

# Ejercicios corte 1-AED

## Docente: Diana Carolina Moreno Chavarro

1. Al inicio de un estudio que buscaba determinar si el ejercicio o los suplementos dietéticos podrían retrasar la pérdida ósea en mujeres adultas, un investigador midió el contenido de mineral en los huesos por absorciometría fotónica. Las medidas, que se tomaron en los huesos radio, húmero y cúbito en los lados dominante y no dominante. Los datos se encuentran en el archivo **Perdi-daOsea.csv**

- a) Obtenga: diagrama de caja para cada variable. Discuta los aspectos más relevantes en los datos.
- b) Obtenga el vector de medias,  $\bar{X}$ , la matriz de covarianzas  $S$  y la matriz de correlaciones  $R$ . Interprete las entradas en la matriz  $R$ .
- c) Calcule la varianza total y la varianza generalizada de los datos.
- d) Calcule la distancia de Mahalanobis entre cada dato y el centroide de los datos (vector de medias). ¿Que observación está a mayor distancia? ¿Que observación está a menor distancia?. Repita el ejercicio usando la distancia euclidiana.

2. La base de datos crabs de la librería MASS (Venables & Ripley 2002) del entorno y lenguaje R, contiene datos de 200 cangrejos de la especie *Leptograpsus Variegatus*, recolectados en la costa de Australia Occidental. La columna sp identifica el color, que puede ser azul (B) o naranja (O). Para cada color se recolectaron 50 cangrejos de cada sexo (columna sex) y se registraron 5 medidas físicas: Tamaño del lóbulo frontal (FL), anchura trasera (RW), la longitud del caparazón (CL), ancho del caparazón (CW) y profundidad del cuerpo (BD). Los investigadores buscaban demostrar que los dos colores podrían ser claramente diferenciados morfológicamente y que en realidad son dos especies diferentes. Los datos se encuentran en el archivo **crabs.csv**

- a) Usando la distancia de Mahalanobis, calculada con  $S_p$  ¿Cuál es la observación de la especie azul más cercana al centroide de la especie naranja?. ¿Cuál es la observación de la especie naranja más cercana al centroide de la especie azul?
- b) Encuentre, si existen, un par de observaciones que bajo la distancia de Mahalanobis sean cercanas mientras que con la distancia euclidiana no lo sean. Explique el resultado.

3. Dada la matriz

$$\begin{pmatrix} 3 & 4 \\ 6 & -2 \\ 3 & 1 \end{pmatrix}$$

- a) Grafique el diagrama de dispersión en  $p = 2$  dimensiones. Localice la media de la muestra en el diagrama.
- b) Dibuje la representación  $n = 3$  dimensional de los datos y trace los vectores de desviación  $y_1 - \bar{x}_1 \mathbf{1}$  y  $y_2 - \bar{x}_2 \mathbf{1}$ .
- c) Dibuje los vectores de desviación en (b) que emanan del origen. Calcule las longitudes de estos vectores y el coseno del ángulo entre ellos. Relacione estas cantidades con  $S$  y  $R$ .
- d) Calcular la varianza muestral generalizada.

4. Suponga que  $X_1$  y  $X_2$  son dos variables aleatorias conjuntamente distribuidas con  $var(X_1) = var(X_2) = \sigma^2$  y  $cor(X_1, X_2) = \rho$ . Sean  $U = X_1 + X_2$  y  $V = X_1 - X_2$ .

- a) Halle  $\text{var}(U)$  y  $\text{var}(V)$ . ¿Bajo qué condiciones las varianzas son cero?
- b) Demuestre que  $U$  y  $V$  son no correlacionadas.