

## MÉTODOS GRÁFICOS EN EL ANÁLISIS MULTIVARIANTE

**Díaz Leno, M. S., Martín Rodríguez, J.  
Tardáguila García, P.**

### INTRODUCCIÓN

El **lenguaje gráfico** es el «conjunto de símbolos y convenios que permiten comunicar una información cuantitativa de la manera más eficiente posible» (GETE-ALONSO y del BARRIO, 1990).

Si nos detenemos un momento a pensar en el mundo que nos rodea vemos como el lenguaje gráfico se utiliza en absolutamente todo nuestro entorno.

Hace ya tiempo que las representaciones gráficas abandonaron las publicaciones especializadas, en las que se utilizan como herramienta de comunicación y análisis de datos estadísticos, para pasar a formar parte de las *herramientas de comunicación social* (televisión, prensa, propaganda...)

La visión es la modalidad sensorial dominante del ser humano; nuestro cerebro está altamente capacitado para el manejo de información visual, siendo capaz de reconocer y procesar imágenes gráficas con una simple inspección ocular.

Existen varios criterios para clasificar los métodos de representación gráfico; pero nosotros nos centraremos en la clasificación en base al *procedimiento gráfico* y la *técnica estadística subyacente*, en el Análisis Multivariante pueden distinguirse claramente dos grandes grupos de técnicas gráficas:

w **Métodos Multivariantes Gráficos** (MMG).

w **Métodos Gráficos Multivariantes** (MGM)

Los **Métodos Multivariantes Gráficos** son potentes herramientas de diagnóstico basadas en el análisis de grandes matrices de datos, que mediante complejos procesos algebraicos asentados sobre métodos numéricos, permiten representar la información del hiperespacio de partida en un subespacio de dimensiones reducidas.

Los **Métodos Gráficos Multivariantes** sólo exigen efectuar una transcripción geométrica de los datos correspondientes a un conjunto de variables, en una representación gráfica.

### 1. MÉTODOS GRÁFICOS UNIVARIANTES MÚLTIPLES

Muchas de las representaciones utilizadas en análisis multivariante no son en sí mismas multidimensionales ya que, a pesar de ser un conjunto de gráficas que forman una representación unitaria, cada una de ellas por separado sólo muestra una dimensión (o a lo sumo dos) de los datos referidos a varias variables o dimensiones. Desde ellas no se puede mostrar una variación común. Son por ello Métodos Gráficos Univariantes Múltiples, más que multivariantes. Entre ellos citaremos los siguientes:

#### 1.1. Stem & leaf

La representación Stem & Leaf es una representación intermedia entre una tabla y un gráfico. Muestra los valores con cifras, aunque su perfil es el de un histograma.

Este tipo de representación se debe a TUKEY (1977).

#### Construcción de un diagrama Steam & Leaf

1º. Se debe escribir a la izquierda de una línea vertical, de arriba hacia abajo, todos los posibles dígitos principales del conjunto de datos.

2º. Luego se representa cada dato a la derecha de la línea, escribiendo sus dígitos secundarios en la fila apropiada.

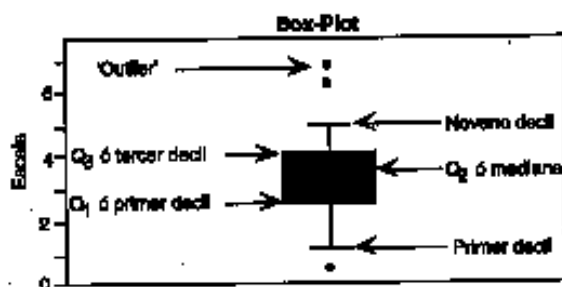
### 1.2. Box-plot

Esta representación gráfica, también debida a TUKEY (1977), puede ser traducida como *Caja con Bigotes* o *Representación Caja*, aunque se conoce usualmente con el nombre de **Box-plot** o bien **Box and Whiskers plot**.

Es un método gráfico simple para resumir la información, proporcionando una rápida impresión de las características más importantes de una distribución.

