

EL PAQUETE ESTADÍSTICO STATA PARA CIENCIAS DE LA SALUD

The statistical package of Health Sciences

JUAN VEGA¹.

RESUMEN

Entre las múltiples utilidades que las tecnologías de la información presentan a los investigadores y profesionales sanitarios se encuentra la de poder disponer de paquetes estadísticos que ayuden a la hora de manejar y analizar datos. Actualmente existe una amplia variedad de herramientas computacionales que permiten realizar estas tareas. El objetivo de este artículo es hacer una descripción del paquete estadístico Stata y las ventajas que éste posee frente a otros programas informáticos en el ámbito de las ciencias de la salud. Se hará hincapié en la potencia y flexibilidad del mismo, su amplia documentación y constante actualización. Además se presentan algunos comandos tanto generales como especializados que pueden utilizarse en el ámbito de las ciencias de la salud.

Palabras claves: Bioestadística, investigación, paquetes computacionales, softwares.

ABSTRACT

Entre las múltiples utilidades que las tecnologías de la información presentan a los investigadores y profesionales sanitarios se encuentra la de poder disponer de paquetes estadísticos que ayuden a la hora de manejar y analizar datos. Actualmente existe una amplia variedad de herramientas computacionales que permiten realizar estas tareas. El objetivo de este artículo es hacer una descripción del paquete estadístico Stata y las ventajas que éste posee frente a otros programas informáticos en el ámbito de las ciencias de la salud. Se hará hincapié en la potencia y flexibilidad del mismo, su amplia documentación y constante actualización. Además se presentan algunos comandos tanto generales como especializados que pueden utilizarse en el ámbito de las ciencias de la salud.

Key Words: Bioestadística, investigación, paquetes computacionales, softwares.

INTRODUCCIÓN

Conceptualmente, un paquete estadístico es un conjunto de programas informáticos específicamente diseñados para el manejo y análisis de datos que permiten resolver problemas de estadística descriptiva e inferencial¹. Este conjunto de programas funciona de manera interrelacionada, de tal forma que es posible pasar de uno a otro sin necesidad de salir del programa principal. En la actualidad existe una amplia variedad de tales paquetes, por ejemplo, Stata, SPSS, R, Minitab, SAS, Statistica, Statgraphics, entre otros.

En los paquetes de análisis de datos es posible trabajar de dos formas: mediante un sistema de menús y ventanas, o mediante la sintaxis, es decir, escribiendo los comandos que debe ejecutar el programa. El trabajo mediante

ventanas conlleva una mayor facilidad de uso pero también implica una mayor rigidez al momento de intentar ejecutar procedimientos que el programa no tenga predefinidos. Por otro lado, el uso de la sintaxis permite mayor flexibilidad ya que es posible incluir cálculos o procedimientos que no vienen programados de manera predeterminada, ampliando las funcionalidades establecidas en las ventanas.

El paquete estadístico de Stata

En la actualidad Stata es un programa ampliamente extendido tanto en ámbito geográfico como por tipo de usuario⁽²⁾ (instituciones académicas, centros de investigación, organismos públicos y empresas). Es utilizado especialmente en biomedicina, epidemiología, ciencias biológicas, ciencias sociales y otras. Este

¹ Escuela de Ingeniería Industrial, Informática y Sistemas. Universidad de Tarapacá, Arica, Chile.

software permite, entre otras funcionalidades, la gestión de datos, el análisis estadístico, el trazado de gráficos y las simulaciones. La popularidad del paquete se explica por la potencia combinada con la sencillez que ofrece para el análisis de datos⁽³⁾.

Uno de los puntos más fuertes de Stata es su flexibilidad en el manejo de datos, siendo muy adecuado para la gestión de grandes volúmenes de información. Permite realizar un elevado número de operaciones que van desde la simple manipulación de los datos hasta la aplicación de complejas técnicas de estimación. Los datos pueden, de una manera muy sencilla, cargarse, fusionarse, añadirse, ordenarse, editarse, generar con ellos nuevas variables y recodificarse en distintos formatos. Adicionalmente, se pueden añadir etiquetas y títulos a las variables, los gráficos, las bases de datos, etc. Hay que destacar tanto su calidad, como la sencillez y la facilidad en la elaboración y edición de gráficos y tablas.

