

ANÁLISIS DE CONGLOMERADOS

DR. JORGE RAÚL PÉREZ GALLARDO



CONACYT



CIMAT

Unidad Aguascalientes

Métodos Multivariados II

Ir de compras...

Un reconocido Centro Comercial de la ciudad que alberga a más de 50 negocios reporta tener una baja afluencia de personas en el último semestre en comparación con el año pasado. Ante esta situación, se programó una reunión con los responsables de los negocios para fijar una estrategia que aumente la afluencia.



Objetivo

Aumentar la frecuencia que las personas van al centro comercial a realizar sus compras



¿Qué hacer?

Ir de compras...



Realizar un estudio sobre las motivaciones que llevan a las personas a ir de compras habitualmente y posteriormente establecer agrupaciones con motivaciones afines y establecer estrategias enfocadas a cada grupo identificado.

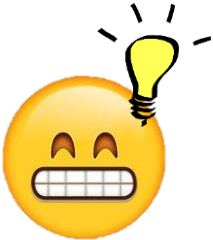
Algunas de las motivaciones pueden ser:

- Es un acto divertido en sí mismo.
- Intento ir poco porque compro compulsivamente y es malo para mí.
- Voy con frecuencia porque aprovecho para cenar fuera con mi pareja.
- Me encanta la aventura de encontrar productos a buen precio.
- No me atrae especialmente, voy por obligación o por necesidad.
- Puedes ahorrar mucho dinero si vas a comprar con frecuencia y estás informado.
- ...

Formar grupos de
personas con
motivaciones similares



Introducción



Aplicar un **Análisis de Conglomerados**, para formar esos grupos.

Análisis de conglomerados o cluster

Técnica multivariada que permite **CLASIFICAR** una muestra de observaciones (individuos u objetos) en un número pequeño de *grupos* o *clústers* mutuamente excluyentes basado en las similitudes que hay entre las observaciones (individuos u objetos de interés). Esto se realiza agrupando a los individuos que son similares **siguiendo algún criterio apropiado**.



Se **DESCONOCE**, al inicio del estudio, el número de grupos y las características de las personas que conforman cada uno de ellos.



Clasificar = Asignar un nuevo objeto u observación en su lugar correspondiente dentro de un conjunto de categorías establecidas.



Un poco de historia...

- ✓ Desde inicios de la civilización la *clasificación* es una actividad básica.
- ✓ Primeros trabajos de clasificación se dan en la Biología al establecer una clasificación de las especies del reino animal (carne roja y los que no).
- ✓ El A. Conglomerados se formalizó con la publicación de “*Principios de Taxonomía Numérica*” en 1963.
- ✓ El rápido crecimiento en años recientes se debe a:
 1. *El desarrollo de las computadoras*
 2. *La importancia fundamental de la clasificación en todos los campos (la ciencia)*



Peter Sneath



Robert Sokal

Una de las habilidades básicas de todo ser vivo consiste en agrupar objetos similares para producir una clasificación.

Brian S. Everitt

Aplicaciones

Los roles más comunes que el A.Conglomerados desempeña son:

- **REDUCCIÓN DE DATOS:** Al hacer frente a un gran volumen de observaciones sin sentido y no agrupadas es necesario contar con procedimientos objetivos para reducir la información mediante agrupaciones o clústers.
- **GENERACION DE HIPOTESIS:** Es probable que se desee desarrollar alguna hipótesis acerca de la naturaleza de un dato o característica de los individuos u objetos bajo estudio.

Ejemplo: Existe la creencia que considerando el ingreso de cada persona es posible establecer segmentos o grupos lógicos de consumidores de bebidas dietéticas.



