.character(c(2:3))]
preg2<-xtabs(FACTOR~ESTRATO + I5_2 , data = subconjunto)[,as.character(c(2:3))]
preg3<-xtabs(FACTOR~ESTRATO + S8_1 , data = subconjunto)[,as.character(c(1:5))]
preg4<-xtabs(FACTOR~ESTRATO + S8_2 , data = subconjunto)[,as.character(c(1:5))]
preg5<-xtabs(FACTOR~ESTRATO + S8_3 , data = subconjunto)[,as.character(c(1:5))]
preg6<-xtabs(FACTOR~ESTRATO + S8_4 , data = subconjunto)[,as.character(c(1:5))]
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

PESOS<- list(c(0,1),
                c(0,1),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0))


TNORESP<- PORC <- vector("list",6)

#Datos Poblacionales

        Ni <- c(397,600,1008,2422)
        ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,6)

for(ii in 1:6)
{
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(pregunta)
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi
```

```
#  print(nri)
#          print(c(sum(nki),nki))

           fi       <- Ni/nki
           preg_c    <- cbind(pregunta,nki,Ni,fi)
           porcentaje <- as.matrix(pregunta/nki)
           Bk       <- rowSums(porcentaje*aj) #Suma Bki
           Dk       <- sum(Wi*Bk)

           IND[ii] <- Dk

           PORC[[ii]] <- porcentaje

#Varianza

           Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

           nr <- nrow(porcentaje)
           nc <- ncol(porcentaje)

           aa <- rep(0,nr)
           for(i in 1:nr)
           for(j in 2:nc)
           for(h in 1:(j-1))
                      aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
           Cov_Bki <- aa*2

           Var_Bk <- sum( ((Var_Bki-Cov_Bki)/ndi) * Wi*Wi )

           VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100

pesos <- c(1/10,1/10,rep(1/5,4))

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100

##
# Minimizando la muestra en cada estrado, teniendo como fijo el tamaño total de la muestra
##


FunMinVar <- function(nnki)
{
           Ni <- c(397,600,1008,2422)
#          ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
           Wi <- Ni/N


#          ndi    <- c(nnki,NNN-sum(nnki))
           ndi    <- c(397,nnki,NNN-397-sum(nnki))

# Este seria el inicio del calculo del ICP, en lugar de leer los archivos CSV
PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-xtabs(FACTOR~ESTRATO + I5_1 , data = subconjunto)[,as.character(c(2:3))]
preg2<-xtabs(FACTOR~ESTRATO + I5_2 , data = subconjunto)[,as.character(c(2:3))]
preg3<-xtabs(FACTOR~ESTRATO + S8_1 , data = subconjunto)[,as.character(c(1:5))]
preg4<-xtabs(FACTOR~ESTRATO + S8_2 , data = subconjunto)[,as.character(c(1:5))]
preg5<-xtabs(FACTOR~ESTRATO + S8_3 , data = subconjunto)[,as.character(c(1:5))]
preg6<-xtabs(FACTOR~ESTRATO + S8_4 , data = subconjunto)[,as.character(c(1:5))]
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

PESOS<- list(c(0,1),
                      c(0,1),
                      c(0,0.25,0.5,0.75,1.0),
                      c(0,0.25,0.5,0.75,1.0),
                      c(0,0.25,0.5,0.75,1.0),
```

```
                  c(0,0.25,0.5,0.75,1.0))


TNORESP<- PORC <- vector("list",6)


#Ponderacion
IND <- VAR <- rep(0,6)

for(ii in 1:6)
{
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(pregunta)
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi

        fi        <- Ni/nki
        preg_c     <- cbind(pregunta,nki,Ni,fi)
        porcentaje <- as.matrix(pregunta/nki)
        Bk         <- rowSums(porcentaje*aj) #Suma Bki
        Dk         <- sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( ((Var_Bki-Cov_Bki)/ndi) * Wi*Wi )

        VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100

# Proceso 2
pesos <- c(1/10,1/10,rep(1/5,4))

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100
        return(Var_Ind*1000000)
}


