

## Usando R – una visión personal

*Edith Seier*

*East Tennessee State University*

*Prof. Emérita de la Universidad Nacional Mayor de San Marcos*

*seier@etsu.edu*

### Resumen

R es un lenguaje de programación gratuito para hacer análisis estadístico y tanto su desarrollo como su uso a nivel mundial siguen creciendo veinte años después de su creación en Nueva Zelanda. Personalmente uso R en aplicaciones, enseñanza e investigación aunque también utilizo otros programas. Este artículo es una revisión retrospectiva de la utilidad que R ha tenido para mí en años recientes, mediante el recuento de algunos de los trabajos que he realizado utilizando R. También se mencionan recursos que existen para el auto-entrenamiento en su uso. El objetivo del artículo es informar a los colegas que aún no utilizan R, o que quieren ampliar su uso, sobre sus principales características, la potencialidad de su uso en diferentes campos de aplicación o actividad y los recursos que existen en la forma de paquetes y literatura para que puedan evaluar la conveniencia que podría tener para ellos emplear R.

### 1.Introducción

R, como un lenguaje para análisis estadístico, fue desarrollado alrededor de 1993 por R. Gentleman y R. Ihaka de la Universidad de Auckland, Nueva Zelanda, y fue inspirado en el lenguaje S desarrollado por J. Chambers de los laboratorios Bell. Actualmente es desarrollado por un equipo de especialistas llamado el “core-group”. Se puede obtener en forma gratuita y legal de <http://www.r-project.org> y existen manuales en diferentes idiomas, incluyendo castellano, a libre disposición para aprender a usarlo. Son veinte años en que el uso de R ha venido ganando terreno, primero en el ambiente académico y luego fuera de él, inclusive en el mundo del análisis financiero. Hay cientos de paquetes escritos en R para realizar análisis específicos que son puestos para uso libre por sus autores. Desde hace varios años se puede acceder a una revista electrónica exclusivamente sobre R aunque también se publican artículos sobre nuevos paquetes en otras revistas estadísticas. Desde el año 2004 hay un congreso internacional anual donde se reúnen sus entusiastas usuarios, en el cual se dan cursillos y charlas sobre R, se intercambian ideas y se presentan trabajos. Para mí, el simple hecho que haya una herramienta tan importante que puede usar cualquier persona en el mundo que tenga acceso a una computadora es una idea maravillosa que reduce los efectos de la diferencia de recursos económicos en el contexto estadístico.

Los estadísticos de mi generación nos hemos visto expuestos en los últimos 40 años a varias versiones de computadoras, sistemas operativos, lenguajes de programación y software estadístico. En los años 70 recuerdo haber escrito programas en Fortran IV para explorar el efecto de transformaciones no lineales de datos en la regresión múltiple usando tarjetas perforadas. Un tiempo después escribí programas también en Fortran para experimentar con métodos del análisis espectral de series de tiempo desde terminales de computadoras grandes que ya no usaban tarjetas sino teclado. Luego vinieron las microcomputadoras a fines de los años 80, con el sistema DOS y luego Windows. Surgieron software como Minitab, SAS, SPSS y MATLAB entre otros. Aun uso frecuentemente Minitab y SAS para el análisis de datos y la enseñanza; pero cada vez me encuentro utilizando R con mayor frecuencia, sobre todo cuando tengo que elaborar un gráfico o análisis no estándar, diseñando un nuevo método y también en la enseñanza. Los estadísticos respondemos a los desafíos que el análisis de datos nos presenta y la experiencia de cada uno de nosotros es diferente. Un artículo escrito por otro estadístico sobre el uso de R podría ser totalmente distinto porque nuestras experiencias talvez serían diferentes. Sin embargo, esta revisión de los trabajos que he hecho utilizando R en los últimos años puede ser útil para otros colegas o estudiantes que estén sopesando empezar o intensificar su propio uso de R. En las secciones siguientes exploraré el uso de R en la investigación, aplicación y enseñanza desde la perspectiva de mis propias experiencias. En la sección final se mencionan algunas fuentes de información y recursos para aprender y aplicar R.

