# USO DE FUENTES ADMINISTRATIVAS PARA LA REDUCCIÓN DE CARGA Y COSTES EN LAS ENCUESTAS ESTRUCTURALES DE EMPRESAS (UFAES)


## Resumen

El uso de fuentes administrativas con fines estadísticos forma parte de la actividad corriente del Instituto Nacional de Estadística español (INE) en diversas áreas. El proyecto UFAES supone un nuevo salto cualitativo en estas actividades, con objetivos orientados a reducir de manera significativa el tamaño muestral de las dos grandes operaciones estructurales de empresas en el INE.

Se trata de potenciar la calidad de las estimaciones a partir de una observación directa completa (en el sentido de haber sido obtenida a partir de la realización de la encuesta con un tamaño muestral como el actual) en el primer año de un bienio, mientras que en el siguiente la observación directa se reducirá compensando estas pérdidas en los tamaños muestrales con las ganancias de eficiencia aportadas por una submuestra suplementaria de la encuesta del año anterior, actualizada con estimaciones e imputaciones asistidas por modelos relativos al cambio en las variables tributarias asociadas a las variables estadísticas objetivo.



**Palabras clave:**
Uso de datos fiscales en encuestas económicas
Reducción de costes y carga estadística a las empresas
Uso de fuentes administrativas con fines estadísticos



**Autores** [1]
**Jorge Saralegui**
**Cristina González**
**Ignacio Arbués.**
**INE**



[1] Los trabajos realizados por los autores contaron con la aportación, indispensable para el proyecto, de expertos responsables de diversas unidades del INE, en particular, Luis Esteban Barbado y otro miembros de su equipo en la unidad de estándares (preparación de formatos de carga e intercambios de información con la AEAT); Gonzalo Parada (diseño de muestras); Antonio González, Carmen Pirla y otros miembros el equipo a cargo de los trabajos de la Encuesta Anual de Servicios (explotación de controles del input y salida de resultados); y Cristina Casaseca (desarrollo informático de estimadores).

# USE OF ADMINISTRATIVE SOURCES TO REDUCE STATISTICAL BURDEN AND COTS IN STRUCTURAL BUSINESS SURVEY (UFAES)


## Abstract

The use of administrative sources with statistical purposes is part of the current activity of the National Statistics Institute (Ine, Spain), in various fields. UFAES project provides a new qualitative impulse to these activities, with objectives oriented to significantly reduce the sample size of the major INE annual structural business surveys.

The core idea is to maintain, for a rotating subpopulation, the estimates from a complete direct observation (in the sense of being obtained from a survey of the same sample size as at present) in the first year of a biennium, while in the following year the sample for direct observation will be drastically reduced in size, compensating the efficiency losses with the gains provided by an additional subsample extracted from the previous year effective sample, updated with model assisted estimates of the change in the tax variables associated with target statistical variables.



## Keywords

Use of tax data in business surveys
Reduction of costs and statistical burden on enterprises
Use of administrative sources for statistical purposes



## Authors [1]

**Jorge Saralegui**
**Cristina González**
**Ignacio Arbués**
**INE (Spain)**



[1] Works carried out by the authors would not have been possible without the contribution of experts from various INE units, in particular: Luis Esteban Barbado and other staff within the Standards Unit (Information flows and institutional relations with the Tax Departement); Gonzalo Parada and Dolores Lorca (sampling design); Antonio González, Carmen Pirla and the rest of the subject matter team of the Annual Structural Survey in the Services Sectors (controls processing and output tables); Cristina Casaseca (EDP, estimators programming).

## I. Introducción

1. En las estrategias de desarrollo de las oficinas centrales de estadística formuladas durante la última década, en Europa y otras partes del mundo, se contempla de manera preeminente el logro de mejoras de eficiencia a través de la reducción de la carga estadística a las empresas y hogares, así como del coste de las operaciones, con compromisos explícitos de analizar el posible redimensionamiento de los tamaños muestrales de las encuestas. Para la consecución de estos objetivos, también contemplados, en el ámbito de la UE, en el Código de Buenas Prácticas y en los planes de modernización del sistema estadístico europeo (Modernisation of European Enterprise and Trade Statistics - MEETS) es necesario establecer programas continuos de actuación y acciones concretas y coordinadas en distintos frentes, como la diversificación y simplificación de las modalidades de recogida de datos, la mejora en los diseños de los soportes y flujos de la información, y muy especialmente el aprovechamiento del potencial que ofrecen las fuentes administrativas para la reducción del coste y la carga objetiva neta soportada por los informantes.

2. En el caso español, obviamente, el uso de fuentes administrativas forma parte de la actividad corriente del Instituto nacional de Estadística en diversas áreas de su actividad, especialmente en estadísticas exhaustivas de origen administrativo, en la formación de marcos e instrumentos de infraestructura estadística, especialmente para los censos, en operaciones de síntesis (contabilidad nacional, sistemas de indicadores), y como apoyo a la depuración, imputación y análisis de datos. Más recientemente, en los últimos años se ha ido ampliando el campo de aprovechamiento directo (microdato) para operaciones por muestreo dirigidas a hogares (Encuestas de salarios, de Población Activa, de Condiciones de Vida, entre otras). A este desarrollo ha ayudado notablemente lo establecido en el convenio de colaboración INE-Agencia Española de la Admón.Tributaria (AEAT) de 2002, así como lo contemplado en la legislación estadística general de la UE y en los reglamentos específicos de la UE sobre la cesión de datos administrativos para el cumplimiento de las obligaciones de suministro de información de los estados miembros.

3. El uso de información contable del Registro Mercantil, las consultas web u otras fuentes utilizadas en las fases de depuración de las encuestas estructurales a empresas, ya está permitiendo una reducción importante de la carga de respuesta en las empresas, al disminuir, en particular, el número de re-contactos para actualización de datos. No obstante, el proyecto que se describe en el documento, con sus dos etapas de desarrollo, sobre simulación (SIMFAES) y uso (UFAES) de fuentes administrativas en encuestas estructurales de empresas, aporta un nuevo salto cualitativo en estas actividades con objetivos orientados a reducir de manera significativa el tamaño muestral de las dos grandes operaciones estructurales de empresas en el INE, la Encuesta Industrial Anual de Empresas (EIAE) y la Encuesta Anual de Servicios (EAS).

4. El documento relata en primer lugar unos antecedentes al planteamiento del proyecto, para pasar en las primeras secciones a la descripción de la idea central del procedimiento de reducción muestral y suplemento con muestra del periodo anterior, actualizada con información externa de origen en la administración fiscal. Seguidamente se definen los conceptos más importantes, para pasar después a las diferentes fases de desarrollo del proyecto, mediante la carga de información externa en el marco, selección de muestras y cálculo de pesos de la muestra integrada, asignación de tipo de variable y procedimientos de

estimación en cada caso. En los apartados finales, se describen los indicadores utilizados en la evaluación de resultados de las simulaciones de Monte Carlo, seguido de unos comentarios a los resultados presentados en anexo.

## II. Antecedentes

5. A lo largo de 2008/09, a raíz de la facilidad de acceso a microdatos tributarios al amparo de lo establecido por los reglamentos europeos y en el convenio de colaboración INE/AEAT, tuvieron lugar los primeros análisis de oportunidad y estudios previos de evaluación y contraste de información procedente de ambas fuentes, concentrados en el marco y muestra de la Encuesta Anual de Servicios (EAS) y el Impuesto de Sociedades (IS), ambos referidos a 2006.

6. Como resultado de estas actividades, se llegó a una serie de conclusiones que confirman que las fuentes tributarias presentan ciertas limitaciones a la hora de satisfacer las necesidades de información estadística que actualmente cubre la EAS, sujeta a reglamento europeo y con fuertes exigencias por parte de usuarios nacionales en cuanto a su calidad y desagregación territorial. Estas limitaciones afectan a la cobertura y al contenido de variables de interés en las declaraciones tributarias, así como a los plazos de disponibilidad de la información, y se pueden resumir como sigue:

- Existencia de variables estructurales objetivo de la encuesta y/o reglamento para las que no existe correspondencia con variables contempladas en la fuente administrativa.
- No cobertura de la subpoblación de empresas de condición jurídica persona física (y otras formas que, de acuerdo a nuestro marco legal, pueden ser asimiladas al concepto de cuasi sociedad, como son las Comunidades de Bienes o las Sociedades Civiles Personales**)** por la fuente administrativa principal (IS) y con limitada cobertura en variables objetivo por otras fuentes (IRPF, IVA, Seguridad social). Esta subpoblación es especialmente significativa en la EAS.
- No cobertura en la fuente tributaria disponible actualmente de los territorios forales (PV y Navarra), por lo que será necesario realizar una acción institucional orientada a extender el procedimiento a estos ámbitos.
- Riesgo de cambios legislativos que perturban la estabilidad de la cantidad/calidad de información aportada por las fuentes externas, y que puedan ser aplicados con escaso margen de maniobra para la adaptación al uso con fines estadísticos.
- Plazos de disponibilidad de la fuente externa relativamente largos, que pueden poner en riesgo las exigencias de puntualidad de la producción estadística.

7. Debido a estas limitaciones, los primeros análisis concluían con la recomendación de mantener anualmente una parte significativa de observación muestral directa, (grosso modo, que como mínimo permitiera alcanzar eficiencia en las estimaciones a nivel nacional), restringiendo las posibilidades de reducción de carga de respuesta y costes a determinados estratos de pequeño o mediano tamaño, y de unidades del ámbito del IS, excluyendo de la reducción muestral a la sub-población de empresarios persona física, por la importancia de la misma en el sector servicios y la escasa cobertura de variables de interés estadístico en la fuente externa para este ámbito.

8. Cabe añadir que si se tiene en cuenta la primera de las limitaciones enumeradas anteriormente, las variables no abastecidas por la fuente administrativa deberán ser estimadas o imputadas a partir del resto de información disponible por observación en la encuesta, con lo cual se exige implícitamente que el tamaño muestral para estimación directa en los diferentes estratos deberá ser suficiente para permitir estas estimaciones con calidad satisfactoria.

Con el fin de superar en lo posible las limitaciones enumeradas anteriormente y alcanzar una reducción significativa de la carga y costes de las encuestas estructurales, se formuló el proyecto que se describe aquí, con una fase previa de simulación mediante la reiteración del procedimiento previsto, aplicado a submuestras de microdatos extraídas de las muestras efectivas de la operación objetivo, ya disponibles.

## III. Simulación del uso de fuentes administrativas en encuestas estructurales de empresas (SIMFAES)

9. Se trata de la etapa de investigación y desarrollo del proyecto UFAES, cuyo objetivo principal es evaluar la eficiencia y sesgo de estimadores asistidos por modelos que utilizan información auxiliar de origen tributario relativa a las mismas unidades y población objetivo y aplicados a submuestras artificiales simuladas obtenidas a partir de los ficheros finales de las encuestas estructurales del INE. Las muestras simuladas serán de tamaño significativamente menor respecto al de las encuestas originales realizadas en mundo real, con el fin de evaluar asimismo la reducción de costes y carga que es posible alcanzar por estos procedimientos.

10. El proyecto UFAES se concibió con un enfoque bienal, primando el aprovechamiento de la componente panel de la información del ámbito tributario, al disponer de información primaria para las mismas unidades en periodos consecutivos. Se trata de potenciar la calidad de las estimaciones a partir de una observación directa completa (en el sentido de estar obtenida a partir de la realización de la encuesta con un tamaño muestral como el actual) en el primer año del bienio, mientras que en el siguiente la observación directa se reducirá de manera significativa , compensando estas pérdidas en los tamaños muestrales con las ganancias de eficiencia aportadas por una submuestra suplementaria del año anterior, actualizada con estimaciones e imputaciones asistidas por modelos relativos al cambio en las variables objetivo.

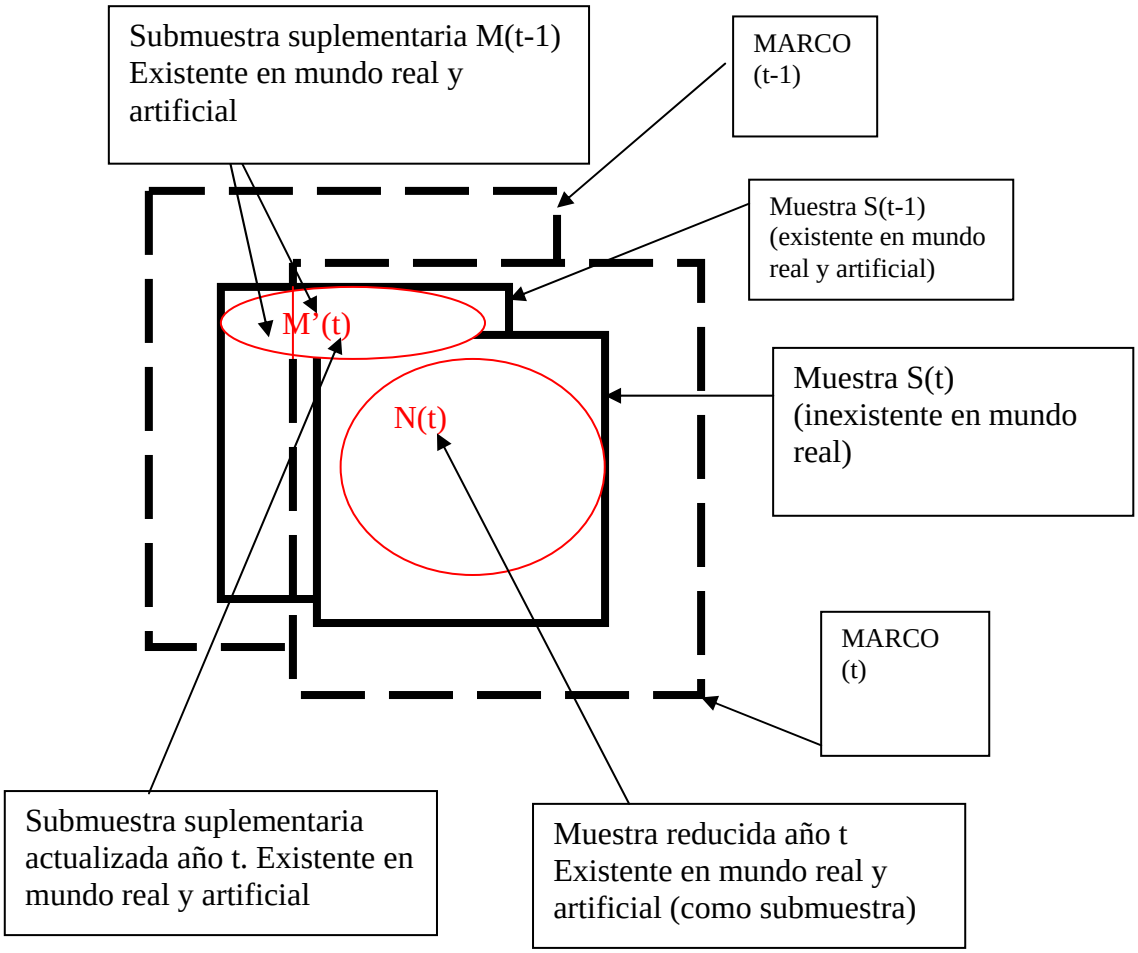
11. Con este enfoque, la simulación se implementó en el periodo bienal para el que ya existía en su momento información final de encuesta (2008-2009), de manera que en el primer año se trabajó con la muestra completa disponible por observación directa tanto en mundo simulado como en mundo real, mientras que en el año siguiente se provocó una reducción artificial del tamaño muestral hasta el límite de la eficiencia para el ámbito nacional, pero que posibilitara las estimaciones indirectas para dominios inferiores, añadiendo una submuestra suplementaria de t-1, actualizada con estimadores del cambio, utilizando por una parte la información de origen administrativo y, por otra, las muestras investigadas por el INE en ambos periodos. Éstas serán la ya mencionada 'completa' del año t-1 y la 'reducida' en el año t, esta última obtenida en el proyecto de simulación, como submuestra artificial de la total efectiva, pero que en mundo real UFAES constituirá la única muestra con recogida directa en ese año.

12. El equipo de proyecto acordó trabajar prioritariamente con la Encuesta Anual de Servicios (EAS), que es la que mayor potencial de reducción muestral presenta, posponiendo el tratamiento de la Encuesta Industrial Anual de Empresas (EIAE) hasta disponer, en su caso, de una visión favorable de los resultados que se vayan alcanzando con la EAS. Las pruebas se realizaron con todo el ámbito de actividades CNAE de la EAS, pero la intención de la fase de implementación UFAES en su primer cuatrienio de referencia en mundo real (2012-2016) es que el concepto 'muestra completa' o 'reducida' se aplique a un subconjunto de actividades (a nivel de divisiones CNAE) solamente, rotando de manera simétrica con el resto de divisiones en el bienio siguiente, de manera que los tamaños muestrales que serán observados directamente en mundo real permanecerán aproximadamente estables en ambos periodos. A este fin se ha utilizado la división natural de la encuesta en la dirigida a las divisiones CNAE del sector comercio (Encuesta Anual de Comercio, EAC), por una parte, y en el resto del sector servicios, por otra. Aunque los trabajos SIMFAES que se describen aquí, ya finalizados, se han realizado con la EAS 08-09, la carga de información externa y otras fases del proyecto incluyeron también las muestras de la Encuesta Industrial Anual de Empresas (EIAE), por lo que las simulaciones de prueba relativas a esta última operación se podrán aplicar en paralelo a la implantación de UFAES en la EAS (sin los sectores del ámbito de Comercio) prevista para el trabajo de campo de 2013 (referencia 2012). Se excluyó de todo el estudio el ámbito de los territorios forales, quedando pendiente la gestión institucional necesaria para obtener la fuente tributaria de origen en los mismos.

## IV. Esquema de Selección de Muestras Artificiales

### C.1. Esquema de selección

13.




14. El esquema anterior representa el diseño previsto por UFAES, pero aplicado en la etapa de simulación. La notación se refiere siempre a muestras efectivas, observadas en trabajo de campo, aunque la selección se realiza en primer lugar como muestras teóricas (no representadas en el gráfico) asociadas a las anteriores, las cuales dan lugar al registro de incidencias (bajas) que es necesario archivar para el cálculo de pesos muestrales de la muestras integradas. En primer lugar se obtendrá una submuestra N(t), a partir de la muestra efectiva final S(t). La muestra teórica que permite obtener la muestra efectiva N(t) se selecciona previamente con la afijación y tamaños necesarios para alcanzar un grado de eficiencia satisfactorio para las variables y estratos requeridos por el reglamento UE de estructura económica de la empresas (SBS), a nivel nacional.

A medida que se avance en la evaluación de los resultados de las simulaciones se podrá optar por reducir los tamaños muestrales por debajo de este nivel estándar. La submuestra artificial efectiva N(t), simulada para esta fase, es la que suministrará información de observación directa en el mundo real en el año t, en el que no existirá lo que aquí se designa como S(t) (en su lugar se dispondrá de la totalidad del marco en t). Para cada registro elemental de N(t) se incorporarán las variables administrativas disponibles para t y t-1.

