Capítulo 4

Análisis de Correspondencias

4.1. Aspectos teóricos

1. Suponer que se estudian dos características en los elementos de un conjunto que pueden darse en los niveles alto, medio y bajo en ambos casos. Si las frecuencias relativas con las que aparecen estos niveles son las mismas para las dos características, indicar la expresión de la representación de las filas y columnas. Emitir conclusiones al respecto.

2. Indicar cómo afecta a la representación de filas y columnas que la tabla de contingencias sea simétrica, es decir que los elementos de la matriz de perfiles (P) sean: $p_{ij} = p_{ji}$.

3. Estudiar el algoritmo expuesto en el libro de *Johnson* & *Wichern* para realizar el análisis de Correspondencias, luego realizar lo siguiente:

   a) Escribir el algoritmo expuesto.

   b) Aplicar el algoritmo en el conjunto de datos, utilizado en clases.

4. Probar que:

   a) $Y = D_r^{-1}\tilde{U}\Lambda = D_r^{-1}(P - rc')D_c^{-1/2}V$

   b) $Z = D_c^{-1}\tilde{V}\Lambda = D_c^{-1}(P - rc')'D_r^{-1/2}U$

donde $\tilde{U}$ y $\tilde{V}$ se hallan definidas en la teoría indicada en el ejercicio anterior.

5. Mostrar que la inercia total puede expresarse por filas o columnas como:

   a) Inercia por filas: $\Phi^2 = \sum_{i=1}^{I} r_i(\tilde{r}_i - c)' D_c^{-1}(\tilde{r}_i - c)$

   b) Inercia por columnas: $\Phi^2 = \sum_{j=1}^{J} c_j(\tilde{c}_j - r)' D_r^{-1}(\tilde{c}_j - r)$

6. Probar que matricialmente la inercia puede escribirse como:

$$\Phi^2 = \frac{\chi^2}{n} = tr\left\{[D_r^{-1/2}(P - rc')D_c^{-1/2}][D_r^{-1/2}(P - rc')D_c^{-1/2}]'\right\}$$

7. Una relación alternativa para calcular la inercia total es:

   a) $\Phi^2 = \sum_{i=1}^{s} \lambda_i - 1$ donde los λ_i son los eigen valores de la matriz $D_r^{-1/2} P D_c^{-1} P' D_r^{-1/2}$ y s es el rango de esa matriz.

   b) $\Phi^2 = \sum_{k=1}^{s} \lambda_k^2$ donde los λ_k son los eigen valores de la matriz $P - rc'$ y s es el rango de esa matriz.

   Ayuda: para demostrar a) y b) partir de la definición dada en el ejercicio anterior.

8. En relación a la inercia en un Análisis de Correspondencias Múltiple, realizar los siguientes:

   a) Mostrar que la inercia en el ACM que utiliza la matriz Binaria está definida como: $\lambda = (2\lambda_I - 1)^2$, donde λ es un eigenvalor basado en la matriz original y λ_I es una eigenvalor basado en la matriz binaria.

   b) Mostrar que la inercias en el ACM que utiliza la matriz de Burt, son los cuadrados de la versión binaria.

4.2. Investigación

Biplots para visualizar unidades muestrales y variables

Un biplot es una representación gráfica de la información en una matriz de datos de $n \times p$ elementos. El término *bi* se refiere a dos tipos de información contenida en una matriz de datos. La información en las filas que corresponde a unidades muestrales y la de las columnas que corresponde a las variables. Cuando hay solo dos variables, los diagramas de dispersión pueden representar ambas informaciones. Sin embargo, cuando existen muchas unidades muestrales medidas en muchas variables se pueden utilizar las estimaciones de las componentes principales asociadas a las unidades muestrales, luego la idea detrás de los biplots es agregar la información acerca de las variables al gráfico de las componentes principales.

Desarrollar la teoría en relación a los biplots, para luego aplicarla en un conjunto de datos de su preferencia.