## **2. Definiendo funciones en R para aplicar nuevos métodos**

Cuando se define una nueva estadística o un nuevo gráfico es conveniente proporcionar al lector una forma de hacer los cálculos necesarios para aplicarlo a sus propios datos de una forma fácil. R es ideal para ello pues siendo gratuito cualquier persona puede usarlo. El conjunto de comandos se puede escribir en forma de una función que luego se puede aplicar a diversos datos fácilmente. Para crear una función simplemente incluimos todos los pasos necesarios para calcular la formula, los ponemos entre llaves y asignamos a la función un nombre de nuestra elección. Los últimos comandos de la función indican el producto final. Qué diferencia hay entre escribir una secuencia de comandos formando una función o simplemente dejarlo como una secuencia de comandos? Los dos tienen aspectos en favor y en contra. Cuando escribimos una función, es suficiente copiarla y pegarla al comienzo de la sesión y podremos aplicarla, cuantas veces queramos, sin embargo, no podemos recuperar los resultados de cálculos intermedios. Por otro lado, si no escribimos los comandos formando una función, tendremos que copiar y pegar los comandos cada vez que deseemos aplicarlos a un nuevo conjunto de datos pero el valor de todos los cálculos intermedios serán accesibles. La decisión de escribir o no una función depende pues de lo que nos sea más conveniente. A continuación mencionamos tres artículos nuestros en los que proporcionamos al lector los comandos en R para aplicar los métodos que en ellos definimos.

**2.1 El Polyplot.-** Este grafico fue definido por Seier y Bonett (2003). Es un gráfico similar al diagrama de cajas o “boxplot”, pero con una mayor cantidad de estadísticas de posición y dispersión, que permite percibir también la curtosis de la distribución al comparar visualmente varias estadísticas de dispersión. El polyplot incluye la media, mediana, los cuartiles, extremos, rango inter-cuartílico, desviación media con respecto a la mediana, desviación standard, y rango. La función polyplot se encuentra en <http://faculty.etsu.edu/seier/Rpolyplot.txt>. Primero se copia la función en la ventana de la sesión de R, luego bastará con aplicarla a un conjunto de datos de una variable. Por ejemplo, aplicada a los datos del pulso (número de latidos por minuto) de 210 estudiantes (los datos están en <http://faculty.etsu.edu/seier/pulserate.txt>) produce el gráfico en la Figura 1. El polyplot da los valores de varias estadísticas de posición adicionales a las que da el boxplot.

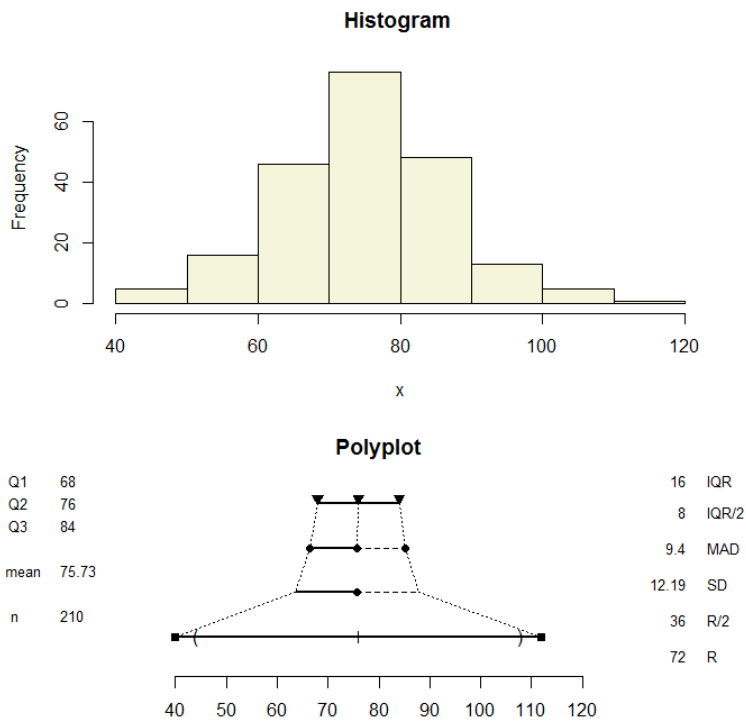


Figura 1 . El Polyplot para el pulso de 210 estudiantes.

**2.2 Un intervalo de confianza para el coeficiente de dispersión.** El coeficiente de dispersión (COD) es una medida de dispersión relativa, alternativa al coeficiente de variación, definida como la desviación media (con respecto a la mediana) dividida por la mediana. Es usada mayormente en el análisis de precios de bienes raíces y en algunos análisis de datos agrícolas. Bonett y Seier (2006) definieron una fórmula, algo larga, para calcular un intervalo de confianza para COD. En <http://faculty.etsu.edu/seier/CODCI.txt> hay una función que implementa dicha fórmula que se puede usar para estimar el COD de una población con un intervalo de confianza calculado con los datos de una muestra.