15. El siguiente paso es construir la submuestra que constituye el suplemento a N(t) y cuyas unidades serán observadas solamente a través de los datos administrativos. Para ello, se obtendrá una submuestra efectiva M(t-1) a partir de una submuestra suplementaria obtenida de la muestra teórica en (t-1), y actualizada con la muestra efectiva S(t-1) cargando el registro completo en t-1. La muestra teórica complementaria en t-1 de la que se deriva la efectiva M se selecciona con una afijación y tamaños que posibiliten que la integración de N(t) y M(t-1), tras el tratamiento de estimación previsto en UFAES, permita satisfacer la demanda nacional e internacional de resultados para dominios inferiores al nacional en el año de referencia t. Reducciones sucesivas de tamaños en la submuestra N (t) llevarán asociados normalmente aumentos en M(t-1).

Se procederá a una actualización de la submuestra efectiva M (t-1) frente el marco del año t causando bajas por el paso del marco t-1 a t, obteniendo la submuestra M'(t). A esta última, se le incorporarán los datos del registro completo de M(t-1) así como las variables disponibles en las fuentes administrativas en t y t-1.

## V. Variables estadísticas centrales y variables administrativas asociadas

16. Una componente esencial del proyecto es la asignación de una tipología a las diferentes variables que intervienen en los estimadores, que permite discernir mediante una codificación específica el proceso de estimación al cual son sometidas las mismas durante las diferentes fases de desarrollo.

Para las variables de origen administrativo, se considera solamente la categoría de ***asociada***, con notación estándar ***a*** en los documentos de proyecto, tratándose de aquellas de origen en las declaraciones administrativas que supuestamente presentan una correlación muy alta con una variable homónima o muy próxima conceptualmente proporcionada por la fuente estadística, mediante encuesta, que se denomina variable central. UFAES utiliza la información de la variable asociada para estimar el cambio en la variable central, por lo que cabe esperar que, en caso de existir, el impacto de la no coincidencia conceptual exacta entre ambas quede en parte amortiguado. Cuando existen variables centrales con variable asociada en varias fuentes administrativas, como es el caso de las diferentes declaraciones de impuestos, se efectuará la carga de información procedente de todas las fuentes donantes, pudiendo figurar más de un valor asociado. Tras la aplicación de los controles de validez de la asociada, se define una asociada principal con un orden de prioridad, que en el caso de las fuentes tributarias se establece en 1) IS; 2) IVA; 3) IRPF.

## VI. Tipología de variables estadísticas: Tipo inicial

17. En cuanto a la tipología de las variables estadísticas (es decir, las que son objeto de observación a través de los instrumentos de recogida de datos de la encuesta) se definen tres niveles, ***tipo inicial, tipo intermedio, y tipo final***, como se detalla en los siguientes apartados. Las tareas de asignación del tipo inicial de variable destacan especialmente por su trascendencia en los resultados del proyecto y, especialmente, por ser de las que absorben mayores costes, en términos de conocimiento experto y documentación. En efecto, lo óptimo en este caso es que esta etapa sea desarrollada por los mejores especialistas en la encuesta, previo estudio en profundidad de todas y cada una de las fuentes administrativas, tareas que, como es habitual en los proyectos de 'rediseño' como UFAES, deberán hacerse compatibles con los trabajos corrientes de desarrollo y explotación de las diversas operaciones en que están implicados

estos mismos expertos, lo cual no siempre es posible y puede tener repercusiones en la marcha de los trabajos, especialmente en el periodo de pruebas. Esta tipología no es cerrada (y esto también añade coste a este tipo de proyectos), puesto que se pueden dar modificaciones en los instrumentos de encuesta y/o en las declaraciones administrativas, de formato o conceptuales, que requerirán estudio detallado para la actualización de estas clasificaciones en cada periodo de ejecución en mundo real.

Como primer resultado de esta fase, se obtiene y documenta apropiadamente la asignación del tipo inicial de variable estadística, que responde a la siguiente clasificación:

**Variables centrales**

18. En los datos de origen administrativo hay variables que no se corresponden con ninguna variable del cuestionario, pero sí con variables que se pueden obtener como combinación de varias de ellas. A estas combinaciones las llamamos variables centrales derivadas. Por otra parte, las variables del cuestionario que entran a formar parte de una de estas combinaciones tienen también consideración de centrales.

Distinguimos, por tanto, entre las variables centrales derivadas y las elementales –todas las demás– y de entre estas últimas, separamos aquellas que forman parte del cálculo de una derivada y las que no.

Por otra parte, también distinguimos aquellas variables centrales que son empleadas para buscar el vecino más próximo, llamadas principales.

Combinando estas distinciones, tenemos la siguiente clasificación:

**CE1**: Central elemental principal sin central derivada, que interviene en cálculos de vecino más próximo, como se ilustra más adelante en los apartados sobre estimadores. Se trata de variables estadísticas centrales, para las que potencialmente existe variable asociada al menos en una de las fuentes administrativas objetivo, adoptando esta categoría independientemente de que, como resultado de la carga de información administrativa, la variable asociada sea aceptada como válida. Además, intervienen en los estimadores asistidos que utilizan información de vecino más próximo.

**CE2**: variable central elemental, sin derivada, que no interviene en la búsqueda del vecino más próximo.

**CD1**: v. central derivada que interviene en el cálculo de vecino próximo. Se trata de variables del cuestionario que se pueden obtener como suma de variables centrales elementales y que, además, intervienen en el cálculo de distancias para determinar el NN.

**CD2**: variable derivada que no interviene en la búsqueda de vecino más próximo;

**CED**: variable central elemental con variable derivada; Se trata de las variables elementales que componen una variable derivada.

En el anexo A3 se documenta la detección de variables centrales en la encuesta anual de servicios (EAS) y sus asociadas en las diferentes fuentes de la AEAT (IS, IVA ordinario, IVA e IRPF en sus diferentes modalidades). A su vez, el anexo A2 (cuadro CD) describe la relación entre variables elementales y sus derivadas.

**Variables No centrales**

19.

**NCDI**: No centrales de estimación directa. Se trata de variables sin asociada en las fuentes externas, y que son objeto de estimación con estimadores del cambio que utilizan como información auxiliar el dato en t-1 y la estimación de las variables centrales (llamada 'explicativa') en t correspondientes a la misma unidad, así como las estimaciones de la muestra (N) observada en t, que interviene en la donación de vecino NN, y en la componente sintética del cambio en los estimadores.

**NCIN**: variables no centrales de imputación indirecta, a partir de variables centrales y no centrales de imputación directa.

**NCID**: variables no centrales de imputación idéntica. Se trata de variables que repiten, por sus características de estabilidad estructural, el dato en t-1.

## VII. Ensayos preparatorios

20. Los trabajos preparatorios fueron iniciados antes de la recepción de los ficheros de la AEAT, con el fin de ganar tiempo en la programación de las técnicas a emplear una vez que se dispusiera de la fuente externa, y se prolongaron durante el periodo Enero-Junio 2011. Dado que existe una solapamiento parcial entre las muestras diseñadas para cada periodo anual, se seleccionó la submuestra común 2007-2008 de la Encuesta de Comercio, para ser considerada en estas primeras simulaciones de prueba como fuente para variables asociadas, que en el escenario real del proyecto serán las abastecidas por la fuente administrativa.

Posteriormente, se realizaron las pruebas con muestras integradas de las dos submuestras del tipo $N(t)$ y $M(t)$, tal como está previsto en la metodología UFAES. Se realizaron m = 1000 simulaciones y se obtuvieron, por tanto, m estimaciones para las diferentes variables centrales y no centrales y según los estimadores estándar propuestos en el proyecto.

21. Se calcularon estimaciones bootstrap EMC y los indicadores de sesgo relativo e impacto del error (IE), que es el cociente entre la estimación del EMC y el total estimado en mundo real del valor de Producción o de Ocupados, según la variable de que se trate. Las replicaciones con reposición bootstrap se realizaron con dos muestras independientes a partir de la muestra original en t, la primera para estimar el sesgo y EMC sin imputación y la segunda para estimar los valores tras la imputación con el cambio (t /t-1), por el procedimiento de sustitución del 20% (correspondiente al tamaño simulado para M) de la muestra, aproximadamente.

Los resultados en forma de largas series de cuadros obtenidos cruzando las diferentes variables y estimadores y conteniendo estimadores del sesgo, impacto del sesgo y sesgo relativo, fueron analizados y almacenados convenientemente.

Aparecían sesgos despreciables para las variables con valores >0 muy frecuentes en ambos periodos y aplicados a submuestras 'comunes' de tamaño significativo (los tamaños de la simulación con submuestra común son muy variables entre estratos). En el resto de casos, pueden aparecer esporádicamente sesgos de mayor nivel y valores extremos, pero cuyos

impactos ('impacto del sesgo') en las variables principales de encuesta son normalmente muy bajos.

Los resultados indican que los valores de *EMC* son altos cuando los errores de muestreo asociados a la variable son altos, pues la simulación 'hereda' el efecto de los errores de muestreo. Por el contrario, los indicadores de *IE* son generalmente muy pequeños. Los resultados en general ponían de manifiesto la escasa importancia de los sesgos atribuibles a la imputación del dato a partir del valor de la variable en t-1 corregida del cambio. Por otra parte, el error medio cuadrático recoge el efecto del error de muestreo en el diseño original para la variable en estudio, por lo que se confirma que el modelo aparentemente no produce una aportación sistemáticamente alta al error.

## VIII. Carga de variables asociadas

22. Ésta es la fase en la que se produce la interacción institucional en UFAES, mediante la cual se efectúa la carga de variables asociadas necesarias para la actualización de la muestra suplementaria M'.
El fichero INE para carga de variables administrativas en la etapa SIMFAES se formó por integración de las muestras teóricas t y t-1 de las encuestas EIAE y EAS 08-09. El microdato de salida contenía los identificadores básicos de cada unidad tipo empresa, comunes con los disponibles en la fuente administrativa, así como identificadores internos de explotación de las encuestas.

A partir de los ficheros auxiliares ASOC-Centrales y CD mencionados anteriormente, se dispone de las identificaciones y ubicaciones detalladas de las variables asociadas en los diversos formularios de las declaraciones fiscales.

Recorriendo los archivos de declaraciones correspondientes a los ejercicios de t y t-1, a cada microdato (empresa) se asignaron los valores de la variable asociada a cada central, presentes en las declaraciones de los dos ejercicios, pues se trata de estimar el cambio de la variable central entre ambos.

Se realizó la carga desde las siguientes fuentes:

- Impuesto de Sociedades (IS).- Modelo 200
- Impuesto de Valor Añadido (IVA), con indicación del tipo: General o Simplificado. Modelo 390 para 2009, Modelos 390 y 392 (grandes empresas) para 2008.
- Impuesto sobre la Renta de las Personas Físicas (IRPF), con indicación del tipo: Estimación Directa o Estimación Objetiva. Modelo D100

La cobertura geográfica es la totalidad del territorio nacional excepto País Vasco y Navarra y no existen restricciones en cuanto a forma legal de la empresa o estratos de tamaño y actividad.

La variable central relacionada con el empleo, es obtenida a partir de fuentes de la seguridad social, a través del Directorio Central de empresas del INE (DIRCE).

### Documentación auxiliar de carga

23. Para facilitar el desarrollo de la integración de ficheros, se proporcionan una serie de normas complementarias para cubrir las particularidades de contenido o cobertura de algunas de las variables asociadas en las declaraciones fiscales, en ambos periodos, que puede ser diferente por cambios en la normativa o en el formato de las declaraciones, especialmente cuando se trata de variables

asociadas correspondientes a variables derivadas centrales, es decir, que se obtienen por suma de elementales.

**Fichero de salida con variables asociadas**

24. El fichero de salida de este proceso contiene información primaria para la ejecución de las distintas fases de tratamiento UFAES. Dado que es de interés para el proyecto evaluar la presencia de una unidad en las diversas fuentes, para cada unidad legal proporcionada (con sus identificadores comunes a la fuente fiscal y tributaria) se extraerá información específica de cada una de ellas.

A cada unidad elemental (variable central*empresa) del fichero de recepción, se asignarán códigos de *presencia o ausencia en todas y cada una* en las fuentes tributarias, en la modalidad de declaración que corresponda, procediendo a la carga correspondiente o figurando en blanco si el código es de ausencia.

El output generará un fichero con el mismo número de unidades proporcionado y enriquecido con la información tributaria, de acuerdo a un diseño de registro propuesto por el INE.

## IX. Controles de recepción de variables asociadas

25. A la recepción de los diferentes data sets con la carga de salida de las diversas fuentes tributarias y para los dos periodos de referencia, es necesario un primer análisis de la información recibida relativa a cada variable estadística de tipo inicial central, es decir, con potencial para disponer de asociadas válidas. Estos controles se podrán adaptar en mundo real en cada ejercicio de ejecución, especialmente si se detectan indicadores de sospecha de algún error en la integración, que necesite un análisis más detallado que el utilizado inicialmente.

Se obtiene una serie de salidas de control y análisis primario, cuyo objetivo principal es dar conformidad a la recepción de variables asociadas, para proceder con las siguientes fases de explotación y desarrollo UFAES. También son útiles para diseñar los controles de validez microdato, previos a la estimación del cambio.

**Controles de Cobertura**

26. Los controles de cobertura se elaboran en primer lugar a niveles agregados, para detectar las frecuencias de presencia de las empresas de la muestra en las diferentes fuentes utilizadas, lo que permite apreciar que los niveles de éxito en la captura desde las diversas fuentes responden a los requisitos de cobertura esperados por UFAES.

Un segundo tipo de control de cobertura se efectúa a nivel mucho más detallado, para cada variable central, mediante el recuento de frecuencias de presencia de la misma en las tras fuentes (ISOC; IVA; IRPF), distinguiendo las que se abastecen de solo una de ellas, o las que están en más de una. Este estudio se realizó para las muestras efectivas S(2008) y S(2009), dando lugar a una larga serie de tablas, que es analizada para confirmar la calidad de la carga en relación con las expectativas del proyecto, así como el peso de cada fuente en la donación de información de variables asociadas. Los estratos y subestratos de análisis se adaptan a los estratos de diseño del tipo actividad *tamaño. Para una variable central determinada, en algunas tabulaciones el control de recepción contempla el concepto de 'estratos CAS' en un año de referencia t (ó t-1), que son aquellos para los que una mayoría (>60%) de unidades tiene presencia de al menos una observación de variable asociada a dicha variable central, en los ficheros

recibidos de la AEAT. Este tipo de controles se repetirá en mundo real para cada par (t; t-1).

**Controles de Contenido**

27. En cuanto a los controles de contenido del microdato recibido aplicados en SIMFAES, para cada estrato CAS (t-1) y cada variable central x, con su asociada a, son del tipo:

| Estrato CAS | $\overline{a}_t^{S(t-1)}$ (1) | $\overline{a}_{t-1}^{S(t-1)}$ (2) | $\overline{a}_t^{S(t)}$ (3) | $\overline{x}_{t-1}^{S(t-1)}$ (4) | $\overline{x}_t^{S(t)}$ (5) | (1)/(2) | (3)/(2) | (5)/(4) |
|---|---|---|---|---|---|---|---|---|
| .... | | | | | | | | |
| | | | | | | | | |
| Total | | | | | | | | |
| | | | | | | | | |

Donde (1) es la media de la variable asociada principal ***a*** con referencia en declaraciones del año t correspondiente a unidades de la muestra S(t-1) y(4) es la media de la variable central x con referencia al año t-1 correspondiente a unidades de la muestra S(t-1). Resto de columnas, análogamente.

Las medias de la variable central o de su asociada principal dentro del estrato se entiende que son calculadas corrigiendo los pesos por falta de respuesta (missing en la asociada).

28. Seguidamente, para cada estrato CAS de una variable central x, se obtienen los cuartiles correspondientes a las distribuciones dentro del estrato de las relaciones de la variable central con su asociada principal (a) y de esta última en varios periodos:

| Estrato CAS | CUARTILES DE LAS DISTRIBUCIONES | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $a_{it}^{S(t-1)} / a_{i(t-1)}^{S(t-1)}$ | | | | $a_{i(t-1)}^{S(t-1)} / x_{i(t-1)}^{S(t-1)}$ | | | | $a_{it}^{S(t)} / x_{it}^{S(t)}$ | | | |
| ... | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 | Q1 | Q2 | Q3 | Q4 |

Los controles de contenido, además de eliminar sospechas de errores en la carga por estratos y variables, permiten afinar la toma de decisiones para diseñar los controles de validez del microdato de la asociada, para pasar, en su caso, a la estimación del cambio en la central.

El análisis de los cuadros de control de la recepción, puede dar lugar a la detección de incidencias especiales observadas en algún subestrato o variables determinados, a tener en cuenta en las fases posteriores de explotación, pudiendo dar lugar a la modificación de algún paso en la estimación, o en un caso extremo a la repetición del proceso de carga, si se trata de error de desarrollo informático o similar.

## X. Control de validez de la asociada y asignación de tipo intermedio de variable estadística

29. Una vez dado el conforme a la recepción llega el momento de efectuar el enlace de las variables asociadas de cada empresa con los datos de las correspondientes variables centrales disponibles en las bases de datos de encuesta en ambos periodos, procediendo a un control de validez de la variable asociada principal, tal como ha sido ya definida. Este control se realiza de manera análoga a un control de outliers de la distribución de asociadas, teniendo en cuenta para su diseño especialmente el cuadro de control de contenido de las distribuciones intercuartílicas por estratos. Es posible de este modo codificar la variable asociada como válida o inválida para la estimación del cambio en la central correspondiente entre t y t-1, produciendo los contadores de frecuencias de éxito en esta fase, para corroborar que los niveles de aceptación son los compatibles con el resto del proceso UFAES.

Tras este control de validez, y a partir del tipo de variable inicial, se procede a asignar una tipología de variable central más elaborada (tipo intermedio de variable) según la codificación siguiente:

**CEAS**: variable central elemental con variable asociada aceptada en t, t-1;

**CEAD**: v. central elemental, sin variable asociada aceptada, pero con variable derivada (suma de v. centrales elementales) asociada, disponible y aceptada en t, t-1;

**CDAS**: central derivada (suma de elementales) con asociadas válidas en t, t-1;

**CEAN/CDAN** (centrales elementales /derivadas sin asociadas en t y/o en t-1);

**CEAF/CDAF** (centrales elementales/derivadas con asociadas, pero inválidas);

Con esta codificación resultante, se genera el fichero Registro Preliminar de Variables Centrales (PREC) que es un instrumento esencial en todo el proceso UFAES. Su cobertura es la muestra completa S(t-1) efectiva, es decir, con registros M'(t-1) y R(t-1), que es el resto de la muestra efectiva S(t-1) que no se integra en la muestra final UFAES (M')U (N). Estos últimos se incluyen por su necesidad de completar el espacio de búsqueda de vecino más próximo para determinar el tipo de variable final.

Contiene, en formato de entrada, solo variables centrales elementales o derivadas. Además de los campos de identificación de la empresa (cuestionario en S(t-1)) y de todas sus variables centrales, sean derivadas o elementales, incluye también el valor de sus asociadas válidas en t y t-1, (vacío para las que no hayan pasado el control de aceptación del cambio), así como identificación de estratos y el valor del dato de la variable central en (t-1).