Ventajas del software

Con respecto a otros paquetes, Stata cuenta con varias ventajas, por ejemplo, posee una interfaz amigable, contiene muchas de las técnicas estadísticas más recientes, se actualiza frecuentemente y posee procedimientos de análisis y métodos gráficos muy poderosos. La principal ventaja sobre sus más cercanos competidores reside en la velocidad del procesamiento de datos⁽²⁾. Aunque también tiene ciertas desventajas, por ejemplo, para algunas personas el manejo de comandos no resulta ser tan amigable. Sin embargo, vale la pena el esfuerzo en el aprendizaje ya que las potencialidades del software aumentan considerablemente.

Stata también se caracteriza por ser una aplicación completa e integrada, es decir, tener todos los elementos en forma conjunta para realizar análisis estadístico y manejo de datos, a diferencia de otros softwares, como

SPSS, que se distribuye mediante módulos separados. Otro aspecto a destacar de Stata es que es muy estable, es decir, raramente se bloquea. Un software estadístico que se está utilizando hoy en día, sobretodo en ciencias básicas, y que también podría ser utilizado en ciencias de la salud es el programa R, el cual si bien es gratuito tiene la desventaja de no poseer soporte técnico y, además por ser intensivo en programación necesita un esfuerzo de aprendizaje aún mayor que Stata.

Otra característica importante de Stata, a diferencia de otros softwares, es que sus procedimientos de cálculo son bien conocidos y están ampliamente documentados. De hecho, Stata se caracteriza por poseer abundante documentación y una completa ayuda en línea. La ayuda de Stata se puede acceder a través de manuales, ayuda en pantalla y mediante su página web. La ayuda en pantalla es muy amplia y proporciona la sintaxis de las órdenes, una serie de opciones relacionadas y ejemplos de utilización. Adicionalmente, existe una variedad de libros que utilizan el programa para ilustrar sus aplicaciones empíricas, algunos libros recomendables en el área de la salud son Carvajal, (2004)⁽⁴⁾; Dupont, (2002)⁽⁵⁾ y Martínez et al., (2014)⁽⁶⁾. Otros libros se pueden encontrar, junto con manuales relacionados, en la página web <http://www.stata.com>.

La página web de Stata incluye, entre otras cosas, respuestas a las preguntas más comunes de los usuarios (FAQ), actualizaciones oficiales del paquete, videotutoriales, información sobre cursos interactivos de Stata por internet y sobre reuniones oficiales de usuarios en distintos lugares del mundo. Cabe destacar que existe una red internacional de usuarios de Stata. También es posible suscribirse a Statalist, un foro dedicado a discusiones sobre Stata así como a discusiones estadísticas. Stata también cuenta con una revista, The Stata Journal (ISI Web of Knowledge) es una publicación trimestral que contiene artículos sobre

estadísticas, análisis de datos, métodos de enseñanza y uso efectivo del software⁽⁷⁾. Además publica artículos revisados junto con notas y comentarios más cortos, columnas regulares, reseñas de libros y otros materiales de interés para investigadores. Su dirección web es <http://www.stata-journal.com/>. Otro aspecto importante en un programa estadístico es la fiabilidad numérica. Esto se relaciona con que pequeñas diferencias en la precisión numérica de los dígitos decimales pueden producir grandes cambios en el resultado de un procedimiento posterior que lo incorpora. En opinión de Rochina y Sanchis, (2002)⁽²⁾ Stata destaca en este campo frente a sus más cercanos competidores. Los resultados de las diversas pruebas realizadas sobre esta materia pueden verse en la web de Stata.

Martínez et al., (2014)⁽⁶⁾ argumentan que ya hace tiempo que Stata está demostrando ser superior en muchos aspectos a SPSS, sin dejar de ser igualmente amigable. En opinión de estos autores, en España especialmente en el ámbito de la epidemiología, casi se ha abandonado SPSS y se usa preferentemente Stata. Por otro lado, una licencia de Stata es más asequible, desde el punto de vista económico, que una licencia de SPSS. Además este último tiene licencias más transitorias que las de Stata.

