6. Con la base de datos *Árabes* realiza un análisis de correspondencias. Explica lo siguiente:

a).Di que perfiles renglón y que perfiles columna son parecidos vía método gráfico.

b).Realiza la tabla completa de todos los tipos de residuos y di que asociaciones parecen importantes.

c).Analiza los resultados arrojados por la función ca().

d).Analiza el biplot de perfiles renglón y perfiles columna en 2D y compáralo con el biplot en 3D.

e).Compara tus resultados con el biplot generado con la función *anacor().*

La base Árabes contiene los porcentajes de hogares con diversas instalaciones y equipos en las áreas árabes de Jerusalén Este.

Las librerías a usar serán las siguientes:

library(ca)

library(ade4)

library(anacor)

library(FactoMineR)

Tabla de contingencia generada a partir de los datos de *Árabes*.

CRISTIAN ARMENIAN JEWISH MOSLEM total.row

toilet 982 972 973 969 3896

kitchen 788 81 656 733 2258

bath 144 176 6 96 422

electricity 862 821 545 747 2975

water 329 303 211 269 1112

radio 73 704 53 605 1435

tv set 46 6 15 34 101

refrigerator 292 263 53 105 713

total.col 3516 3326 2512 3558 12912

1. *Análisis de Perfiles Renglón vía método gráfico.*



En la gráfica de correspondencia bidimensional referente a los perfiles renglón podemos notar que los siguientes perfiles muestran características similares:

-"electricity" y "water"

-"bath" y "refrigerator"

Notemos lo siguiente: "kitchen" y "tv set" muestran perfiles similares para todas las regiones excepto para "JEWISH". De igual forma, "bath" y "radio" describen perfiles parecidos excepto para "CRISTIAN". Cabe destacar que el perfil de la variable "toilet" es muy similar en todas las regiones, esto se interpreta como que, independientemente de la región en la que se encuentren los individuos, el baño es eventualmente requerido. Estadísticamente hablando, la variable "toilet" no ayudaría a caracterizar a las variables de región, ya que la frecuencia del uso de esta variable es muy similar para todas las regiones.

"electricity" y "water"



Los perfiles renglón de "electricity" y "water" son demasiado parecidos.



El barplot muestra que la diferencia entre ambos perfiles es casi insignificante para las cuatro regiones.

De acuerdo a lo observado anteriormente podemos suponer que las variables "electricity" y "water" explican casi lo mismo respecto a las variables de región, es decir, existe una gran similitud entre los perfiles de "electricity" y los de "water".

"bath" y "refrigerator"



Los perfiles de "bath" y de "refrigerator" son similares, sin embargo, no lo son tanto como en la comparación anterior. De hecho, los perfiles ubicados en CRISTIAN y MOSLEM difieren bastante.



Después de todo estos perfiles no son tan parecidos.

*Análisis de Perfiles Columna vía método gráfico.*



En la gráfica de correspondencia bidimensional referente a los perfiles columna podemos notar que los perfiles de región son muy similares en algunas características pero también difieren en otras. Por ejemplo: Todas los perfiles de las regiones coinciden en las variables "electricity" y "water" como era de esperarse. Sin embargo, los perfiles de CRISTIAN y JEWISH son muy distintos a los de ARMENIAN Y MOSLEM respecto a la variable "radio" aunque ellos sean muy parecidos entre sí.

Desde nuestro punto de vista, los perfiles de "CRISTIAN" y "JEWISH" son los más similares, ya que sólo difieren cuando los perfiles se encuentran con respecto a la variable "toilet".

"CRISTIAN" y "JEWISH"





Los perfiles de "CRISTIAN" y "JEWISH" no son muy parecidos, sobre todo respecto a las variables "toilet" y "refrigerator", sin embargo, guardan una cercanía razonable.

b). Análisis de residuos.

