ARTICULOS ORIGINALES
Rev Chil Salud Pública 2013;
Vol 17 (2): 117-123

Representación gráfica de información multivariante. Aplicación al sistema de salud de Chile (2010)

Irene Schiattino
Escuela de Salud Pública
Facultad de Medicina,
Universidad de Chile.
ischiattino@med.uchile.cl

Claudio Silva
Escuela de Salud Pública
Facultad de Medicina,
Universidad de Chile.

RESUMEN

Objetivo: Divulgar entre los investigadores en salud pública algunas propuestas de gráficos estadísticos multivariantes, tales como las caras de Chernoff, los gráficos de estrella, las curvas de Andrews y gráficos de radar.

Material y método: Los gráficos estadísticos multivariantes propuestos son aplicados para describir los Sistemas de Salud Públicos de Chile, año 2010, (n=29) en base a información cuantitativa de dimensión p=7; ellos son procesados usando software estadísticos standard como R, STATA y SAS.

Resultados: La descripción de los 29 Sistemas de Salud Públicos de Chile en base a información cuantitativa de dimensión siete, representados en cada uno de los gráficos multivariantes exhibidos, permite determinar conglomerados, estándares y tendencias.

Conclusiones: La disponibilidad de eficientes software estadísticos sugiere complementar las indispensables estrategias de análisis exploratorio de nuestra información con representaciones gráficas adecuadas al contexto multidimensional que pueden ayudar a alcanzar una mejor comprensión del problema de interés. La síntesis gráfica generada por la aplicación de gráficos estadísticos multivariantes para describir los Sistemas de Salud Públicos de Chile complementa nuestro conocimiento de este aspecto de la realidad nacional, puede sugerir algunas hipótesis, refutar otras y ayudar en la interpretación de resultados complejos. Naturalmente esto es extensible a muchas otras situaciones de interés para el investigador salubrista.

Palabras clave: Investigación en Salud Pública. Representación gráfica. Información multivariante.

Graphical representation of multivariate data, applied to the Chilean Healthcare System

ABSTRACT

Objective : To inform public health researches about some proposal of multivariate statistical graphics, such as Chernoff faces, star charts, Andrews curves and radar graphs. Its use is now possible thanks to recent software development.

Material and method: Proposed multivariate statistical graphics are used to describe the Systems of Public Health of Chile, year 2010 (n = 29), using quantitative information with dimension p = 7; these graphics are processed using standard statistical software such as R, STATA and SAS.

Results: The description of the 29 systems of public health of Chile based on quantitative information from dimension seven, represented in each of multivariate graphs, allow to determine clusters, standards and trends.

Conclusions: The availability of efficient statistical software indispensable suggests to complement strategies of exploratory analysis of our data with appropriate graphical representations of multidimensional context that can help us to achieve a better understanding of the problem of interest. The graphical summary generated by the application of multivariate statistical graphs to describe Public Health Systems of Chile complements our knowledge of this aspect of national life, may suggest hypotheses, refute others and assist in the interpretation of complex results. Naturally this can be extended to many other situations of interest to the health care professional.

Key words: Public Health research. Graphics representation, Multivariate information.

INTRODUCCIÓN

La investigación en Salud Pública es esencialmente observacional; por lo tanto siempre estamos enfrentando situaciones de estudio multidimensionales, referidas a información naturalmente compleja. Familiarizarnos con tal información requiere un análisis exploratorio de datos exhaustivo, no sólo uni y bidimensional, sino multidimensional.

En décadas recientes se ha desarrollado una gran variedad de procedimientos gráficos para representar datos multivariantes (Ver referencias), que la disponibilidad de eficientes software estadísticos sugiere aplicar para complementar las estrategias básicas uni y bidimensionales de análisis exploratorio de la información. Representaciones gráficas novedosas pueden ayudarnos a alcanzar una mejor comprensión del problema de interés.

