Centro de Investigación en Matemáticas (CIMAT). Unidad Monterrey

Estadística Multivariada
Tarea 5: Analisis de Correspondencia

Marcelo Alberto Sanchez Zaragoza

28 de mayo de 2021

1. Problema 1

Otra forma de derivar los resultados del análisis de correspondencia simple es encontrando una matriz $\hat{P}$ de dimensión rxs con rango reducido $t < min(r, s)$ que aproxime P minimizando el criterio de mínimos cuadrados ponderados:

$$tr\{D_r^{-1/2}(P - \hat{P})D_c^{-1}(P - \hat{P})^t D_r^{-1/2}\}$$

Usando el teorema de Eckart-Young, encuentre la matrix $\hat{P}$ que arroje la mejor aproximación de rango reducido de P en este sentido. Muestre que la mejor aproximación de rango 1 de P es la solución trivial $\hat{P} = rc^t$.

Solución

Podemos partir de la descomposición de valores singulares y apoyandonos del teorema de Eckart-Young, donde nos dice que $\hat{Z}$ es la siguiente:

$$D_r^{-1/2} P D_c^{-1/2} = \sum_{i=1}^{t} \lambda_i^{1/2} u_i v_i^t$$

Si observamos v_i son los vectores propios de ZZ^t y u_i son los vectores propios correspodientes a $Z^t Z$, ambos asociados al valor propio λ_i. Ahora $\lambda_i^{1/2}$ son los valores singulares no nulos de las dos matrices anteriores y donde $t = minr, s$. Así tenemos que se cumple:

$$D_r^{-1/2} P D_c^{-1/2} v_i = \lambda_i^{1/2} u_i$$
$$u_i^t D_r^{-1/2} P D_c^{-1/2} = \lambda_i^{1/2} v_i^t$$

y

$$|(D_r^{-1/2} P D_c^{-1/2} v_i)(D_r^{-1/2} P D_c^{-1/2} v_i)^t - \lambda_i I| = 0$$

para i=1,...,t, observemos que la anterior expresión tiene determinante igual a cero por lo que la aproximación a P es la siguiente:

$$\hat{P} = D_r^{-1/2} \hat{Z} D_c^{-1/2} = \sum_{i=1}^{t} \lambda_i^{1/2} (D_r^{1/2} u_i)(D_c^{1/2} v_i)^t$$

Para la segunda demostración que nos piden partimos de $u_1 = D_r^{1/2} 1_r$ y $v_1 = D_c^{1/2} 1_s$, donde 1_r es un vector columna y 1_s es un vector renglón. Ahora regresando a la expresión $D_r^{-1/2} P D_c^{-1/2} v_i$ tenemos:

$$\begin{aligned} u_1^t (D_r^{-1/2} P D_c^{-1/2}) &= (D_r^{1/2} 1_q)^t (D_r^{1/2} P D_c^{-1/2}) \\ &= 1_q P D_c^{-1/2} = c^t D_c^{-1/2} \\ &= [\sqrt{c_1}, ..., \sqrt{c_q}] = (D_c^{1/2} 1_q)^t = v_1 \end{aligned}$$

tambien tenemos:

$$\begin{aligned} D_r^{-1/2} P D_c^{-1/2} v_1 &= (D_r^{-1/2} P D_c^{-1/2})(D_c^{1/2} 1_q) \\ &= D_r^{-1/2} P 1_q = D_r^{-1/2} r \\ &= \begin{pmatrix} \sqrt{r_1} \\ \sqrt{r_2} \\ \dots \\ \sqrt{r_q} \end{pmatrix} = D_r^{1/2} 1_q = u_1 \end{aligned}$$

Así tenemos que $(u_1, v_1) = (D_r^{1/2} 1_q, D_c^{1/2} 1_q)$ son vectores singulares asociados con valor singular $\lambda_1 = 1$. Para cualquier matriz P, el término común de expasión es:

$$D_r^{1/2} u_1 v_1^t D_c^{1/2} = D_r 1_q 1_q^t D_c = r c^t$$

2. Problema 2

El conjunto de datos mundodes representa 91 países en los que se han observado 6 variables, Razón de natalidad, Razón de mortalidad, mortalidad infantil, esperanza de vida en hombres, esperanza de vida en mujeres y PNB per cápita. Del conjunto de datos se ha tomado la esperanza de vida de hombres y de mujeres. Se han formado cuatro categorías tanto para la mujer como para el hombre. Se denotan por $M1$ y $H1$ a las esperanzas entre menos de 41 años a 50 años, $M2$ y $H2$, de 51 a 60 años, $M3$ y $H3$, de 61 a 70 años, y $M4$ y $H4$, para entre 71 a más de 80. La siguiente tabla de contingencia muestra las frecuencias de cada grupo.

