

07 MAY 1997



ISSN 0328-1175

NOTAS SOBRE POLITICA ESTADISTICA EN EL MUNDO

JULIO 1996 JULIO 1996
Nº25
JULIO 1996 JULIO 1996

"NOTAS SOBRE POLITICA ESTADISTICA EN EL MUNDO" quiere mostrar, a todos los que participan o se interesan en la estadística oficial, temas que capturan la atención y preocupación de quienes tienen responsabilidad directa en la concepción, organización y conducción de las actividades estadísticas oficiales de países y organismos internacionales.

EN ESTE NUMERO:

● AMERICA

Canadá: La confidencialidad de los microdatos de encuestas.

● EUROPA

Cooperación estadística para la ribera sur del mar Mediterráneo.

"NOTAS SOBRE POLITICA ESTADISTICA EN EL MUNDO" es una publicación periódica que se edita en Buenos Aires, Argentina, bajo la responsabilidad de la Dirección del Instituto Nacional de Estadística y Censos (INDEC). Responde al propósito de dar a conocer aspectos de la estadística de distintos países relacionados con la organización, el funcionamiento, las prioridades y otros aspectos en los que se reflejan las respectivas políticas estadísticas. Está dirigida a profesionales y técnicos que actúan en el Sistema Estadístico Nacional (SEN), a autoridades y funcionarios públicos relacionados con la estadística oficial, a representantes de los medios, a empresarios y a usuarios en general. La selección de temas no representa apoyo ni prioridad por parte de las autoridades del INDEC. Puede ser reproducida libremente, con mención de la fuente.

INDEC

CANADA: LA CONFIDENCIALIDAD DE LOS MICRODATOS DE ENCUESTAS

El compromiso de mantener la confidencialidad de los datos individuales, que la estadística oficial asume con los informantes, es motivo de permanente preocupación de los responsables de aquélla. Tal preocupación es muy pertinente. Hay que tener en cuenta que la confidencialidad está relacionada, en gran medida, con la veracidad de las declaraciones individuales que proveen los informantes a partir de las cuales se elabora la información estadística.

En ocasión de censos nacionales, para asegurar la confidencialidad se aplican normas basadas en recomendaciones emanadas de los organismos internacionales. Los países más adelantados en materia de estadística oficial, dedican especiales esfuerzos al desarrollo de modelos teóricos y procedimientos técnicos aplicables a distintos casos de estudio. La Oficina de Estadística de Québec (provincia de Canadá), desarrolló un método para medir el riesgo de divulgar datos individuales de encuestas, que fué publicado en mayo de 1995.

Para dar cumplimiento a la ley de protección de la vida privada vigente en Canadá, es fundamental disponer de procedimientos técnico-prácticos a aplicar en la difusión de archivos de datos individuales provenientes de encuestas.

Algunas medidas para mejorar la prestación de los servicios de la Oficina de Estadística de Québec (Canadá), hicieron posible la financiación y realización de numerosas encuestas a pedido. Los demandantes de tales investigaciones, a posteriori de los relevamientos, comenzaron a pedir archivos de microdatos a los efectos de llevar a cabo, ellos mismos, análisis complementarios. Ante esta situación la Oficina encaró la puesta a punto de instrumentos que le permitieran, en primer lugar, evaluar el riesgo de divulgación de datos confidenciales y, en segundo lugar, enmascarar los datos individuales si el riesgo de divulgación era muy alto. Dado que los procedimientos para enmascarar disminuyen el contenido informativo de un archivo, deben ser usados con cautela para maximizar la información a difundir, asegurando que el riesgo de divulgación sea aceptable. Por tanto, es esencial que la medición del riesgo sea confiable y se apoye sobre bases lo más objetivas posible.

El documento propone medir el riesgo a partir de una modelización del proceso de divulgación en la que intervienen tres conceptos fundamentales: **divulgación inferencial, variables confidenciales y variables identificadoras.**

■ DIVULGACION INFERENCIAL:

Se produce cuando un usuario puede, gracias a datos difundidos, aumentar su información sobre ciertas características de un determinado individuo.

■ VARIABLES CONFIDENCIALES:

Cuando un individuo, cuyos datos están incluidos en un archivo, no desea que los valores de algunas variables correspondientes a sí mismo, sean difundidos, esas variables son llamadas **variables confidenciales, en ese archivo.**

■ VARIABLES IDENTIFICADORAS:

Si un usuario puede, potencialmente, conocer a priori el valor de una determinada variable, correspondiente a un determinado individuo, esa variable es llamada **variable identificadora, en ese archivo.**

■ CELULA:

Es el nivel de información más detallado al que el usuario pudo haber tenido acceso a priori.

Los tres conceptos - divulgación inferencial, variables confidenciales, variables identificadoras - permiten modelizar el proceso de divulgación de información confidencial.