**2.3 La función B para distancias a la frontera.** Los métodos de estadística espacial se encuentran implementados mayormente en unos pocos programas comerciales y en diversos paquetes de R. En el análisis de patrones de puntos las distancias de los puntos a su vecino más cercano reciben gran importancia, sin embargo esa no es la única distancia que tiene interés. Recientemente definimos (Joyner, Ross & Seier, 2013) una función B que es la función de distribución de la variable ‘distancia de los puntos a la frontera’. La función  $\hat{B}$ , calculada con los datos, es comparada con la función B deducida asumiendo una distribución espacial totalmente aleatoria. El objetivo del análisis es detectar si la distribución de puntos en el mapa es similar a lo que se puede esperar de una distribución espacial totalmente aleatoria o si por el contrario tienden a estar más cerca de las fronteras del mapa o mas bien lejos de las fronteras y mas cerca del centro. En el apéndice del artículo que publicamos (Joyner, Ross & Seier, 2013) incluimos los comandos en R para poder producir este tipo de gráficos, incluidas las bandas alrededor de la función teórica para poder evaluar si la función estimada o empírica está lo suficientemente lejos de la teórica como para pensar que es realmente diferente.

En otros trabajos de investigación he utilizado R para hacer simulaciones que permiten comparar la potencia del test que uno define con otros tests existentes. Es sabido que R no es muy veloz en la ejecución de algunos bucles (‘loops’) o iteraciones pero algunas de esas iteraciones se pueden sustituir por cálculos planteados en forma matricial.

Cuando se escribe un libro de texto, o un artículo utilizando latex, como es común en estadística matemática, R es ideal para producir los gráficos pues estos se pueden guardar como archivos de tipo ‘postscript’ (.ps o .eps) que se insertan muy fácilmente en el texto escrito en latex. R permite hacer no solo gráficos estadísticos bastante complejos sino también dibujar figuras geométricas y graficar funciones. Además R permite efectuar cálculos matemáticos de diversa índole.

### **3. Utilizando R para aplicaciones en temas especializados**

R es particularmente útil en aplicaciones que requieren procesos iterativos y facilidades gráficas. También hay paquetes en R que hacen análisis en tópicos incluidos solo en unos pocos software comerciales, tal es el caso de la bioinformática y el análisis espacial entre otros.

**3.1 Análisis que requieren repetir un cálculo muchas veces-** Cuando enseño el método de análisis de varianza (ANOVA) de un criterio me gusta incluir un ejemplo de la aplicación de este método en un ámbito aun relativamente moderno, el análisis de datos de micro-arreglos (‘microarrays’). Los datos de un micro-arreglo consisten en la intensidad de la expresión de miles de genes. Si se tienen micro-arreglos, digamos de pacientes con tres o más variedades diferentes de una enfermedad, luego de normalizar los datos uno de los primeros análisis estadísticos es proceder a identificar los genes que están más diferencialmente expresados entre los diferentes grupos. Para ello se aplica ANOVA a los datos de cada gen y se ordenan los

genes respecto al valor de la estadística F, en orden decreciente. En los primeros lugares quedarán los genes que se expresan en forma mas diferenciada entre los diferentes grupos y que por tanto son candidatos potenciales a ser considerados en un método de diagnóstico. Si sólo hay dos grupos, obviamente se aplica el test t de student en lugar de ANOVA. El problema es que hay miles de genes y hacer tantos ANOVAs con un software interactivo sería cansador y tomaría mucho tiempo. Asumiendo que el vector f tiene los subíndices que indican a que grupo pertenece cada paciente, que m es el número de genes y que mfe es la matriz de datos de intensidad en que cada columna corresponde a un paciente y cada fila a un gen, el siguiente comando calculará en segundos el ANOVA para cada uno de los genes:

```
for (i in 1:m)
{
stat[i]<-oneway.test( mfe[i,]~f,var.equal=T)$statistic
pv[i]<-oneway.test(mfe[i,]~f,var.equal=T)$p.value
}
```