Esta representación, consiste básicamente en una caja dispuesta verticalmente que «encierra» el recorrido intercuartílico; es decir, la línea inferior indica el primer cuartil (25 percentil), y la línea superior señala el tercer cuartil (75 percentil). Otra línea intermedia marca el 50 percentil o Mediana así como su posición relativa en relación al rango intercuartílico.

Los Whiskewrs (bigotes), o líneas que se prolongan verticalmente, marcan los valores extremos en algunos casos. En grandes conjuntos de datos, marcan el 10 y 90 e incluso 5 y 95 percentil.



Usualmente, los «bigotes» indican el 10 y 90 percentil, siendo señalados los puntos extremos o aberrantes («Outliers»), con pequeños círculos o estrellas.

### 1.3. Diagramas de dispersión.

Un **Diagrama de puntos**, más conocido como **Diagrama de Dispersión**, es un método simple pero eficiente para ilustrar un determinado comportamiento o bien analizar una distribución en particular; su finalidad puede ser la de poner de manifiesto una relación entre variables, analizar proximidades entre individuos y/o poblaciones, localizar outliers...

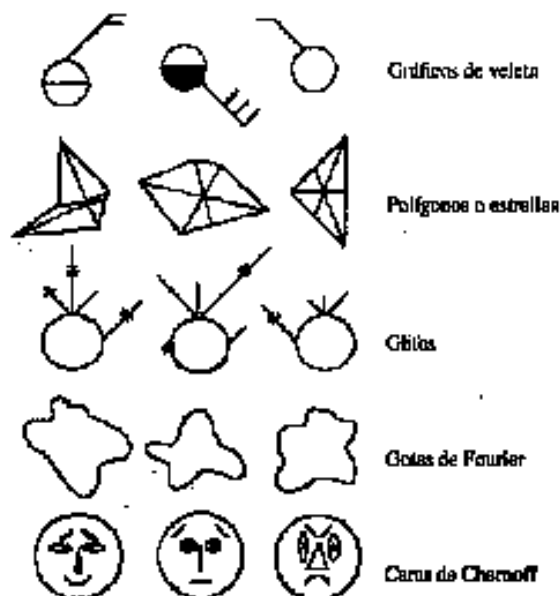
## 2. MÉTODOS GRÁFICOS MULTIVARIANTES (MGM)

¿Cómo podemos representar gráficamente valores de más de tres variables en una representación gráfica?

Cuando los datos tienen más de dos o tres dimensiones, la representación gráfica se hace complicada: las dimensiones del plano no son capaces de acoger un mayor número de variables que lo que acoge la representación cartesiana convencional, o la tridimensional; por lo tanto se requiere otro tipo de representación.

Existen diferentes métodos para representar datos multivariantes, prácticamente tantos como autores se han ocupado del tema. No se realizará por ello una revisión exhaustiva ni una descripción detallada de los MGM; solamente se enumerarán algunas de las técnicas existentes en la actualidad, y únicamente se expondrán con cierto detalle las más importantes y actuales que permiten la representación gráfica de entidades definidas por varias variables, como son las **Caras de Chernoff** (CHERNOFF, 1973), y las **Curvas de Andrews** (ANDREWS, 1972), o las Gotas de Fourier.

La práctica totalidad de los Métodos Gráficos Multivariantes que se conocen, no son otra cosa que «constructos pictóricos», (también conocidos como gráficos pictoriales o figurativos, o simplemente «Iconos»), formados por elementos geométricos (puntos, rectas curvas, triángulos, círculos...) que varían en función de los valores que toman las variables en los individuos a los que representan.



### Interpretación de las representaciones gráficas multivariantes

Cada elemento de un icono no puede ser convertido al valor numérico; las transformaciones que en general suelen realizarse, son lo suficientemente complejas para

que nosotros podamos interpretar esos valores mentalmente con la simple observación visual de los mismos.

La correcta interpretación de estos métodos consiste, fundamentalmente, en buscar gráficos similares. Entidades con similares valores para las variables tendrán formas parecidas; y entidades con diferentes valores, presentarán formas diferentes.

