

¿Por qué comprar un programa estadístico si existe R?

CHRISTIAN SALAS ✉

*School of Forestry and Environmental Studies, Yale University, USA.
Departamento de Ciencias Forestales, Universidad de La Frontera, Temuco, Chile.*

RESUMEN. La estadística es una ciencia aliada a la investigación científica. Los científicos que trabajan en ecología, recursos naturales e ingeniería comúnmente emplean programas de computación para realizar análisis estadísticos. En este trabajo se revisan brevemente dos de los programas estadísticos más usados en estudios ecológicos, SPSS y SAS, y se comparan con el software estadístico R. Sobre la base de este análisis, se propone el uso de R en ciencias ecológicas e ingeniería en Latinoamérica y en países del tercer mundo en general, porque ofrece el uso gratuito de un software de primer nivel, así como también un mayor control de los análisis conducidos, extensa documentación, y un ambiente de programación desarrollado para aplicaciones estadísticas y con capacidad para ser empleado en otras áreas cuantitativas de diversas disciplinas.

[Palabras clave: estadística, computación, ecología estadística, estadística ambiental, biometría, software libre, Latinoamérica, docencia de estadística]

ABSTRACT. **Why purchase commercial statistical software if there is R?:** Statistics is used in all scientific disciplines. Researches on ecology, natural resources, and engineering use statistical software packages for conducting their statistical analysis. We briefly review two statistical software packages most often used in ecological and engineering studies, SPSS and SAS, and compare them with the free statistical software R. We recommend the use of R for problems in ecology and engineering in Latin America and third world countries, not only because it is a free, top shelf statistical software, but also because it offers a greater control of how each procedure is performed, thorough and widely available documentation, and a computing environment both suitable for statistics as well as for many quantitative areas in several disciplines.

[Keywords: statistics, computing, statistical ecology, environmental statistics, biometrics, free software, Latin America, statistical teaching]

INTRODUCCIÓN

La estadística es una ciencia aliada de la investigación científica. Nadie discute el gran aporte a la ciencia moderna de la teoría de la relatividad de Einstein, la mecánica cuántica y la teoría de la evolución de Darwin. Sin embargo, pocos parecen reconocer el gran aporte de la estadística en el último siglo. Ne-

yman (1955) proclamó a la estadística como “la sirvienta para todas las ciencias”, ejemplificando su importancia en el estudio de la existencia y evolución, medicina, psicología, industria y astronomía. La estadística no sólo ha contribuido enormemente al desarrollo de las anteriores, sino también en ingeniería, agricultura y silvicultura. Por ejemplo, Gregoire & Köhl (2000) comentan sobre el amplio uso

✉ School of Forestry and Environmental Studies,
Yale University, 360 Prospect Street, New
Haven, CT 06511, USA. Tel.: +1(203)432-9398.
Fax: +1(203)432-3809.
christian.salas@yale.edu

Recibido: 18 de marzo de 2008; Fin de arbitraje: 13 de mayo
de 2008; Revisión recibida: 14 de mayo de 2008; Aceptado: 19
de junio de 2008

de la estadística en el desarrollo de las ciencias forestales. Mediante la estadística se evalúan cuantitativamente hipótesis de investigación, se desarrollan modelos predictivos, se estiman parámetros y se analizan experimentos, entre otras aplicaciones. El análisis de regresión es uno de los métodos estadísticos más empleados en varias disciplinas, mientras que los métodos multivariados gozan de popularidad entre ecólogos. Los investigadores que trabajan en disciplinas aplicadas como las ciencias agrícolas y forestales, así como también en disciplinas que requieren de un intensivo trabajo en laboratorio (e.g., microbiología y análisis químicos) emplean diversos modelos estadísticos para el análisis de sus experimentos (nótese que los llamados “diseños experimentales” son, en realidad, modelos estadísticos).