Met <- c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN")

NNN  <- 511
nnki <- c(20,20)

FunMinVar(nnki)

#Funcion de optimizazcion continua para el tamanio de muestra
#toma como maximo la poblacion (Ni) y los minimos son introducidos(lower=...)
```

```
sal  <- optim(nnki,FunMinVar,method=Met[4],lower=c(10,10),upper=Ni[2:3])

kk <- c(397,sal$par,NNN-397-sum(sal$par))
kk
sum(kk)
```

v) Método BFGS para asignación de tamaño de muestra por estrato utilizando la varianza del IAT

```
PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-xtabs(FACTOR*VENTAS~ESTRATO + P1_1EST , data = subconjunto)[,as.character(c(1:5))]
preg2<-xtabs(FACTOR*ACTIVO~ESTRATO + P1_2EST , data = subconjunto)[,as.character(c(1:5))]
preg3<-xtabs(FACTOR*VENTAS~ESTRATO + D2_2EST , data = subconjunto)[,as.character(c(1:5))]
preg4<-xtabs(FACTOR*VENTAS~ESTRATO + D2_4EST , data = subconjunto)[,as.character(c(1:5))]
preg5<-xtabs(FACTOR*PERSONAL~ESTRATO + P6_1EST , data = subconjunto)[,as.character(c(1:5))]
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0))

PORC <- vector("list",5)

#Datos Poblacionales

        Ni <- c(397,599,1008,2423)
  ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

for(ii in 1:5)
{
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(pregunta)

        porcentaje <- as.matrix(pregunta/nki)
        Bk        <- porcentaje%*%a
        Dk        <- sum(Wi*Bk)

        IND[ii] <- Dk
        PORC[[ii]] <- porcentaje

#Varianza
        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))
        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)
        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        VAR[ii] <- Var_Bk
}
```

```
IND*100
sqrt(VAR)*100

pesos <- rep(1/5,5)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100


##
# Minimizando la muestra en cada estrado, teniendo como fijo el tamaño total de la muestra
##


FunMinVar <- function(nnki)
{
	Ni <- c(397,600,1008,2422)
  N  <- sum(Ni)
	Wi <- Ni/N


	ndi     <- c(397,nnki,NNN-397-sum(nnki))

PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-xtabs(FACTOR*VENTAS~ESTRATO + P1_1EST , data = subconjunto)[,as.character(c(1:5))]
preg2<-xtabs(FACTOR*ACTIVO~ESTRATO + P1_2EST , data = subconjunto)[,as.character(c(1:5))]
preg3<-xtabs(FACTOR*VENTAS~ESTRATO + D2_2EST , data = subconjunto)[,as.character(c(1:5))]
preg4<-xtabs(FACTOR*VENTAS~ESTRATO + D2_4EST , data = subconjunto)[,as.character(c(1:5))]
preg5<-xtabs(FACTOR*PERSONAL~ESTRATO + P6_1EST , data = subconjunto)[,as.character(c(1:5))]
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
		c(0,0.25,0.5,0.75,1.0),
		c(0,0.25,0.5,0.75,1.0),
		c(0,0.25,0.5,0.75,1.0),
		c(0,0.25,0.5,0.75,1.0))

PORC <- vector("list",5)


#Ponderacion
IND <- VAR <- rep(0,5)

for(ii in 1:5)
{
	pregunta <- PREG[[ii]]
	a        <- PESOS[[ii]]

	aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

	nki        <- rowSums(pregunta)

	porcentaje <- as.matrix(pregunta/nki)
	Bk         <- porcentaje%*%a
	Dk         <- sum(Wi*Bk)

	IND[ii] <- Dk
	PORC[[ii]] <- porcentaje

#Varianza
	Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))
	nr <- nrow(porcentaje)
	nc <- ncol(porcentaje)
	aa <- rep(0,nr)
	for(i in 1:nr)
	for(j in 2:nc)
	for(h in 1:(j-1))
		aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
	Cov_Bki <- aa*2
```