Aplicaciones

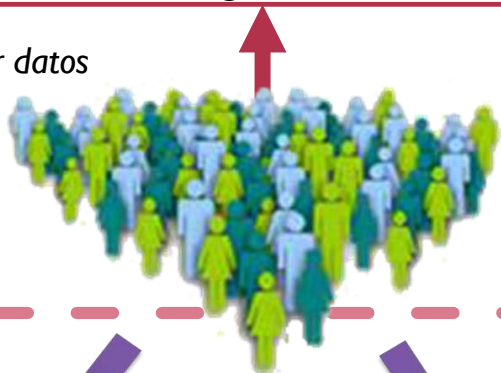
- ❑ **Biología:** Taxonomía
- ❑ **Psicología:** Q-análisis
- ❑ **Ingeniería:** Reconocimiento de patrones, caracterización de proveedores
- ❑ **Salud:** Zona de atención en desastres
- ❑ **Economía:** Perfiles sociodemográficos
- ❑ **Mercadotecnia:** Segmentación de mercado



Esquema general de aplicación

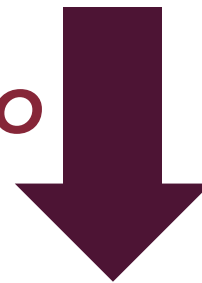
Nacionalidad, edad, nivel socioeconómico, propósito del viaje, actividades realizadas, transporte utilizado, gasto realizado, ...

P0. Extraer datos

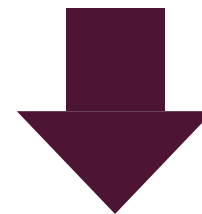


P1. Establecer una regla de asociación

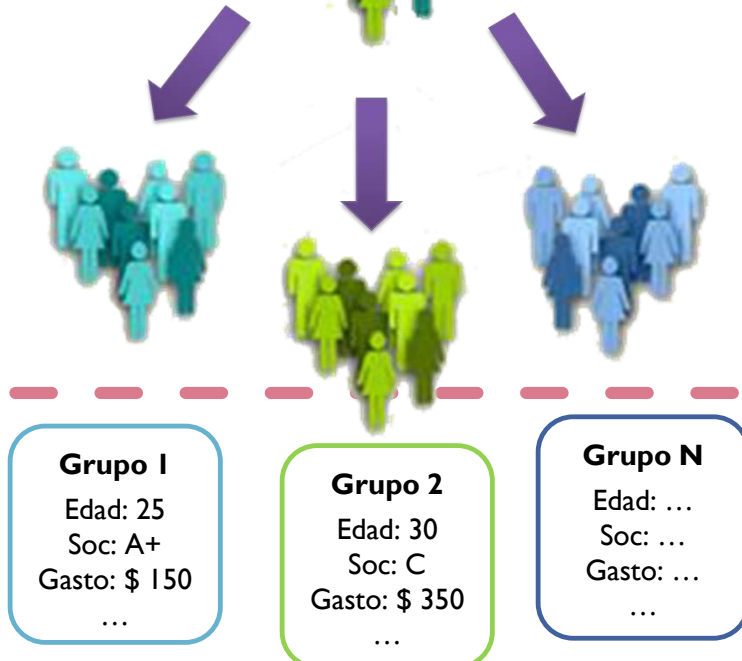
ALGORITMO



P2. Clasificar en los grupos a los individuos (u observaciones)



P3. Caracterizar a los individuos (u observaciones) que conforman los grupos



Análisis de Conglomerados

Objetivo:

Considerando los datos de la **MUESTRA** sobre los objetos bajo estudio (personas, productos, ideas, etc), **CLASIFICAR** en un **CONJUNTO REDUCIDO DE GRUPOS** (conglomerados, clústers o clases) dichos objetos.

