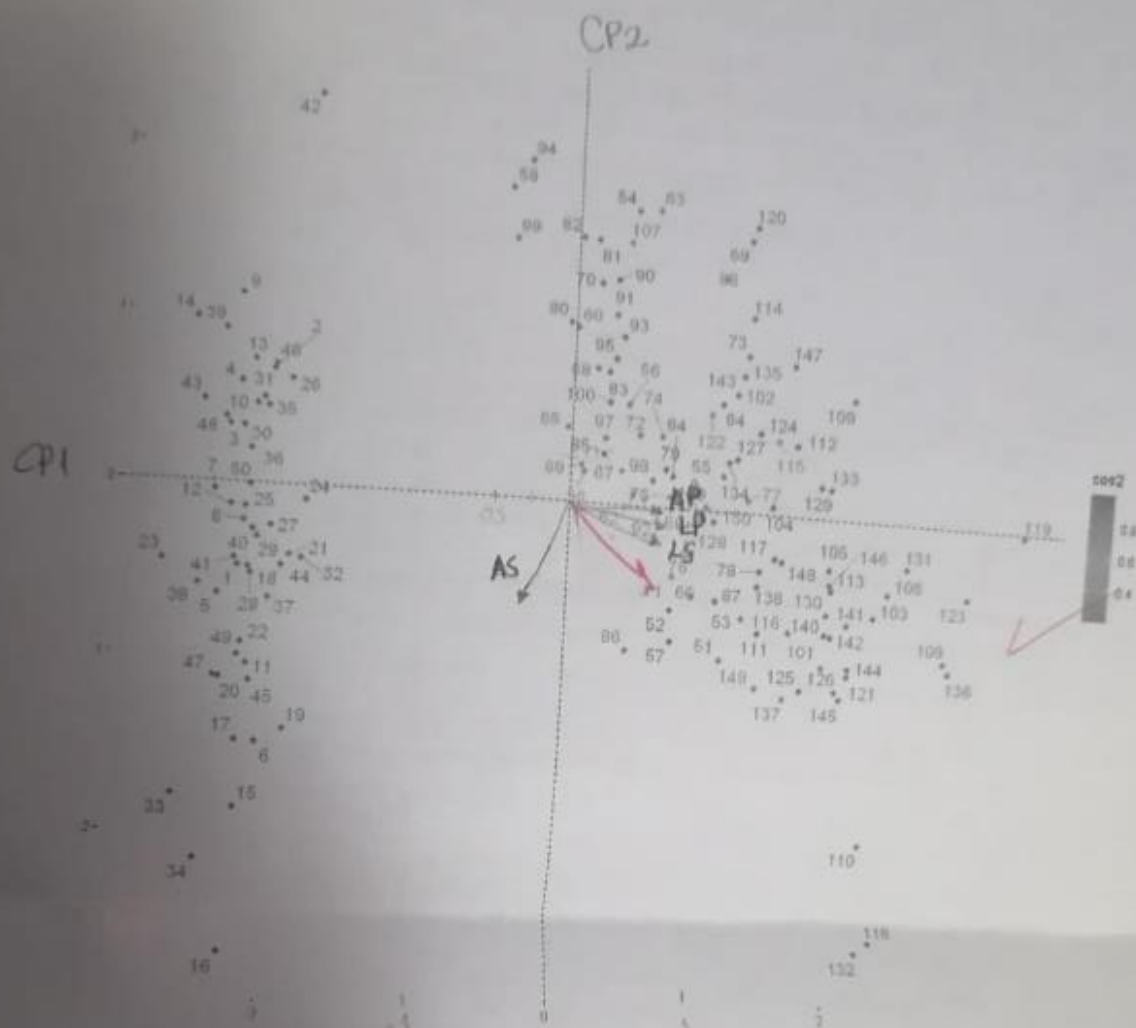


1. El mercado de automóviles chileno es uno de los más diversificados del mundo, con más de 60 marcas diferentes. Se requiere construir un Data Warehouse (DW) para una automotora que relacione los tipos de autos y sus clientes, los cuales se encuentran divididos en clases de preferencias como CarLovers, PadredeFamilia, Funcional, Empoderada, etc. Los automóviles por su parte, se clasifican en marcas agrupadas en: Europeos, Americanos, Asiáticos, etc. Que por su vez contienen características como: sedan, citycar, hushback, etc. En este DW también se debe considerar el lugar de venta: puesto que la automotora se encuentra en todo el país. Este DW permitirá analizar: costos, ventas y descuentos. Identifique dimensiones y las medidas, realice un esquema o esquemas gráficos de hipercubos en tres dimensiones y presente un ejemplo de dato atómico. (1.5)

La base de datos *Iris*, del repositorio de la Universidad de Irvine California, clasifica diferentes tipos de flores de lirios, de acuerdo a cuatro variables simples: largo y ancho de sépalos y pétalos de la flor. Se realiza un análisis de componentes principales, para extraer solo las dos primeras. Los resultados se muestran a continuación:

	CP1	CP2	CP3	CP4
Valor propio	2,8824	0,9140	0,1468	0,0207
Largo sépalo	0,5211	-0,3774	0,7196	0,2613
Ancho sépalo	-0.2693	-0.9233	-0.2444	-0.1235
Largo pétalo	0.5804	-0.0245	-0.1421	-0.8014
Ancho pétalo	0.5648	-0.0669	-0.6343	0.5236

- Determine el porcentaje de pérdida de la información, en un análisis de dos componentes.
- En la gráfica de las dos primeras componentes, sobre los lirios, grafique las variables originales en el plano (primera componente como la abscisa y la segunda como ordenada).
- Interprete cada una de las dos primeras componentes, con un solo concepto cada una.
- Interprete una separación en grupos y caracterícelos (dos o tres grupos máximo).



3. Realice un dendrograma usando el método de agrupamiento jerárquico, de la Clasificación Internacional de Industrias Uniforme (CIIU), la cual esta descrita de acuerdo a una métrica de disimilaridad, que tiene una escala de 0 a 300 donde 0 es el mismo tipo de empresa y 300 las más diferentes, entre sí. Los tipos de empresas a agrupar son: A: Agrícola, C: Manufacturo, DE: Suministros (agua, luz gas etc), F: Construcción, Comercio, H: Transporte. Obtenga dos y tres o cuatro, grupos de acuerdo a la siguiente matriz de disimilaridades. (2)

	A	C	DE	F	G
H	200	200	200	150	100
G	200	200	200	150	0
F	200	200	200	0	
DE	140	140	0		
C	50	0			

4. - Diga cuál es el elemento estadístico, sobre el cual se trabaja, para realizar la separación en el análisis discriminante de Fisher. (1,5)
- Que condición debe cumplir la distribución de probabilidades para ser separable por el análisis de Fisher.
 - Que significa el valor lambda en el análisis de Fisher.
 - Explique el método de los K-vecinos ponderado. (1)