```
            Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

            VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100


# Proceso 2
pesos <- rep(1/5,5)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100

            return(Var_Ind*1000000)
}


Met <- c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN")

NNN  <- 511
nnki <- c(20,20)

FunMinVar(nnki)

#Funcion de optimizazcion continua para el tamanio de muestra
#toma como maximo la poblacion (Ni) y los minimos son introducidos(lower=...)

sal  <- optim(nnki,FunMinVar,method=Met[4],lower=c(10,10),upper=Ni[2:3])

kk <- c(397,sal$par,NNN-397-sum(sal$par))
kk
sum(kk)
FunMinVar(sal$par)
```

vi) Método BFGS para asignación de tamaño de muestra por estrato utilizando la varianza del IPM

```
PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & D2_1EST==j ]*VENTAS[ESTRATO==i & D2_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & I3_2AEST==j ]*VENTAS[ESTRATO==i & I3_2AEST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & I3_1EST==j ]*VENTAS[ESTRATO==i & I3_1EST==j ])


PREG <- list(preg1,preg2,preg3,preg4,preg5)
```

```
})


PESOS<- list(c(0,0.25,0.5,0.75,1.0),
              c(0,0.25,0.5,0.75,1.0),
              c(0,0.25,0.5,0.75,1.0),
              c(0,0.25,0.5,0.75,1.0),
              c(0,0.25,0.5,0.75,1.0))


PORC <- vector("list",5)

#Datos Poblacionales

       Ni <- c(397,599,1008,2423)
  ndi <- c(397, 41, 35, 38)
       N  <- sum(Ni)
       Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

for(ii in 1:5)
{
       pregunta <- PREG[[ii]]
       a        <- PESOS[[ii]]

       aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

       nki       <- rowSums(pregunta)


       fi        <- Ni/nki
       preg_c     <- cbind(pregunta,nki,Ni,fi)
       porcentaje <- as.matrix(pregunta/nki)
       Bk         <- rowSums(porcentaje*aj) #Suma Bki     porcentaje%*%a
       Dk         <- sum(Wi*Bk)

       IND[ii] <- Dk

       PORC[[ii]] <- porcentaje

#Varianza

       Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

       nr <- nrow(porcentaje)
       nc <- ncol(porcentaje)

       aa <- rep(0,nr)
       for(i in 1:nr)
       for(j in 2:nc)
       for(h in 1:(j-1))
              aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
       Cov_Bki <- aa*2

       Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

       VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100

pesos <- c(.3,.25,.2,.15,.1)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100
```

```
##
# Minimizando la muestra en cada estrado, teniendo como fijo el tamano total de la muestra
##


FunMinVar <- function(nnki)
{
		Ni <- c(397,600,1008,2422)
#		ndi <- c(397, 41, 35, 38)
  N  <- sum(Ni)
		Wi <- Ni/N


		ndi     <- c(397,nnki,NNN-397-sum(nnki))

PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & D2_1EST==j ]*VENTAS[ESTRATO==i & D2_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & I3_2AEST==j ]*VENTAS[ESTRATO==i & I3_2AEST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & I3_1EST==j ]*VENTAS[ESTRATO==i & I3_1EST==j ])


PREG <- list(preg1,preg2,preg3,preg4,preg5)

})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
			c(0,0.25,0.5,0.75,1.0),
			c(0,0.25,0.5,0.75,1.0),
			c(0,0.25,0.5,0.75,1.0),
			c(0,0.25,0.5,0.75,1.0))

PORC <- vector("list",5)


#Ponderacion
IND <- VAR <- rep(0,5)

for(ii in 1:5)
{
		pregunta <- PREG[[ii]]
		a       <- PESOS[[ii]]

		aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

		nki       <- rowSums(pregunta)


		fi        <- Ni/nki
		preg_c     <- cbind(pregunta,nki,Ni,fi)
		porcentaje <- as.matrix(pregunta/nki)
		Bk         <- rowSums(porcentaje*aj) #Suma Bki     porcentaje%*%a
		Dk         <- sum(Wi*Bk)
```