El uso de un programa de computación (i.e., software) estadístico es importante tanto en la ciencia básica como en la aplicada (e.g., ejercicio profesional). En la práctica, tanto investigadores como profesionales emplean algún programa estadístico para realizar pruebas de hipótesis, ajustes de modelos y análisis de diseños experimentales complejos. Muchas veces deben analizar grandes bases de datos y una gran cantidad de variables. Microsoft Excel® es una planilla de cálculo ampliamente usada debido a que es parte de la instalación típica de computadores con sistema operativo Microsoft Windows®, y con la ayuda de algunos “add-ins” puede también ejecutar algunos procedimientos estadísticos (Zhu & Kuljaca 2002). Sin embargo, su uso en análisis estadístico sigue siendo bastante limitado. Además, se ha mostrado la baja calidad de los procedimientos estadísticos de M. Excel® (McCullough & Wilson 1999, 2002, 2005). Algunas investigaciones emplean análisis bastante específicos que son realizados en programas estadísticos pequeños, diseñados exclusivamente para tales fines, y a los que denominaré “software-tarea-específicos”. Por ejemplo, el programa SPPA (“Spatial Point Pattern Analysis”), que se usa para calcular la función de Ripley (Ripley 1977), la cual es empleada en estudios de estadística espacial como el de Haase et al. (1996) en matorrales y el de Salas et al. (2006) en bosques. Aunque

este tipo de programas son específicos, se pueden usar en investigación y ofrecen lo que se necesita para un trabajo puntual. Sin embargo, su uso está limitado a un tipo de análisis y, por lo tanto, sólo son de interés para un pequeño abanico de usuarios. En consecuencia, este trabajo se centrará en programas estadísticos genéricos que permiten ejecutar una variada gama de procedimientos, y no se abordarán los software-tarea-específicos.

En este artículo se revisan y comparan características generales de dos programas estadísticos comerciales de amplio uso en ecología (SPSS y SAS) con el programa estadístico libre R. El objetivo del presente trabajo es aportar a la discusión con respecto a los programas estadísticos empleados en ciencias ecológicas y a los recursos económicos necesarios para su uso tanto en investigación como en docencia universitaria. Se advierte al lector que el autor no tiene afiliación con ninguno de los tres programas estadísticos analizados.

PROGRAMAS ESTADÍSTICOS

Existen varios programas estadísticos genéricos. Sin embargo, sólo nombraré a los que se citan generalmente en artículos en revistas científicas latinoamericanas (SPSS, Stata, Systat y SAS), y luego introduciré una nueva alternativa, el programa R. Dado que SPSS, Stata y Systat poseen estructuras similares (aunque con diferencias en sus procedimientos), sólo consideraré SPSS por su mayor popularidad.

SPSS (SPSS Inc. 2007) es un software lanzado al mercado en 1968. Originalmente se desarrolló para las ciencias sociales, por lo que ofrece un uso sencillo de las opciones, acceso rápido a datos y procedimientos, generación de salidas y gráficos. SPSS es un programa con una interfaz gráfica de usuario (término denominado en computación, “GUI”) amigable, y sólo a través de ésta se accede a sus opciones (e.g., abrir los datos y ejecutar cálculos) mediante el uso de los botones de la interfaz gráfica.

SAS (SAS Institute Inc. 2007) ha sido por largos años el software más utilizado en la

comunidad estadística y, por lo tanto, también se ha propagado su uso entre investigadores de diferentes disciplinas. SAS, a diferencia de SPSS, es un programa que requiere el ingreso de comandos (i.e., sintaxis) para ejecutar gran parte de sus rutinas y opciones. Por lo tanto, necesita del conocimiento de la sintaxis antes de su uso. SAS ha llegado a ser el programa estándar empleado en ensayos clínicos y por la industria farmacéutica en los Estados Unidos.