«Es posible modelizar el proceso de divulgación de la siguiente manera: si un usuario quisiera conocer ciertas características de un individuo determinado, representadas por las variables confidenciales, podría, gracias a la información que dispone a priori de variables identificadoras, determinar un subgrupo de la población (una célula), a la que ese individuo pertenece. De este modo, si la información confidencial poseída a priori por el usuario se representa *con la distribución en la población de cada variable confidencial*, la información confidencial a posteriori aportada por la difusión del archivo corresponderá, de la misma manera, *a la distribución de cada variable confidencial en la célula a la que pertenece el individuo investigado*».

Se puede definir el riesgo de divulgación como: *una medida del incremento de la información confidencial aportado por la difusión*. Hay que calcular, por un lado, el riesgo para cada individuo del archivo y, por otro, el riesgo para cada variable confidencial. Valorizado el riesgo a nivel individual, los sectores técnicos de la oficina de estadística pueden localizar los elementos problemáticos del archivo y proceder a enmascarar los datos.

Cuando se trata de archivos voluminosos, no es posible considerar separadamente cada observación para decidir si se va a aceptar o no que los datos individuales sean difundidos. Hay que trabajar con un criterio de decisión que tome en cuenta una medida agregada del riesgo para cada variable confidencial y hay que fijar también un límite de riesgo, más allá del cual se considerará que el riesgo de divulgación es demasiado alto. Este es un tema sobre el que no hay abundante bibliografía.

Desde el punto de vista práctico, a partir de la teoría de la información se propone una medida basada en la entropía de la distribución de una variable confidencial, para cada célula de variables identificadoras. A todos los

individuos que pertenecen a una misma célula se les asigna el mismo valor de riesgo. En el caso de datos de encuestas no se sabe con certeza si el individuo investigado fue o no seleccionado para la muestra. Por tanto, para cada célula, la medida del riesgo a usar debe estar basada en la distribución estimada de una variable confidencial. El método de medir el riesgo que propone el trabajo debe ser considerado esencialmente como un marco de evaluación, que puede necesitar ajustes en función de la especificidad de los datos analizados.

La última parte del trabajo considera la aplicación del método a archivos provenientes de datos obtenidos por encuestas, casos en los que el problema analizado, la divulgación de datos confidenciales, adquiere características particulares.

Para más información:

En la Biblioteca del INDEC:

Bureau de la Statistique du Québec. Direction de la méthodologie. Guillet, Michel. Une approche générale de la mesure du risque de divulgation lors de la diffusion de microdonnées. mayo 1995.

COOPERACION ESTADISTICA PARA LA RIBERA SUR DEL MAR MEDITERRANEO

Organizado por EUROSTAT y el INE de España, se llevó a cabo el «Seminario Euromediterráneo de Cooperación Estadística», en Valencia a mediados de diciembre de 1995. Estaban invitadas las oficinas nacionales de estadística de los quince países miembros de la Unión Europea y de doce estados de la ribera sur del Mediterráneo.

Los objetivos de esta convocatoria mediterránea eran establecer las prioridades y las formas de la cooperación estadística en la zona que se emprendería durante el primer semestre de 1996, así como considerar las propuestas para el Seminario de Roma que se realizaría en el mes de junio en la capital italiana.

En opinión de un alto jefe del INE de España, el seminario de Valencia «ha sido la primera acción de la Conferencia de Barcelona en el ámbito estadístico, estableciendo las bases para su próxima reunión que se celebrará en Italia, en el mes de junio de 1996.» La discusión fué activa y se organizó alrededor de documentos de síntesis preparados por Eurostat. «Aspectos como la formación profesional de los estadísticos y la mejora de los instrumentos de armonización estadística, con una referencia explícita a los Sistemas de Cuentas Nacionales, son campos sobre los que se va a actuar en un futuro próximo».

Para más información: Fuentes Estadísticas N°10, nov.1995, pag 27, N°11, dic.1995, pag 27, N° 13, febr.1996, pag 13.

NOTAS es coordinada por Lelia Boeri de Cervetto. Dirigir toda correspondencia o comunicación a: INDEC-NOTAS, Av. Julio A. Roca 609 (1067) Buenos Aires, ARGENTINA. **Tel:** (541)349-9613, (541)349-9642. **Fax:** (541)775-2632, (541)349-9520, (541)342-7631. **E-Mail:** notas@indec.mecon.ar Oficina de distribución: Julio A. Roca 615 Tel: 349-9662

NOTAS SOBRE POLITICA ESTADISTICA EN EL MUNDO

INDEC

INSTITUTO NACIONAL DE
ESTADISTICA Y CENSOS

AV. PRESIDENTE JULIO A. ROCA 609
1067 BUENOS AIRES
ARGENTINA