## XI. Asignación de tipo final de variable central

30. A partir del fichero PREC ya es posible asignar el tipo de variable final en M' que va a condicionar el estimador a utilizar en las diferentes variables del cuestionario.

Este tipo final de variable se obtiene a partir del tipo intermedio y del resultado de la búsqueda de vecino más próximo (NN3, en terminología UFAES) en el conjunto M'(t-1) U R(t-1), para recepción del cambio en la variable asociada válida del donante.

Es decir, la búsqueda de vecino más próximo se realiza en el ámbito de la muestra completa en t-1, para la que existen todas las variables de cuestionario, de las que se seleccionan para componer el vector de búsqueda (ver significado del acrónimo en anexo ASOC-Centrales): CIFNEG. OCUPR4. COMPTOT. EFINAL GPERSAL. ADQTOT. OTROEXT (OTROEX en sociedades). Se trata de variables disponibles para todos los registros de S(t-1), sean de M' o de Resto. El significado de los acrónimos se puede consultar en los anexos.

Se determinan ámbitos disjuntos de búsqueda del vecino NN3 según el estrato y la condición jurídica (persona física y resto), por la diferente disponibilidad de asociadas en cada caso, y se fija un máximo de receptores por donante.

31. Tras la captura, en su caso, del vecino más próximo como donante del cambio en la asociada para aquellas variables centrales sin asociada válida original, se archiva el dato de identificación del donante, la fuente de la asociada válida y sus valores en t, t-1. Esto permite ya asignar el tipo final de variable central, que utiliza la misma codificación que el tipo intermedio sin más que añadir una V solamente a los tipos de variables con asociada inválida o inexistente originalmente pero con asociada válida procedente de donante NN3.

También se establece una tipología a nivel de cuestionario en función del tipo de variables centrales disponibles en el mismo, para ser tenido en cuenta en el momento de la estimación de sus variables no asociadas. Una explotación de frecuencias del tipo final de variable, según cada variable central elemental o derivada, se puede consultar en el anexo Tipo Final-Centrales.

## XII. Estimación del cambio en variables de la submuestra suplementaria M'(t)

32. Una vez determinado el tipo de variable central final, y disponible el valor de sus asociadas validadas del cambio entre los dos periodos del bienio, se procede en primer lugar a la estimación de las variables centrales elementales y derivadas con asociada válida, para después proceder a estimar las centrales elementales con derivada válida, las no centrales de estimación directa y no centrales de estimación indirecta.

Para ello se construye el fichero principal de estimación RE cuya cobertura es la unión de M'(t-1) y N(t), es decir, la muestra efectiva final UFAES (t). Para M', incluye, junto con la información generada en el desarrollo anterior del registro preliminar PREC, las indicaciones sobre el tipo de estimador a utilizar en cada tipología de variable final que se proporciona en documentación auxiliar para desarrollo. En cuanto a los registros N, incluyen su dato para cada variable en t, así como el peso de diseño de cada registro N(t), necesarios para realizar estimaciones de totales y medias por estratos, limitadas a esta submuestra.

El estimador de las variables centrales elementales y centrales derivadas se realiza con un estimador del cambio en la asociada validada (ASOC):

ASOC (Conserva el cambio en la asociada)

Caso centrales asociadas;
a: valores asociados fuente administrativa, válidos;
x: valores de la variable estadística en t; t-
1;==========

$$\left|a_{i(t-1)}\right| \geq 10(mil€) \,\&\, (sig(x_{t-1}) = sig(a_{t-1}) = sig(a_t)) \Rightarrow x_{i(t)} = x_{i(t-1)} \frac{a_{it}}{a_{i(t-1)}};$$

EN OTRO CASO:

$$sig(x_{t-1}) = sig(x_{i(t-1)} + (a_{i(t)} - a_{i(t-1)})) \Rightarrow x_{i(t)} = (x_{i(t-1)} + a_{i(t)} - a_{i(t-1)});$$
$$sig(x_{t-1}) \neq sig(x_{i(t-1)} + (a_{i(t)} - a_{i(t-1)})) \Rightarrow x_{i(t)} = x_{i(t-1)};$$

33. Para el resto de variables (centrales elementales con derivada asociada y no centrales de estimación directa), los estimadores principales son del tipo combinado con una componente de vecino más próximo y otra de conservación de estructura del valor en t-1. Los estimadores generales, que tienen algunas excepciones reflejadas en documentación auxiliar, son el CEUNN/CCENN, que se formulan a continuación. Otras variables no centrales de estimación indirecta o de estimación idéntica (que mantienen exactamente el dato t-1 de nivel o de estructura % respecto a su explicativa) son estimadas una vez realizada la estimación principal, en la fase inmediatamente anterior a la integración del fichero UFAES final M'U N:

**CEUNN**

$$\hat{y}_{iht} = ((1 - a) * \hat{x}_{ih(t)} * \left( \frac{y_{hjt}}{x_{hjt}} \right) + a * y_{ih(t-1)} \left( \frac{\hat{x}_{hit}}{x_{hi(t-1)}} \right)$$

y variable no central, o central no asociada, real en t-1, imputada en t;
x: variable central con asociada válida ;
t-1; t; años de referencia; h estrato ;
i unidad a imputar ;
j unidad vecino más próximo según procedimiento NN1 o NN2
0<= a <=1; Valor para el caso general: ½

Variables componentes del vector distancia (ver significado de los acrónimos en anexo):

NN1: CIFNEG; GPERSAL; OCUPR4; COMPTOT;

NN2 utilizado para variables adqtot, ventoto y sus componentes: se incluye ADQTOT y se excluye GPERSAL;

El estrato de búsqueda de donantes se construye como agregación parcial de estratos de diseño para lograr un número de donantes que se considera suficiente.Aquí, la definición de vecino más próximo (NN1/NN2 en notación UFAES) es diferente al NN3, dado que en este caso se trata de búsqueda en t, es decir, que deberán ser variables centrales con asociada (en NN3 se trata de variables estadísticas en t-1). En anexo figura el algoritmo genérico de búsqueda (NN).

En paralelo a la primera realización UFAES en 2013 se seguirá investigando el potencial de estimadores asistidos por modelos de regresión multivariante, utilizando como explicativas las variables centrales con asociadas válidas. Dada

la necesidad de implantar UFAES lo antes posible, por el ahorro tan significativo en los presupuestos de trabajo de campo del INE para 2013 (referencia 2012), ha sido necesario posponer el diagnóstico sobre la posible introducción de estimadores con estas técnicas en siguientes ciclos de UFAES.

**CCENN**

$$\hat{x}_{iht} = (CN)^{CCE}_{iht} * \left( a * \frac{x_{hi(t-1)}}{CN_{hi(t-1)}} + (1-a) * \frac{x_{hjt}}{CN_{hjt}} \right)$$

Estimador para variables centrales CEAN/CEAF/CDAF/CDAN (sin asociada) en registros cuya cifra de negocios (variable central principal) tampoco está asociada a la fuente tributaria. Conserva el cambio en el estrato con efecto del vecino más próximo NN.i unidad a imputar ;j, vecino más próximo de la unidad i en t

x, variable CEAN/CEAF/CDAF/CDAN, real en t-1; imputada en t; CN: Cifra de negocios, conocida en t-1 para i, j, pero solamente disponible en t para la unidad j. Se estima para la unidad i en t con el estimador CCE, que es el estimador del cambio en las medias del estrato UFAES.

t-1; t; años de referencia; h estrato ;

## XIII. Reiteración de muestras artificiales SIMFAES

34. Los trabajos de simulación del proyecto de uso de fuentes administrativas en las encuestas estructurales de empresas (SIMFAES), incluidos todos los trabajos preparatorios, fueron ejecutados por el equipo de trabajo a lo largo de un año aproximadamente entre 2011-2012, a tiempo parcial.

Se obtuvieron 4 pares de submuestras simuladas de la Encuesta Anual de Servicios del INE (EAS) en el bienio de referencia 2008-2009 con el esquema UFAES. Cada muestra consiste en la unión de una submuestra representativa a nivel nacional extraída de la muestra teórica EAS 2009, y reducida por las incidencias de trabajo de campo (muestra efectiva) a la muestra N(2009), y una submuestra extraída de la muestra teórica EAS 2008, actualizada a M (2009) por incidencias de 2008 y posteriormente a M'(2009) por bajas en el marco entre 2008 y 2009. En cada una de las submuestras se mantienen las incidencias de cambio de estrato tanto en M(t-1) como en N(t), con el fin de calcular los pesos de la muestra integrada MUN para estimaciones UFAES, como se describe a continuación.

**Selección de muestras N(t)**

35. La muestra global de la EAS-09 fue de unas 139.000 empresas, de las cuales cerca de 10.000 eran de Navarra y País Vasco, excluidas de la simulación. Se calcularon tamaños muestrales para lograr estimaciones a nivel nacional, resultando un tamaño global aproximado de 89.000 unidades. Descontando todas las de Navarra y País Vasco, se seleccionaron 4 submuestras aleatorias que se han cruzado con la muestra efectiva global, resultando submuestras efectivas de tamaños muestrales variables: 59.502; 59.549; 59.528 y 59.502, que en media corresponden con los tamaños aproximados que cabe esperar en mundo real, es

decir, una reducción media de un -35,9% sobre la muestra teórica EAS. Dado que la implementación en el primer año UFAES (2013 con referencia 2012) se hará sobre una subpoblación de aproximadamente la mitad del marco del sector servicios (actividades EAS excluido comercio, div: 45 46 y 47) el ahorro neto directo esperado en mundo real en ese primer año de realización se estima en unos 600.000 €, un -18% aproximadamente del presupuesto EAS del año anterior (antes de UFAES). Si se contabiliza el ahorro indirecto debido a que UFAES facilita la no externalización de otras operaciones de trabajos de campo, la reducción de coste en 2013 llegará a ser de cerca de un millón de euros.

**Muestras M'(t)**

36. Se obtuvieron 4 submuestras (las llamadas M'(t)) de aproximadamente 39.000 unidades cada una, obtenidas de la siguiente manera:

Se obtuvieron los tamaños por estrato de la muestra suplementaria M(08), con un incremento aproximado del 15%, para tener en cuenta su posterior actualización.

De la muestra teórica de la EAS-08, se seleccionan 4 submuestras M(08), con coordinación negativa con cada submuestra N(09). Luego se cruzan con la muestra efectiva S(08), y posteriormente se actualizan con el Dirce 09, eliminando las bajas, obteniendo las 4 submuestras finales M'(09), con tamaños de 39.075 ;39.021 ;39.133 ;y 39.084 empresas.

En la serie de gráficos 1 se puede observar la distribución de la muestra por divisiones según su composición en M y N. El grafico 1.3 permite visualizar también la presencia de unidades >20 ocupados en la muestra M, introducidos como prueba experimental. Dada su repercusión en algunos indicadores de evaluación SIMFAES y su escaso número, no se incluirán en la submuestra M real.

## XIV. Cálculo de Pesos SIMFAES

37. Las bases de datos de entrega con las muestras reiteradas SIMFAES contienen la información a utilizar para el cálculo de los pesos necesarios para realizar estimaciones tanto con la muestra MUN como con la N solamente (necesario para estimadores que usan el cambio en las medias). El procedimiento para llegar a los mismos se formula a continuación, referido al escenario teórico de simulación SIMFAES EAS2008-EAS2009. No obstante, en aras de su generalización, se establece la notación como en el caso de *mundo real*. La diferencia fundamental es que en el primer caso, se dispone de una muestra S(t) de la que se extrae la submuestra SIMFAES N(t), mientras que en el caso *real,* solo existe N(t), y no existe S(t).

A este fin, para evitar confusiones, se utiliza la símbolo P (para el marco) y p (para la muestra, sea SIMFAES *o real* en el caso más general) en el año de referencia t, con el fin de diferenciarla de la notación S(t) utilizada en esquemas anteriores para referirse a la muestra original de 2009 de la que se extrae la submuestra simulada N(t).

**Cálculo del peso SIMFAES de la muestra integrada N ∪ M':**

38. Para las empresas de N, que es la muestra SIMFAES de t (EAS09), se considera el estrato inicial y final observados en t (EAS09) y, para las empresas de M', que es la submuestra efectiva SIMFAES en t-1 (EAS08), ya corregida de

bajas en el marco de 2009, el estrato inicial y final que tuvieron en t-1 (EAS08), independientemente del estrato al que pertenezca en t (2009).

Para el cálculo del factor SIMFAES en las unidades que han cambiado de estrato, siendo su origen (selección) en el estrato h, en N (t) y M'(t-1) (estas últimas, independientemente del estrato al que pertenezcan una vez imputado en t) usamos:

$$f_{cc} = \frac{P_h(t)}{p_h(t) + m_h^{'}(t-1)};$$

siendo el numerador el tamaño del marco en t (09) y el denominador la suma del tamaño teórico de M'(t-1) (antes de bajas y cambios de estrato pero después del contraste con el marco en t) y de N(t).

Para las que no han cambiado de estrato en N(t) o en M(t-1) (estas últimas, independientemente del estrato al que pertenezcan una vez imputado en t) usamos:

$$f_{sc} = \frac{P_h(t)\left(1 - \frac{b_h(t) + b_h^{'}(t-1)}{p_h(t) + m_h^{'}(t-1)}\right) - \sum_{k \neq h} \frac{P_h(t)}{p_h(t) + m_h^{'}(t-1)}(p_h^k + m_h^{'k})}{p_h^* + m_h^{'(*)}},$$

donde

$b_h$ , $b_h^{'}$ son respectivamente el número de empresas, duplicadas o cerradas temporalmente en el estrato h de la muestra total SIMFAES N(t) en t (2009) o en M'(t-1), esta última la submuestra (teórica) SIMFAES en t-1 (EAS08), después de contraste con el marco.

$p_h^k$ es el número de empresas seleccionadas en la muestra SIMFAES en t (2009) en el estrato h y que realmente pertenecen al estrato k.

$m_h^{'k}$ es el número de empresas seleccionadas en la muestra SIMFAES en t-1 (2008) corregida de contraste con el marco en t, en el estrato h y que realmente pertenecían al estrato k en t-1 (2008) independientemente al estrato de llegada SIMFAES en t.

$p_h^*$ es el número de empresas de la muestra efectiva en el estrato h que no ha cambiado de estrato en N(t)

$m_h^{'(*)}$ es el número de empresas de la muestra efectiva M'(t-1) que no han cambiado de estrato en t-1, independientemente de que su estrato actual (h) en t sea diferente al de origen (y llegada) en t-1.

**Calculo del peso SIMFAES de la submuestra sin integrar N (t):**

39. Para las empresas de N (submuestra efectiva SIMFAES en t) se considera el estrato inicial y final que tuvieron en t (EAS09).

Para el cálculo del factor SIMFAES en las unidades que han cambiado de estrato, siendo su origen en el estrato h, usamos:

$$f_{cc} = \frac{P_h}{p_h}$$

y para las que no han cambiado de estrato en N(t) con origen/destino en h usamos:

$$f_{sc} = \frac{P_h\left(1-\frac{b_h}{p_h}\right)-\sum_{k\neq h}\frac{P_h}{p_h}(p_h^k)}{p_h^*},$$

donde

$b_h$ es el número de empresas, duplicadas o cerradas temporalmente en el estrato h de la muestra SIMFAES teórica N(t) de t (EAS09).

$p_h^k$ es el número de empresas seleccionadas en la muestra total N(t) (en t (2009) en el estrato h y que realmente pertenecen al estrato k.

$p_h^*$ es el número de empresas de la muestra en el estrato h que no ha cambiado de estrato en N(t).

## XV. Resultados de la Simulación

40. Como se ha ido apuntando en los apartados anteriores, los diferentes pasos de explotación, y muy especialmente los controles de cobertura y contenido realizados a la recepción de datos externos, producen salidas de resultados para evaluación y ajuste, en su caso, de las siguientes fases.

En cuanto al análisis de resultados propiamente dicho, es bien conocida la dificultad de evaluación de resultados de la estimaciones obtenidas mediante muestras reiteradas para el mismo periodo de encuesta, por el problema de cómo discernir qué parte de la diferencia observada en las estimaciones es efecto del sesgo introducido por el modelo de simulación, del sesgo de la fuente externa, de la propia variabilidad del muestreo repetido o del sesgo debido a defectos del marco e incidencias del trabajo de campo. Superar estas incertidumbres en el caso SIMFAES exige asumir en particular la validez de las aproximaciones para estimar el efecto conjunto de la varianza y sesgo aportado de la muestra repetida N, para lo cual se concedió cierta prioridad en la evaluación al indicador del error medio cuadrático (EMC) de la muestra N calculado con las cuatro estimaciones simuladas de cada variable, dado que se trata de la parte de la muestra final UFAES no contaminada por el modelo.

41. La evaluación se realizó en dos fases:

i) Una evaluación a partir de un conjunto de indicadores del Sesgo y EMC respecto al dato objetivo (EAS 09) obtenidos por procedimientos de Monte Carlo a partir de las cuatro muestras simuladas, para un número limitado de variables principales (centrales) de la estructura contable de la empresa.

**ii)** Un análisis detallado, relativo a un mayor número de variables, concordante con los envíos EAS para satisfacer los requerimientos del reglamento UE de estadísticas de estructura económica de las empresas SBS), que se realiza ya con el fichero final integrado con todas las variables de encuesta. Las salidas de este análisis, a su vez, se actualizan reiterando la explotación una vez detectado y corregido, en su caso, el impacto de la presencia de outliers en la submuestra M

(en la submuestra N se supone neutralizado este efecto por el tratamiento habitual de preparación de ficheros finales de la EAS). Este análisis detallado se ha realizado por parsimonia solamente con una muestra simulada, por lo que ya no se utilizan los estadísticos Monte Carlo obtenidos con las cuatro muestras simuladas, sino otros indicadores que se describen seguidamente.

**Indicadores para evaluación SIMFAES:**

42

B: Indicadores de Sesgo;

EMC: Indicadores de Error Medio Cuadrático

M U N: Muestras SIMFAES integrada; 4 reiteraciones.

N: Submuestra obtenida en recogida real T. de Campo en t; id.

M: submuestra suplementaria UFAES; id.

EAS09; EAS08; Eas10: valores estimados de la encuesta principal EAS

$$\hat{B}_{M\cup N}\% = \frac{100}{4\hat{X}_{\acute{E}AS09}}\sum_{S}^{4}\hat{X}_{´S(M\cup N)} - \hat{X}_{EAS09} = \frac{100}{\hat{X}_{\acute{E}AS09}}(\overline{\hat{X}}_{(M\cup N)} - \hat{X}_{EAS09});$$

$$E\hat{M}C_{M\cup N}\% = \frac{100}{\hat{X}_{\acute{E}AS09}}\sqrt{\frac{1}{4}\sum_{S}^{4}(\hat{X}_{´S(M\cup N)} - \hat{X}_{EAS09})^2};$$

$$E\hat{M}C_{N}\% = \frac{100}{\hat{X}_{\acute{E}AS09}}\sqrt{\frac{1}{4}\sum_{S}^{4}(\hat{X}_{´S(N)} - \hat{X}_{EAS09})^2}$$

$$DifTV_{09/08} = (\frac{\overline{\hat{X}}_{(M\cup N)}}{\hat{X}_{\acute{E}AS08}} - \frac{\hat{X}_{EAS09}}{\hat{X}_{EAS08}})*100 = \frac{\hat{B}_{(M\cup N)}}{\hat{X}_{\acute{E}AS08}}\cdots*100;$$

$$\hat{B}^{rel1}_{(M\cup N)} = \frac{\hat{B}_{(M\cup N)}}{E\hat{M}C_{N}};$$

$$E\hat{M}C^{rel1}(\%) = \frac{E\hat{M}C_{M\cup N}}{\hat{X}_{EAS09}}*100$$

$$E\hat{M}C^{rel2} = \frac{E\hat{M}C_{M\cup N}}{E\hat{M}C_{N}};(3)$$

**Indicadores para el análisis detallado a nivel de clases y comunidad autónoma**

43. Para el análisis por clases y comunidades autónomas se han utilizado los indicadores que siguen a continuación, aunque obviamente pueden ser objeto de adecuación en cada periodo de ejecución real UFAES.