```
            IND[ii] <- Dk

            PORC[[ii]] <- porcentaje

#Varianza

            Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

            nr <- nrow(porcentaje)
            nc <- ncol(porcentaje)

            aa <- rep(0,nr)
            for(i in 1:nr)
            for(j in 2:nc)
            for(h in 1:(j-1))
                        aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
            Cov_Bki <- aa*2

            Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

            VAR[ii] <- Var_Bk
}


IND*100
sqrt(VAR)*100


pesos <- c(.3,.25,.2,.15,.1)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100

            return(Var_Ind*1000000)
}


Met <- c("Nelder-Mead", "BFGS", "CG", "L-BFGS-B", "SANN")

NNN  <- 511
nnki <- c(20,20)

FunMinVar(nnki)

#Funcion de optimizazcion continua para el tamanio de muestra
#toma como maximo la poblacion (Ni) y los minimos son introducidos(lower=...)

sal  <- optim(nnki,FunMinVar,method=Met[4],lower=c(10,10),upper=Ni[2:3])

kk <- c(397,sal$par,NNN-397-sum(sal$par))
kk
sum(kk)
FunMinVar(sal$par)
```

vii) Proceso de simulación de muestras calculando las varianzas de los indicadores (ICP, IAT, IPM), estadísticos básicos e histogramas

```
selemu<-function(n1=300,n2=10,n3=10,n4=10)
{
c(sample(unique(bases1232$NOEMP[bases1232$ESTRATO==1]),n1),sample(unique(bases1232$NOEMP[bases1232
$ESTRATO==2]),n2),
sample(unique(bases1232$NOEMP[bases1232$ESTRATO==3]),n3),sample(unique(bases1232$NOEMP[bases1232$
ESTRATO==4]),n4))
}


cuantos<-1000
```

```
Bk3<-matrix(0,nrow=cuantos,ncol=3)
for (jj in 1:cuantos)
{
#empresas<-selemu(397,5,18,91)
empresas<-selemu(397,18,38,77)
muestra<-which(bases1232$NOEMP %in% empresas)

Bk2<-with(bases<-subset(bases1232[muestra,],MES=="2906"),
 {
 Ni <- c(397,600,1008,2422)
 ndi<-table(bases$ESTRATO)
# print(nrow(subconjunto))
 FACTOR<-rep(0,nrow(bases))
  FACTOR<-
ifelse(ESTRATO==1,Ni[1]/ndi[1],ifelse(ESTRATO==2,Ni[2]/ndi[2],ifelse(ESTRATO==3,Ni[3]/ndi[3],Ni[3]/ndi[3])))
############## ICP  #######################
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-matrix(0,nrow=4,ncol=2)
dimnames(preg1)<-list(as.character(1:4),as.character(2:3))
preg2<-preg1
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg6<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & I5_1==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & I5_2==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & S8_1==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & S8_2==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & S8_3==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg6[i,j]<-sum(FACTOR[ESTRATO==i & S8_4==j ])
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

PESOS<- list(c(0,1),
                        c(0,1),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0),
                        c(0,0.25,0.5,0.75,1.0))

print(jj)
TNORESP<- PORC <- vector("list",6)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,6)

#print(PREG)
Bk1<-numeric()
for (ii in 1:6)
  {
        pregunta <- PREG[[ii]]
        a        <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)
```

```
#Indice

		nki	<- rowSums(pregunta)
  nri	<- ndi-nki
  TNORESP[[ii]] <- nri/ndi
#  print(nri)
#		print(c(sum(nki),nki))

		fi	<- Ni/nki
		preg_c	<- cbind(pregunta,nki,Ni,fi)
		porcentaje <- as.matrix(pregunta/nki)
		Bk	<- rowSums(porcentaje*aj) #Suma Bki
		Dk	<- sum(Wi*Bk)

		IND[ii] <- Dk

		PORC[[ii]] <- porcentaje

#Varianza

		Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

		nr <- nrow(porcentaje)
		nc <- ncol(porcentaje)

		aa <- rep(0,nr)
		for(i in 1:nr)
		for(j in 2:nc)
		for(h in 1:(j-1))
			aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
		Cov_Bki <- aa*2

		Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

		VAR[ii] <- Var_Bk
  }

IND*100
sqrt(VAR)*100

# Proceso 1
IND5 <- c(mean(IND[1:2]),IND[3:6])
mean(IND5)*100
VAR5 <- c( sum(VAR[1:2])/4,VAR[3:6])
sqrt(sum(VAR5)/25)*100

# Proceso 2
pesos <- c(1/10,1/10,rep(1/5,4))

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
SD_ICP<-sqrt(Var_Ind)*100

############### TERMINA ICP ##################
 ############## IAT ###############
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_2EST==j ]*ACTIVO[ESTRATO==i & P1_2EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & D2_2EST==j ]*VENTAS[ESTRATO==i & D2_2EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
```

```
    preg4[i,j]<-sum(FACTOR[ESTRATO==i & D2_4EST==j ]*VENTAS[ESTRATO==i & D2_4EST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
    preg5[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0))

cat(jj)
TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

#print(PREG)
Bk1<-numeric()
for (ii in 1:5)
 {
        pregunta <- PREG[[ii]]
        a        <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki        <- rowSums(pregunta)
 nri        <- ndi-nki
 TNORESP[[ii]] <- nri/ndi
#  print(nri)
#       print(c(sum(nki),nki))

        fi         <- Ni/nki
        preg_c     <- cbind(pregunta,nki,Ni,fi)
        porcentaje <- as.matrix(pregunta/nki)
        Bk         <- rowSums(porcentaje*aj) #Suma Bki
        Dk         <- sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        if (is.nan(Var_Bk))  print(porcentaje)
        VAR[ii] <- Var_Bk
 }


IND*100
sqrt(VAR)*100
```

```
# Proceso 2
pesos <- rep(1/5,5)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
cat("\t",Var_Ind,ndi,VAR,"\n")
SD_IAT<-sqrt(Var_Ind)*100
################### TERMINA IAT ####################
################### IPM ############################
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & D2_1EST==j ]*VENTAS[ESTRATO==i & D2_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & I3_2AEST==j ]*VENTAS[ESTRATO==i & I3_2AEST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & I3_1EST==j ]*VENTAS[ESTRATO==i & I3_1EST==j ])



PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
               c(0,0.25,0.5,0.75,1.0),
               c(0,0.25,0.5,0.75,1.0),
               c(0,0.25,0.5,0.75,1.0),
               c(0,0.25,0.5,0.75,1.0))

print(jj)
TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

#print(PREG)
Bk1<-numeric()
for (ii in 1:5)
  {
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(pregunta)
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi
#  print(nri)
#       print(c(sum(nki),nki))

        fi       <- Ni/nki
        preg_c    <- cbind(pregunta,nki,Ni,fi)
```

```
            porcentaje <- as.matrix(pregunta/nki)
            Bk          <- rowSums(porcentaje*aj) #Suma Bki
            Dk          <- sum(Wi*Bk)

            IND[ii] <- Dk

            PORC[[ii]] <- porcentaje

#Varianza

            Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

            nr <- nrow(porcentaje)
            nc <- ncol(porcentaje)

            aa <- rep(0,nr)
            for(i in 1:nr)
            for(j in 2:nc)
            for(h in 1:(j-1))
                        aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
            Cov_Bki <- aa*2

            Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

            VAR[ii] <- Var_Bk
  }


IND*100
sqrt(VAR)*100


# Proceso 2
pesos <- c(.3,.25,.2,.15,.1)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
SD_IPM<-sqrt(Var_Ind)*100

################## TERMINA IPM ###################

c(SD_ICP,SD_IAT,SD_IPM)

#  print(Bk1)
 })
 Bk3[jj,]<-Bk2
}

colnames(Bk3)<-c('ICP','IAT','IPM')
Bk3
summary(Bk3)
par(mfrow=c(2,2))
hist(Bk3[,1],main='Desviación Estandar ICP')
hist(Bk3[,2],main='Desviación Estandar IAT')
hist(Bk3[,3],main='Desviación Estandar IPM')
par(mfrow=c(1,1))