Veamos, de manera simplificada, alguno de estos métodos.

### 2.1. Polígonos o Estrellas

Determinan perfiles configurados por segmentos que parten de un origen común, y cuya longitud corresponde al valor que —para cada entidad— toma la variable a la cual dicho segmento representa.

### 2.2. Gotas op Manchas de Fourier

Se trata de otro método de representación gráfica multivariante.

Esta está determinada por la función de fourier:

$$f(t) = \frac{y_1}{2\pi} + y_2 \sin(t) = y_3 \cos(t) + y_4 \sin(2t) + y_5 \cos(2t)$$

donde  $y$  es una variable  $p$ -dimensional y  $t$  varía desde  $-3,14$  hasta  $3,14$ .

### 2.3. Caras de Chernoff

El principio heurístico de las caras de Chernoff está basado en la capacidad del hombre para reconocer, comparar y agrupar caras.

Originalmente, las caras estaban determinadas por 18 parámetros tales como la longitud de la nariz, el tamaño de los ojos o la curvatura de la boca (CHERNOFF, 1973). Los 18 parámetros o facciones que originalmente determinaban el dibujo de las Caras de CHERNOFF.

### 2.4. Curvas de Andrews

ANDREWS (1972) propone una técnica muy simple, conocida como «Curvas de Andrews», para obtener una representación visual de datos multivariantes, donde cada punto es representado en una suma de funciones trigonométricas.

$$f(t) = \frac{x_1}{2\pi} + y_2 \sin(t) = x_3 \cos(t) + x_4 \sin(2t) + x_5 \cos(2t)$$

para valores de  $t$  comprendidos entre  $-\pi < t < \pi$

#### Interpretación

ANDREWS (1972), destaca el hecho de que cada curva recoge y resume información de todas las variables que afectan al individuo, de manera que las diferencias entre individuos para alguna de ellas debe resultar manifiesta en alguna parte de las respectivas curvas.

La conservación de las distintas euclídeas originales para las curvas permite, según el autor, afirmar que dos curvas similares a lo largo del recorrido de  $t$  corresponden a individuos próximos.

## BIBLIOGRAFÍA

- Anuario de Estadística Agraria (1990). Publicado por el Ministerio de Agricultura, Pesca y Alimentación. Datos correspondientes a 1987.
- Avila-Zarza, C. (1993). *Métodos Gráficos Multivariantes y su Aplicación en las Ciencias de la Vida*. Trabajo de Grado. Dpto. de Estadística y Matemática Aplicadas. Universidad de Salamanca.
- Alonso, G. (1982). 491 Nuevos métodos de representación gráfica de datos multivariantes». *Publicaciones de Bioestadística y Biomatematica*. 5. Eunibar. Universidad de Barcelona.
- Andrews, D. F. (1972). «Plots for high dimensional data». *Biometrika*, 28: 125-136.
- Chernoff, H. (1973). «The use of faces to represent points in  $k$ -dimensional space graphically». *Journal of the american Statistical Association*, 68: 361-368.
- Chernoff, H. (1982). «Chernoff faces». *encyclopedia of Statistical Sciences*. I: 436-438. John Wiley & Sons. New York.
- Everit, B. S. (1993). *Cluster Analysis*. (3rd. ed.). Halsted Press. New York.
- Gete-Alonso, J.C. y Barrio, V. del (1990). *Lenguaje Gráfico*. Alhambra. Madrid.
- Hamilton, L.C. (1990). *Modern Data Analysis. A First Course in Applied Statistics*. Brooks/Cole Publishing Company. California.
- Tukey, J.W. (1977). *Exploratory Data Analysis*. M.A. Addison-Wesley. Reading.
- Vicente, S.; Galindo, M.P.; Vicente-Villardón, J.L.; Martín, A.; Barrera, I. y Fernández-Gómez, M.J. (1993). «Análisis gráfico y descripción estructural de la variabilidad de cultivos en Castilla-León». *Investigación Agraria*, Vol 8(3):315-329.