En esta línea de pensamiento docente-estadístico, el objetivo de este trabajo es presentar una aplicación local de algunas de las propuestas de gráficos estadísticos multivariantes, tales como Caras de Chernoff, Gráficos de estrella, Gráficos de radar y Curvas de Andrews. Estos gráficos en general responden a ¿qué variables son las dominantes en cada unidad de observación elegida? ¿se pueden establecer similitudes entre estas unidades? ¿se pueden establecer grupos ("clusters")? ¿existen valores atípicos?. En la actualidad podemos encontrar variadas aplicaciones de estos gráficos, en el ámbito de percepción de consumidores, calidad de servicio, actitud hacia políticas de protección del medio ambiente, clasificación de productos, ensayos clínicos, ciencias sociales, etc. En este trabajo, son aplicados para describir los Sistemas de Salud Públicos de Chile (n=29) en base a información cuantitativa de dimensión p=7, Año 2010.

Los métodos gráficos descritos son implementados principalmente en los software estadísticos R, STATA y SAS.

METODOLOGÍA

El Sistema Público de Salud Chileno cubre el territorio nacional y está estructurado en 29 organizaciones denominadas Servicios de Salud. La información estadística sobre el funcionamiento de cada uno de ellos es llevada por el Ministerio de Salud y se actualiza constantemente en la página web del Departamento de Estadísticas e Información de Salud de ese ministerio (www.deis.cl). De esa fuente seleccionamos 7 dimensiones: Dotación, Días Cama Disponibles, Días Cama Ocupados, Días de Estada, Promedio Camas Disponibles, Número de Egresos y Número de Egresos Fallecidos para cada uno de los servicios (año 2010). Un dendrograma clásico evidencia la heterogeneidad de estos 29 "vectores". (Ver Figura 1). Las variables fueron estandarizadas para evitar que una variable "domine" el cálculo de distancias y el proceso de conglomeración debido a que sus valores son (por un tema de escala) miles de veces mayores que los de las otras variable.

RESULTADOS

Caras de Chernoff

Chernoff (1971, 1973) propuso el uso de una cara o rostro humano para representar puntos en k dimensiones, asignando a cada unidad de observación una cara de manera que la posición, longitud y forma de cada una de los componentes faciales (ojos, cejas, nariz, boca, orejas, pelo, etc.) refleja el comportamiento de una de las variables que intervienen en el estudio. El rango de variabilidad se establece de manera que la estructura global

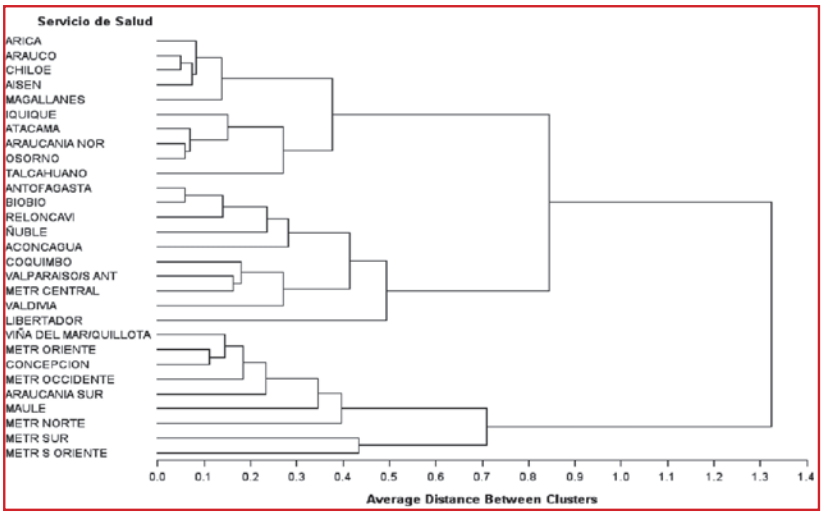

Figura 1. Dendograma: 29 Servicios de Salud del Sistema Público Chileno, Año 2010

mantenga las características básicas de una cara. La asignación de cada variable a los párametros faciales es arbitraria.

El algoritmo que los software utilizan en la actualidad fue desarrollado por Flury (1980), Schupbach (1987) y Friendly (1991).