Asig_ipm<-Bk3
```

viii) Simulación de muestras calculando las estimaciones totales de los indicadores (ICP, IAT, IPM), estadísticos básicos e histogramas de las diferencias con respecto a las cifras publicadas por INEGI para Junio 2009

```
selemu<-function(n1=300,n2=10,n3=10,n4=10)
{
c(sample(unique(bases1232$NOEMP[bases1232$ESTRATO==1]),n1),sample(unique(bases1232$NOEMP[bases1232
$ESTRATO==2]),n2),
sample(unique(bases1232$NOEMP[bases1232$ESTRATO==3]),n3),sample(unique(bases1232$NOEMP[bases1232$
ESTRATO==4]),n4))
```

```
}

cuantos<-1000
Bk3<-matrix(0,nrow=cuantos,ncol=3)
for (jj in 1:cuantos)
{
empresas<-selemu(397,41,35,38)
muestra<-which(bases1232$NOEMP %in% empresas)

Bk2<-with(bases<-subset(bases1232[muestra,],MES=="2906"),
 {
 Ni <- c(397,600,1008,2422)
 ndi<-table(bases$ESTRATO)
# print(nrow(subconjunto))
 FACTOR<-rep(0,nrow(bases))
  FACTOR<-
ifelse(ESTRATO==1,Ni[1]/ndi[1],ifelse(ESTRATO==2,Ni[2]/ndi[2],ifelse(ESTRATO==3,Ni[3]/ndi[3],Ni[3]/ndi[3])))
############### ICP  ########################
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg1<-matrix(0,nrow=4,ncol=2)
dimnames(preg1)<-list(as.character(1:4),as.character(2:3))
preg2<-preg1
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg6<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & I5_1==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & I5_2==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & S8_1==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & S8_2==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & S8_3==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg6[i,j]<-sum(FACTOR[ESTRATO==i & S8_4==j ])
PREG <- list(preg1,preg2,preg3,preg4,preg5,preg6)
})

PESOS<- list(c(0,1),
                c(0,1),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0))

print(jj)
TNORESP<- PORC <- vector("list",6)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,6)

Bk1<-numeric()
for (ii in 1:6)
  {
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)
```

```
#Indice

        nki     <- rowSums(pregunta)
 nri     <- ndi-nki
 TNORESP[[ii]] <- nri/ndi

        fi      <- Ni/nki
        preg_c    <- cbind(pregunta,nki,Ni,fi)
        porcentaje <- as.matrix(pregunta/nki)
        Bk      <- rowSums(porcentaje*aj) #Suma Bki
        Dk      <- sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        VAR[ii] <- Var_Bk
 }

IND*100
sqrt(VAR)*100

# Proceso
pesos <- c(1/10,1/10,rep(1/5,4))

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
ICP<-Indice*100

############### TERMINA ICP ##################
 ############## IAT ###############
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_2EST==j ]*ACTIVO[ESTRATO==i & P1_2EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & D2_2EST==j ]*VENTAS[ESTRATO==i & D2_2EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & D2_4EST==j ]*VENTAS[ESTRATO==i & D2_4EST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])
PREG <- list(preg1,preg2,preg3,preg4,preg5)
})
```

```
PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                 c(0,0.25,0.5,0.75,1.0),
                 c(0,0.25,0.5,0.75,1.0),
                 c(0,0.25,0.5,0.75,1.0),
                 c(0,0.25,0.5,0.75,1.0))

cat(jj)
TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

#print(PREG)
Bk1<-numeric()
for (ii in 1:5)
 {
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki      <- rowSums(pregunta)
  nri      <- ndi-nki
  TNORESP[[ii]] <- nri/ndi
#  print(nri)
#       print(c(sum(nki),nki))

        fi       <- Ni/nki
        preg_c    <- cbind(pregunta,nki,Ni,fi)
        porcentaje <- as.matrix(pregunta/nki)
        Bk       <- rowSums(porcentaje*aj) #Suma Bki
        Dk       <- sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))

        nr <- nrow(porcentaje)
        nc <- ncol(porcentaje)

        aa <- rep(0,nr)
        for(i in 1:nr)
        for(j in 2:nc)
        for(h in 1:(j-1))
                aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
        Cov_Bki <- aa*2

        Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

        if (is.nan(Var_Bk))  print(porcentaje)
        VAR[ii] <- Var_Bk
 }


IND*100
sqrt(VAR)*100

# Proceso
pesos <- rep(1/5,5)

Indice <- sum(pesos*IND)
Indice*10

Var_Ind <- sum(pesos*pesos*VAR)
cat("\t",Var_Ind,ndi,VAR,"\n")
```

```
IAT<-Indice*100
################### TERMINA IAT ####################
################## IPM ###########################
 PREG<-with(subconjunto<-subset(bases,MES=="2906"),{
preg3<-matrix(0,nrow=4,ncol=5)
dimnames(preg3)<-list(as.character(1:4),as.character(1:5))
preg1<-preg2<-preg4<-preg5<-preg3

 for (i in dimnames(preg1)[[1]])
   for (j in dimnames(preg1)[[2]])
     preg1[i,j]<-sum(FACTOR[ESTRATO==i & D2_1EST==j ]*VENTAS[ESTRATO==i & D2_1EST==j ])

 for (i in dimnames(preg2)[[1]])
   for (j in dimnames(preg2)[[2]])
     preg2[i,j]<-sum(FACTOR[ESTRATO==i & P1_1EST==j ]*VENTAS[ESTRATO==i & P1_1EST==j ])

 for (i in dimnames(preg3)[[1]])
   for (j in dimnames(preg3)[[2]])
     preg3[i,j]<-sum(FACTOR[ESTRATO==i & P6_1EST==j ]*PERSONAL[ESTRATO==i & P6_1EST==j ])

 for (i in dimnames(preg4)[[1]])
   for (j in dimnames(preg4)[[2]])
     preg4[i,j]<-sum(FACTOR[ESTRATO==i & I3_2AEST==j ]*VENTAS[ESTRATO==i & I3_2AEST==j ])

 for (i in dimnames(preg5)[[1]])
   for (j in dimnames(preg5)[[2]])
     preg5[i,j]<-sum(FACTOR[ESTRATO==i & I3_1EST==j ]*VENTAS[ESTRATO==i & I3_1EST==j ])


PREG <- list(preg1,preg2,preg3,preg4,preg5)
})

PESOS<- list(c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0),
                c(0,0.25,0.5,0.75,1.0))

print(jj)
TNORESP<- PORC <- vector("list",5)

#Datos Poblacionales
N  <- sum(Ni)
        Wi <- Ni/N

#Ponderacion
IND <- VAR <- rep(0,5)

Bk1<-numeric()
for (ii in 1:5)
  {
        pregunta <- PREG[[ii]]
        a       <- PESOS[[ii]]

        aj <- matrix(a,nrow=4,ncol=length(a),byrow=TRUE)

#Indice

        nki       <- rowSums(pregunta)
  nri       <- ndi-nki
  TNORESP[[ii]] <- nri/ndi

        fi        <- Ni/nki
        preg_c     <- cbind(pregunta,nki,Ni,fi)
        porcentaje <- as.matrix(pregunta/nki)
        Bk         <- rowSums(porcentaje*aj) #Suma Bki
        Dk         <- sum(Wi*Bk)

        IND[ii] <- Dk

        PORC[[ii]] <- porcentaje

#Varianza

        Var_Bki<-rowSums((aj^2)*(porcentaje)*(1-porcentaje))
```

```
            nr <- nrow(porcentaje)
            nc <- ncol(porcentaje)

            aa <- rep(0,nr)
            for(i in 1:nr)
            for(j in 2:nc)
            for(h in 1:(j-1))
                        aa[i] <- aa[i] + a[j]*a[h]*porcentaje[i,j]*porcentaje[i,h]
            Cov_Bki <- aa*2

