

TAREA NÚMERO 7

- Cada día de atraso en implicará una pérdida de 10 puntos.
 - Las tareas son estrictamente de carácter individual, tareas iguales se les asignará cero puntos.
 - En nombre del archivo debe tener el siguiente formato: `TareaX_nombre_apellido.pdf` y/o `TareaX_nombre_apellido.html`.
1. [20 puntos] Complete todas las demostraciones que quedaron pendientes en la presentación de la clase.
 2. [20 puntos] Construya un ejemplo pequeño *a pie* con 4 variables categóricas que permita verificar el Teorema 2 visto en clase. Además, compare con el resultado de la función `MCA(...)` del paquete `FactoMineR`.
 3. [20 puntos] Este ejercicio se refiere al **Principio de Equivalencia Distribucional** del AFC: La presencia de modalidades de una variable asociadas a un mismo fenómeno, y por tanto fuertemente correlacionadas, se traduce en perfiles casi iguales, o equivalentemente, por una proporcionalidad aproximada entre filas o entre columnas de la tabla cruzada T . Si calculáramos las distancias entre modalidades a partir de la tabla T , con la fórmula de la distancia euclídea usual, las modalidades correlacionadas de una variable aportarían varias veces la misma información a las distancias entre las modalidades de la otra variable. Corregimos este sesgo, calculando la distancia entre modalidades a partir de sus perfiles, con la fórmula χ^2 , la cual goza de la propiedad de equivalencia distribucional que se enuncia así:
Equivalencia Distribucional: supongamos que dos perfiles columna son iguales, es decir: $\frac{f_{is}}{f_{\bullet s}} = \frac{f_{ih}}{f_{\bullet h}} \forall i$. Sea la matriz \tilde{V} definida a partir de V (Tabla de Frecuencias Relativas) por sustitución de las columnas s y h por la columna suma de ambas: $(v_{1s} + v_{1h}, \dots, v_{ns} + v_{nh})^t$. Entonces $\chi^2(p_{f_l}, p_{f_r}) = \chi^2(\tilde{p}_{f_l}, \tilde{p}_{f_r}) \forall l, r$, donde \tilde{p}_{f_l} es el perfil de la modalidad l calculado con la tabla \tilde{V} .
Demuestre la propiedad anterior. Además, enuncie y demuestre la propiedad análoga intercambiando los roles de los perfiles fila y columna.
En la práctica esta propiedad se presenta en forma aproximada, lo cual significa que si dos modalidades de una variable son aproximadamente correlacionadas, el fusionarlas deja las distancias entre los perfiles de las modalidades de la otra variable, aproximadamente iguales. Sin duda la propiedad de equivalencia distribucional es la razón más importante que motiva el uso de la distancia χ^2 .
 4. [20 puntos] Usando las variables categóricas de la tabla de datos `poison` del paquete `FactoMineR` realice un AFCM con la función `MCA(...)` del paquete `FactoMineR`. Incluya e interprete al menos: 1) Gráfico de varianzas explicadas por los ejes 2) Gráfico de variables 3) Gráfico de variables con modalidades 4) el Biplot 5) $\cos^2(x)$ de modalidades y variables y 6) contribuciones de variables y modalidades a la formación de ejes.

5. [20 puntos] Sea $B = (b_{ij})$ la matriz de Burt de tamaño $m \times m$ definida como $X^t X$, donde X se define como sigue:

- Sea X_p el código disyuntivo completo de la variable cualitativa p -ésima, para $p = 1, \dots, v$.
- X_p tiene n filas y m_p columnas, para $p = 1, \dots, v$.
- Se denota por $X = [X_1 | \dots | X_p | \dots | X_v]$ la concatenación de las v matrices de códigos disyuntivos completos.
- X tiene dimensión $n \times m$, con $m = \sum_{p=1}^v m_p$.

Pruebe que:

- El total de la fila i de la matriz B es $\sum_{j=1}^m b_{ij} = vx_{\bullet i}$.
- El total de la matriz de Burt B es $v^2 n$, es decir, $\sum_{i=1}^m \sum_{j=1}^m b_{ij} = v^2 n$.
- El perfil fila de la matriz de Burt B es el centro de gravedad de los perfiles fila de los individuos (filas de X) que poseen la modalidad i .