R (Ihaka & Gentleman 1996; R Development Core Team 2007) es un programa estadístico y un lenguaje de programación de uso libre, de distribución gratuita y de código abierto (i.e., el código fuente del programa está disponible para los usuarios), desarrollado como un gran proyecto colaborativo de estadísticos de diversos países y disciplinas. R también es un programa basado sobre comandos, en el que se puede acceder a todos los procedimientos y opciones a través de sintaxis computacional. Fue oficialmente presentado en 1997 y es un software libre que se rige por la licencia general pública ("General Public License" o GPL) de la fundación de software libre ("Free Software Foundation" o GNU, <http://www.gnu.org/>). R es muy similar al programa estadístico S-plus (el cual no es gratuito y es distribuido por Insightful Corporation), ya que la implementación base y semántica de ambos son derivados de un lenguaje estadístico llamado S y de un lenguaje llamado Scheme (Ihaka & Gentleman 1996). Las diferencias entre R y S-plus radican en el léxico empleado, en el código para modelar y en otros aspectos técnicos computacionales que escapan al alcance de este artículo, pero que pueden ser revisados en Hornik (2008). De todas maneras, la mayoría de los comandos de R funcionan en S-plus, y viceversa.

Es importante hacer notar que tanto SPSS como SAS son programas comerciales y, por lo tanto, tienen una orientación y administración diferentes a las de R, y además se enfocan en aquellos mercados y usuarios que les proporcionan los mayores beneficios.

COMPARACIÓN GENERAL

Dado que diferentes programas implementan distintos algoritmos para llevar a cabo los mismos tipos de análisis, los usuarios se benefician de una comparación entre los programas más usados. Se han realizado comparaciones de cálculos para los procedimientos (e.g., regresión y experimentos factoriales, entre otros) implementados por algunos programas estadísticos (Okunade et al. 1993; McCullough 1999; Zhu & Kuljaca 2002). Sin embargo, estos son bastante específicos y se circunscriben a aspectos puntuales. En este contexto, se presenta una comparación general sobre la base de una serie de aspectos (Tabla 1).

a. Amigabilidad con el usuario. SPSS es bastante amigable para el usuario, ya que permite acceder a todas las opciones mediante un menú de funciones. Por su parte, SAS y R requieren conocer la sintaxis y/o los comandos antes de ejecutar un procedimiento, lo cual los hace poco amigables para aquellos usuarios no familiarizados con la programación computacional o con poco interés por aprender una cantidad de instrucciones. Para ejecutar R no es necesario utilizar el menú de funciones y para ejecutar SAS, además de requerir sintaxis es necesario utilizar los botones de la interfaz gráfica (e.g., se escribe la sintaxis para ajustar un modelo mediante el procedimiento de regresión, "PROC REG", y luego se debe accionar el botón "run" para ajustar el modelo).

No obstante, tanto en SAS como en R existen nuevas aplicaciones desarrolladas para facilitar el uso de los programas. En las últimas versiones de SAS se han implementado las utilidades "Insight" y "Analyst" que permiten realizar algunos tipos de análisis accionando botones sin necesidad de conocer la sintaxis. De igual forma, para R existe "Rcommander" (Fox 2005), que permite similares funciones. Sin embargo, R y SAS son sistemas basados en comandos, por lo que el empleo de las utilidades GUI de estos programas puede resultar complicado para el usuario ya que debe acceder a diferentes menús antes de ejecutar un procedimiento específico, difi-

Tabla 1: Comparación de aspectos generales entre los programas estadísticos SPSS, SAS y R.**Table 1.** Comparison of general features of the statistical software SPSS, SAS, and R.

Aspecto	Programa estadístico		
	SPSS	SAS	R
Amigabilidad con el usuario	Excelente	Baja-Regular	Baja-Regular
Manipulación de datos	Baja	Buena	Buena
Calidad de gráficos	Regular	Buena-Excelente	Excelente
Control de procesos	Baja	Excelente	Excelente
Costo	U\$S 1500	U\$S 7200	Gratis
Código fuente disponible	No	No	Sí
Variedad análisis estadísticos	Buena	Buena-Excelente	Excelente
Documentación	Excelente	Buena	Buena-Excelente
Soporte técnico	Bueno	Bueno	Bajo
Sistema operativo	Windows®	Windows®	Windows®
		Macintosh®	Macintosh®
		Linux	Linux

