

# Segundo Parcial Análisis Inteligente de Datos

Maestría en Minería de Datos-UBA

2019

**Ejercicio 1** Se desea estudiar las características de la leche de distintos mamíferos. En el archivo **leches.xlsx** se encuentran los porcentajes de proteína, agua, grasa y lactosa de la leche de 20 mamíferos. Se pide responder brevemente a las siguientes preguntas.

1. Considere una clusterización decidiendo criteriosamente el número de clusters.
2. Justifique la selección del método y de la distancia.
3. Indique si ha estandarizado las variables.
4. Explique las características distintivas de los clusters formados.
5. Testee la igualdad de los vectores medios entre dos cualesquiera de los clusters armados.

**Ejercicio 2** Se tomaron cuatro muestras al azar para cinco marcas de cerveza elegidas también en forma aleatoria dentro de las marcas de mayor venta en el país. A ellas se les determinó el valor calórico expresado en calorías por 100 ml, y se desea saber si las marcas difieren en cuanto a la variable de respuesta analizada.

Marca	Valor Calórico			
A	31.5	32	32.7	30.9
B	32.3	31.9	33	31.7
C	28.8	27.6	29.1	27.7
D	24.5	25.3	24.9	26.1
E	25.4	26.3	23.9	24.1

1. Plantear el modelo de análisis y supuestos.
2. Testear al hipótesis con un nivel de confianza del 95% de que las marcas no difieren en el valor calórico.
3. Si las dos primeras marcas corresponden a cervezas blancas y las siguientes tres a cervezas negras, plantear un contraste que permita ver si existen diferencias en el valor calórico entre ambos tipos de cervezas.
4. Mediante el test de tukey analizar si hay diferencias dentro de los tipos de cerveza.
5. Hallar un intervalo de confianza del 95% para la diferencia entre las marcas con mayor y menor valor calórico.

**Ejercicio 3** En el archivo **vehiculos.xlsx** se registraron 40 accidentes categorizados por su gravedad

- Antigüedad del vehículo.
- Edad del conductor.
- Potencia del vehículo.

1. Comparar mediante un test de Hotelling los vectores medios de los grupos.

2. Analizar el cumplimiento de los supuestos para el análisis discriminante lineal.
3. Realizar análisis discriminante lineal, cuadrático o robusto según corresponda, explicando cuál eligió y por qué.
4. Evaluar de dos maneras distintas el poder discriminante de la técnica elegida.