D1) E= ABS(0,5*(EAS08+EAS10)/EAS09-1) (Nota: estima la raíz del EMC de EAS 09 en una subpoblación, a partir de una sola muestra UFAES).

D2) (MUN-EAS09)/0,5*(ABS(N-EAS09)+E*EAS09). (Nota: El denominador estima el EMC del valor objetivo EAS 09)

D3) (MUN-EAS09)/ABS(N-EAS09). Idem

D4) Dif. Tasa (MUN-N)/EAS08 %

D5) REPERCUSIONES % de la diferencia (MUN-EAS09)/EAS08 en la clase CNAE, en el cambio en la División.

**Control de Outliers**

44. Al igual que en las fases de explotación rutinarias de la encuesta principal, también es necesario realizar un control de outliers UFAES, pero en la muestra estimada M'(t-1) solamente, dado que este control, en mundo real, ya habrá sido aplicado a N(t) antes de la aplicación del procedimiento UFAES. Este control, para SIMFAES, se realiza solo en una de las muestras simuladas, en un escenario *como si* el dato EAS09 no fuera conocido.

La detección de outliers y su tratamiento -usualmente corrigiendo sus pesos- darán lugar a un re-examen de los indicadores detallados por clases, y así sucesivamente hasta la estabilidad.

Los indicadores de detección principales, por clases y comunidades autónomas son:

O1) Diferencia de Tasas (MUN-N) 09/08 %

O2) Repercusión % de la Diferencia (MUN-N)/08 del cambio en la clase respecto al cambio en la división.

O3) Dif. Relativa (MUN-N)/08, respecto a la media de valores absolutos de esas diferencias en la división. El denominador es un estimador de la raíz del EMC en la división.

O4) Dif tasa MUN-N %

## XVI. Comentarios a los resultados

45. Los gráficos del anexo son una muestra de un conjunto más amplio que se obtuvieron para todas las variables centrales y centrales más importantes, a nivel nacional y, de manera más limitada, por comunidades autónomas. Para su interpretación se puede consultar la CNAE 2009 por secciones y divisiones del ámbito EAS en anexo A.5).

Los gráficos de la serie 1, que ya han sido comentados en la sección relativa al muestreo, son necesarios para comprender el comportamiento de los datos en los diferentes colectivos, dada la importancia del tamaño muestral relativo de M' y de N para la interpretación del impacto del modelo en los mismos. Los de la serie 2 representan los intervalos de aceptación simétricos definidos por

1+ (-) EMC(MUN)^0,5 /EAS09, por una parte, y de 1+ (-) EMC(N)^0,5/EAS09, por otra, con el EMC según se han definido anteriormente, calculado con las cuatro simulaciones. Incluye una línea que representa el dato Media (MUN)/EAS09 por lo que su distancia al eje =1 es un indicador del sesgo.

Los gráficos de la serie 3.1-3.4 representan la relación Sesgo MUN/EMC(N) y los de la serie 4.1-4.4 representan la relación entre EMC(MUN) y EMC(N) que conviene analizar de manera conjunta con la serie 2.

Se observa en términos generales que las clases con sesgos elevados se mantienen normalmente dentro del intervalo definido por el EMC de la muestra N (muestra real, sin componente estimada UFAES). Las excepciones encontradas hasta el momento responden a la presencia de alguna observación muy influyente del microdato asociado o debido a que se presentan en clases y/o en comunidades autónomas con muestras muy pequeñas para las que es de suponer que el EMC de N obtenido en SIMFAES tampoco sea una estimación muy eficiente.

Los gráficos de la serie 5 presentan el sesgo y la estimación de EMC(N), en relación con el peso de la muestra M en la integrada MUN, dentro de cada subpoblación, con el fin de detectar si existe un efecto a medida que aumenta el tamaño en % de la submuestra M (diámetro de la burbuja) en la muestra integrada. Los de la serie 6 ilustran también este efecto en los EMC de la muestrea N y de MUN. Las series se elaboran según subconjuntos determinados por el peso en la división (comercio) o en la sección (resto), y para todos los ámbitos.

Una vez más, las anomalías observadas aquí no parecen depender del peso de M en la muestra MUN de la clase CNAE de que se trate, lo cual nos indica de nuevo la gran importancia del impacto en los sesgos SIMFAES de los EMC de Monte Carlo de la propia muestra N, ajenos a la componente M.

## XVII. Conclusiones y Recomendaciones

46. El procedimiento UFAES está preparado para ejecución en la EAS 2012 (trabajo de campo en 2013), al estar a punto todos los desarrollo necesarios.

Existen algunas incertidumbres a la vista de algunas diferencias en las estimaciones simuladas respecto al dato objetivo (EAS09) que se observan en colectivos muy detallados, aunque es altamente probable que se deban principal o exclusivamente a la propia variabilidad poblacional y a los sesgos por incidencias de marco y del trabajo de campo, que se ponen de manifiesto al repetir muestras simuladas, en las que se detecta una aportación muy significativa a la dispersión de las estimaciones procedente de la componente real N.

47. Es recomendable no obstante no diseñar la muestra UFAES 2012 con tamaños de M tan elevados como en SIMFAES, utilizados de modo experimental, y excluir de la selección en M, al menos en los primeros periodos anuales UFAES, las empresas con más de 20 ocupados, por su escasa repercusión en la disminución de carga global y costes y porque este estrato, que en las simulaciones SIMFAES tenía una cierta presencia en algunas divisiones de la submuestra M (gráfico 1.3), ha introducido un extra de variabilidad en las simulaciones.

Es necesario continuar con la investigación en estimadores estocásticos, con menor componente determinística que los usados en SIMFAES09, como los que proporcionan métodos de imputación multivariante como el IveWare (Universidad de Michigan), utilizados en la encuesta de condiciones de vida en el INE, aunque se prevén dificultades de implantación en este tipo de modelos, con un escenario de gran número de variables monetarias relativas a la estructura contable de las empresas.

Asimismo, se recomienda realizar pruebas de simulación con la Encuesta Anual Industrial de Empresas, para su posible introducción a las técnicas UFAES. Es de esperar, no obstante, que la reducción muestral alcanzable en esa operación no sea de la envergadura de la que es posible aplicar en la EAS.

## Referencias/References


Brackstone G.J. (1987), Issues in the use of Administrative Records.

EC (2011), Regulation concerning structural business statistics.

INE (2011), Encuesta Anual de Servicios. Metodología (Annual Business Survey on the Services Sectors. Methodology). www.ine.es .

Saralegui, J. (2003), Integration of external data from Tax and Public Accounts in the Central Business Register. 54th ISI Session (Berlin).

Saralegui, J. (2004), Use of tax data for sample design under confidentiality restrictions. Conference Q2004. Mainz (Germany).

Eurostat (2012), Framework for the Integration of Business Statistics (FRIBS).

Eurostat (2011), Daniel Lewis on behalf of work package 3 of the ESSnet on the Use of Administrative and Accounts Data in Business Statistics. Using administrative data to estimate survey variables not directly available from administrative sources.


.

# ANEXO

# A1) NN.

## Criterio de búsqueda del vecino más próximo

Para la unidad *i* del estrato de vecindad *h* contenido en la división *d*, denotamos por $\vec{x}_{dhit}$ al vector que contiene las variables que queremos usar para elegir al NN.

Primero calculamos la matriz de covarianza

$$C_d = \frac{1}{n_d} \sum_{h,i} \left(\vec{x}_{hit} - \vec{\mu}_d\right)\left(\vec{x}_{hit} - \vec{\mu}_d\right)^T$$

donde

$$\vec{\mu}_d = \frac{1}{n_d} \sum_{h,i} \vec{x}_{hit} \ .$$

La matriz de covarianza la calculamos empleando la división entera en la que está contenido el estrato porque al haber estratos pequeños, el estimador con el estrato sería poco robusto.

Entonces, para cada *i* en el conjunto de receptores del estrato de vecindad *h* y *j* en el conjunto de donantes (también en el estrato de vecindad *h*), calculamos

$$dist_{ij} = \left(\vec{x}_{dhit} - \vec{x}_{dhjt}\right)^T C_d^{-1} \left(\vec{x}_{dhit} - \vec{x}_{dhjt}\right)$$

Elegimos como NN para *i* el *j* que hace mínimo $dist_{ij}$ .

# A2) CD
# Correspondencia de variables centrales derivadas y elementales. Personas jurídicas

| **Variables derivadas de las personas jurídicas** | | |
|---|---|---|
| **NOMBRE SAS** | **DESCRIPCIÓN** | **CALCULO** |
| **CONSUMO** | Consumo de mercaderías, mat. primas, aprovisionamientos, etc. | COMPTOT-EXMERF-EXMATF-EXAPRF+EXMERP+EXMATP+EXAPRP+DETMER |
| **ADQTOT** | Adquisiciones inmovilizado material | ADQTERR_1+ADQINF_1+ADQEDIF_1+ADQINS_1+ADQTRAN_1+ADQMAQ_1+ADQOTR_1+ADINTERR+ADINEDIF+ADININS |
| **VENTOTO** | Ventas inmovilizado material | VENTERR_1+VENINF_1+VENEDIF_1+VENINS_1+VENTRAN_1+VENMAQ_1+VENOTR_1+VEINTERR+VEINEDIF+VEININS |
| **OTROEX** | Otros gastos de explotación | GASTOT*+GASTGES+PERDVA+IMPRODI+OTROIMP |
| **EPPIO** | Existencias a principios de año | EXMERP+EXMATP+EXAPRP+EXPRODP |
| **EFINAL** | Existencias a final de año | EXMERF+EXMATF+EXAPRF+EXPRODF |

* COMTOT y GASTOT son variables del cuestionario pero, a su vez, son sumas de otras variables del cuestionario que se detallan a continuación:
COMPOT=(COMPMER+COMPMAT+COMPAPR+COMPSUBC)
GASTOT=(GASTINV+GASTARR+GASTREP+GASTPROF+GASTRANS+GASTBAN+GASTPUBL+GASTSUM+GASTOTRO)

# A2) CD
# Correspondencia de variables derivadas y elementales centrales. Personas físicas

| **Variables derivadas de las personas físicas** | | |
|---|---|---|
| **NOMBRE SAS** | **DESCRIPCIÓN** | **CALCULO** |
| **CONSUMO** | Consumo de mercaderías, materias primas, aprovisionamientos, etc. | COMPTOT-EXMERF-EXMATF-EXAPRF-EXPRODF+EXMERP+EXMATP+EXAPRP+EXPRODP |
| **ADQTOT** | Adquisiciones inmovilizado material | ADQTERR_1+ADQINF_1+ADQEDIF_1+ADQINS_1+ADQTRAN_1+ADQMAQ_1+ADQOTR_1+ADINTERR+ADINEDIF+ADININS |
| **VENTOTO** | Ventas inmovilizado material | VENTERR_1+VENINF_1+VENEDIF_1+VENINS_1+VENTRAN_1+VENMAQ_1+VENOTR_1+VEINTERR+VEINEDIF+VEININS |
| **IMPUES** | Tributos fiscalmente deducibles | IMPRODI+OTROIMP |
| **OTROING** | Otros ingresos | EXPINT+EXPLSUB1+EXPLSUB2+EXPLOTRO+INGPROV |
| **OTROSGP** | Otros gastos de personal | GPERIND+GPERPEN+GPEROTR |
| **OTROEXT** | Otros servicios exteriores | GASTINV+GASTRANS+GASTPRI+GASTBAN+GASTPUBL+GASTOTRO |

# A3) *ASOC-CENTRALES*
## Correspondencia de variables centrales y asociadas en las declaraciones tributarias
## Impuesto de sociedades

| | I.S 08 | EAS08 | EAS09 |
|---|---|---|---|
| **Importe neto de la cifra de negocios.** | [255] | CIFNEG | CIFNEG |
| **Trabajos realizados por la empresa para su activo.** | [259] | EXPINT | EXPINT |
| **Aprovisionamientos.** | [260] | COMPTOT-VARTOT+VARPROD+DETMER | COMPTOT-VARTOT+VARPROD+DETMER |
| **Ingresos accesorios y otros de gestión corriente.** | [266] | EXPLOTRO | EXPLOTRO |
| **Subvenciones de explotación incorporadas al resultado del ejercicio.** | [269] | EXPLSUB1 | EXPLSUB1 |
| **Gastos de personal** | [270] | GPERTOT | GPERTOT |
| **Sueldos, salarios y asimilados** | [271] | GPERSAL | GPERSAL |
| **Indemnizaciones** | [273] | GPERIND | GPERIND |
| **Seguridad Social a cargo de la empresa** | [274] | GPERSS | GPERSS |
| **Retribuciones a largo plazo mediante sistemas de aportación o prestación definida** | [275] | GPERPEN | GPERPEN |
| **Retribuciones mediante instrumentos de patrimonio** | [276] | GPERPAT | GPERPAT |
| **Otros gastos sociales** | [277] | GPEROTR | GPEROTR |
| **Otros gastos de explotación** | [279] | GASTOT+GASTGES+PERDVA+IMPRODI+OTROIMP | GASTOT+GASTGES+PERDVA+IMPRODI+OTROIMP |
| **Amortizaciones del inmovilizado** | [284] | DOTTOT | DOTTOT |
| **Imputación de subvenciones de inmovilizado no financiero y otras** | [285] | EXPLOSUB2 | EXPLOSUB2 |
| **Exceso de provisiones** | [286] | INGPROV | INGPROV |
| **Deterioro y resultado por enajenaciones del inmovilizado** | [287] | DETINM | DETINM |
| **Otros resultados** | [295] | ---------- | OTRESU |
| **Resultado de explotación** | [296] | RESEXPL | RESEXPL |
| **Ingr. financiero+ Dif. pos. cambio** | [320]+[321]+[297]+[312]*(+) | INGFIN | INGFIN |
| **Gastos financieros y gastos asimilados+ Dif. Neg. cambio** | [322]+[323]+[305] +[309]+ [312]*(-) | GASTFIN | GASTFIN |
| **Resultado de la cuenta de pérdidas y ganancias** | [500] | RESERJER | RESERJER |
| **Existencias a ppio. de año** | [138] | EXMERP+EXMATP+EXAPRP+EXPRODP | EXMERP+EXMATP+EXAPRP+EXPRODP |
| **Existencias a final de año** | [138_AA] | EXMERF+EXMATF+EXAPRF+EXPRODF | EXMERF+EXMATF+EXAPRF+EXPRODF |
| **Personal asalariado fijo(30/09)** | [041](*) | REFIHC+REFIMC+REFIHP+REFIMP | REFIHC+REFIMC+REFIHP+REFIMP |
| **Personal asalariado no fijo (30/09)** | [042](*) | REVHC+REVMC+REVHP+REVMP | REVHC+REVMC+REVHP+REVMP |

## IRPF

| | I.R.P.F | | | | | EAS08 | EAS09 |
|---|---|---|---|---|---|---|---|
| | E.D.N | E.D.N | E.D.S | E.D.S | E.O | | |
| | 2008 | 2009 | 2008 | 2009 | | | |
| **Ingresos de explotación** | [106] | [106] | [106] | [106] | -- | CIFNEG | CIFNEG |
| **Otros ingresos** | [107] | [107] | [107] | [107] | -- | EXPINT+EXPLSUB1+EXPLSUB2+EXPLOTRO+INGPROV | EXPINT+EXPLSUB1+EXPLSUB2+EXPLOTRO+INGPROV |
| **Consumos de explotación** | [110] | [110] | [110] | [110] | -- | COMPTOT-EXMERF-EXMATF-EXAPRF-EXPRODF+EXMERP+EXMATP+EXAPRP+EXPRODP | COMPTOT-EXMERF-EXMATF-EXAPRF-EXPRODF+EXMERP+EXMATP+EXAPRP+EXPRODP |
| **Sueldos y salarios** | [111] | [111] | [111] | [111] | -- | GPERSAL | GPERSAL |
| **Seguridad Social a cargo de la empresa** | [112] | [112] | [112] | [112] | -- | GPERSS+NGPERSS | GPERSS+NGPERSS |
| **Otros gastos de personal** | [113] | [113] | [113] | [113] | -- | GPERIND+GPERPEN+GPEROTR | GPERIND+GPERPEN+GPEROTR |
| **Arrendamientos y cánones** | [114] | [114] | [114] | [114] | -- | GASTARR | GASTARR |
| **Reparaciones y conservación** | [115] | [115] | [115] | [115] | -- | GASTREP | GASTREP |
| **Servicios de profesionales independientes** | [116] | [116] | [116] | [116] | -- | GASTPROF | GASTPROF |
| **Suministros** | [117] | -- | [117] | | | GASTSUM | --- |
| **Otros servicios exteriores** | [118] | [117] | [118] | [117] | -- | GASTINV+GASTRANS+GASTPRI+GASTBAN+GASTPUBL+(GASTOTRO) | GASTINV+GASTRANS+GASTPRI+GASTBAN+GASTPUBL+GASTSUM+(GASTOTRO) |
| **Tributos fiscalmente deducibles** | [119] | [118] | [119] | [118] | -- | IMPRODI+OTROIMP | IMPRODI+OTROIMP |
| **Gastos financieros** | [120] | [119] | [120] | [119] | -- | GASTFIN | GASTFIN |
| **Amortizaciones: Dotaciones del ejercicio fiscalmente deducibles** | [121] | [120] | [121] | [120] | -- | DOTTOT | DOTTOT |
| **Pérdidas por deterioro del valor de los elementos patrimoniales: importes deducibles** | -- | [121] | -- | [121] | -- | --- | DETINM+PERDVA |
| **Personal asalariado** | -- | -- | -- | -- | Modulo1 | OCUPR1+...+OCUPR4/4 | OCUPR1+...+OCUPR4/4 |
| **Personal no asalariado** | -- | -- | -- | -- | Modulo2 | OCUPNR1+...+OCUPNR4/4 | OCUPNR1+...+OCUPNR4/4 |