cultando también la reproducción de análisis estadísticos. Aunque el uso de las utilidades GUI hace más amigable el software, el usuario pierde el control de lo que requiere del programa (ver punto d), limitando además la flexibilidad de los análisis y la posibilidad de respaldar y registrar lo que se ha hecho. Claro está, existen varias clases de usuarios: hay quienes prefieren el control total de cada proceso ejecutado y quienes tan sólo buscan los resultados, sin importarles el proceso de cálculo. De todas maneras, tanto SAS como R permiten la adaptación al uso tanto de unos como de otros.

b. Manipulación de datos. Todos los programas permiten leer datos en una gran variedad de formatos estándares (e.g., ASCII, txt y dat), y de algunos otros específicos. Los tres programas permiten abrir datos en archivos de tipo M. Excel®, lo cual resulta atractivo para muchos usuarios.

Tanto SPSS como SAS permiten abrir el archivo de datos en una ventana aparte, lo cual puede ser importante para algunos usuarios. R muestra los datos, o una porción de ellos, según se requiera, aunque empleando la función “fix” también es posible mostrar y editar los datos en una ventana. Debido a que tanto

SAS como R emplean sintaxis y procedimientos que afectan a los datos en una memoria virtual, cualquier modificación en los datos (e.g., crear una variable que no existe en el archivo) no es guardada físicamente en el archivo (a menos que se requiera al finalizar una sesión). Es decir, uno puede incorporar transformaciones a un archivo de datos, dejando intacto el archivo original. Esta es una gran ventaja ya que por muchos análisis que se realicen, siempre es posible partir de un único archivo de datos, evitando la duplicación de versiones. SPSS, en cambio, trabaja físicamente sobre el archivo, modificándolo cada vez que se realizan análisis. Debe notarse que es posible evitar esto, pero debe seleccionarse en alguna de las diversas ventanas existentes. Además, SPSS no es tan versátil en la manipulación de un gran número de archivos con datos, a comparación de SAS o de R.

c. Calidad de gráficos. Una representación gráfica permite mostrar más eficientemente resultados y también ayuda a su entendimiento. SPSS ofrece una serie de gráficos tipo que si bien pueden modificarse en su formato (e.g., leyendas y color), son difíciles de personalizar. Aunque tanto SAS como R permiten el diseño personalizado de gráficos, SAS requiere del uso de diferentes rutinas (o paquetes), mien-

tras que las sintaxis de R son más sencillas (o “planas”) y no requieren de una gran cantidad de paquetes. R también ofrece una amplia gama de formatos en los cuales los gráficos pueden ser exportados, sin necesidad de mayor sintaxis. Finalmente, a título subjetivo, la calidad visual de un gráfico en R parecería ser superior a la de SAS y a la de SPSS.

d. Control de procesos. Los procedimientos estadísticos usan una serie de algoritmos que poseen diferentes variantes. Cuando un usuario no los conoce en profundidad, normalmente el programa usa variantes predefinidas de estos algoritmos. En todos los programas, estas definiciones pueden ser especificadas. En SPSS, sin embargo, resulta complejo cambiarlas (sólo están disponibles algunas básicas). Dado que SAS y R requieren sintaxis, también permiten un mayor control de los procedimientos estadísticos a ser ejecutados. Por ejemplo, al ajustar modelos no lineales puede especificarse el empleo de derivadas analíticas o numéricas. R es más flexible por ser de código abierto; un usuario puede usar las funciones programadas en el software como también escribir funciones propias de manera sencilla.