La Figura 2 presenta el gráfico de Caras de Chernoff de los Sevicios de Salud agrupados en seis cluster según lo sugerido en el Dendograma, obtenido mediante estandarización clásica de sus variables (Ver Figura 1). Cada una de las caras representa un Servicio de Salud y las variables fue-

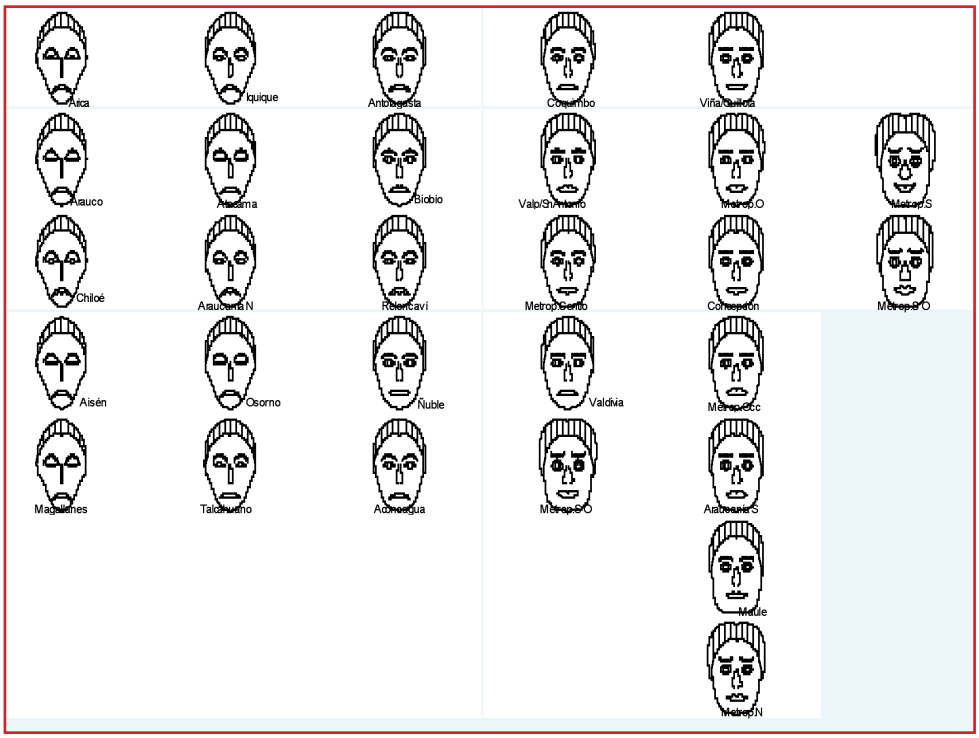

Figura 2. Caras de Chernoff : 29 Servicios de Salud del Sistema Público Chileno, Año 2010.

ron asignadas a los rasgos físicos como se indica: fline=Dotación, isize=Días Cama Disponibles, bcurv=Días Cama Ocupados, hupper=Días de Estada, hdark=Dias de estada, nose=Promedio Camas Disponibles, msize=Número de Egresos, mcurv=Número de egresos Fallecidos.

En el software STATA se aplicó primero el comando chernoff con las especificaciones de los rasgos físicos, seguido de la opción saveall para grabar las caras y a continuación se aplicó el comando graph combine para definir los cluster sugeridos por el dendograma.

Gráficos de Estrella

Los gráficos de estrella (Chambers, 1983), al igual que las Caras de Chernoff, asignan a cada unidad de observación una estrella con tantos rayos o ejes (igualmente espaciados y que confluyen en un centro geométrico) como variables queramos representar. Las longitudes de los rayos son proporcionales a los valores de las variables en la observación asociada a la estrella. Los extremos de los rayos se conectan con segmentos de líneas rectas para formar una estrella. De esta forma, podemos agrupar las observaciones según las similitudes que presentan las estrellas. En todas las estrellas se usa siempre el mismo rayo o eje para representar la misma variable. El eje j en una estrella i depende de xij en valor absoluto o relativo). Para facilitar la inspección visual que nos permita distinguir entre estrellas, se suelen representar características similares en rayos próximos.

Algunos software ordenan en función de la correlación de cada variable con la primera componente principal.

La Figura 3 presenta el Gráfico de Estrellas de los Servicios de Salud Chilenos, cada una de las estrellas representa un Servicio de Salud y las variables (rayos o ejes de cada estrella) fueron asignadas como se indica en la figura. Se aplicó el comando graph7 del STATA y la opción star, adicionalmente se aplicó la opción select para obtener los respectivos clusters.

