



**PEP1**  
**Análisis de Datos**

Prof: Max Chacón  
Noviembre 2015

1. Se requiere construir un Data Warehouse (DW) para una importadora, la cual está orientada al estudio de clientes, los cuales son clasificados solo de acuerdo a su localización geográfica y tamaño, como: personas individuales, pequeñas empresas, medianas y grandes. Este DW permitirá a la importadora analizar: costos, ventas y descuentos. Identifique dimensiones y las medidas, realice un esquema gráfico y presente un ejemplo de dato atómico. (1,5)

2. Para entrenar un modelo de regresión se utilizó validación cruzada de 10 dobleces y se obtienen los siguientes errores en los conjuntos de entrenamiento y test.

	1	2	3	4	5	6	7	8	9	10
Entre %	3,4	3,2	2,9	3,1	3,8	3	2,95	3	3,1	3,3
Test %	4,2	4,6	5	4,7	4,1	4,9	5	4,9	4,5	4,3

Determine que precisión se le entregará al usuario del proceso y cual será el modelo presentado para hacer las predicciones. (1,0)

3. El Barómetro Merco consiste en una clasificación de las 50 empresas con más prestigio, en función de su puntuación en las siguientes variables:

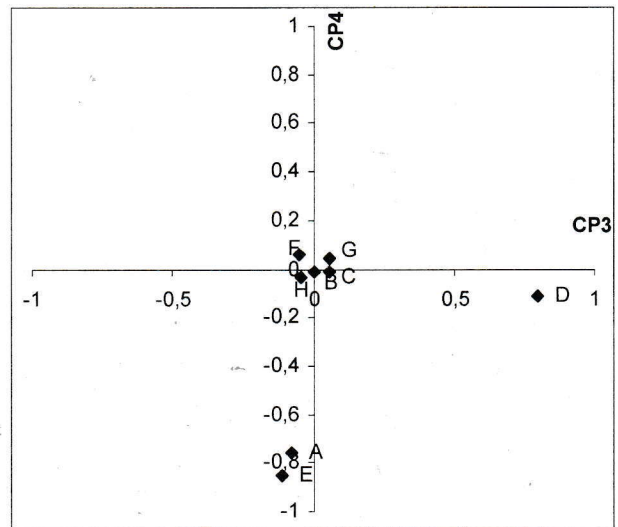
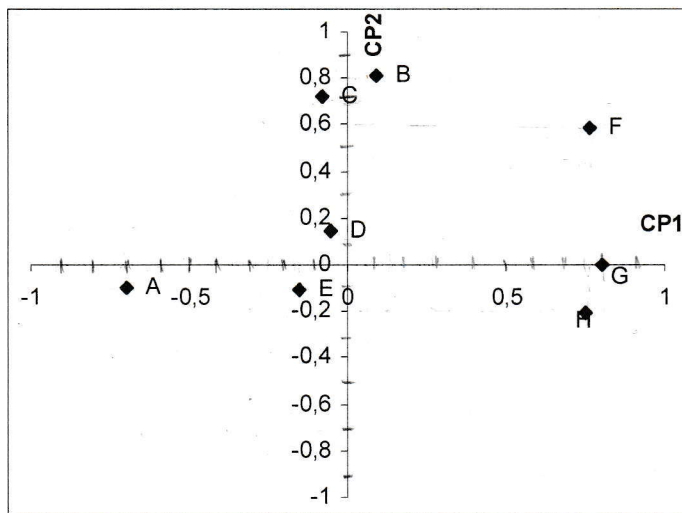
- REF: Resultados económico-financieros.
- CPS: Calidad producto/servicio.
- CCCL: Cultura corporativa y calidad laboral.
- ERSC: Ética y responsabilidad social corporativa.
- DGPI: Dimensión global y presencia internacional.
- IDI: Investigación, desarrollo e innovación.

Los valores y vectores característicos – normalizados - son los siguientes:

	CP1	CP2	CP3	CP4
$\lambda$	2,15	1,32	1,18	0,77
REF	-0,015	0,743	0,296	0,25
CPS	0,053	0,066	0,855	-0,215
CCCL	0,582	-0,187	0,038	0,439
ERSC	0,626	-0,101	0,207	0,216
DGPI	0,626	-0,101	0,107	0,216
IDI	0,245	0,629	-0,405	-0,036

De las 50 empresas se pueden concentrar en 8 (A,B...H) grupos fundamentales, estas proyecciones se muestran en las dos primeras componentes y en la tercera y cuarta componentes.

- Determine el porcentaje de información explicada con 2, 3 y 4 componentes.
- Intente la caracterización de cada componente hasta la cuarta componente.
- Según la caracterización realizada identifique cada grupo de empresas (A,B, ...,H).



4. Realice un dendrograma usando el método de agrupamiento jerárquico, tomando como individuos los grupos de empresas mostrados en el gráfico de las dos primeras componentes principales, del problema anterior. Indique cuales representantes de las empresas formarían dos y tres nuevos grupos. (1,5)

(2)