

## DIDÁCTICA DE LA ESTADÍSTICA CON R

**Alejandro Galindo Alba**, *Departamento de Análisis Económico y Economía Política. Universidad de Sevilla.*

### RESUMEN.

La estadística cada vez está tomando un papel más relevante en el desarrollo de la sociedad moderna. La crisis bursátil, las encuestas electorales, el manejo y clasificación de la información, la ciencia del dato o el Big Data son solo algunos ejemplos para entender la necesidad de tener una sólida cultura estadística para poder analizar nuestro entorno desde un punto de vista crítico y fundamentado.

Por todo ello, el objetivo del siguiente taller es introducir una propuesta para la mejora de la didáctica de la estadística que nos permita manejar un gran volumen de datos reales, evitar el uso de la calculadora, la visualización de gráficos y el análisis crítico de los resultados a través de un software libre destinado hasta hoy a Estudios Superiores.

**Bloque temático:** Más recursos y menos cuentas.

**Nivel educativo:** Educación Secundaria Obligatoria y Post-Obligatoria.

### 1. INTRODUCCIÓN.

El objetivo de este taller es presentar una nueva herramienta para aquellos docentes que deseen introducir una opción de mejora en la didáctica de las matemáticas; particularmente en la estadística.

Conseguiremos:

- Evitar el tedioso trabajo con la calculadora.
- Trabajar con gran cantidad de datos.
- Trabajar con datos reales.
- Facilitar la visualización de los datos mediante gráficas.
- Fomentar la reflexión y el análisis crítico.
- Eliminar el gasto en software.

Para ello trabajaremos con R; un potente software gratuito que nos permitirá relacionar la utilidad de la estadística y la realidad del día a día.

A la vez, al eliminar los gastos en software también eliminamos una posible brecha económico-social. Este software no necesita ser instalado en ordenadores último modelo, ya que funciona en ordenadores menos actualizados realizando de igual manera sus funciones. Las rentas familiares no son un problema para que un alumno pueda instalar el software a casa.

## 2. LA ESTADÍSTICA Y R.

Como señalaron Barriuso, Gómez, Haro, & Parreño en 2013, los alumnos adquieren las habilidades en el cálculo, pero no comprenden, en determinados casos, el sentido de lo que aprenden. Quizás les falta manipular los conceptos, es decir, verlos desde distintos ángulos para descubrir las relaciones que se dan entre ellos, para adquirir una comprensión más sólida de la estadística.

La tecnología y determinado software favorecen esta manipulación y la simulación de experimentos con mayor rapidez y fiabilidad. Si además este software es gratuito, facilitamos la accesibilidad al mismo tanto por parte de las administraciones como del alumnado, suponiendo un importante ahorro y eliminando a su vez posibles barreras económicas.

Por otro lado, acercar al alumno a este tipo de herramientas sin tener que esperar a llegar a estudios superiores les facilitará su manejo y comprensión en el futuro.

Para su demostración o inicio en su utilización, lo más interesante será aplicarlo a la estadística de las matemáticas de primero de Bachillerato; en ella se trata la estadística descriptiva bidimensional, la cual nos da bastante juego a la hora de plantear problemas, de ver la utilidad que tiene la estadística y de fomentar la visualización de los resultados mediante gráficas. Llevar esa materia a su aplicación en R nos permitiría trabajar con una gran variedad de datos, estudiar sus distribuciones unidimensionales, bidimensionales, distribuciones condicionadas, correlación de variables, regresión, estimación... y todo ello utilizando datos reales.

Desde un punto de vista más práctico, los alumnos podrían extraer datos reales de la página del Instituto Nacional de Estadística, volcarlos en R y utilizar las herramientas del programa para sacar conclusiones y demostrar la utilidad de su uso desde temprana edad.