Gráfico de Radar

Se asemeja a un gráfico de estrella, pero se construye en base a UNA estrella para cada variable en estudio y no para cada unidad de observación; las unidades de observación identifican rayos concéntricos. Los valores individuales de cada variable, debidamente

Estandarizados, se representan sobre los rayos recién mencionados y la conexión de ellos genera la estrella de esa variables. Los gráficos de radar son un tipo de gráfico de estrellas, consiste en un conjunto de rayos a partir de un centro, determinados por una variable categórica seleccionada previamente y un conjunto de estrellas concéntricas definidas por el comportamiento de cada una variables numéricas en un espacio específico de datos. El espacio determinado es establecido por la unidad de referencia de la variable en todos los subgrupos. La longitud de los datos de un determinado rayo es proporcional a la magnitud de la variable para el subgrupo con respecto al máximo de la variable a través de todos los subconjuntos. La Figura 4 presenta el gráfico de radar de los Servicios de Salud Chilenos. Se aplicó el comando radar del STATA, con la variable que representa los respectivos Servicios de Salud y las siete variables numéricas previamente estandarizadas.

Curvas de Andrews

Andrews (1972) propuso un método basado en la transformación de Fourier para representar datos multivariantes en dos dimensiones. Cada punto k dimensional es representado por una función

$$f_x(t)=X_1/\sqrt{2}+X_2 sent+X_3 cost+X_4 sent2+X_5 cost2+X_6 sent3+X_7 cost3+...$$

en un rango de $-\prod < t < \prod$ para permitir la representación de los datos y $X = X_1, X_2, ...X_k$ representa el vector de observaciones. Se puede demostrar que esta representación preserva las distancias euclídeas, en el sentido de que a dos elementos de la población con valores similares de las variables corresponden curvas próximas. Cada variable de cada observación es representado por una componente individual en la suma de la transformada de Fourier. La magnitud de cada variable de un sujeto o elemento particular afecta la frecuencia, la amplitud y la periodicidad de la función, dando una representación única para cada sujeto.

La Figura 5 presenta las Curvas de Andrews para cada uno de los Servicios de Salud Chilenos, cada una de las funciones representa el comportamiento de las dimensiones estudiadas. Cada sección de esta figura muestra como se agrupan estas gráficas replicando los conglomerados iniciales. Se programaron las funciones y se aplicó el comando graph twoway del STATA.