## IVA 2008

| *IVA 390* | **IVA 390 ordinario** | **IVA 390 simplificado** | **IVA 392 Gran Empresa** | **EAS08** | *IVA 392* |
|---|---|---|---|---|---|
| **10. Volumen de operaciones. Operaciones realizadas en el ejercicio. Total volumen de operaciones** | [108] | [108] | [108] | CIFNEG | **9. Volumen de operaciones. Operaciones realizadas en el ejercicio. Total volumen de operaciones** |
| **10. Volumen de operaciones. Operaciones realizadas en el ejercicio. Entregas de bienes inmuebles y operaciones financieras no habituales + Entregas de bienes de inversión** | [106]+[107] | ---- | [106]+[107] | VENTOTO | **9. Volumen de operaciones. Operaciones realizadas en el ejercicio. Entregas de bienes inmuebles y operaciones financieras no habituales + Entregas de bienes de inversión** |
| **5. Operaciones realizadas en régimen general. IVA deducible. Total bases imponibles bienes y servicios, importaciones, adquisiciones intracomunitarias corrientes** | [48]+[52]+[56] | ---- | [48]+[52]+[56] | COMPTOT | **5. Operaciones realizadas en régimen general. Deducciones. Total bases imponibles bienes y servicios, importaciones, adquisiciones intracomunitarias corrientes** |
| **5. Operaciones realizadas en régimen general. IVA deducible. Total bases imponibles operaciones interiores, importaciones, adquisiciones intracomunitarias de bienes de inversión** | [50]+[54]+[58] | ---- | [50]+[54]+[58] | ADQTOT | **5. Operaciones realizadas en régimen general. Deducciones. Total bases imponibles operaciones interiores, importaciones, adquisiciones intracomunitarias de bienes de inversión** |
| **6. Operaciones realizadas en régimen simplificado. A. Nº de unidades de módulo (Personal empleado)** | | Módulo1 (A) | ---- | OCUPT1+...+OCUPT4 | |
| **6. Operaciones realizadas en régimen simplificado. Cuotas soportadas operaciones corrientes (D)** | | apdo. 6 (D) | ---- | COMPTOT+GASTOT | |

## IVA 2009

| | IVA<br>390 (ordinario) | IVA<br>390 (simplificado) | EAS09 |
|---|---|---|---|
| **10. Volumen de operaciones. Operaciones realizadas en el ejercicio. Total volumen de operaciones** | [108] | [108] | CIFNEG |
| **10. Volumen de operaciones. Operaciones realizadas en el ejercicio. Entregas de bienes inmuebles y operaciones financieras no habituales + Entregas de bienes de inversión** | [106]+[107] | ---- | VENTOTO |
| **5. Operaciones realizadas en régimen general. IVA deducible. Total bases imponibles bienes y servicios, importaciones, adquisiciones intracomunitarias corrientes** | [48]+[512]+[52]+[56] | ---- | COMPTOT |
| **5. Operaciones realizadas en régimen general. IVA deducible. Total bases imponibles operaciones interiores, importaciones, adquisiciones intracomunitarias de bienes de inversión** | [50]+[520]+[54]+[58] | ---- | ADQTOT |
| **6. Operaciones realizadas en régimen simplificado. A. Nº de unidades de módulo (Personal empleado)** | | Módulo1 (A) | OCUPT1+...+OCUPT4 |
| **6. Operaciones realizadas en régimen simplificado. Cuotas soportadas operaciones corrientes (D)** | | apdo. 6 (D) | COMPTOT+GASTOT |

# A4) REC._ TIPO FINAL DE VBLE CENTRAL ELEMENTAL O DERIVADA

### TIPO FINAL DE VARIABLE ; TOTALES MUESTRA

| COD_VBLE | CEAF | CEAN | CDAF | CDAN | CDAS | CDASV | CEAS | CEASV | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| ADQTOT | 0 | 0 | 614 | 2.760 | 22.746 | 12.848 | 0 | 0 | 38.968 |
| CIFNEG | 267 | 2.209 | 0 | 0 | 0 | 0 | 30.928 | 5.566 | 38.970 |
| CONSUMO | 0 | 0 | 339 | 2.466 | 27.502 | 8.661 | 0 | 0 | 38.968 |
| DETINM | 534 | 334 | 0 | 0 | 0 | 0 | 20.234 | 2.533 | 23.635 |
| DOTTOT | 1.424 | 2.762 | 0 | 0 | 0 | 0 | 28.436 | 6.348 | 38.970 |
| EFINAL | 0 | 0 | 144 | 14.261 | 19.330 | 5.233 | 0 | 0 | 38.968 |
| EPPIO | 0 | 0 | 2 | 0 | 38.968 | 0 | 0 | 0 | 38.970 |
| EXAPRP | 0 | 0 | 0 | 2 | 0 | 0 | 38.968 | 0 | 38.970 |
| EXMATP | 0 | 0 | 0 | 2 | 0 | 0 | 38.968 | 0 | 38.970 |
| EXMERP | 0 | 0 | 0 | 2 | 0 | 0 | 38.968 | 0 | 38.970 |
| EXPRODP | 0 | 0 | 0 | 2 | 0 | 0 | 38.968 | 0 | 38.970 |
| GASTFIN | 796 | 2.589 | 0 | 0 | 0 | 0 | 29.575 | 6.010 | 38.970 |
| GPERPAT | 11 | 288 | 0 | 0 | 0 | 0 | 21.705 | 1.631 | 23.635 |
| GPERSAL | 113 | 2.466 | 0 | 0 | 0 | 0 | 30.375 | 6.016 | 38.970 |
| GPERSS | 1.304 | 2.909 | 0 | 0 | 0 | 0 | 28.866 | 5.891 | 38.970 |
| IMPUES | 2 | 0 | 62 | 2.840 | 9.668 | 2.765 | 0 | 0 | 15.337 |
| INGFIN | 606 | 336 | 0 | 0 | 0 | 0 | 20.425 | 2.268 | 23.635 |
| OCUPR4 | 0 | 0 | 339 | 2.466 | 27.502 | 8.661 | 0 | 0 | 38.968 |
| OTROEX | 0 | 0 | 697 | 375 | 19.633 | 2.928 | 0 | 0 | 23.633 |
| OTROEXT | 0 | 0 | 961 | 2.984 | 8.387 | 3.003 | 0 | 0 | 15.335 |
| OTROING | 2 | 0 | 172 | 2.276 | 9.415 | 3.472 | 0 | 0 | 15.337 |
| OTROSGP | 2 | 0 | 88 | 2.865 | 9.587 | 2.795 | 0 | 0 | 15.337 |
| RESEJER | 1.340 | 418 | 0 | 0 | 0 | 0 | 18.784 | 3.091 | 23.633 |
| RESEXPL | 2.355 | 520 | 0 | 0 | 0 | 0 | 17.193 | 3.565 | 23.633 |
| VENTOTO | 0 | 0 | 180 | 2.760 | 27.814 | 8.214 | 0 | 0 | 38.968 |

### TIPO FINAL DE VARIABLE ; % MUESTRA M '(2009)

| COD_VBLE | CEAF | CEAN | CDAF | CDAN | CDAS | CDASV | CEAS | CEASV | TOTAL |
|---|---|---|---|---|---|---|---|---|---|
| ADQTOT | 0,00 | 0,00 | 1,58 | 7,08 | 58,37 | 32,97 | 0,00 | 0,00 | 100 |
| CIFNEG | 0,69 | 5,67 | 0,00 | 0,00 | 0,00 | 0,00 | 79,36 | 14,28 | 100 |
| CONSUMO | 0,00 | 0,00 | 0,87 | 6,33 | 70,58 | 22,23 | 0,00 | 0,00 | 100 |
| DETINM | 2,26 | 1,41 | 0,00 | 0,00 | 0,00 | 0,00 | 85,61 | 10,72 | 100 |
| DOTTOT | 3,65 | 7,09 | 0,00 | 0,00 | 0,00 | 0,00 | 72,97 | 16,29 | 100 |
| EFINAL | 0,00 | 0,00 | 0,37 | 36,60 | 49,60 | 13,43 | 0,00 | 0,00 | 100 |
| EPPIO | 0,00 | 0,00 | 0,01 | 0,00 | 99,99 | 0,00 | 0,00 | 0,00 | 100 |
| EXAPRP | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 99,99 | 0,00 | 100 |
| EXMATP | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 99,99 | 0,00 | 100 |
| EXMERP | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 99,99 | 0,00 | 100 |
| EXPRODP | 0,00 | 0,00 | 0,00 | 0,01 | 0,00 | 0,00 | 99,99 | 0,00 | 100 |
| GASTFIN | 2,04 | 6,64 | 0,00 | 0,00 | 0,00 | 0,00 | 75,89 | 15,42 | 100 |
| GPERPAT | 0,05 | 1,22 | 0,00 | 0,00 | 0,00 | 0,00 | 91,83 | 6,90 | 100 |
| GPERSAL | 0,29 | 6,33 | 0,00 | 0,00 | 0,00 | 0,00 | 77,94 | 15,44 | 100 |
| GPERSS | 3,35 | 7,46 | 0,00 | 0,00 | 0,00 | 0,00 | 74,07 | 15,12 | 100 |
| IMPUES | 0,01 | 0,00 | 0,40 | 18,52 | 63,04 | 18,03 | 0,00 | 0,00 | 100 |
| INGFIN | 2,56 | 1,42 | 0,00 | 0,00 | 0,00 | 0,00 | 86,42 | 9,60 | 100 |
| OCUPR4 | 0,00 | 0,00 | 0,87 | 6,33 | 70,58 | 22,23 | 0,00 | 0,00 | 100 |
| OTROEX | 0,00 | 0,00 | 2,95 | 1,59 | 83,07 | 12,39 | 0,00 | 0,00 | 100 |
| OTROEXT | 0,00 | 0,00 | 6,27 | 19,46 | 54,69 | 19,58 | 0,00 | 0,00 | 100 |
| OTROING | 0,01 | 0,00 | 1,12 | 14,84 | 61,39 | 22,64 | 0,00 | 0,00 | 100 |
| OTROSGP | 0,01 | 0,00 | 0,57 | 18,68 | 62,51 | 18,22 | 0,00 | 0,00 | 100 |
| RESEJER | 5,67 | 1,77 | 0,00 | 0,00 | 0,00 | 0,00 | 79,48 | 13,08 | 100 |
| RESEXPL | 9,96 | 2,20 | 0,00 | 0,00 | 0,00 | 0,00 | 72,75 | 15,08 | 100 |
| VENTOTO | 0,00 | 0,00 | 0,46 | 7,08 | 71,38 | 21,08 | 0,00 | 0,00 | 100 |

## A5) CNAE 2009

### G; Comercio al por mayor y al por menor; reparación de vehículos de motor y motocicletas

45; Venta y reparación de vehículos de motor y motocicletas
46; Comercio al por mayor e intermediarios, excepto de vehículos de motor y motocicletas
47; Comercio al por menor, excepto vehículos de motor y motocicletas

### H; Transporte y almacenamiento

49; Transporte terrestre y por tubería
50; Transporte marítimo y por vías navegables interiores
51; Transporte aéreo
52; Almacenamiento y actividades anexas al transporte
53; Actividades postales y de correos

### I; Hostelería

55; Servicios de alojamiento
56; Servicios de comidas y bebidas

### J; Información y comunicaciones

58; Edición
59; Actividades cinematográficas, de vídeo y de programas de televisión, grabación de sonido y edición musical
60; Actividades de programación y emisión de radio y televisión
61; Telecomunicaciones
62; Programación, consultoría y otras actividades relacionadas con la informática
63; Servicios de información

### L; Actividades inmobiliarias

68; Actividades inmobiliarias

## M; Actividades profesionales, científicas y técnicas

69; Actividades jurídicas y de contabilidad
70; Actividades de las sedes centrales; actividades de consultoría de gestión empresarial
71; Servicios técnicos de arquitectura e ingeniería; ensayos y análisis técnicos
72; Investigación y desarrollo
73; Publicidad y estudios de mercado
74; Otras actividades profesionales, científicas y técnicas
75; Actividades veterinarias

### N; Actividades administrativas y servicios auxiliares

77; Actividades de alquiler
78; Actividades relacionadas con el empleo
79; Actividades de agencias de viajes, operadores turísticos, servicios de reservas y actividades relacionadas con los mismos
80; Actividades de seguridad e investigación
81; Servicios a edificios y actividades de jardinería
82; Actividades administrativas de oficina y otras actividades auxiliares a las empresas

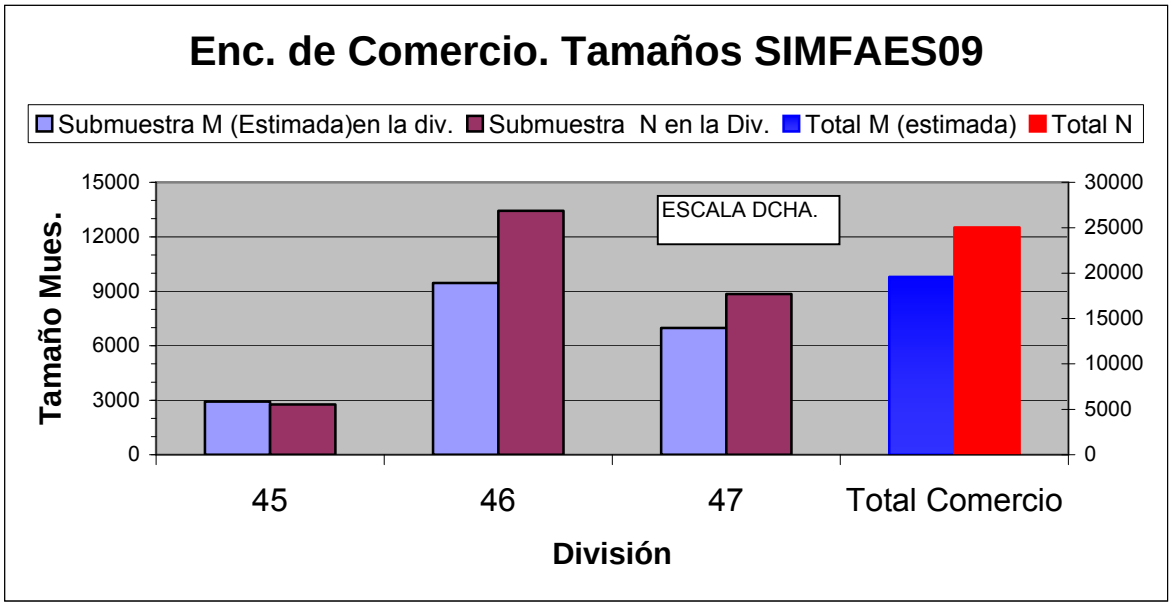
## R; Actividades artísticas, recreativas y de entretenimiento

90; Actividades de creación, artísticas y espectáculos
91; Actividades de bibliotecas, archivos, museos y otras actividades culturales
92; Actividades de juegos de azar y apuestas
93; Actividades deportivas, recreativas y de entretenimiento

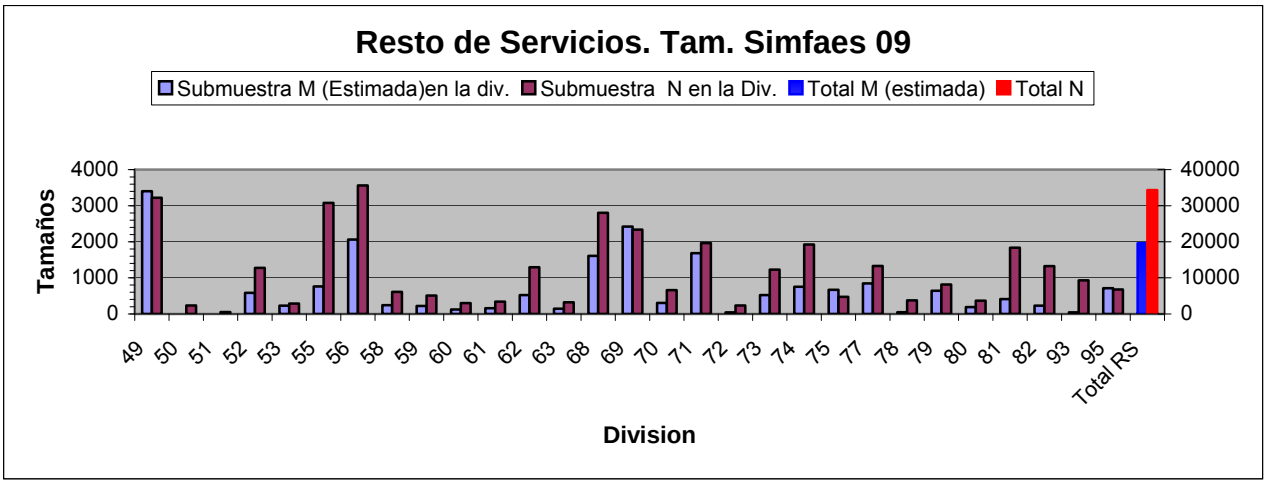
## S; Otros servicios; de los cuales sólo llevamos:

95; Reparación de ordenadores, efectos personales y artículos de uso doméstico

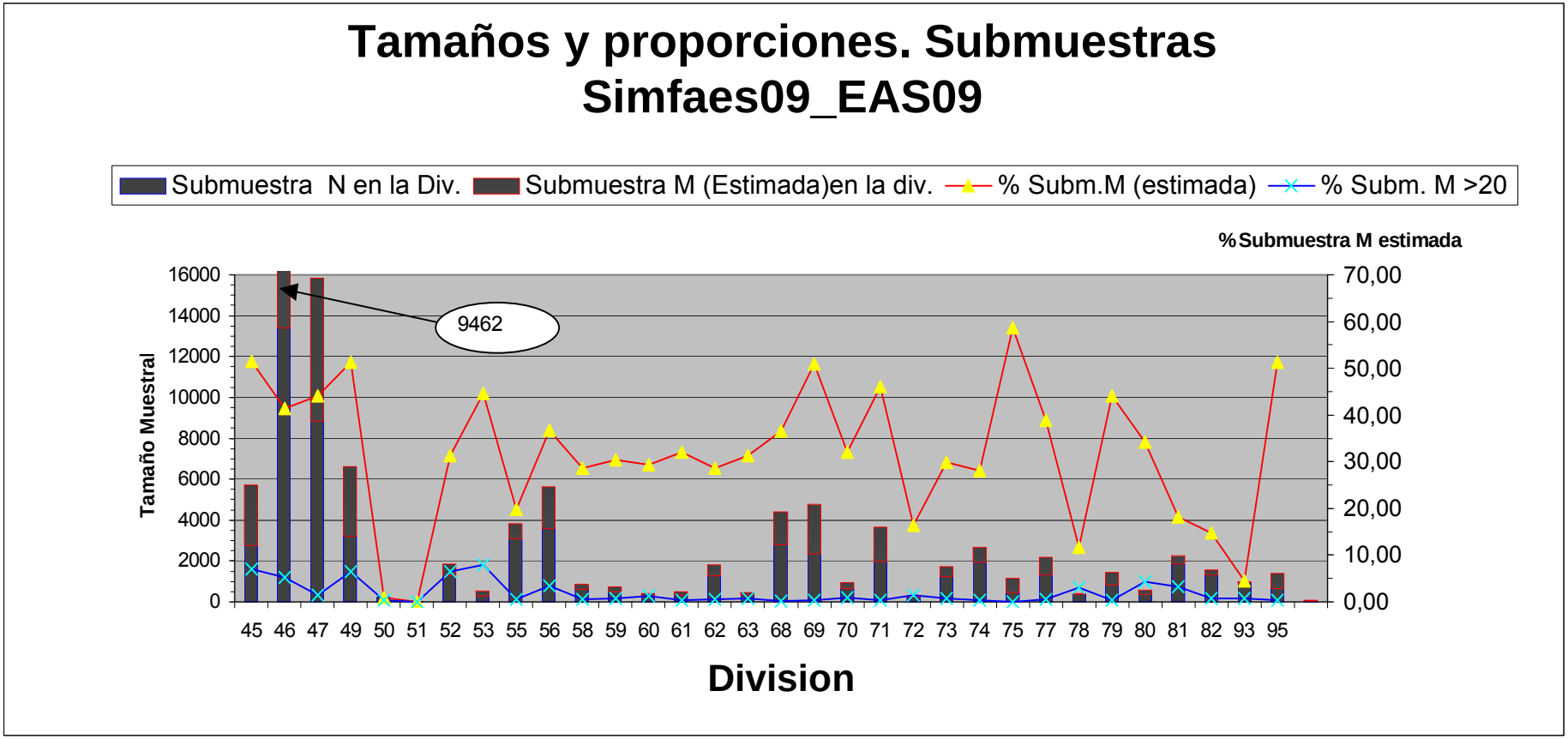
**G 1.1**


Enc. de Comercio. Tamaños SIMFAES09
Submuestra M (Estimada)en la div.
Submuestra N en la Div.
Total M (estimada)
Total N
Tamaño Mues.
15000
12000
9000
6000
3000
0
30000
25000
20000
15000
10000
5000
0
ESCALA DCHA.
45
46
47
Total Comercio
División