Tanto SPSS y SAS ofrecen generalmente una gran cantidad de salidas completas (“outputs”) para un procedimiento estadístico cualquiera. En cambio, R ofrece como salidas sólo aspectos básicos y, en el caso de que el usuario necesite más detalles, debe solicitarlos especialmente. Contar con demasiadas salidas puede provocar errores de análisis en aquellos usuarios con escasos conocimientos estadísticos (Searle 1989). En este sentido parece acertada la filosofía de R, que sólo muestra lo solicitado por el usuario al momento de ejecutar un procedimiento. Así, esa multiplicidad de estadísticos que aparecen en las salidas de los otros programas no “distrae” al usuario de R, quien puede centrarse en lo que más conoce, le interesa y entiende.

e. Costo. Cada empresa que desarrolla programas estadísticos ofrece diferentes costos de adquisición de sus productos en función del tipo de licencia (e.g., personal, empresa, educación y otros). No obstante, sólo se indi-

can aquí los costos para un usuario particular (i.e., licencia personal) como base de comparación. SPSS cuesta U\$S 1599 (SPSS Inc. 2007), un valor promedio a levemente alto entre los programas estadísticos de similar categoría, y corresponde a una licencia perpetua (i.e., puede ser empleado de por vida) de la última versión disponible (la número 16.0). Si el usuario desea actualizar la versión que adquirió, debe pagar un adicional de U\$S 400.

SAS es el software estadístico más caro del mercado. Cuesta U\$S 7200 (com. pers. SAS, New Haven, CT, USA, diciembre de 2007). Este valor corresponde sólo a una licencia anual para la versión del año en curso. El costo de actualización anual es de U\$S 2100, aproximadamente un 30% del valor de adquisición.

R es totalmente gratuito y está disponible en Internet, al igual que cada versión nueva. No es necesario pagar ni por obtener el software ni por actualizarlo. La instalación típica de R consiste de una serie de paquetes, y cada paquete nuevo que ha sido desarrollado es publicado en Internet, desde donde se puede obtener y agregar a R. A diferencia de SPSS y de SAS, la distribución de R es actualizable, pudiéndosele agregar paquetes adicionales por separado. Si bien el costo no debería ser el único aspecto a considerar en una comparación entre programas, la gratuidad de R es claramente ventajosa. A pesar de que para los usuarios particulares el costo es decisivo, para instituciones y empresas puede ser no tan importante porque las licencias corporativas (si fueran calculados luego en una base individual) son bastante menores tanto para SPSS como para SAS.

f. Variedad de análisis estadísticos. SPSS ofrece un amplio rango de procedimientos estadísticos, que probablemente cubren gran parte de los que se utilizan en ecología e ingeniería. Sin embargo, cuando se requiere de mayores especificaciones en los procedimientos (e.g., el ajuste de un modelo no lineal no converge y es necesario definir el rango posible de valores para los parámetros del modelo), SPSS no ofrece mucha versatilidad. SAS es un software de sólido desarrollo y en

donde la empresa invierte cerca del 20% de sus utilidades en investigación (SAS 2007). Por lo tanto, la variedad de procedimientos implementados es bastante amplia. R ha sido desarrollado por estadísticos que trabajan en diferentes instituciones a nivel mundial y, por lo tanto, implementa algoritmos modernos y robustos. Además, un número importante de paquetes están continuamente siendo desarrollados y puestos a disposición en Internet para su instalación. Esto implica, también, la disponibilidad de una gama amplia de procedimientos de primer nivel. Por ejemplo, el paquete nlme para ajustar modelos lineales y no-lineales de efectos mixtos en R, es explicado en detalle en el libro de Pinheiro & Bates (2000), es un referente en el tema. Así también, para aquellos usuarios que prefieren software-tarea-específicos, el desarrollo por expertos de paquetes en diferentes disciplinas permite la existencia de paquetes específicos, como el geoR para análisis geoestadístico en R.

La renovación e implementación de nuevos procedimientos en R es relativamente rápida. Frecuentemente aparecen nuevos procedimientos y/o paquetes en desarrollo y en revisión, los cuales después pueden ser obtenidos a través de Internet e instalados directamente. En cambio, SAS y SPSS demorarán años en implementar nuevos procedimientos, requiriendo necesariamente una nueva versión del software. Sin embargo, los paquetes de R no están garantizados, y son mejorados a medida que los usuarios encuentren problemas y los desarrolladores actualizan los paquetes. Por su parte, SPSS y SAS, al ser programas comerciales, deberían ofrecer paquetes más depurados.