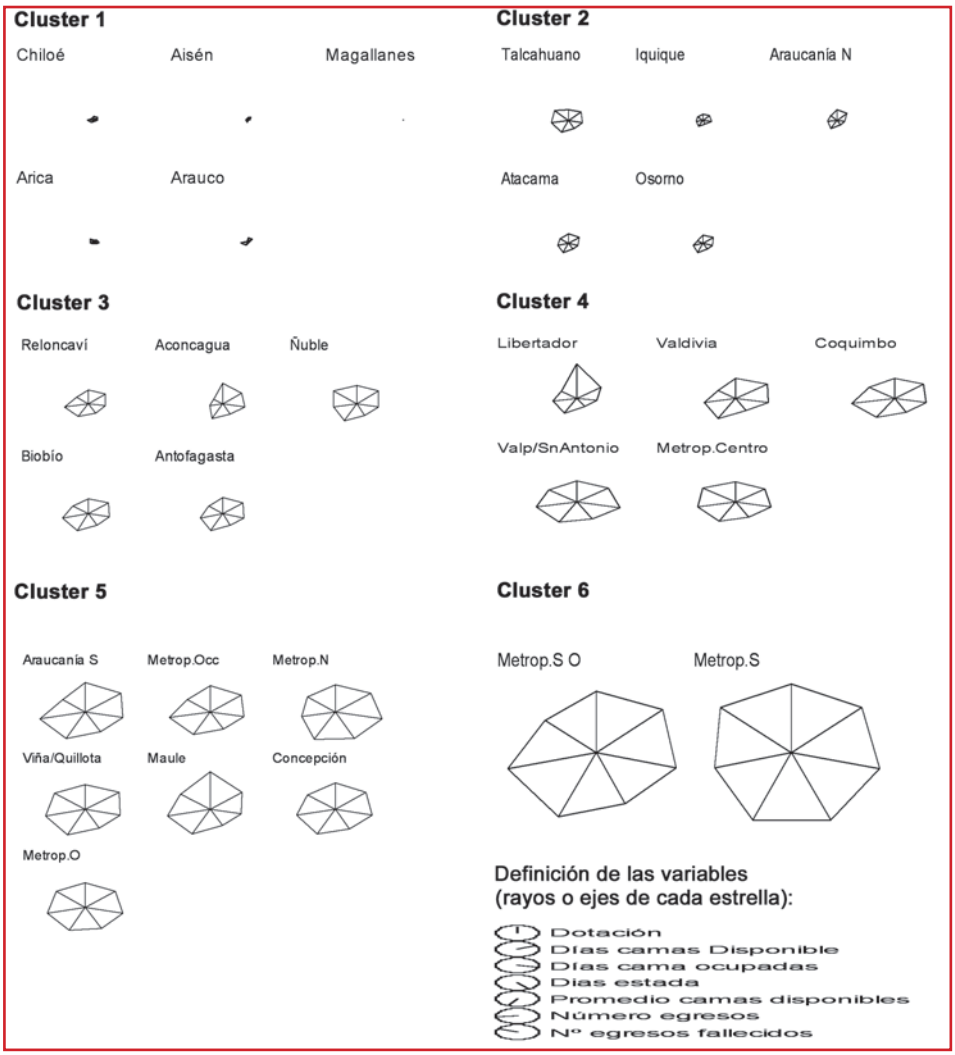

Figura 3. Gráficos de Estrellas: 29 Servicios de Salud del Sistema Público Chileno, Año 2010

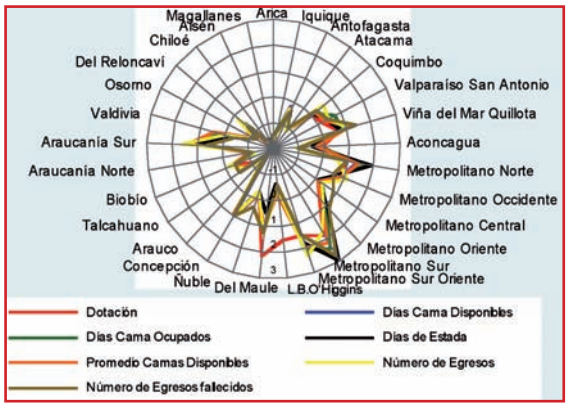

Figura 4. Gráficos de Radar: 29 Servicios de Salud del Sistema Público Chileno, Año 2010

