

INVESTIGACIÓN REPRODUCIBLE

Caso de estudio:

Asignación de cupos 2015



 Wilmer González
wilmer.a.gonzalez@ucv.ve

Las investigaciones reproducibles son aquellas diseñadas para garantizar su capacidad de replicación bajo cualquier condición, siguiendo, uno de los pilares fundamentales del método científico que es la reproducibilidad. Sin embargo esto no siempre se cumple cuando se presenta un artículo, dado que dicha publicación no representa la investigación per se, sino las conclusiones del estudio y algunos recursos adicionales; y al tratar de repetir los experimentos a partir de los artículos, los lectores logran reproducir tan sólo 8 de cada de 18 artículos. [1] *John P A Loannidis et al (Repeatibility of published microarray gene expression analyses)*

Se entiende como Reproducibilidad a la capacidad de repetir un experimento, en cualquier lugar y por cualquier

persona. Dentro de las ventajas de investigaciones reproducibles, se tienen las siguientes:

- Posibilidad de confirmación de los resultados por parte del lector.
- Facilidad de evolución de la investigación debido a la integración de nuevos colaboradores.
- Aumento de las referencias que hacen interesantes las investigaciones reproducibles.

Es importante resaltar la importancia de estos aspectos particularmente en estos tiempos en los cuales la comunidad

científica global sufre una crisis de credibilidad, entre otras razones por la alta velocidad en la que se realizan las investigaciones que no permiten auditar por completo una propuesta.

En los casos de estudios estadísticos, la investigación reproducible consiste en la descripción del análisis y el entorno computacional requerido para repetir total y exactamente los resultados y métodos propuestos. Los recursos adjuntos a este tipo de investigaciones

incluyen:

- Datos crudos: Los datos iniciales provenientes de fuentes oficiales.
- Código asociado a los procesos realizados.
- Coordenadas de las fuentes de los datos.

Algunas alternativas para cumplir estos requisitos:

- **Research Compendia:** Permite crear una colección de recursos asociados a una investigación incluyendo el código utilizado en el análisis de los datos.

<http://researchcompendia.org/>

- **Generación de documentos dinámicos con R y Sweave (paquete de R):** Proporciona un marco flexible

para mezclar texto y código R para la generación automática de documentos. Un archivo fuente simple contiene el texto de documentación y el código R, los cuales son entrelazados dentro de un documento final que contiene el texto de documentación junto con el código R y/o la salida del código (texto, gráficos). <https://www.statistik.lmu.de/~leisch/Sweave/>

■ **DataJoy:** Esta última herramienta es la más reciente (Mayo 2015) y mediante una plataforma en la nube permite ejecutar código con un editor de R y Python. Además de permitir diferentes niveles de permiso, así como la inclusión de colaboradores, esta herramienta permite publicar código que sea de libre edición y lectura, pudiendo entonces cualquier lector reproducir los análisis y fragmentos de código ejecutados para llegar a las conclusiones de una investigación.

■ Otras herramientas provistas por la Comunidad Científica, pueden ser ubicadas en: <http://reproducibleresearch.net/links/>

Caso de estudio: OPSU 2015.

Por las mencionadas ventajas de la investigación reproducible en conjunción con el interés de analizar los datos de la asignación de cupos universitarios en la UCV para el periodo 2015 por parte de la Oficina de Planificación del Sector Universitario (OPUSU), se plantea un repositorio base de los recursos necesarios para analizar dicho proceso.

Inicialmente tomaremos en cuenta:

■ “resoluciones y recomendaciones tomadas por el Consejo Nacional de Universidades en sesión extraordinaria celebrada el día 09 de junio de 2015”. Disponible en <http://www.spcnu.gob.ve/resolucionescnu/2015/>

[ResolucionN%C2%B0501JUNIO \(EXTRAORD\)2015.pdf](#)

■ Lista de asignados por la OPSU a la Universidad Central de Venezuela. Disponible en http://ingreso.opsu.gob.ve/lista_asignados/0000000106-a.pdf?d=D

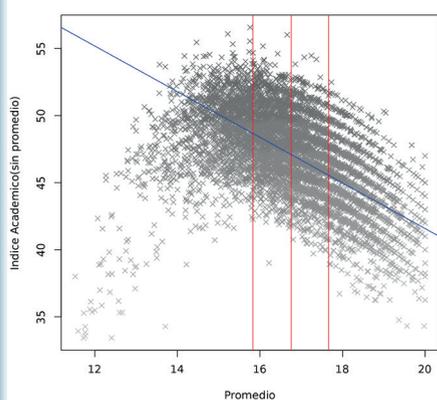
Estos recursos también se encuentran en el proyecto definido en la plataforma DataJoy

■ <https://www.getdatajoy.com/project/56290ca2c9b330470e580cd5> además de un fragmento de código propuesto para el uso público en el análisis de la asignación de cupos por parte de la Oficina de Planificación del Sector Universitario en el año 2015.

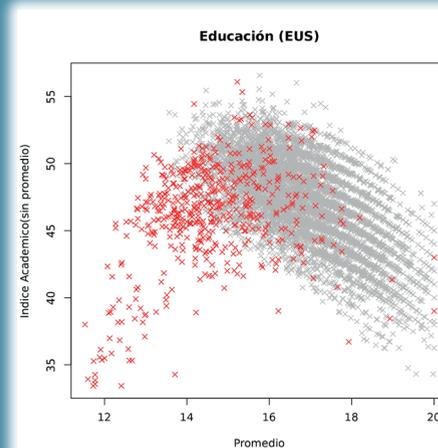


Por ejemplo en la siguiente visualización:

Se puede apreciar:



- La influencia decreciente que genera el resto de las variables contempladas en el índice académico.
- La existencia de observaciones por encima de 50 puntos en el resto de las variables, de donde se deduce que el índice académico está evaluado en más de 100 puntos.
- Observe que el único programa que difiere en un comportamiento descendente en el resto de las variables es el de Educación (EUS):



Los recursos necesarios para reproducir las apreciaciones presentadas se hicieron públicos y disponibles para invitar a la comunidad en general, no sólo a realizar nuevos análisis sobre los datos mencionados, sino a ejecutarlos de una manera verificable y reproducible. ■