Stata en las ciencias de la salud

Stata es desarrollado por Stata-Corporation y está disponible en las siguientes versiones: Stata/IC para bases de datos de tamaño medio, Stata/SE para bases de datos grandes y Stata/MP para mayores volúmenes de datos con análisis más complejos⁽⁸⁾. Su interfaz es fácil de utilizar, consiste en menús con las opciones de análisis y procesamiento de datos, y adicionalmente tiene la opción de comandos, lo que como ya se ha mencionado otorga mayor versatilidad. Al abrir Stata se observa una pantalla dividida en cuatro paneles. Estos contienen los resultados, una lista de variables, los comandos actuales y la historia reciente de los comandos, Figura 1. La sintaxis de comandos es lógica, clara y predecible⁽²⁾. La estructura

típica de los comandos está compuesta de los siguientes elementos:

comando variables [if expression], [opciones]

El nombre del comando indica la acción que se quiere realizar. A continuación se debe escribir la o las variables a utilizar. En forma opcional es posible indicar una condición sobre variables (tal como un filtro). Finalmente se pueden escribir opciones que otorgan mayor detalles a las acciones a ejecutar⁽⁹⁾.

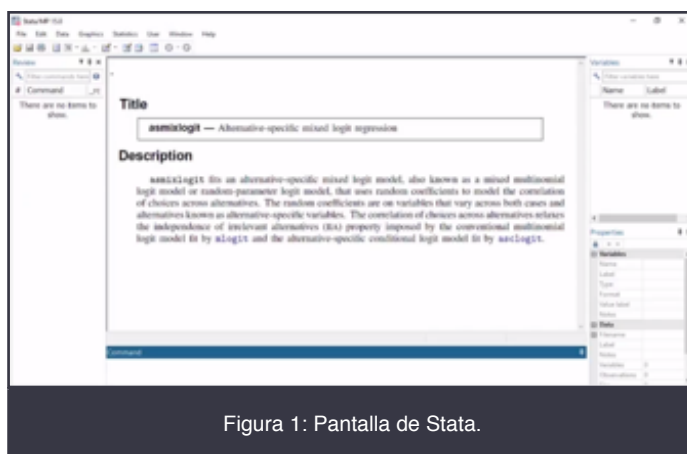


Figura 1: Pantalla de Stata.

Stata cuenta con una gran cantidad de comandos que se pueden utilizar en ciencias de la salud. La Tabla 1 presenta algunos ejemplos de comandos con una breve descripción de su funcionalidad. Los doce primeros comandos (en celeste) son de propósito general, es decir, también pueden ser aplicados en otras áreas de conocimiento. Estos comandos corresponden a procedimientos relacionados con estadística descriptiva, intervalos de confianza, pruebas de hipótesis, regresión, etc.

Los últimos ocho comandos (en verde) son especializados y se aplican principalmente en las ciencias de la salud. Permiten realizar análisis estadísticos en estudios de supervivencia, de casos y controles, de cohortes, etc. Cada uno de estos comandos posee además diversas opciones que otorgan gran capacidad a los procedimientos

Tabla 1. Ejemplos de comandos en Stata que se pueden utilizar en ciencias de la salud

Nombre del Comando	Descripción
<i>generate</i>	Permite generar nuevas variables
<i>recode</i>	Recodifica variables
<i>summarize</i>	Principales estadísticos de las variables como la media, desviación estándar y percentiles
<i>tabulate</i>	Cuadro de frecuencias, tablas de contingencia con estadísticos de asociación
<i>graph bar</i>	Crea gráficos de barra
<i>histogram</i>	Crea histogramas
<i>ci</i>	Calcula intervalos de confianza y errores estándar
<i>ttest</i>	Realiza Test t de una muestra
<i>oneway</i>	Análisis de varianza de una vía
<i>regress</i>	Realiza regresiones lineales
<i>logit</i>	Realiza regresiones logísticas
<i>glm</i>	Modelos lineales generalizados
<i>cs</i>	Calcula Riesgo relativo y su intervalo de confianza
<i>cc</i>	Calcula Odd ratio y su intervalo de confianza para estudios de caso-control
<i>tabroc</i>	Dibuja la curva receiver operating characteristic (ROC) para comparar pruebas diagnósticas
<i>stcox</i>	Permite realizar Regresión de Cox
<i>poisson</i>	Permite realizar Regresión de Poisson
<i>stset</i>	Permite realizar estudios de supervivencia
<i>sts graph</i>	Gráfica curvas de Kaplan-Meier
<i>ir</i>	Calcula tasas de incidencia en estudios de cohorte

Fuente: Elaboración propia.

de cálculo que se pueden realizar.

