

APLICACIÓN DE MÉTODOS MULTIVARIANTES PARA DETECCIÓN DE DUPLICADOS

P. Botella-Rocamora¹ y M. A. Martínez-Beneito²

¹Departament d'Estadística i Investigació Operativa. Universitat de València

²Área de Epidemiología. Direcció General de Salut Pública.

Conselleria de Sanitat. Generalitat Valenciana

Los registros de enfermedades contienen información sobre casos diagnosticados. A partir del tratamiento de los datos disponibles se obtiene información en forma de Indicadores de Salud. El control de duplicados en los registros resulta muy importante para la calidad de los Indicadores calculados a partir de los datos del registro para evitar sobre o supra-estimación en los mismos. Los registros deben contener información que permita la identificación del paciente, pues un mismo paciente puede ser recogido por diferentes fuentes de información. Es habitual incluir datos como SIP, DNI que permiten identificar a los pacientes de forma directa, aunque la cumplimentación de los mismos en algunos casos no es la deseable. La información que siempre está disponible es el nombre y los apellidos del paciente, así como la fecha de nacimiento en algunos casos. Esta información tiene el problema añadido de los habituales errores de tecleo y diferentes transcripciones para los mismos nombres y el cálculo aproximado de la fecha de nacimiento a partir de la edad del paciente. Ante cada nuevo paciente, no es posible comparar las características en forma de variables de este paciente con todos y cada uno de los ya introducidos en el registro. En este trabajo presentamos un método que, aplicando métodos estadísticos multivariantes, permite la detección de duplicados incluso con escasez de información aprovechando la similitud entre las cadenas de nombres y apellidos parecidos.