Y estos comandos ordenarán los genes en orden decreciente de acuerdo al valor de la estadística F y mostrarán en pantalla los 30 (o cualquier otro número que uno prefiera) genes que se expresan en forma más diferenciada:

```
y<-data.frame(stat,pv);
y1<-y[order(pv,decreasing=F),];
y1[1:30,]
```

Actualmente estoy participando en un proyecto de investigación en biología en el cual los alumnos realizan numerosos experimentos y observaciones durante el verano y tienen que analizar los datos. En uno de los casos teníamos 240 archivos de datos que queríamos analizar en la misma forma para comparar los resultados. Luego de discutir los gráficos y análisis que queremos producir para cada experimento, lo más efectivo fue escribir un conjunto de comandos en R para producir el producto deseado. Así solo había que cambiar el nombre del archivo de datos cada vez para producir los gráficos y cálculos en forma automática en lugar de hacerlo en forma interactiva con un software tradicional.

**3.2 Aplicaciones del análisis espacial.** En el área de estadística espacial si bien existe algo de software comercial especializado, la mayor parte de los análisis se pueden hacer y se hacen usando diversos paquetes en R. Información muy completa respecto a estos paquetes se puede encontrar en <http://cran.r-project.org/web/views/Spatial.html>. A modo de ejemplo incluimos la Figura 2 preparada utilizando funciones de los paquetes *maps* y *spatstat* en R. Una vez que se ha cargado el paquete *maps*, para lograr el mapa basta con escribir `map("world","Peru")`.

En <http://earthquake.usgs.gov/earthquakes/eqarchives/epic/> encontramos las coordenadas de donde ocurrieron sismos de magnitud 5 o mayor magnitud desde 1973, entre los 83 a 68 grados de longitud oeste y 0 a 18 grados de latitud sur. Esos puntos fueron graficados sobre el mapa

usando el comando 'points'. Luego de crear un objeto 'ppp' con el paquete *spatstat*, se logra el gráfico de densidad con la función `plot(density())`. El gráfico de densidad ilustra que hay una mayor concentración de sismos en la zona sur del país, especialmente en la costa. La función G, graficada en la Figura 2, es la función de distribución de las distancias al vecino más cercano, distancias calculadas con el comando `dnn` del paquete *spatstat*. La función G, en el último gráfico de la Figura 2, indica que las distancias cortas son mucho más frecuentes confirmando así que los sismos no ocurren en forma totalmente aleatoria en cualquier zona del país sino que forman conglomerados cuando se considera un período largo de observación.

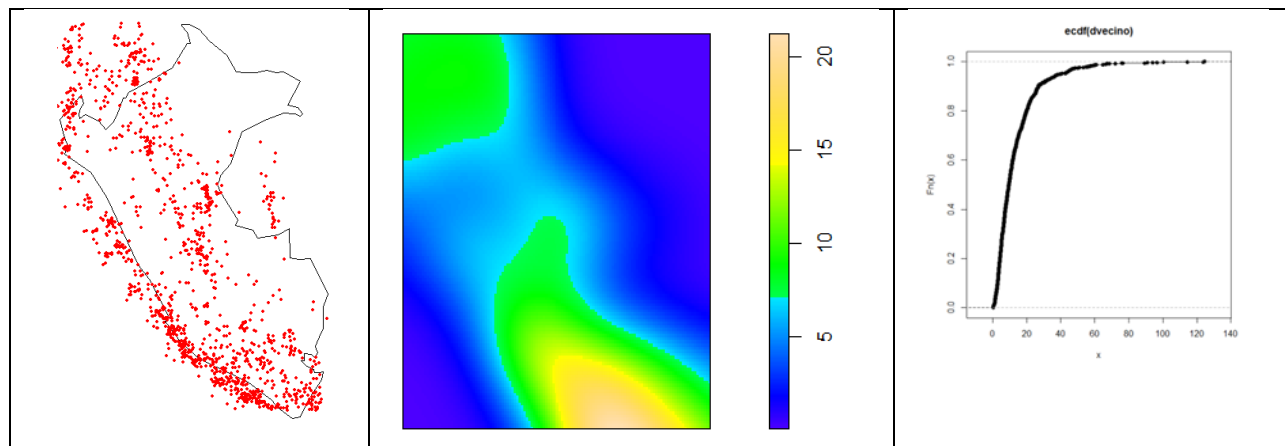


Figura 2. Sismos de magnitud  $\geq 5$  desde 1973, densidad y función G.