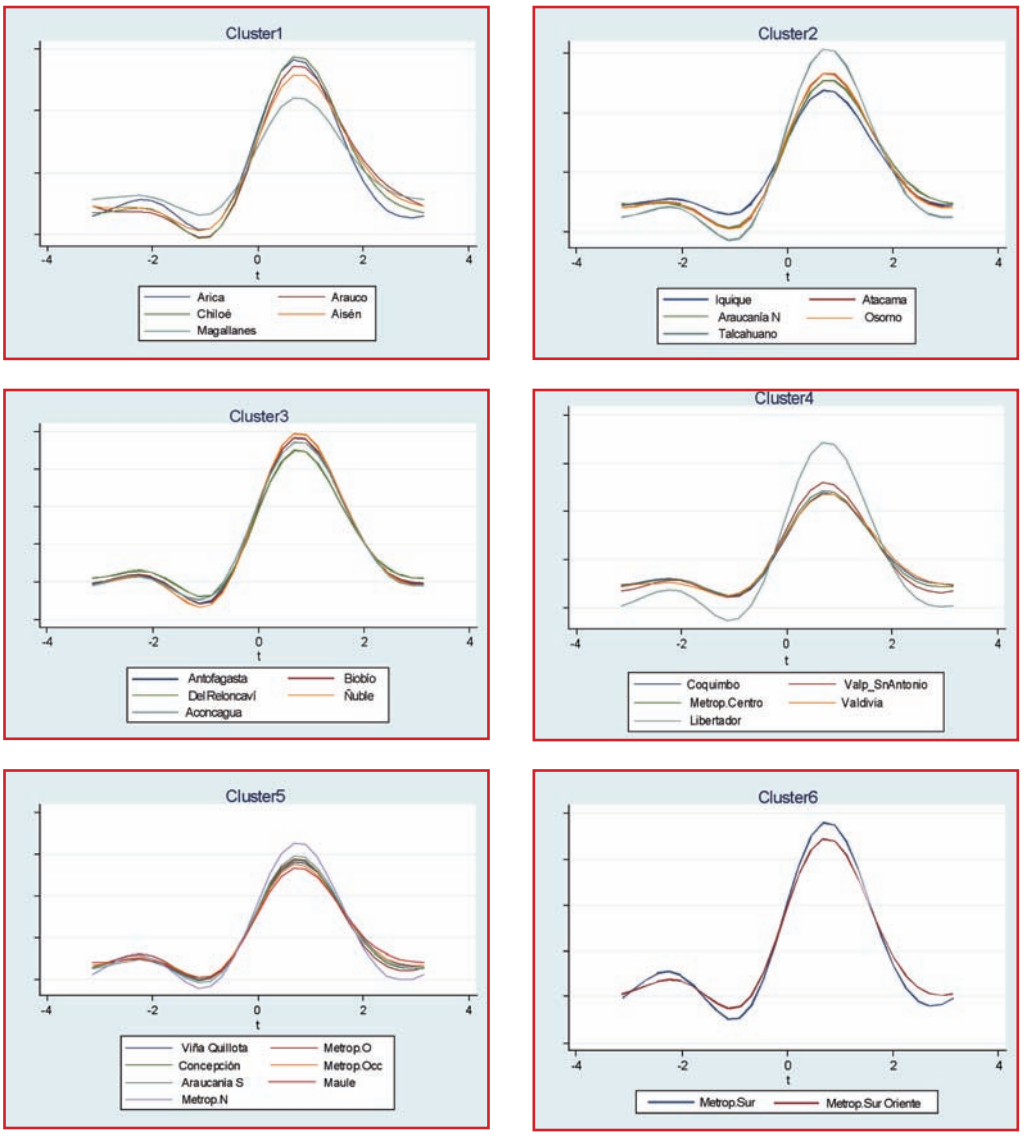

Figura 5. Curvas de Andrew's: 29 Servicios de Salud del Sistema Público Chileno, Año 2010.

CONCLUSIONES

La información multivariante en términos de distancias (euclidianas o no) y su uso para detectar agrupaciones de casos en conglomerados, es familiar en los cursos clásicos de Estadística y, de alguna manera, llega al público general. Sin embargo, es innegable que la individualidad de cada atributo desaparece, a menudo, en este enfoque. En cambio, los métodos gráficos ilustrados aquí, respetan, en distinta medida, esa individualidad y son una ayuda en la interpretación de los resultados, determinando clusters, estándares y tendencias.

La descripción de los 29 Sistemas de Salud Públicos de Chile en base a información cuantitativa de dimensión siete, representados en cada uno de los gráficos multivariantes, corresponde a una realidad en Chile y resulta motivante que la síntesis gráfica generada por cada método sea consistente con esa percepción: para nuestros alumnos resulta motivador que análisis necesariamente sofisticados en términos matemáticos, tengan una cara

amistosa, palpable, real y, a veces, hasta cómica. Debemos, además, destacar que la inquietud por representar adecuadamente información multivariante se mantiene vigente y busca explotar las posibilidades tecnológicas más recientes. Por ejemplo, la Organización para la Cooperación y el Desarrollo Económicos (OECD) ha creado un "Better Life Index" que reúne información de cada uno de sus 36 países miembros sobre 11 índices socio-económicos (que a su vez resumen datos sobre 96 variables). La representación gráfica multivariante propuesta es muy interesante, colorida y animada, pero escapa a las posibilidades de un texto clásico; sugerimos visitar http://www.oecdbetterlifeindex.org/.

REFERENCIAS

1. Chernoff H. The Use of Faces to Represent Points in K-Dimensional Space Graphically. Journal of the American Statistical Association. 1973; 68(342): 361-368.
2. Friendly M. Statistical Graphics for Multivariate Data. Proceedings of the SAS User's Group International Conference. 1991;16, 1157-1162
3. Toit S.H.C, Steyn A.G.W, Stumpf R.H. Graphical Exploratory Data Analysis. 1986; pp 54-104. New York: Springer-Verlag
4. Raciborski R.. Graphical representation of multivariate data using Chernoff faces. The Stata Journal. 2009; 9 (3): 374-387.
5. Shi-Tao Y, Glaxo S, King P, PA. Using Radar Chart to display Clinical Data. NESUG18 (Northeast SAS Users Group Conference Proceeding).
6. [Online: www.nesug.org/conference/general-conference.htm]
7. Stata Graphics. Reference manual. Release 12 (2011). A Stata Press Publication Stata Corp LP College Station, Texas.