Existen diferencias entre SAS y R en cuanto a la variedad de análisis estadísticos, aunque sólo a escala detallada. Por ejemplo, se podría decir que SAS posee una leve ventaja en modelos mixtos vs. R, ya que ofrece la opción de elegir diferentes distribuciones de probabilidad para los parámetros aleatorios. R, en cambio, actualmente sólo ofrece la opción de emplear una distribución normal. Dado que diferentes personas generan paquetes para R, y a pesar de que existe una cierta estandarización al respecto, las mismas funciones computacio-

nales están presentes en diferentes paquetes. Esto implica que el uso del programa no está 100% optimizado. Sin embargo, R ha sido y es desarrollado gracias a un trabajo colaborativo importante. El veloz avance ocurrido en los últimos tres años hace pensar que se implementarán mejoras.

g. Documentación y soporte de ayuda. Todos los programas ofrecen documentación, tanto manuales de usuario como libros con aplicaciones. Sin embargo, SPSS ofrece una documentación fácil de usar y de entender, quizás debido a que fue originalmente diseñado para las ciencias sociales, en las que la formación cuantitativa no es generalmente muy profunda. Siguiendo el estilo colaborativo de R, la comunidad científica usuaria de R ha sido especialmente generosa al producir manuales y diversos documentos gratuitos.

Una ventaja de SPSS y de SAS es el soporte (e.g., servicio al cliente), a través del cual es posible indicar problemas de ejecución en ciertos procedimientos y, por lo tanto, obtener el respaldo técnico de las respectivas empresas. Por otra parte, para R no existe un respaldo formal de una empresa con respecto a todos sus paquetes, rutinas y funcionamiento general. Es decir, R no tiene ninguna garantía legal y el usuario asume cualquier potencial problema causado por su uso (esto es definido en detalle en la licencia GPL). La falta de un responsable legal de R podría ser una desventaja para empresas que piensan emplearlo. Sin embargo, no debería ofrecer mayores problemas para usuarios individuales o instituciones de investigación. De todas maneras, gracias al trabajo colaborativo mencionado más arriba, los potenciales problemas en algún paquete son también mejorados, aunque eso depende de la voluntad y el esfuerzo del creador de dicho paquete. Finalmente, para los tres programas existen foros en Internet donde se plantean los problemas relacionados con los respectivos programas y las técnicas de análisis estadísticos empleadas, y donde los usuarios independientes publican sus soluciones, siendo una excelente alternativa de ayuda gratuita.

h. Sistemas operativos. A pesar de que el sistema operativo (S.O.) Microsoft Windows® está ampliamente difundido, existe una gran cantidad de usuarios que usan otros sistemas operativos. Los tres programas analizados están implementados para Windows®. Si bien tanto SPSS como SAS pueden funcionar en el S.O. Linux, su configuración es compleja. SPSS también puede ejecutarse en Macintosh®. R es el único que funciona de manera estable e íntegra en los tres sistemas operativos de mayor uso. La versatilidad de plataformas donde R puede ser instalado ofrece una ventaja para los diferentes usuarios en distintas disciplinas.

USO DE PROGRAMAS ESTADÍSTICOS EN DOCENCIA E INVESTIGACIÓN

La elección de un software estadístico en ecología y disciplinas afines (e.g., ciencias forestales, agrícolas y ambientales), y en ingeniería, normalmente depende de la formación de los usuarios, como así también de si será empleado en docencia o en investigación. En la docencia en ciencias biológicas-sociales, la tendencia es emplear programas que permitan ejecutar los procedimientos en la forma más sencilla posible, evitando que el alumno se confunda con demasiados detalles de programación (como los necesarios para SAS y R). En este contexto, los usuarios preferirían el uso de programas con GUI amigables. Por otra parte, en la docencia de disciplinas con preparación en matemática y programación computacional, si se fomentara el empleo de programas estadísticos basados sobre sintaxis desde los estudios de pre-grado, se ganaría un mejor entendimiento del tema (e.g., para poder programar el ajuste de algún modelo es necesario primero saber el modelo que se va a ajustar) y la resolución más fácil de los problemas del área. En cualquier caso, y aunque no con todas las facilidades de SPSS, tanto SAS como R poseen utilidades GUI que los transforman en programas basados sobre el uso de botones.