#### 4. Usando R en la enseñanza de la estadística.

Aparte de enseñar a los alumnos a usar R dentro de los cursos de estadística para hacer gráficos y análisis básicos, R me es útil en el aula entre otras cosas para enseñar temas que no están en los paquetes habituales y para ilustrar ideas ligadas a la distribución de estadísticas mediante simulación.

**4.1 Para enseñar métodos de inferencia de cómputo intensivo.** - En secciones especiales de los cursos introductorios y cursos de estadística aplicada me gusta incluir el test de permutaciones para comparar dos grupos y el test de permutación para datos apareados. Para construir intervalos de confianza uso el bootstrap y ese intervalo se utiliza para docimar hipótesis para la media (u otro parámetro) de una población. Encuentro que R es ideal para los tests de permutación y el bootstrap. Aunque ya existen paquetes para hacer estos cálculos, yo prefiero escribir mis propios programas porque así puedo explicar a los alumnos lo que estamos haciendo con la computadora. En el curso introductorio utilizo actividades táctiles con fichas para que los alumnos entiendan fácilmente el proceso de agrupaciones aleatorias para el test de permutación o la toma de muestras con reemplazamiento del mismo tamaño de la muestra

original para aplicar el bootstrap. Luego les proporciono el código en R que hace cosas similares a las experiencias táctiles que hemos realizado en clase. Los programas que utilizamos en clase están a libre disposición en <http://faculty.etsu.edu/seier/RcommCh3.txt>.

Como ejemplo del test de permutación para datos apareados usamos el clásico ejemplo de los datos del experimento de Darwin con plantas de maíz para estudiar los efectos de la endogamia, utilizando 15 pares de plantas, para comparar la estatura de plantas que fueron producidas por auto-fertilización y por fertilización cruzada. Cabe señalar que este es el ejemplo que utilizó R.A. Fisher cuando definió el test de permutaciones como una forma de validar los resultados de aplicar el test t a ese mismo conjunto de datos. La hipótesis nula es  $H_0: \mu_d = 0$ . Hemos escrito un programa que calcula las diferencias para cada par de datos y la media de esas diferencias para los datos originales ( $\bar{d} = 2.62$ ) y también para cada una de todas las posibles permutaciones. Si hay  $n$  pares de datos, el número de permutaciones es  $2^n$ . Para el ejemplo de los datos del experimento de Darwin  $n=15$  y por tanto hay  $2^{15} = 32768$  posibles permutaciones. R hace todos esos cálculos en un instante. La Figura 3 muestra la distribución de la media de las diferencias considerando todas las posibles 32768 permutaciones. En 1726 de esos 32768 casos la media de las diferencias es igual o mayor que 2.62 o menor que -2.62, por lo que el p-valor es 0.05267334

Esta es la versión exacta del test que se puede aplicar cuando las muestras no son muy grandes. Para muestras mas grandes hemos preparado una versión aproximada del test, que trabaja no con todas las posibles permutaciones ya que se puede exceder la capacidad del software, sino con un numero grande (por ejemplo 10000) de permutaciones aleatorias no necesariamente todas diferentes y que produce un p-valor aproximado.

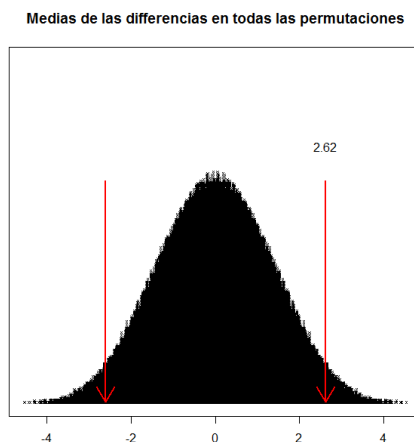


Figura 3. Distribución empírica lograda por permutaciones para datos apareados

**4.2 Para explorar distribuciones de estadísticas usando simulación.** Antes de tratar teóricamente la distribución de algunas estadísticas podemos querer explorar la forma de tales distribuciones usando simulación. Un colega quería motivar a sus alumnos para estudiar la distribución de los cuartiles y mediana y como un primer paso quería mostrar a sus alumnos como los cuartiles y la mediana varían de muestra a muestra de una misma distribución. Me dijo: “quiero ver una película del boxplot”. Escribí los siguientes comandos para hacer esa especie de video. La idea es simplemente generar una muestra aleatoria de un cierto tamaño, digamos  $n=100$ , de una distribución normal (u otra distribución de nuestra elección) y graficar el boxplot. Para repetir esta acción muchas veces, se usa un bucle con el comando ‘for’ y un contador

```
for(i in 1:1000) {x<-rnorm(100) ; boxplot(x)}
```