            Var_Bk <- sum( (Var_Bki-Cov_Bki)/ndi * Wi*Wi )

            VAR[ii] <- Var_Bk
  }


IND*100
sqrt(VAR)*100


# Proceso 2
pesos <- c(.3,.25,.2,.15,.1)

Indice <- sum(pesos*IND)
Indice*100

Var_Ind <- sum(pesos*pesos*VAR)
sqrt(Var_Ind)*100
IPM<-Indice*100
################## TERMINA IPM ###################

c(ICP,IAT,IPM)

 })
 Bk3[jj,]<-Bk2
}

colnames(Bk3)<-c('ICP','IAT','IPM')
Bk3
#Diferencia de las estimaciones con los datos publicados
Dif<-matrix(c(Bk3[,1]-41.38271,Bk3[,2]-50.2259,Bk3[,3]-49.32136),ncol=3,nrow=cuantos)
colnames(Dif)<-c('ICP','IAT','IPM')

summary(Dif)
colSums((abs(Dif)>4)/cuantos)*100
par(mfrow=c(2,2))
hist(Dif[,1],main='Diferencia ICP')
hist(Dif[,2],main='Diferencia IAT')
hist(Dif[,3],main='Diferencia IPM')
par(mfrow=c(1,1))
```

ix) Graficas de las varianzas de los indicadores

```
load("1000Optim.RData")
compara<-function(x,...)
{
nc<-ncol(x)
minx<-min(x)
maxx<-max(x)
#plot.new()
colores<-c("blue","green","brown","orange")
leg.txt<-colnames(x)
plot(ecdf(x[,1]),xlim=c(minx,maxx),col.points='blue',...)
for (i in (2:nc))
   plot(ecdf(x[,i]),col.points =colores[i],add = TRUE)
   legend("bottomright", leg.txt, col=colores , pch=1, title="icp")
}
compara(cbind(Actual=Actual[,1],Asignacion1=Asigna1[,1],Asignacion2=Asigna2[,1],Asignacion3=Asigna3[,1]),main="IC
P")

compara(cbind(Actual=Actual[,2],Asignacion1=Asigna1[,2],Asignacion2=Asigna2[,2],Asignacion3=Asigna3[,2]),main="IA
T")

compara(cbind(Actual=Actual[,3],Asignacion1=Asigna1[,3],Asignacion2=Asigna2[,3],Asignacion3=Asigna3[,3]),main="IP
M")
```

Bibliografía

Arthanari, T.S. and Dodge Yadolah (1981). *Mathematical Programming in Statistics*. A Wiley-Interscience Publication John Wiley and Sons, Inc.

Cochran, W. G. 1977. *Técnicas de muestreo.* C.E.C.S.A. México.

Khuri, A.I. and Cornell, J.A. (1987). *Response Surface: Designs Analysis.* Marcel Dekker, Inc. New York.

Lohr, Sh. L. 2000. *Muestreo: Diseño y Análisis.* Internacional Thomson Editores.

Prawda Witenberg, J. (1981). *Métodos y modelos de investigación de operaciones.*Editorial Limusa.

Prékopa A. (1978). *"The use of stochastic programming for the solution of the some problems in statistics and probability".* Technical Summary report 1983. University of Wisconsin-Madison.

Rao, S.S. 1978-79. *Optimization theory and applications.* Wiley Eastern Limited.
Ríos, S., Ríos Insua, S., y Ríos Insua, M.J. (1989). *Procesos de decisión Multicriterio.*EUDEMAUNIVERSIDAD Manuales.

Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S., and Asok, C. (1984). *Sampling theory of surveys with application.* State University Press.

M. Bazaraa, H. D. Sherali and C. M. Shetty, Nonlinear Programming: Theory and Algorithms, Second Edition, John Wiley and Sons, 1993.

D. G. Luenberger, Linear and Nonlinear Programming, Second Edition, John Wiley and Sons, 1984.