En investigación, los usuarios normalmente poseen mayor nivel de conocimientos estadísticos. En esta área pareciera que el empleo

de software basado sobre sintaxis es más apropiado. Nótese también que dado que R es un lenguaje de programación, permite su uso en una variedad de problemas que no son necesariamente estadísticos, como por ejemplo, optimización y modelación matemática. Otra ventaja de R, tanto en ciencia básica como aplicada, es que puede ser empleado independientemente de la institución del usuario. Con otros programas no gratuitos, el tiempo invertido en aprenderlos no es capitalizado cuando el usuario debe trasladarse a otra institución que no posee dicho software. En este contexto, existe un número importante y creciente de centros académicos y de investigación que emplean R.

La documentación de un software computacional es muy importante para saber realmente qué está calculando cada procedimiento pre-programado en un software (Searle 1989). En países latinoamericanos y del tercer mundo en general, donde el acceso a libros y literatura actualizada es muchas veces complejo, la posibilidad de contar con acceso a documentación gratuita es una fortaleza. En este sentido, el uso de R ofrece una ventaja. De todas maneras, es recomendable la compra de literatura (sobre todo para R) dado que la disponible gratuitamente por lo general no es la que mejor satisface las necesidades y requerimientos de los usuarios.

En Latinoamérica, el control de uso de programas legales (i.e., copias permitidas) es muy débil, y la piratería de software es un problema comúnmente aceptado. La tasa de piratería de software en Latinoamérica alcanza el 66%, una de las más altas del mundo. Países como Venezuela, El Salvador, Bolivia y Paraguay están ubicados entre los 20 países con mayores tasas de piratería en el mundo, con porcentajes que se ubican entre 82% y 86% (Business Software Alliance 2006). De acuerdo al mismo estudio, Chile y Argentina presentan tasas menores aunque igualmente altas en términos globales, con porcentajes de 68% y 75%, respectivamente. Incluso a nivel universitario, dado los elevados costos de algunos de los programas usados, se enseña empleando software sin licencia, que también es distribuido a los estudiantes para su práctica personal. Con

este proceder, es difícil disminuir la piratería. Aún peor, y considerando que a través de la docencia no tan solo se entrega conocimiento sino que se establecen principios éticos y morales, el uso de software sin licencia no debería ocurrir en centros académicos. Cuando los proyectos de investigación cuentan con fondos, se destinan grandes sumas de dinero a la adquisición de software estadístico. Estos fondos podrían ser destinados a otros ítems si se emplea un software estadístico gratuito. El uso de un programa estadístico de excelente nivel y gratuito como R ayudaría a la enseñanza en las universidades, además de permitir a los estudiantes seguir empleando el mismo programa en su futuro ejercicio profesional, sin necesidad de invertir dinero en programas estadísticos y con la ventaja de aprovechar los conocimientos técnicos adquiridos previamente.

CONCLUSIONES

Debido a que SAS y R son programas basados en comandos, permiten al usuario un mayor control de los procedimientos ejecutados, en comparación con SPSS. Además de la clara ventaja del costo cero de R versus los otros programas, las salidas de procesos que ofrece R son concisas y dejan al usuario la opción de solicitar un mayor nivel de detalle, favorecen una mejor práctica en el uso de la estadística y evitan la tentación de tratar de interpretar todos los estadísticos que aparecen en una salida. En resumen, la gran versatilidad de los procedimientos estadísticos disponibles (así como los tarea-específicos), la capacidad de producir gráficos de calidad y la amplia documentación gratuita, entre otros aspectos, hacen de R un excelente programa estadístico para ser usado en docencia e investigación. La gratuidad de R además, permite no solo transmitir el uso de un software legal, sino también acceder libremente a un programa de alta calidad. Por otra parte, la transparencia en la construcción de R permite un mayor control del proceso de generación de conocimiento por parte de los usuarios.