### 2.1. ORDENADORES Y ENSEÑANZA DE LA ESTADÍSTICA.

Los ordenadores han jugado un papel fundamental en el desarrollo de la Estadística; tanto por facilitar el acceso a ella, como por proporcionar diferentes software sin los cuales hoy día sería inimaginable la realización de análisis de datos.

Esta herramienta ha facilitado el trabajo con mayor número de datos reduciendo las horas de estudio gracias a su velocidad. Esto, permite a los alumnos investigar otros aspectos de los estudios estadísticos, como la recolección y planificación de la muestra, el diseño experimental y el análisis e interpretación de los resultados; es decir, tal como señala Batanero en 2001, estas herramientas permiten que los estudiantes establezcan una relación con la estadística de la misma manera que lo podría hacer un estadístico profesional. Además, se acerca al alumno al manejo de la informática en general a través de procesadores de texto y hojas de cálculo.

## 3. MATERIALES.

Cada participante deberá disponer de su propio portátil durante el taller. Así mismo, deberá tener instalado los paquetes R y R-Studio en su ordenador.

En el taller se trabajará con la versión R-Studio para facilitar el manejo de las ventanas.

Se comienza el taller explicando brevemente que herramientas de R que se van a utilizar y se introducirán las actividades prácticas.

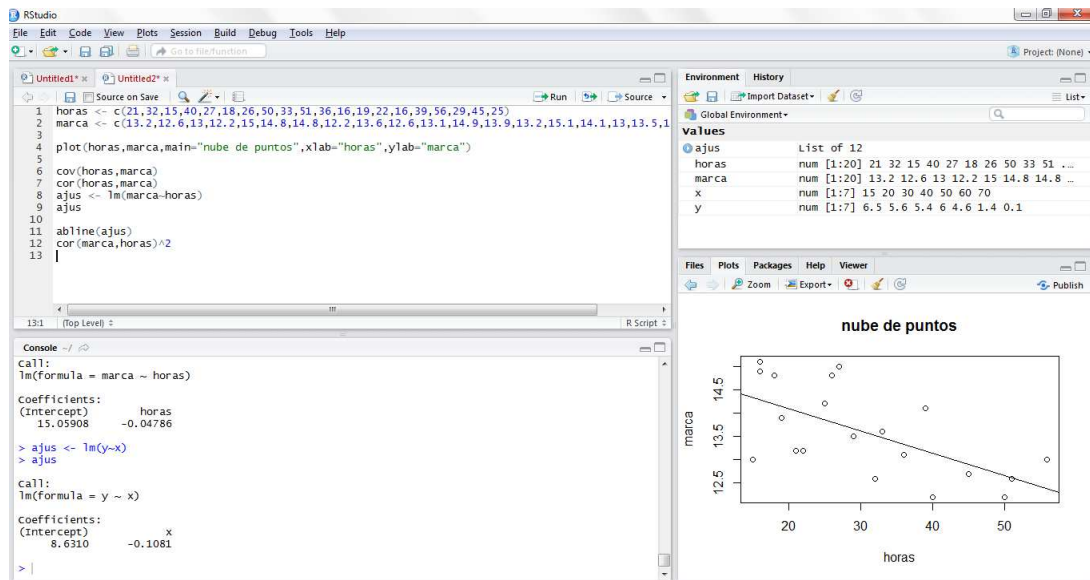


Figura 1. Pantalla principal de R-Studio.

### 3.1. COMITÉ ORGANIZADOR.

Para que el taller se lleve a cabo de manera satisfactoria el comité debería facilitar:

- Aula con mesa y silla para cada participante.
- Proyector.
- Pizarra (No es necesariamente indispensable pero si conveniente).
- Enchufes o regletas para todos los participantes.
- Conexión a internet (No es necesariamente indispensable pero sí muy conveniente).

### 3.2. ASISTENTES.

Material conveniente por parte de los asistentes:

- Portátil. Sería conveniente tener ya R y R-Studio instalados. De no ser así podríamos instalarlo en algún hueco (previamente al taller, para no restarle tiempo) a aquellos compañeros que no lo tengan.
- Cuaderno y bolígrafo.