Si escribimos esa línea en R veremos aparecer rápidamente 1000 boxplots uno tras otro. Obviamente podríamos cambiar la media y desviación standard de la distribución (para que fueran por ejemplo  $\mu=200$  y  $\sigma=50$ ), poner un color a diagrama y también mantener el eje vertical fijo de 0 a 400. Para ello añadimos algunas especificaciones al comando anterior:

```
for(i in 1:1000) {x<-rnorm(100,200,50) ; boxplot(x,col='red',ylim=c(0,400))}
```

Si copiamos este comando en R, veremos generarse 1000 diagramas de caja (o ‘boxplots’) para muestras de la misma distribución y apreciaremos la variabilidad entre muestras. Se puede también almacenar los valores de las estadísticas usadas en la construcción del boxplot para luego apreciar su distribución empírica. Para ello tendríamos primero que crear lugares para almacenar esas estadísticas y luego graficarlas usando histogramas. Los comandos siguientes realizarán producirán el video del boxplot y los histogramas de las distribuciones empíricas logradas por simulación de la Figura 4.

```
## creamos los espacios de almacenamiento
```

```
minimos<-numeric(1000);maximos<-numeric(1000);cuartiles1<-numeric(1000);
```

```
medianas<-numeric(1000); cuartiles3<-numeric(1000) ;
```

```
## generamos las muestras
```

```
## calculamos y almacenamos las estadísticas y graficamos el boxplot
```

```
for(i in 1:1000) {
```

```
x<-rnorm(100,200,50) ; minimos[i]<-min(x) ; maximos[i]<-max(x) ; medianas[i]<-median(x);
```

```
cuartiles1[i]<-quantile(x,0.25); cuartiles3[i]<-quantile(x,0.75);
```

```
boxplot(x,col='red',ylim=c(0,400))}
```

```
win.graph()
```

```
## preparamos los graficos con las distribuciones simuladas
```

```
par(mfcol=c(2,3))
```



```

hist(minimos,col= 'red')
hist(maximos,col= 'red' )
hist(cuartiles1,col= 'blue')
hist(cuartiles3,col= 'blue')
hist(medianas,col= 'darkgreen')

```

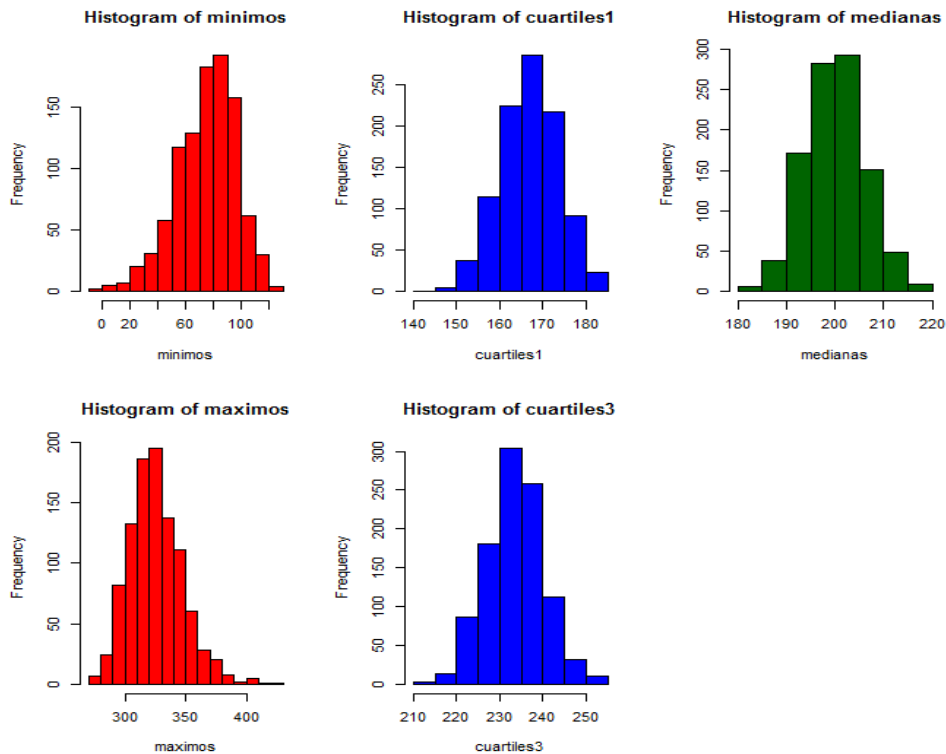


Figura 4. Distribuciones empíricas para cinco estadísticas, obtenidas por simulación