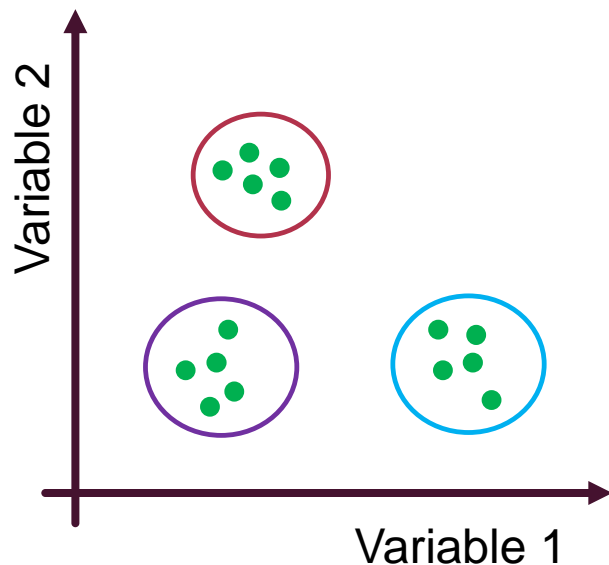
- Es comparable con el **Análisis Factorial** en el sentido de que busca **evaluar la estructura** de los datos de la muestra. Pero difiere en:
 - ❑ El AF busca la **agrupación de variables**, y
 - ❑ El AF hace la agrupación basado en **patrones de variación** (correlaciones) en datos mientras que el A. Conglomerados agrupa mediante **proximidades** (distancia).

Las agrupaciones resultantes deben exhibir una **alta homogeneidad interna** (dentro del grupo los individuos tienen características muy similares) y una **muy alta heterogeneidad externa** (entre individuos de grupos diferentes se aprecian diferencias significativas)

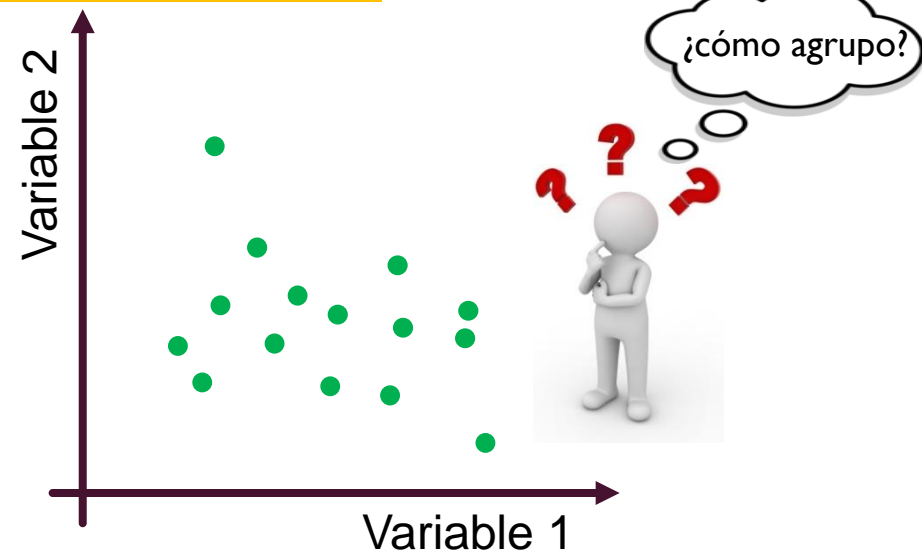
Postulado

Dada una colección de n objetos, individuos, animales, etc., los cuales se describen por un conjunto de p características o variables; producir una **DIVISIÓN ÚTIL** en cierto número de clases. El número de clases y sus características serán determinadas por el analista.

IDEAL



SITUACIÓN REAL



¿Si se agrega otra(s) variable(s) será más evidente el agrupamiento?

Clasificación de técnicas

Todos los métodos comparten dos pasos principales:

1. *La selección de la medida de proximidad (similitud o distancia)*
2. *La selección del algoritmo de agrupamiento*

Se distinguen dos grandes categorías:

- ❑ **Jerárquicos**: Agrupan clústers para formar uno nuevo o bien separar alguno existente de tal forma que minimice alguna función de distancia o maximice alguna medida de similitud.
- ❑ **No jerárquicos** (*partitivos o de optimización*): Realizan una sola partición de los individuos en K grupos. Se debe especificar *a priori* los grupos que deben ser formados. Difiere de la categoría anterior ya que trabaja con la matriz de los datos originales y no requiere una conversión en matriz de distancia o similitud.

Clasificación de técnicas