AGRADECIMIENTOS

A Dylan Craven (Yale University, USA), Gabriel Mancilla (Universidad de Chile) y Salvador Gezan (Rothamsted Research, UK) por sus comentarios en un primer borrador del presente trabajo. La detallada revisión de los evaluadores anónimos permitió mejorar el artículo. Cualquier error remanente en el artículo es responsabilidad del autor.

BIBLIOGRAFÍA

- BUSINESS SOFTWARE ALLIANCE. 2006. *Fourth Annual BSA and IDC Global Software Piracy Study*. Washington DC, USA. 17 p.
- FOX, J. 2005. The R Commander: A basic-statistics graphical user interface to R. *Journal of Statistical Software* 14(9):42.
- GREGOIRE, TG & M KÖHL. 2000. Editorial: Statistical ecology and forest biometry. *Environmental and Ecological Statistics* 7:213-216.
- HAASE, P; FI PUGNAIRE; SC CLARK & LD INCOLL. 1996. Spatial patterns in a two-tiered semi-arid shrubland in southeastern Spain. *Journal of Vegetation Science* 7:527-534.
- HORNIK, H. 2008. *The R FAQ*. 118 p. <http://CRAN.R-project.org/doc/FAQ/R-FAQ.html> [consultado el 13 de Mayo, 2008].
- IHAKA, R & R GENTLEMAN. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5(3):299-314.
- McCULLOUGH, BD. 1999. Assessing the reliability of statistical software: Part II. *The American Statistician* 53(2):149-159.
- McCULLOUGH, BD & B WILSON. 1999. On the accuracy of statistical procedures in Microsoft Excel 97. *Computational Statistics & Data Analysis* 31(1):27-37.
- McCULLOUGH, BD & B WILSON. 2002. On the accuracy of statistical procedures in Microsoft Excel 2000 and Excel XP. *Computational Statistics & Data Analysis* 40(4):713-721.
- McCULLOUGH, BD & B WILSON. 2005. On the accuracy of statistical procedures in Microsoft Excel 2003. *Computational Statistics & Data Analysis* 49(4):1244-1252.
- NEYMAN, J. 1955. Statistics - Servant of all sciences. *Science* 122(3166):401-406.
- OKUNADE, AA; CF CHANG & RD EVANS. 1993. Comparative analysis of regression output

- summary statistics in common statistical packages. *The American Statistician* **47**(4):298-303.
- PINHEIRO, JC & DM BATES. 2000. *Mixed-effects models in S and Splus*. Springer-Verlag, New York, USA. 528 p.
- R DEVELOPMENT CORE TEAM. 2007. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- RIPLEY, BD. 1977. Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, B*. **39**(2):172-212.
- SALAS, C; V LEMAY; P NÚÑEZ; P PACHECO & A ESPINOSA. 2006. Spatial patterns in an old-growth *Nothofagus obliqua* forest in south-central Chile. *Forest Ecology and Management* **231**(1-3):38-46.
- SAS INSTITUTE INC. 2007. *SAS*. Cary, NC, USA. <http://www.sas.com>.
- SAS. 2007. *SAS overview*. <http://www.sas.com/corporate/overview/index.html> [consultado el 14 de Diciembre, 2007].
- SEARLE, SR. 1989. Statistical computing packages: Some words of caution. *The American Statistician* **43**(4):189-190.
- SPSS INC. 2007. *SPSS*. Chicago, IL, USA. <http://www.spss.com> [consultado el 14 de Diciembre, 2007].
- ZHU, X & O KULJACA. 2002. A short preview of free statistical software packages for teaching statistics to industrial technology majors. *Journal of Industrial Technology* **21**(2):1-6.