Realiza proyecciones por filas, por columnas y conjuntas de filas y columnas. Comprobar que en la proyección por filas las categorías están claramente separadas y que en el caso del hombre, las dos últimas categorías están muy cercanas. Comprobar en la proyección conjunta la cercanía de las categorías H3 con M3 y M4.

Solución

Al realizar la tabla de correspondencia encontramos nuestra matriz F:

$$F = \begin{pmatrix} 0{,}10989 & 0 & 0 & 0 \\ 0{,}076923 & 0{,}131868 & 0 & 0 \\ 0 & 0{,}054945 & 0{,}164835 & 0 \\ 0 & 0 & 0{,}25274 & 0{,}20879 \end{pmatrix}$$

Igual encontramos las respectivas matrices D_r y D_c, son las siguientes:

$$D_r = \begin{pmatrix} 0{,}10989 & 0 & 0 & 0 \\ 0 & 0{,}2087 & 0 & 0 \\ 0 & 0 & 0{,}2197 & 0 \\ 0 & 0 & 0 & 0{,}46153 \end{pmatrix}$$

$$D_c = \begin{pmatrix} 0{,}1868 & 0 & 0 & 0 \\ 0 & 0{,}18681 & 0 & 0 \\ 0 & 0 & 0{,}41758 & 0 \\ 0 & 0 & 0 & 0{,}20879 \end{pmatrix}$$

Ahora encontramos nuestra matriz Z, la cual nos ayudara a encontrar las respectivas representaciones, la matriz es:

$$Z = \begin{pmatrix} 0{,}76696 & 0 & 0 & 0 \\ 0{,}38949 & 0{,}6676 & 0 & 0 \\ 0 & 0{,}27116 & 0{,}54107 & 0 \\ 0 & 0 & 0{,}57572 & 0{,}67259 \end{pmatrix}$$

Una vez que tenemos nuestra matriz Z buscamos las respectivas representaciones de las mujeres y hombres(renglones y columnas). Las representaciones se colocaron en forma de matriz y son las siguientes:

$$C_r = \begin{pmatrix} 1{,}67076 & -1{,}2143 \\ 1{,}13439 & 0{,}6776 \\ -0{,}2986 & 0{,}5446 \\ -0{,}7687 & -0{,}27676 \end{pmatrix}; \quad C_c = \begin{pmatrix} 1{,}55642 & -0{,}72703 \\ 0{,}76528 & 1{,}0664 \\ -0{,}62603 & 0{,}07928 \\ -0{,}82524 & -0{,}46227 \end{pmatrix}$$

Las representación de las mujeres(renglones) se muestra en la figura 2.1 y la representación de los hombres(columnas) es muestra en la figura 2.2

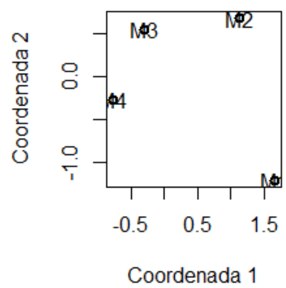

Figura 2.1: Proyección por mujeres(filas)

Se observa en la proyección de las mujeres que las categorias están claramente separadas y en el caso de la proyección de los hombres las dos ultimas categorian se encuentran muy cerca.

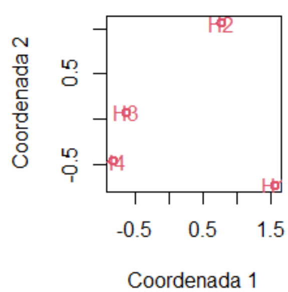

Figura 2.2: Proyección por hombres(columnas)

Finalmente en la figura 2.3 mostramos la proyección conjunta, se observa que las categorias $M3$, $M4$ y $H3$ se encuentran cerca.

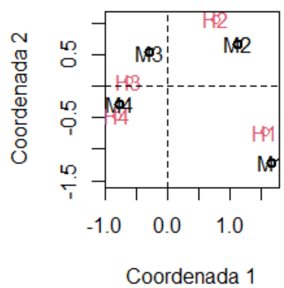

Figura 2.3: Proyección conjunta

Se anexa el código que se ocupo para obtener las matrices y representaciones anteriores.