## 4. ACTIVIDADES.

A continuación, para no extenderme en demasía y adaptarme a las normas de presentación de talleres, presento algunas pinceladas de ciertas actividades que se realizarán durante el taller.

### 4.1. ACTIVIDAD 1.

Esta actividad estará destinada a tomar contacto con la herramienta. Se entiende que los alumnos habrán recibido alguna clase previa de teoría y que aunque aún es pronto para dominar toda la terminología, si le resultarán familiares términos como regresión, covarianza, correlación, etc.

El objetivo es que los alumnos vayan haciendo el ejercicio a la vez que el profesor, deteniéndonos en cada paso y analizando los resultados obtenidos.

*Se midió el contenido de oxígeno, variable  $Y$ , a diversas profundidades, variable  $X$ , en el lago Wörthersee de Austria, obteniéndose los siguientes datos, en miligramos por litro.*

*¿Podrías establecer una relación entre la profundidad del lago y el oxígeno?*

$X$	15	20	30	40	50	60	70
$Y$	6,5	5,6	5,4	6	4,6	1,4	0,1

### 4.2. ACTIVIDAD 2.

En esta tarea trabajaremos con un gran conjunto de datos perteneciente a R conocido como Fisher's iris data. Este recoge para una muestra de 150 flores de tres especies diferentes de iris (iris setosa, versicolor y virginica) junto a sus variables longitud y anchura de sépalo (cm.) y longitud y anchura de pétalo (cm.).

	Sepal.Length	Sepal.width	Petal.Length	Petal.width	Species
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3.0	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
...	...	...	...	...	...
146	6.7	3.0	5.2	2.3	virginica
147	6.3	2.5	5.0	1.9	virginica
148	6.5	3.0	5.2	2.0	virginica
149	6.2	3.4	5.4	2.3	virginica
150	5.9	3.0	5.1	1.8	virginica

### 4.3. ACTIVIDAD 3. RECUPERAR DATOS DEL INE.

El primer paso de todo análisis estadístico es la obtención de los datos. Sin este paso no se puede hacer o se hace mal todo lo que se haga posteriormente no será útil.

Como la página del I.N.E. facilita gran cantidad de datos, veremos un ejemplo de descarga de datos desde ella:

---

*Obtener los datos mensuales del I.P.C. general desde enero de 2002 hasta septiembre de 2014.*

#### **4.4. ACTIVIDAD 4.**

*El conjunto de datos trees, incluido en R, proporciona las medidas de la circunferencia en base (inches), altura (ft) y volumen (cubic ft) de 31 cerezos negros recién talados.*

*Realice un estudio acerca de la posible relación entre las variables presentando un informe con las conclusiones.*

## **REFERENCIAS.**

BARRIUSO, J.M., GÓMEZ, V., HARO, M.J., & PARREÑO, F. (2013). Introducción a la estadística con R. *Revista Suma*. 72(Marzo 2013), pp. 17-30.

BATANERO, C. (2002). Estadística y didáctica de la matemática: Relaciones, problemas y aportaciones mutuas. En C. Penalva, G. Torregrosa y J. Valls (Eds.), *Aportaciones de la didáctica de la matemática a diferentes perfiles profesionales* (pp. 95-120). Alicante: Universidad de Alicante.

BATANERO, C. (2001). *Didáctica de la Estadística*. Granada: Universidad de Granada.

BOE, (2015). Real Decreto 1105/2014, de 26 de diciembre, por el que se establece el currículo básico de la Educación Secundaria Obligatoria y del Bachillerato.

FEBRERO, M., GALEANO, P., GONZÁLEZ, J. & PATEIRO, B. (2012). *Prácticas de Estadística con R*. Compostela: Universidad de Santiago de Compostela.

GARCÍA, A. (2010). *Estadística básica con R*. Madrid: UNED.