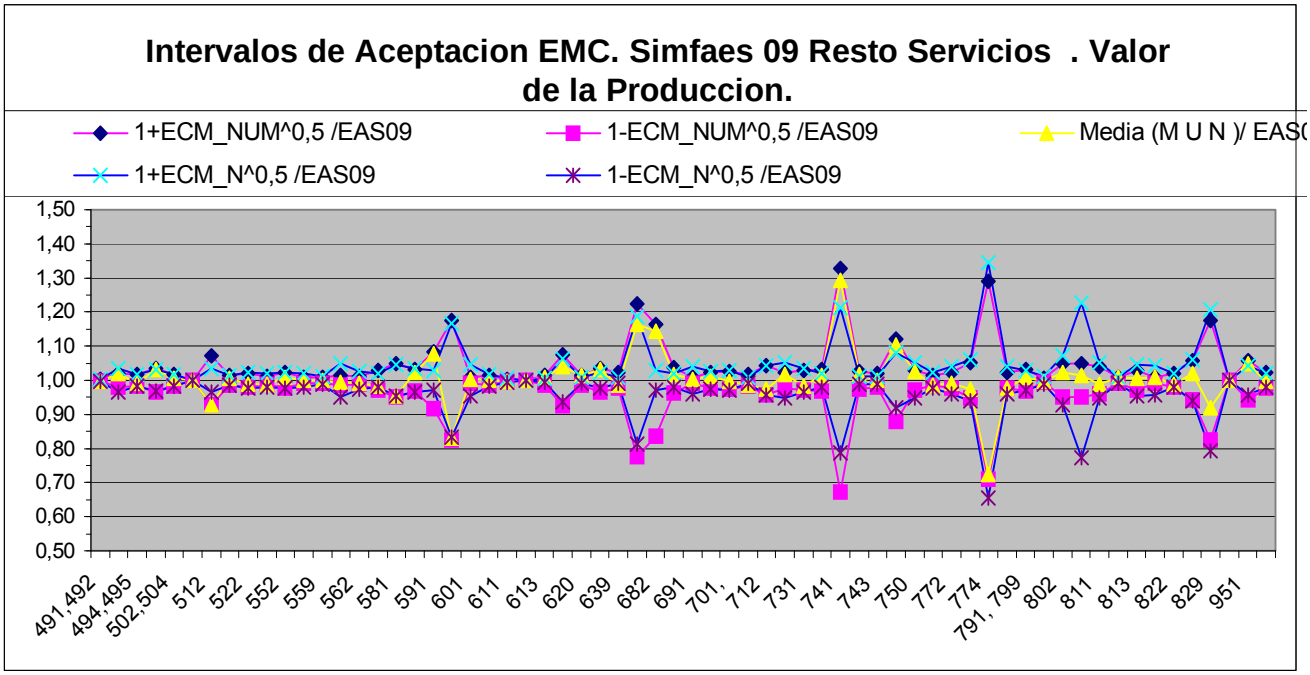
G 1.2


Resto de Servicios. Tam. Simfaes 09
Submuestra M (Estimada)en la div.
Submuestra N en la Div.
Total M (estimada)
Total N
Tamaños
4000
3000
2000
1000
0
40000
30000
20000
10000
0
49
50
51
52
53
55
56
58
59
60
61
62
63
68
69
70
71
72
73
74
75
77
78
79
80
81
82
93
95
Total RS
Division

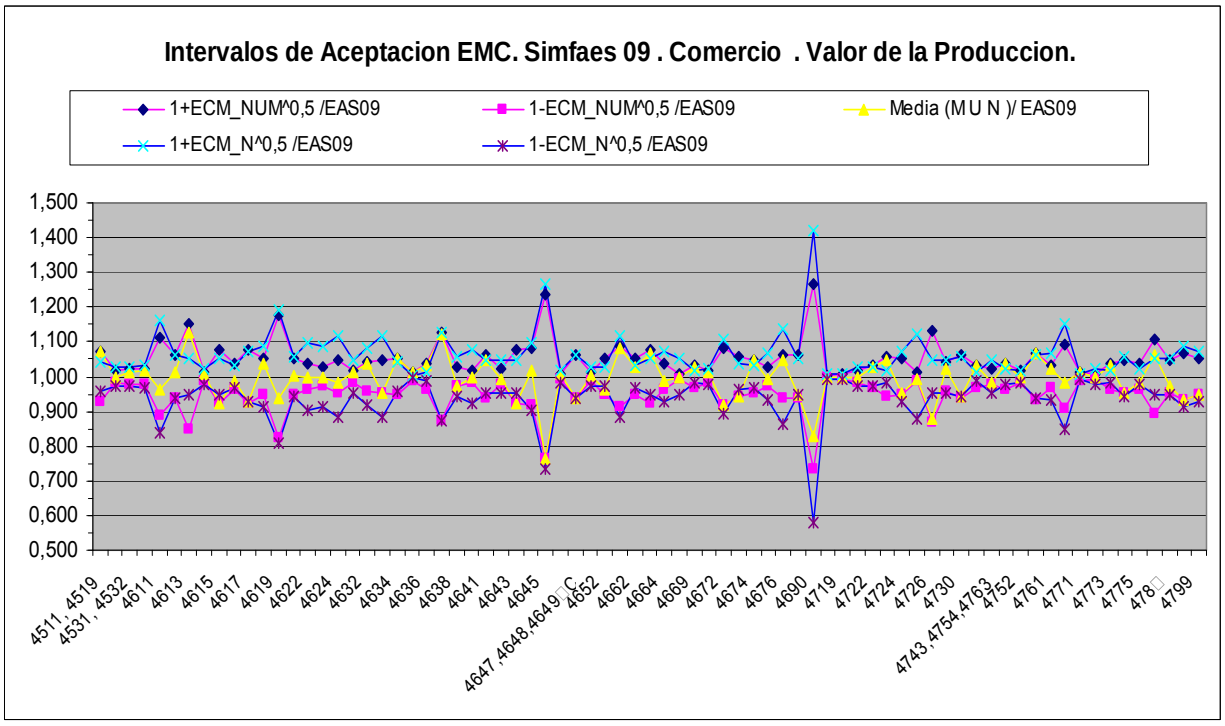
**G 1.3**


Tamaños y proporciones. Submuestras
Simfaes09_EAS09
Submuestra N en la Div.
Submuestra M (Estimada)en la div.
% Subm.M (estimada)
% Subm. M >20
%Submuestra M estimada
Tamaño Muestral
9462
16000
14000
12000
10000
8000
6000
4000
2000
0
70,00
60,00
50,00
40,00
30,00
20,00
10,00
0,00
45 46 47 49 50 51 52 53 55 56 58 59 60 61 62 63 68 69 70 71 72 73 74 75 77 78 79 80 81 82 93 95
Division

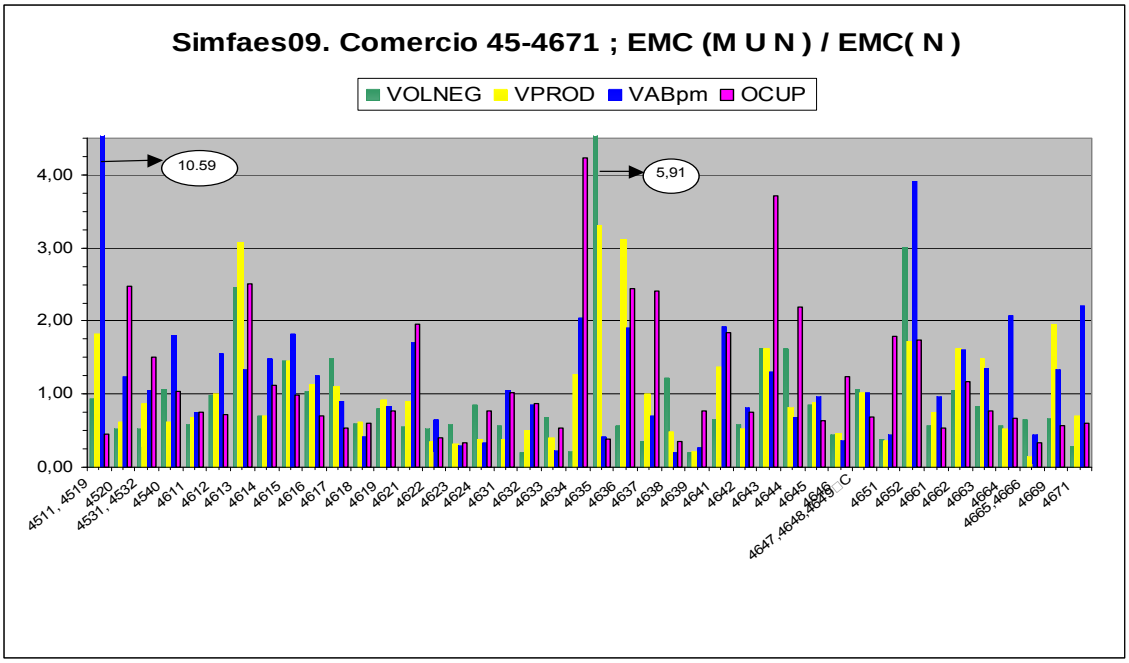
**G 2.1**


Intervalos de Aceptacion EMC. Simfaes 09 Resto Servicios . Valor de la Produccion.
1+ECM_NUM^0,5 /EAS09
1-ECM_NUM^0,5 /EAS09
Media (M U N )/ EAS(
1+ECM_N^0,5 /EAS09
1-ECM_N^0,5 /EAS09
1,50
1,40
1,30
1,20
1,10
1,00
0,90
0,80
0,70
0,60
0,50
491,492
494,495
502,504
512
522
552
559
562
581
591
601
611
613
620
639
682
691
701,
712
731
741
743
750
772
774
791, 799
802
811
813
822
829
951

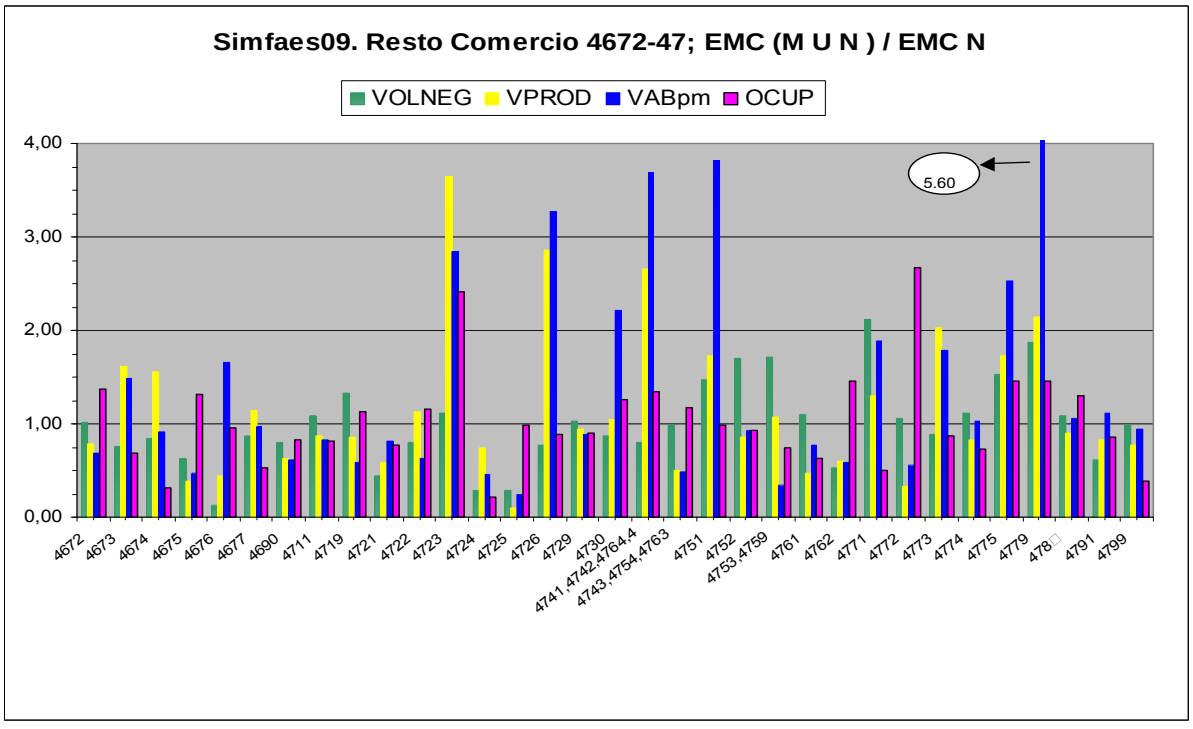
**G 2-2**


Intervalos de Aceptacion EMC. Simfaes 09 . Comercio . Valor de la Produccion.
1+ECM_NUM^0,5 /EAS09
1-ECM_NUM^0,5 /EAS09
Media (M U N )/ EAS09
1+ECM_N^0,5 /EAS09
1-ECM_N^0,5 /EAS09
1,500
1,400
1,300
1,200
1,100
1,000
0,900
0,800
0,700
0,600
0,500
4511, 4519
4531, 4532
4611
4613
4615
4617
4619
4622
4624
4632
4634
4636
4638
4641
4643
4645
4647,4648,4649
4652
4662
4664
4669
4672
4674
4676
4690
4719
4722
4724
4726
4730
4743,4754,4763
4752
4761
4771
4773
4775
4799

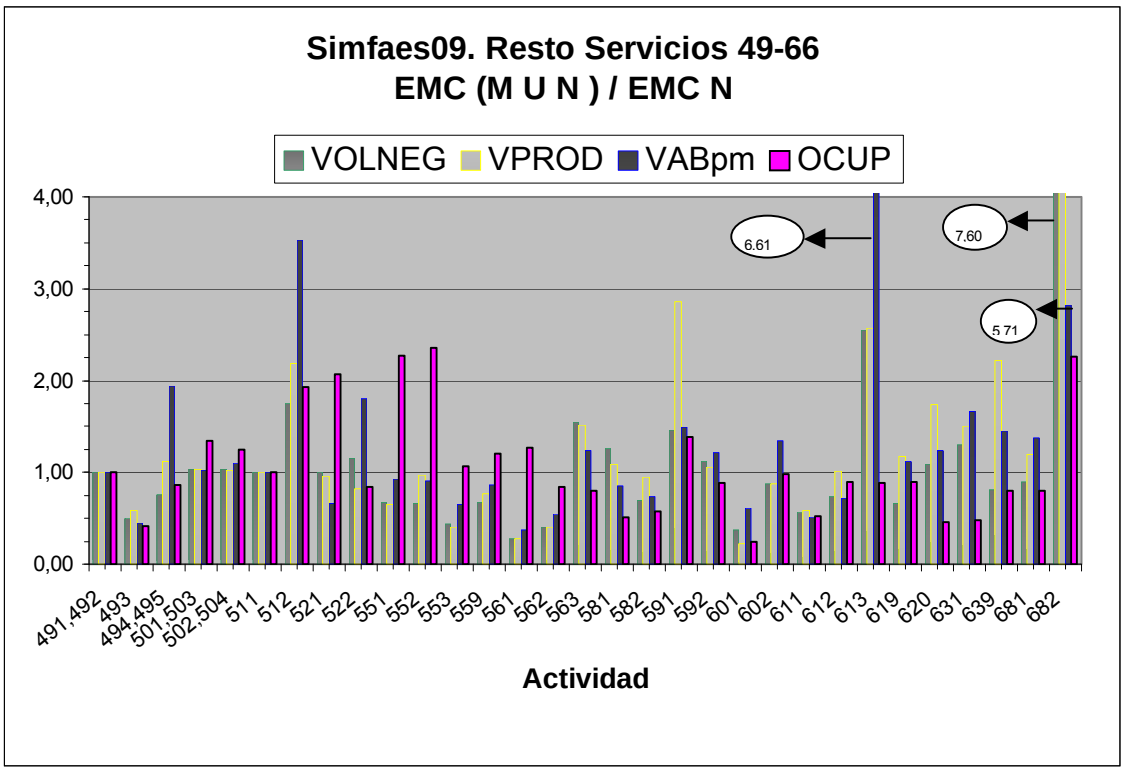
G 3.1


Simfaes09. Comercio 45-4671 ; EMC (M U N ) / EMC( N )
VOLNEG
VPROD
VABpm
OCUP
10.59
5,91
4,00
3,00
2,00
1,00
0,00

