

VALIDACIÓN DE FICHEROS ADMINISTRATIVOS DE CARÁCTER ECONOMICO

Matías Ledesma Sereno

EUSTAT

m-ledesma@eustat.eus

Ander Iparraguirre García

EUSTAT

a-iparraguirregarcia@eustat.eus

PALABRAS CLAVE

Validación, Edición, Validate R-package

RESUMEN

Una fase del proceso de producción estadística que requiere muchos recursos es la validación de los datos recogidos. La estandarización de los procesos de validación puede ayudar a hacer más eficiente esta etapa de la producción. Actualmente, en EUSTAT se recogen ficheros administrativos de distintas fuentes (Impuesto de Sociedades, Registro Mercantil,...) los cuales deben ser validados. Se va a aplicar el paquete de R “Validate” desarrollado por Mark van der Loo, Edwin de Jonge y Paul Hsieh. El objetivo es implementar un procedimiento estándar de validar todos los ficheros económicos provenientes de distintas fuentes.

El paquete “Validate” se utiliza para validar e identificar los registros fallidos mediante las reglas establecidas previamente. Para ello se ofrece una cantidad significativa de funciones, las cuales tienen como propósitos facilitar esta labor. Las reglas designadas atacan a cada registro del fichero y resumen si esta se cumple, si no se cumple, si hay valores perdidos, si hay errores o si hay alguna advertencia. La gran ventaja de este paquete es su sintaxis simple y su fácil implementación. Se complementa igualmente con operadores lógicos de R lo cual facilita la escritura de las reglas.

El objetivo de este proyecto es la estandarización e industrialización de la validación de datos. Es decir, diagnosticar la calidad del fichero introduciendo únicamente un input de reglas y el fichero al que ataca.