Otro beneficio de utilizar los comandos es permitir al usuario construir sus propios programas o rutinas, con el tipo de análisis que frecuentemente realiza y que responden a sus necesidades particulares, permitiendo extender las funciones propias de Stata. Además estas rutinas pueden aplicarse de manera automática, sin necesidad de escribir las operaciones varias veces o seguir una secuencia específica de ventanas, con el consiguiente ahorro de tiempo y esfuerzo. Asimismo Stata se retroalimenta continuamente con las rutinas y los comandos escritos por los miembros de la red internacional de usuarios que posee. Estos comandos y rutinas desarrolladas están disponibles

(en muchos casos sin costo) para todos los usuarios de Stata. Esta extensa comunidad conformada está aportando a las funciones nuevas rutinas constantemente⁽¹⁰⁾.

CONCLUSIONES

En la actualidad, el uso de programas computacionales para administrar datos se ha tornado una herramienta indispensable para el trabajo de investigación. Stata es un programa estadístico muy potente que permite manejar, analizar y representar gráficamente datos. Si se toma el tiempo de aprender a programar procedimientos, se logrará gran flexibilidad y versatilidad en la utilización del

software.

Otro elemento a destacar de Stata es que se desarrolla y actualiza rápidamente tanto de un modo formal como informal y que proporciona una excelente asistencia al usuario, que proviene tanto de la propia empresa como de su comunidad de usuarios. Si a todas estas ventajas le añadimos el hecho de ser un programa de bajo costo en comparación con los demás paquetes estadísticos y que además proporciona precios asequibles para la comunidad académica, podemos concluir que Stata se convierte en un programa sumamente útil para la investigación aplicada. Y, específicamente, debido a que posee una gran cantidad de funciones y procedimientos tanto generales como específicos que se relacionan con las ciencias de la salud se concluye que este paquete es altamente recomendable para esta área de investigación.

REFERENCIAS

1. De Ramírez, J. y Buitrago, J. (2007). Paquetes tecnológicos para el tratamiento de datos en investigación en educación matemática. Paradigma. vol.28 n°1.
2. Rochina, M. y Sanchis, J. (2002). Revista de Economía Aplicada, vol. 10, n° 28, pp. 163-174.
3. Mestre, R. Stata, Un paquete sencillo pero potente.

- Revista de Economía Aplicada. vol2. n°5, pp. 163-172.
4. Carvajal, Y. (2004). Administración y exploración de datos en Salud Pública usando Stata. Facultad de Medicina. Universidad de Chile.
5. Dupont, W. (2002). Statistical Modeling for Biomedical Researchers. A Simple Introduction to the Analysis of Complex Data. Cambridge University Press.
6. Martínez, M., Sánchez-Villegas, A., Toledo, E. y Faulin, J. (2014). Bioestadística amigable. Elsevier.
7. Stata-journal.com [Internet]. [Consultado el 18 de Julio de 2017]. Disponible en: <http://www.stata-journal.com/>
8. Stata.com [Internet]. [Consultado el 18 de Julio de 2017]. Disponible en: <http://www.stata.com/products/which-stata-is-right-for-me/>.
9. Vela, F. (2009). El proceso de enseñanza-aprendizaje del análisis de regresión lineal con Stata. Universidad Autónoma Metropolitana Xochimilco. División de Ciencias Sociales y Humanidades.
10. Morales, D. (2014). ¿Stata o R? Blog Economía Aplicada. [Internet]. [Consultado el 18 de Julio de 2017]. Disponible en: <https://moraleseconomia.blogspot.cl/2014/07/stata-o-r.html>

ACEPTACIÓN Y CORRESPONDENCIA

Correo Autor:

jvegav@uta.cl

Fecha de Aceptación: 01 de Agosto del 2017

Fecha de Publicación: 30 de Septiembre del 2017