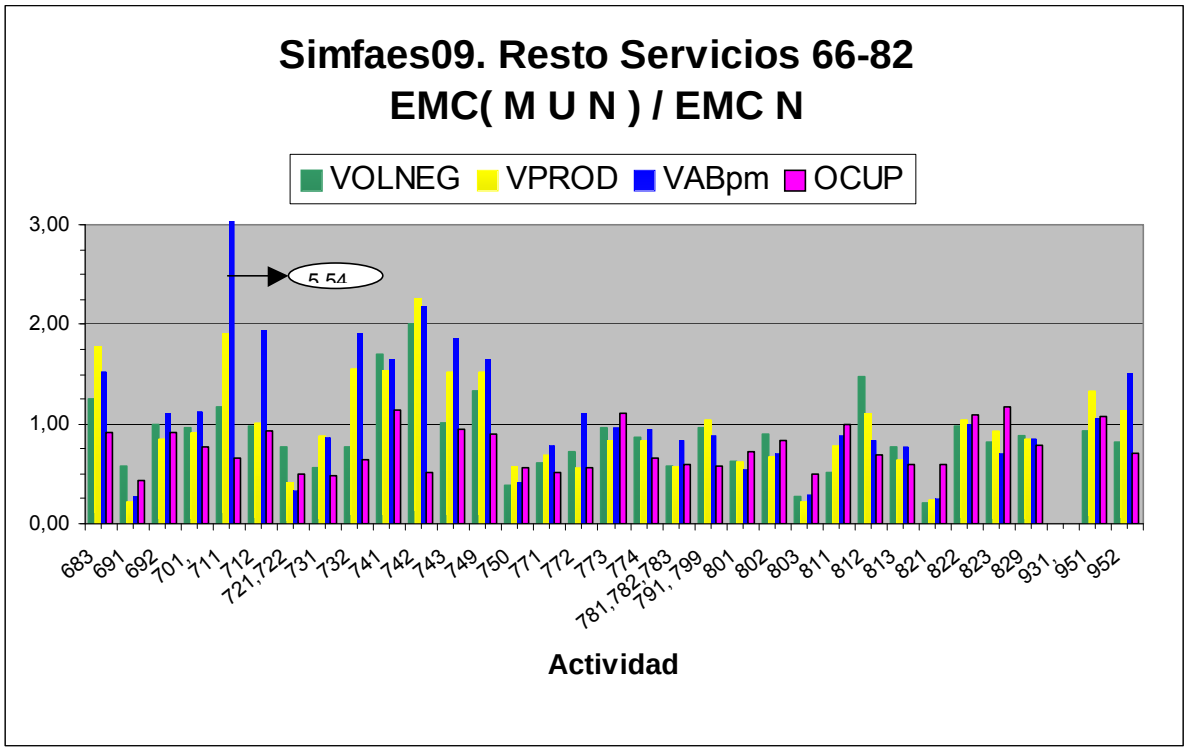
**G 3.2**


Simfaes09. Resto Comercio 4672-47; EMC (M U N ) / EMC N
VOLNEG
VPROD
VABpm
OCUP
5.60
4,00
3,00
2,00
1,00
0,00

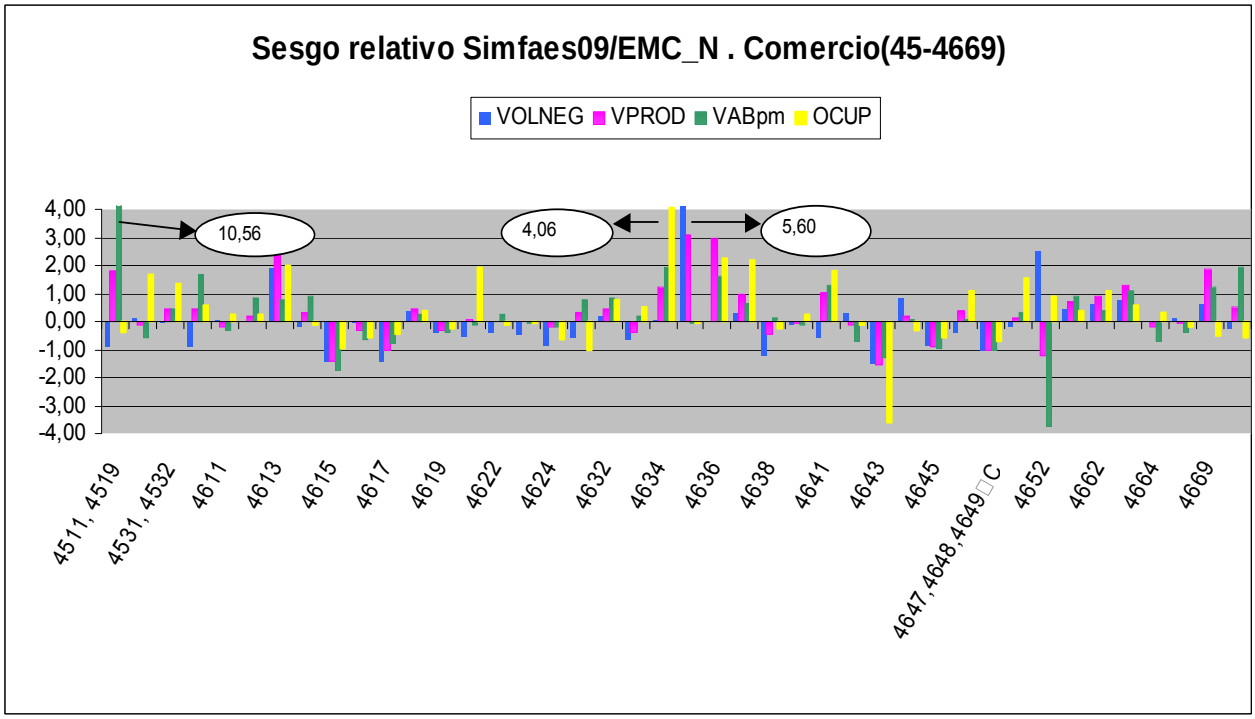
**G 3.3**


Simfaes09. Resto Servicios 49-66
EMC (M U N ) / EMC N
VOLNEG
VPROD
VABpm
OCUP
4,00
3,00
2,00
1,00
0,00
6.61
7.60
5.71
491,492
493
494,495
501,503
502,504
511
512
521
522
551
552
553
559
561
562
563
581
582
591
592
601
602
611
612
613
619
620
631
639
681
682
Actividad

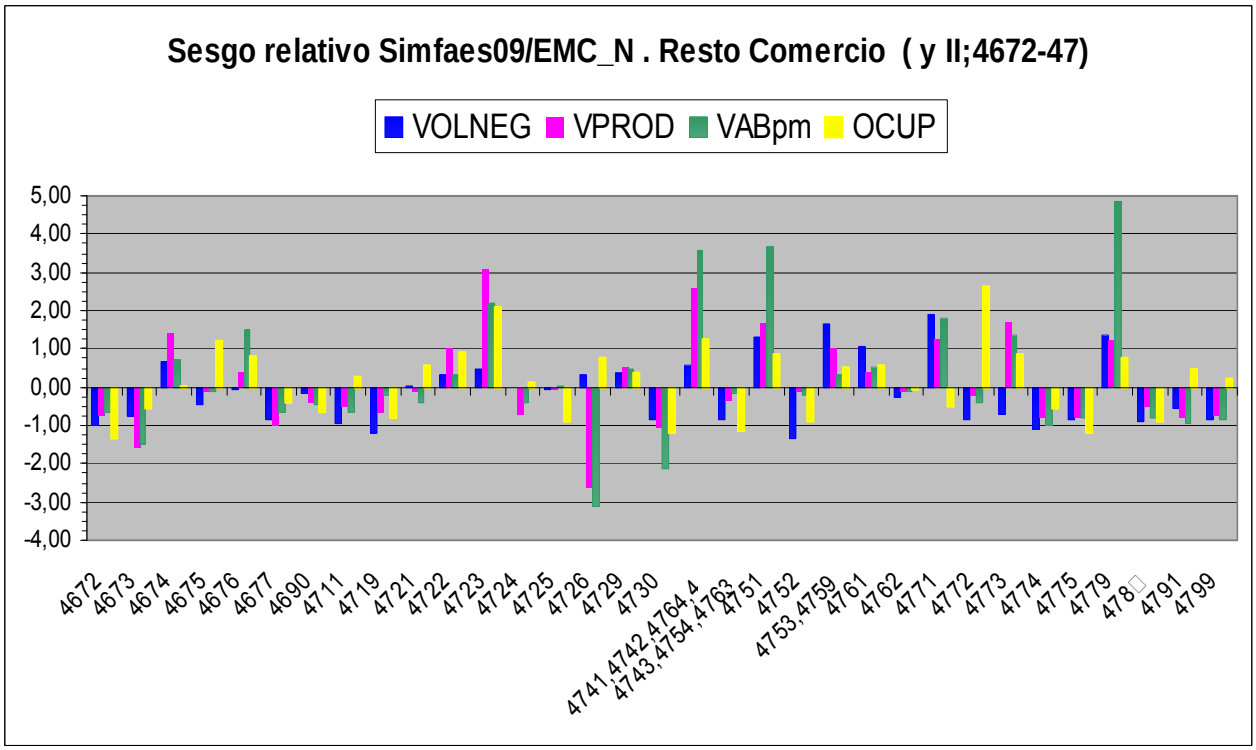
**G 3.4**


Simfaes09. Resto Servicios 66-82
EMC( M U N ) / EMC N
VOLNEG
VPROD
VABpm
OCUP
3,00
2,00
1,00
0,00
5.54
683
691
692
701
711
712
721,722
731
732
741
742
743
749
750
771
772
773
774
781,782,783
791, 799
801
802
803
811
812
813
821
822
823
829
931
951
952
Actividad

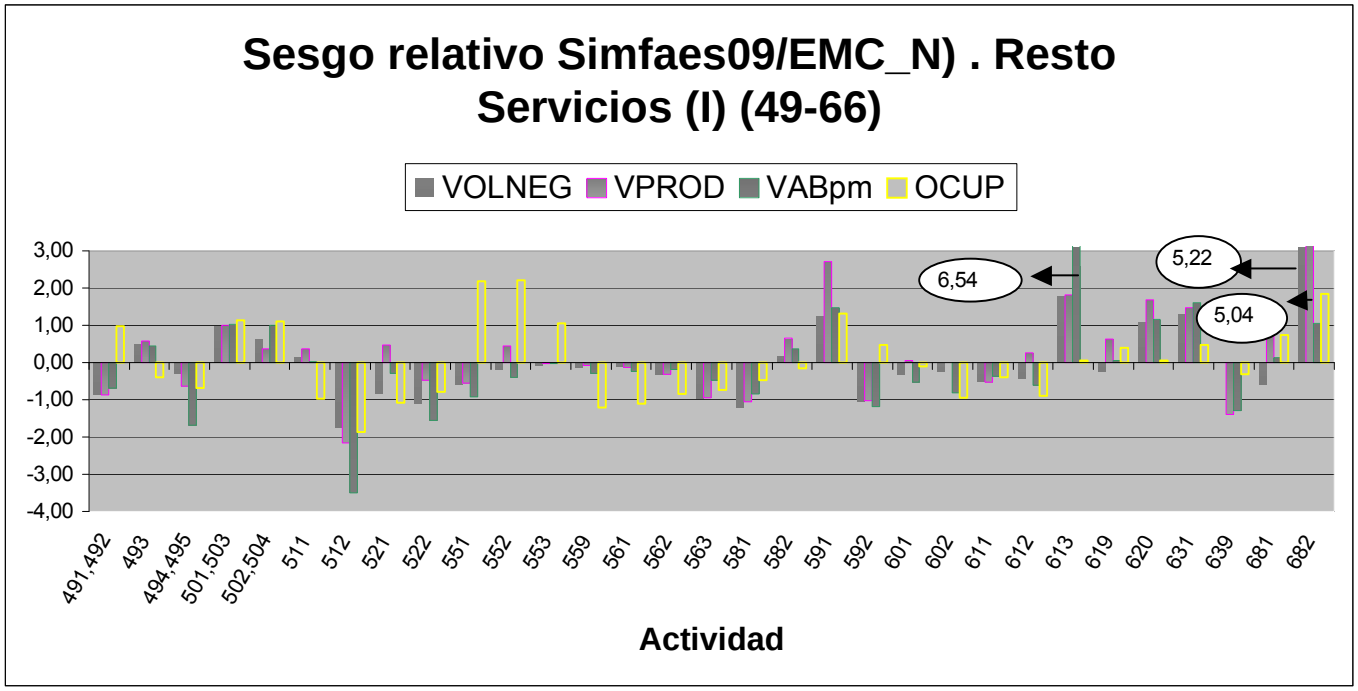
**G 4.1**


Sesgo relativo Simfaes09/EMC_N . Comercio(45-4669)
VOLNEG
VPROD
VABpm
OCUP
10,56
4,06
5,60
4,00
3,00
2,00
1,00
0,00
-1,00
-2,00
-3,00
-4,00
4511, 4519
4531, 4532
4611
4613
4615
4617
4619
4622
4624
4632
4634
4636
4638
4641
4643
4645
4647,4648,4649
C
4652
4662
4664
4669

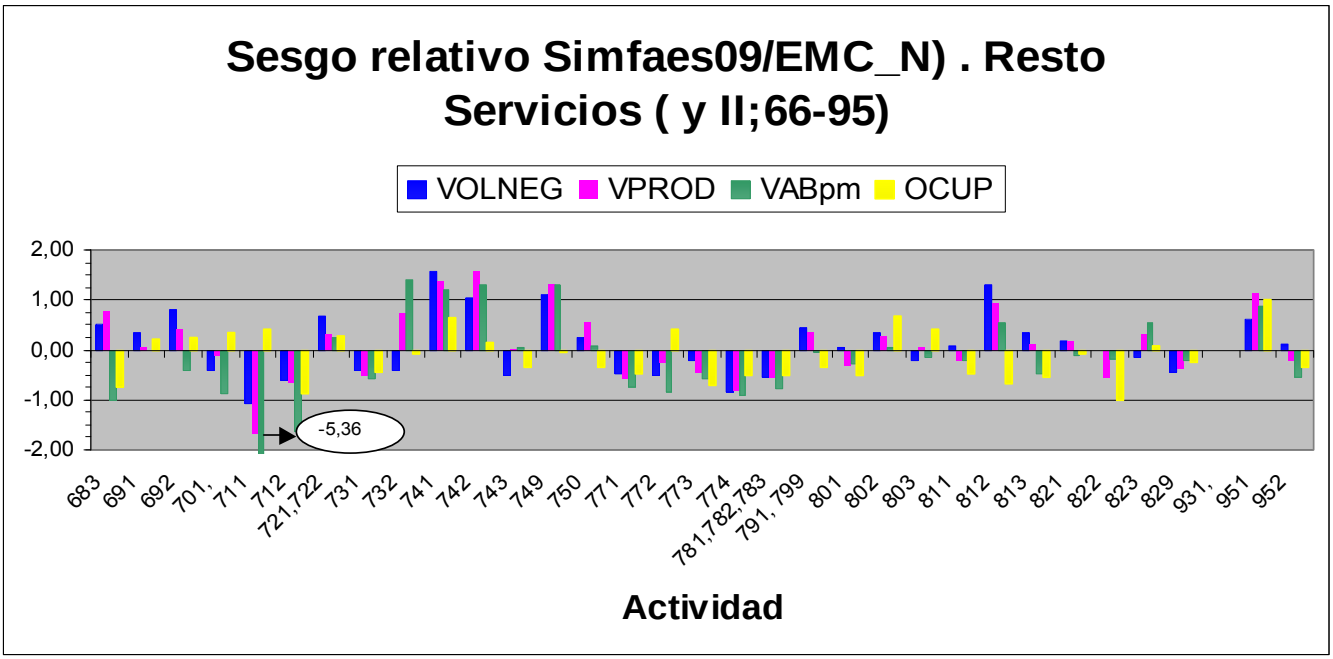
**G 4.2**


Sesgo relativo Simfaes09/EMC_N . Resto Comercio ( y II;4672-47)
VOLNEG
VPROD
VABpm
OCUP
5,00
4,00
3,00
2,00
1,00
0,00
-1,00
-2,00
-3,00
-4,00
4672
4673
4674
4675
4676
4677
4690
4711
4719
4721
4722
4723
4724
4725
4726
4729
4730
4741,4742,4764,4
4743,4754,4763
4751
4752
4753,4759
4761
4762
4771
4772
4773
4774
4775
4779
478
4791
4799

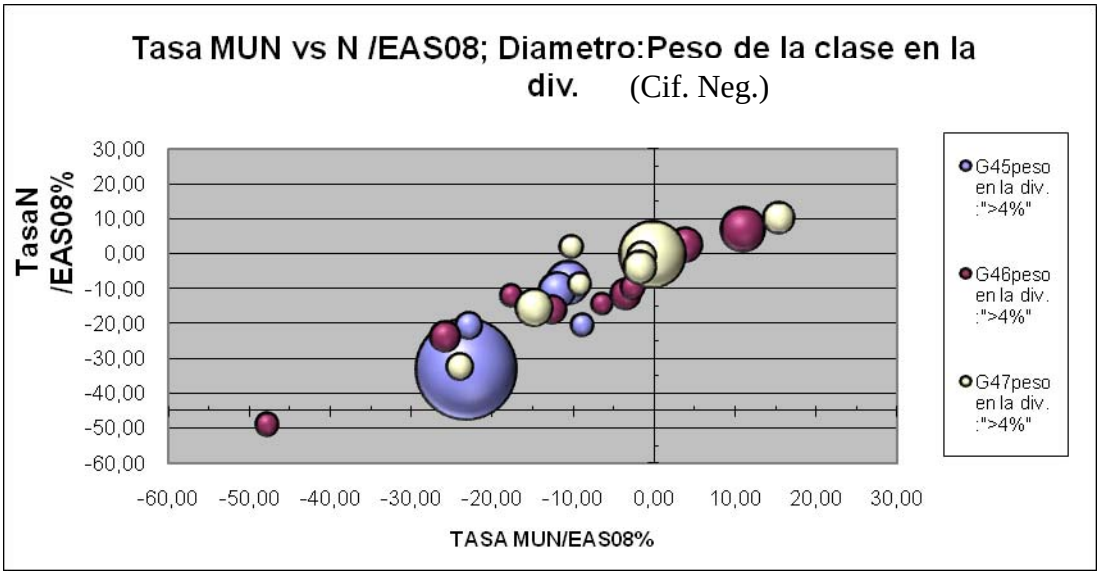
**G 4.3**


Sesgo relativo Simfaes09/EMC_N) . Resto Servicios (I) (49-66)
VOLNEG
VPRODD
VABpm
OCUP
3,00
2,00
1,00
0,00
-1,00
-2,00
-3,00
-4,00
6,54
5,22
5,04
491,492
493
494,495
501,503
502,504
511
512
521
522
551
552
553
559
561
562
563
581
582
591
592
601
602
611
612
613
619
620
631
639
681
682
Actividad

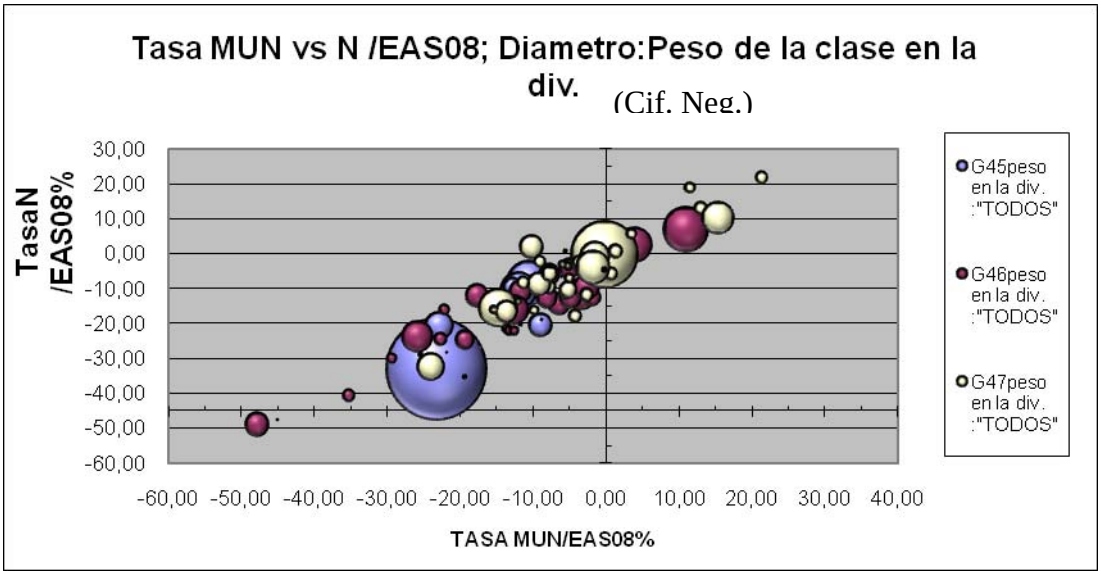
**G 4.4**


Sesgo relativo Simfaes09/EMC_N) . Resto Servicios ( y II;66-95)
VOLNEG
VPROD
VABpm
OCUP
2,00
1,00
0,00
-1,00
-2,00
-5,36
683
691
692
701,
711
712
721,722
731
732
741
742
743
749
750
771
772
773
774
781,782,783
791,799
801
802
803
811
812
813
821
822
823
829
931,
951
952
Actividad