Hay muchas variaciones que se pueden hacer a esta experiencia. Por ese motivo decidí asignar, como ejercicio para los alumnos de computación estadística, escribir los comandos con la finalidad de crear una secuencia de gráficos o video dando rienda suelta a su creatividad. Una colega que me ayudaba en el curso decidió usar la idea para motivar a un grupo de niñas de segundo de secundaria a las que quería mostrar la idea de programación y me cuenta que le funcionó muy bien pues el gráfico dinámico del boxplot motivó a las niñas para querer aprender algo de programación con la idea de crear gráficos prácticamente como un juego.

### 4.3 En la enseñanza del curso de Series de tiempo

La parte de modelos ARIMA se encuentra en la mayor parte del software comercial para estadística en general y en los especializados en series de tiempo. Sin embargo, lo mismo no es necesariamente cierto en cuanto al análisis espectral. Con ese fin, hace muchos años escribía

mis programas en Fortran, luego en Matlab pero hace ya varios años he migrado hacia el R. Se pueden hacer las mismas cosas que se hacen con Matlab con la ventaja que todos los estudiantes pueden instalar R gratuitamente en su computadora. R tiene la transformada rápida de Fourier, lo que permite escribir nuestros propios programas para calcular el periodograma si uno así lo desea, pero también tiene ya incorporadas convenientes funciones. Existen varios paquetes para series de tiempo que van desde métodos bastante flexibles y relativamente recientes de descomposición a temas más sofisticados como los modelos ARFIMA y GARCH que antes solo se podían encontrar en software comercial muy especializado. Personalmente me gusta escribir mis propios comandos en R, cuando es posible, para mostrar a los alumnos los pasos necesarios por ejemplo para generar una serie de tiempo en base a varias armónicas o calcular un periodograma, pero ya hay muchos paquetes en R que hacen excelente trabajo y se pueden usar en clase o para la ejecución de tesis.

El 'task view' preparado por R. Hyndman <http://cran.r-project.org/web/views/TimeSeries.html> ofrece una visión de los recursos existentes en R para esta área. Un artículo interesante sobre el uso de R en la enseñanza de series de tiempo, Hodgess (2004), se puede encontrar en <http://www.amstat.org/publications/jse/v12n3/hodgess.html>

## **5. Recursos para aprender y enseñar a usar R**

Actualmente R ya se está enseñando y utilizando en la mayor parte de las universidades en el Perú y también se dan cursos de extensión, pero uno puede luego seguir aprendiendo por su cuenta de acuerdo a sus necesidades. Personalmente yo no aprendo nuevos tópicos de uso de software mediante el estudio detallado de un manual sino tratando de resolver problemas concretos; una estrategia algo caótica pero que me ha sido efectiva. En el caso de R es conveniente aprender lo básico primero y luego buscar proyectos específicos de acuerdo al interés de cada uno. Cuando hay algo que quiero hacer en R empiezo a buscar los comandos necesarios para hacerlo. Hay mucha información que la gente comparte generosamente en la internet. Claro que el tener algunos libros sobre el tema también ayuda, pero como ya mencionamos, en el portal de R se pueden encontrar manuales.

La motivación para aprender R en base a un caso concreto es particularmente cierto cuando se trata de gráficos. Si tengo en mente un gráfico que quiero producir para un artículo, trato de encontrar la forma de hacerlo en R y en ese proceso aprendo nuevos comandos. Por ejemplo, para poner varios gráficos en una sola figura como en la Figura 4 utilizo `par(mfcol=c(2,3))`, es decir uno crea un panel con dos filas y tres columnas de gráficos. Sin embargo, actualmente necesito hacer un panel más complejo en que no todas las particiones son del mismo tamaño, entonces trato de averiguar si hay una función más flexible, que este caso es la `split.screen` y procedo a estudiarla para lograr el tipo de gráfico que deseo hacer. En una oportunidad necesité incluir letras griegas en un gráfico y busqué en la internet la forma de hacerlo, por

ejemplo para incluir el símbolo  $\beta$  se escribe `expression(beta)`. Es de esa manera que he aprendido la mayor parte de las cosas que hago en R y que luego comparto con mis alumnos en el momento apropiado.

Si uno desea enterarse de todos los recursos que existen en R para un área específica, es conveniente entrar a <http://cran.r-project.org/web/views/>. Las “task views” son recuentos bastante completos escritos por especialistas en cada área. Hay una larga lista de temas para los cuales existen “task-views”, incluidos temas que tienen mucho interés actual tales como análisis de encuestas de diseño complejo, minería de datos, o estadística espacial. En el portal de R se puede acceder a una revista electrónica <http://journal.r-project.org/archive/2013-1/> donde hay artículos acerca de mejoras al lenguaje en sí y también sobre nuevos paquetes en R. Los creadores de algunos paquetes publican artículos explicando el respectivo paquete en la revista de software estadístico, la cual es electrónica y gratuita y se encuentra en <http://www.jstatsoft.org/>

R tiene ya incorporados varios conjuntos de datos que se pueden utilizar como ejemplos. Para ver una lista de los datos disponibles, escribir `data()` y se abrirá una ventana con la lista de datos. Varios otros conjuntos de datos vienen en paquetes adicionales.

En cuanto a la enseñanza de R, lo hago de dos modos. En los cursos de estadística que enseño les explico a mis alumnos como aplicar el método que estamos discutiendo usando R. Tenemos acceso también a software comercial e igualmente aprendemos a usarlo, pero pienso que el ponerlos en contacto con un lenguaje potente y gratuito como R es algo apropiado. Por otro lado, enseño un curso de computación estadística en el pre-grado en el cual la mitad del curso está dedicado a R y la otra mitad a SAS. En este curso el énfasis está en el manejo de archivos de datos y programación incluyendo procesos iterativos, uso de expresiones condicionales, simulaciones, y preparación de funciones.

Aunque este artículo refleja las experiencias de alguien que se dedica al trabajo universitario, R tiene mucho que ofrecer también al estadístico que mayormente trabaja en aplicaciones de la estadística. Hay paquetes en R para hacer análisis de encuestas de diseño complejo, minería de datos, estadística espacial, o aplicaciones a finanzas. En <http://cran.r-project.org/web/views/> se puede ver una larga lista de temas para los cuales existen “task-views”, es decir referencias muy completas de los recursos en R para cada uno de esos temas.

## **Conclusiones**

R es un software gratuito, algo que lo hace especialmente atractivo para su uso y enseñanza tanto en análisis sencillos como más especializados. Nuevos paquetes escritos en R para aplicaciones específicas aparecen continuamente y la información sobre cómo usarlos se puede encontrar fácilmente en la internet. R es casi indispensable en áreas como el análisis espacial y ofrece muchas herramientas para el análisis de series de tiempo. Es especialmente útil para producir gráficos con una gran cantidad de información, cuando queremos hacer un análisis ad-hoc que no se encuentra en el software comercial

o cuando algunos análisis se tienen que repetir para muchos conjuntos de datos. Puede ser una herramienta útil en las manos del docente o investigador y también para el estadístico que trabaja mayormente en aplicaciones. Existe mucha información sobre el uso de R, lo cual facilita el autoaprendizaje.

### **Referencias**

Bonett, D.G. ,Seier, E. (2006) Confidence Interval for a Coefficient of Dispersion in Non-normal Distributions. *Biometrical Journal*, Vol 48 (1) pp 144-148

Hodgess, E. M. (2004) A Computer Evolution in Teaching Undergraduate Time Series . *Journal of Statistics Education* Volume 12, Number 3 [www.amstat.org/publications/jse/v12n3/hodgess.html](http://www.amstat.org/publications/jse/v12n3/hodgess.html)

Joyner, M., Ross, C. , Seier, E. (2013) Distance to the Border in Spatial Point Patterns. *Spatial Statistics* Vol 6 pp 24-40

Seier, E., Bonett, D.G. (2011) A polyplot for visualizing location, spread, skewness and kurtosis. *The American Statistician*- Vol. 65, No. 4, 258-261