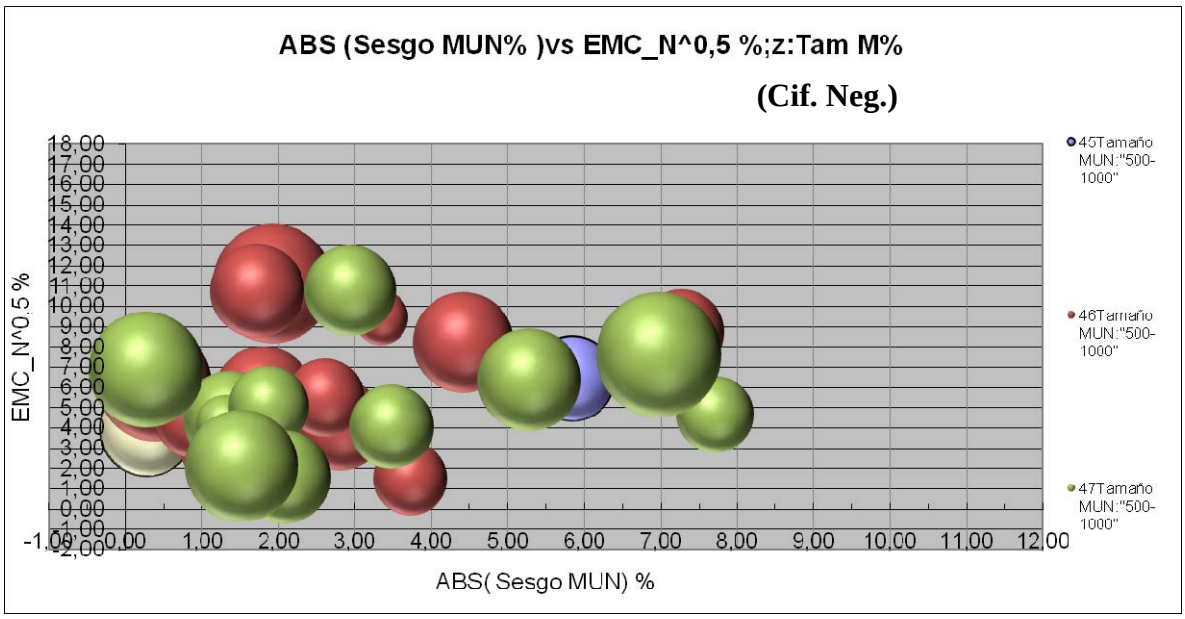
**G 5.1**


Tasa MUN vs N /EAS08; Diametro:Peso de la clase en la div. (Cif. Neg.)
TasaN /EAS08%
30,00
20,00
10,00
0,00
-10,00
-20,00
-30,00
-40,00
-50,00
-60,00
-60,00 -50,00 -40,00 -30,00 -20,00 -10,00 0,00 10,00 20,00 30,00
TASA MUN/EAS08%
G45peso en la div. :">4%"
G46peso en la div. :">4%"
G47peso en la div. :">4%"

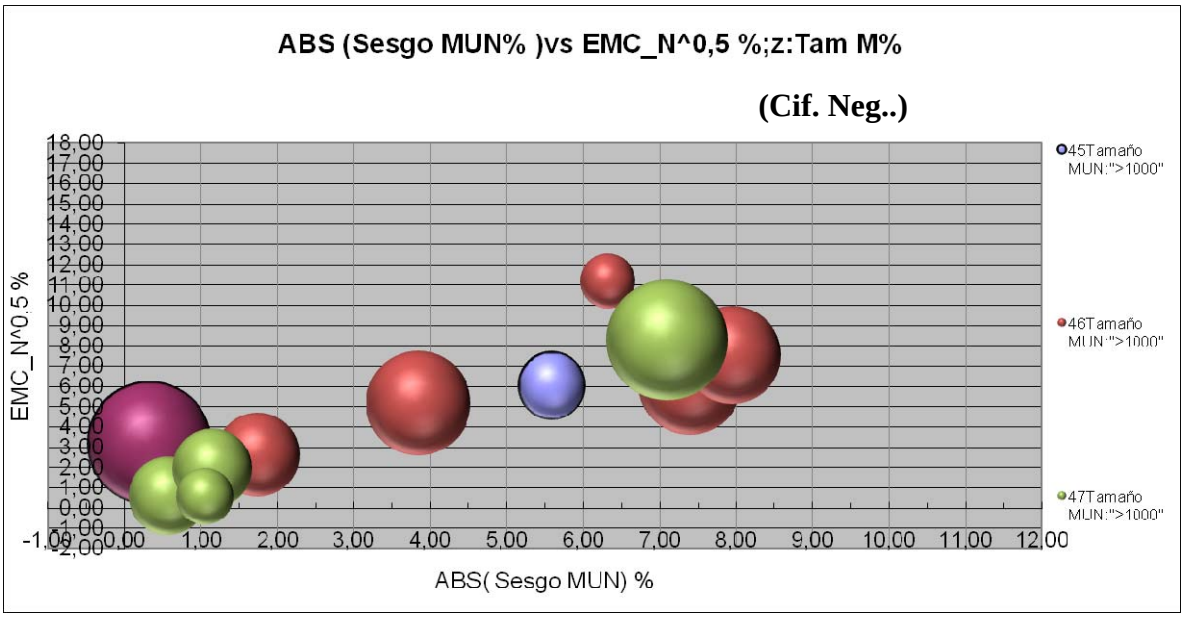
**G 5.2**


Tasa MUN vs N /EAS08; Diametro:Peso de la clase en la div. (Cif. Neg.)
TasaN /EAS08%
30,00
20,00
10,00
0,00
-10,00
-20,00
-30,00
-40,00
-50,00
-60,00
-60,00 -50,00 -40,00 -30,00 -20,00 -10,00 0,00 10,00 20,00 30,00 40,00
TASA MUN/EAS08%
G45peso en la div. :"TODOS"
G46peso en la div. :"TODOS"
G47peso en la div. :"TODOS"

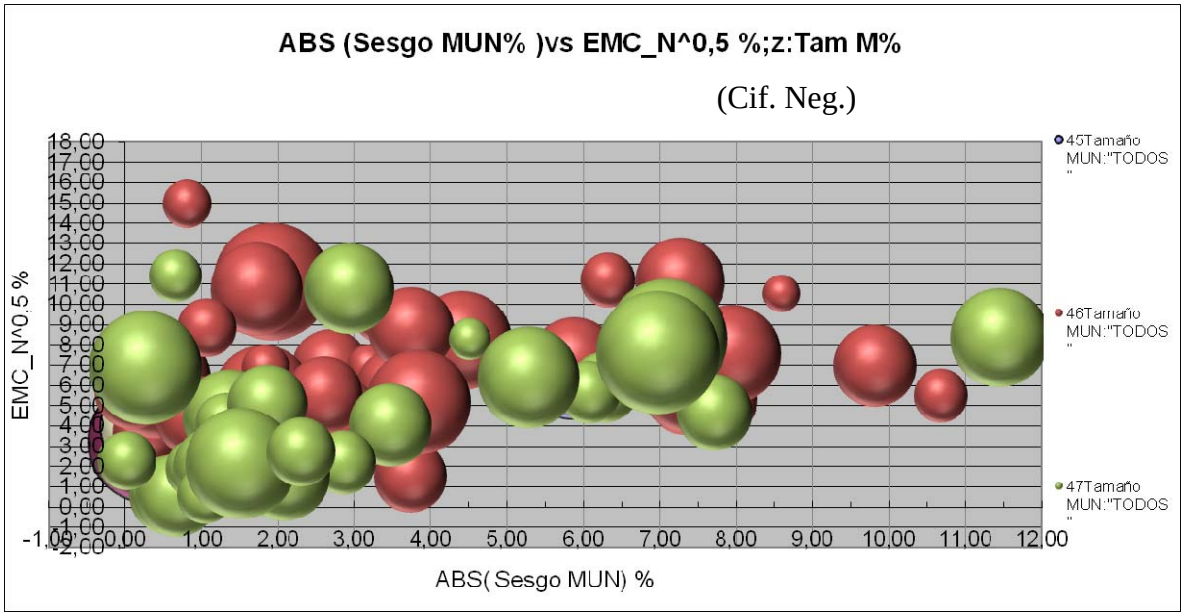
**G 6.1**


ABS (Sesgo MUN% )vs EMC_N^0,5 %;z:Tam M%
(Cif. Neg.)
EMC_N^0,5 %
ABS( Sesgo MUN) %
45Tamaño MUN:"500-1000"
46Tamaño MUN:"500-1000"
47Tamaño MUN:"500-1000"

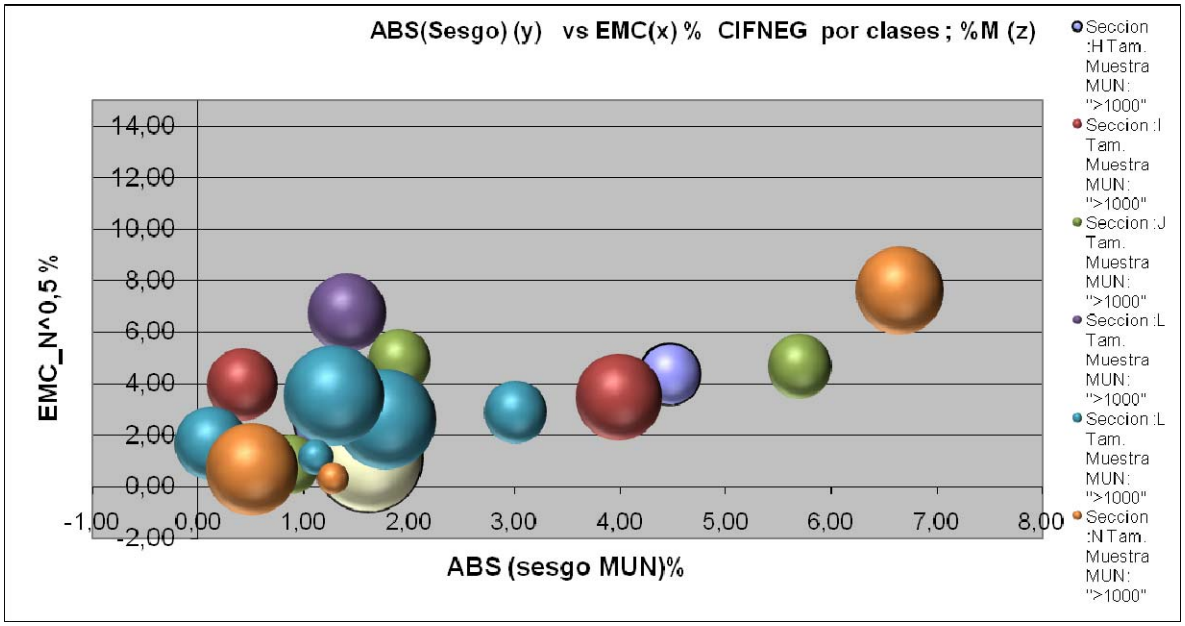
**G 6.2**


ABS (Sesgo MUN% )vs EMC_N^0,5 %;z:Tam M%
(Cif. Neg..)
EMC_N^0,5 %
ABS( Sesgo MUN) %
45Tamaño MUN:">1000"
46Tamaño MUN:">1000"
47Tamaño MUN:">1000"

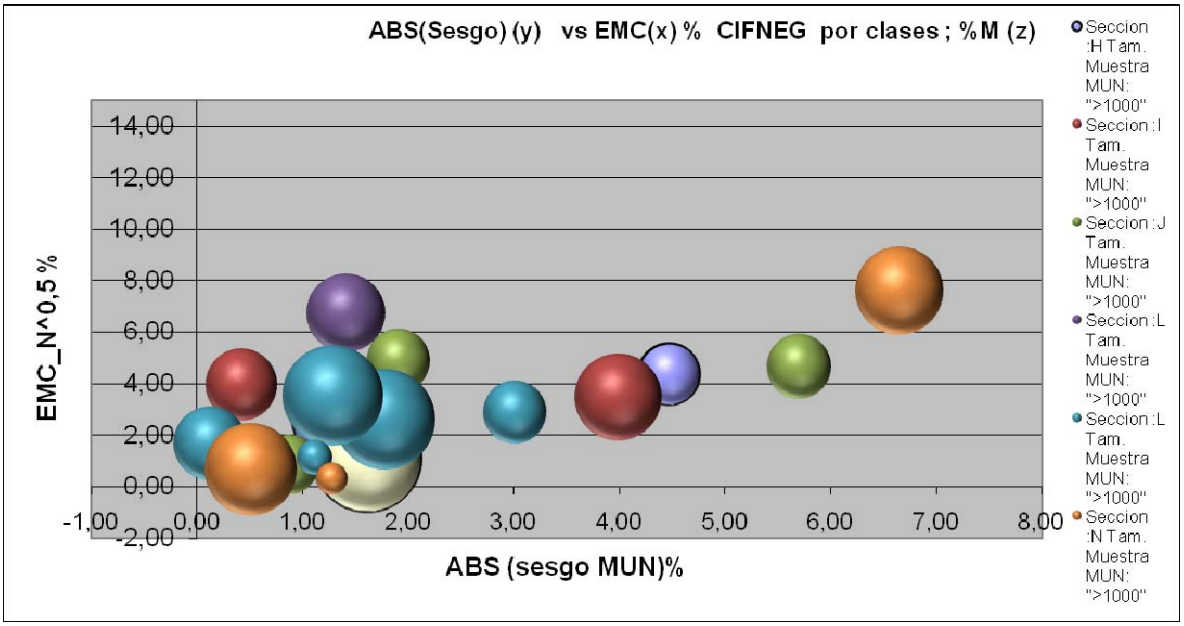
**G 6.3**


ABS (Sesgo MUN% )vs EMC_N^0,5 %;z:Tam M%
(Cif. Neg.)
EMC_N^0,5 %
ABS( Sesgo MUN) %
45Tamaño MUN:"TODOS "
46Tamaño MUN:"TODOS "
47Tamaño MUN:"TODOS "

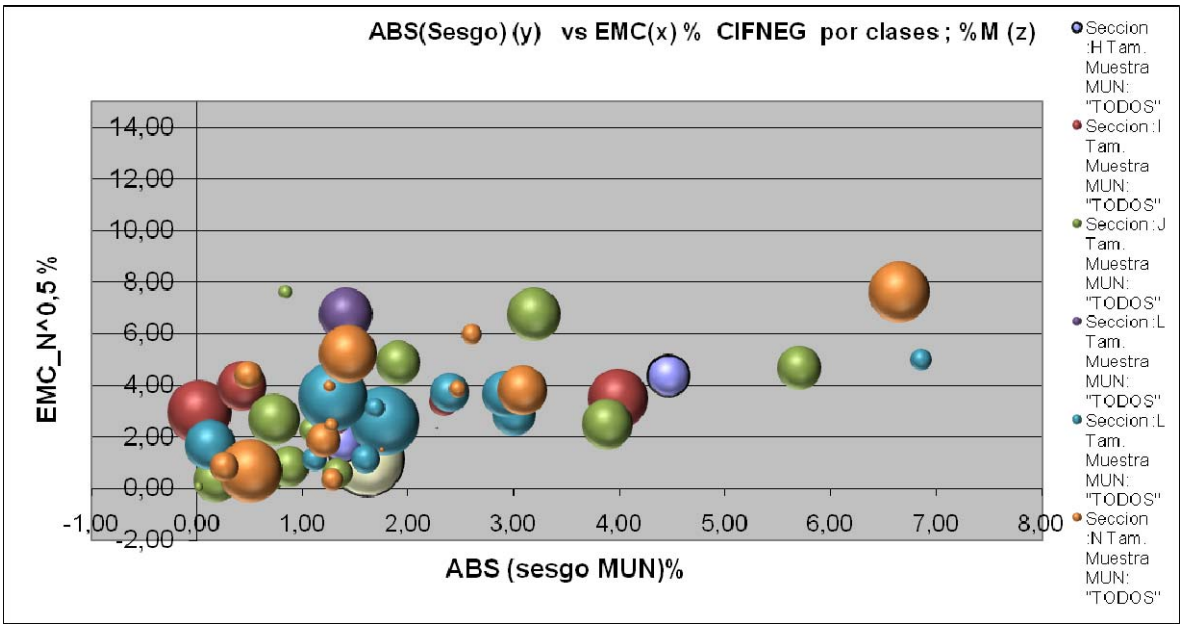
**G 6.4**


ABS(Sesgo) (y) vs EMC(x) % CIFNEG por clases ; %M (z)
EMC_N^0,5 %
ABS (sesgo MUN)%
Seccion :H Tam. Muestra MUN: ">1000"
Seccion :I Tam. Muestra MUN: ">1000"
Seccion :J Tam. Muestra MUN: ">1000"
Seccion :L Tam. Muestra MUN: ">1000"
Seccion :L Tam. Muestra MUN: ">1000"
Seccion :N Tam. Muestra MUN: ">1000"

**G 6.5**


ABS(Sesgo) (y) vs EMC(x) % CIFNEG por clases ; %M (z)
EMC_N^0,5 %
ABS (sesgo MUN)%
14,00
12,00
10,00
8,00
6,00
4,00
2,00
0,00
-2,00
-1,00
0,00
1,00
2,00
3,00
4,00
5,00
6,00
7,00
8,00
Seccion :H Tam. Muestra MUN: ">1000"
Seccion :I Tam. Muestra MUN: ">1000"
Seccion :J Tam. Muestra MUN: ">1000"
Seccion :L Tam. Muestra MUN: ">1000"
Seccion :L Tam. Muestra MUN: ">1000"
Seccion :N Tam. Muestra MUN: ">1000"

**G 6.6**


ABS(Sesgo) (y) vs EMC(x) % CIFNEG por clases ; %M (z)
EMC_N^0,5 %
ABS (sesgo MUN)%
14,00
12,00
10,00
8,00
6,00
4,00
2,00
0,00
-2,00
-1,00
0,00
1,00
2,00
3,00
4,00
5,00
6,00
7,00
8,00
Seccion :H Tam. Muestra MUN: "TODOS"
Seccion :I Tam. Muestra MUN: "TODOS"
Seccion :J Tam. Muestra MUN: "TODOS"
Seccion :L Tam. Muestra MUN: "TODOS"
Seccion :L Tam. Muestra MUN: "TODOS"
Seccion :N Tam. Muestra MUN: "TODOS"