Pearson's Chi-squared test

data: X

X-squared = 1895.642, df = 21, p-value < 2.2e-16

$Tabla.completa

CRISTIAN ARMENIAN JEWISH MOSLEM Total

toilet 982 972 973 969 3896

kitchen 788 81 656 733 2258

bath 144 176 6 96 422

electricity 862 821 545 747 2975

water 329 303 211 269 1112

radio 73 704 53 605 1435

tv set 46 6 15 34 101

refrigerator 292 263 53 105 713

Total 3516 3326 2512 3558 12912

$Expected

CRISTIAN ARMENIAN JEWISH MOSLEM

toilet 1060.89963 1003.57001 757.95787 1073.57249

kitchen 614.86431 581.63786 439.28872 622.20911

bath 114.91264 108.70291 82.09913 116.28532

electricity 810.10688 766.32977 578.77943 819.78392

water 302.80297 286.43990 216.33705 306.42007

radio 390.75743 369.64142 279.17596 395.42519

tv set 27.50279 26.01657 19.64932 27.83132

refrigerator 194.15335 183.66156 138.71252 196.47258

$Chi

CRISTIAN ARMENIAN JEWISH MOSLEM

toilet 5.867804 0.9931202 61.0101434 10.185997

kitchen 48.752165 430.9180711 106.9086793 19.727487

bath 7.362763 41.6630794 70.5376269 3.538659

electricity 3.324125 3.9001929 1.9714762 6.462068

water 2.266438 0.9573976 0.1316654 4.569746

radio 258.395051 302.4435442 183.2377149 111.074368

tv set 12.440442 15.4003071 1.1000973 1.367259

refrigerator 49.311371 34.2727624 52.9630314 42.587283

$r.crudos

CRISTIAN ARMENIAN JEWISH MOSLEM

toilet -78.89963 -31.57001 215.042131 -104.57249

kitchen 173.13569 -500.63786 216.711276 110.79089

bath 29.08736 67.29709 -76.099133 -20.28532

electricity 51.89312 54.67023 -33.779430 -72.78392

water 26.19703 16.56010 -5.337051 -37.42007

radio -317.75743 334.35858 -226.175960 209.57481

tv set 18.49721 -20.01657 -4.649318 6.16868

refrigerator 97.84665 79.33844 -85.712515 -91.47258

$residuos

CRISTIAN ARMENIAN JEWISH MOSLEM

toilet -2.422355 -0.9965542 7.8108990 -3.191551

kitchen 6.982275 -20.7585662 10.3396653 4.441564

bath 2.713441 6.4546944 -8.3986682 -1.881132

electricity 1.823218 1.9748906 -1.4040927 -2.542060

water 1.505469 0.9784670 -0.3628573 -2.137696

radio -16.074671 17.3909041 -13.5365326 10.539183

tv set 3.527101 -3.9243225 -1.0488552 1.169298

refrigerator 7.022206 5.8542944 -7.2775704 -6.525893

$ajustados

CRISTIAN ARMENIAN JEWISH MOSLEM

toilet -3.398231 -1.384104 10.4152793 -4.487347

kitchen 9.010785 -26.522583 12.6831458 5.744793

bath 3.234158 7.616742 -9.5149399 -2.247155

electricity 2.436313 2.612704 -1.7833845 -3.404497

water 1.846090 1.187900 -0.4229332 -2.627241

radio -19.987106 21.408329 -15.9981620 13.133714

tv set 4.150959 -4.572440 -1.1732779 1.379205

refrigerator 8.469025 6.990162 -8.3425948 -7.888105

De acuerdo a la prueba Ji-cuadrada tenemos lo siguiente:

Un estadístico de 1895.64, con 21 grados de libertad y p-value < 2.2e-16

Recordemos que para esta prueba tenemos el siguiente contraste:

Ho: Independencia entre renglones y columnas

vs

Ha: Al menos existe asociación entre algún renglón y alguna columna.

Ho también se puede interpretar como la inexistencia de asociación entre renglones y columnas. Como nuestro p-value es muy cercano a cero, entonces nos encontramos en área de rechazo, es decir, rechazamos la hipótesis de no asociación, entonces existe asociación entre al menos un renglón y una columna.

Notemos lo siguiente:

Al comparar los valores de la tabla con los valores esperados, podemos observar que difieren en gran medida los siguientes variables:

-Los valores de "radio"

-Los valores de "kitchen"

Esta diferencia evidencia el grado de no independencia entre las variables.

En la matriz de distancias Ji-Cuadrada, encontramos valores muy altos para la variable "radio” con respecto a las distintas variables de región, también las variables "kitchen" y "refrigerator" muestran valores elevados con respecto a las variables de región. Recordemos que la distancia Ji-cuadrada representa la "lejanía" entre los perfiles renglón o columna de su perfil medio. Por otro lado, la Ji-cuadrada es una medida del nivel de asociación global de las variables en estudio.

Analicemos los residuos crudos:

La variable "kitchen" cuenta con residuos crudos muy elevados con respecto a las diferentes regiones, cabe destacar que "kitchen" y "ARMENIAN" cuentan con el residuo crudo más grande pero negativo, esto significa que que hay una fuerte asociación negativa con esta variable. Por otro lado la variable "radio" también cuenta con residuos crudos elevados, negativos con CRISTIAN y JEWISH y positivos con las otras regiones. En cuanto a los residuos y los residuos ajustados tenemos que las variables "radio" y "kitchen" siguen aportando valores altos.

Recordemos que entre más grande sea el residuo más fuerte es la asociación entre las variables, esto lo podemos interpretar como que tanto la variable "radio", como la variable "kitchen" conservan una importante asociación con las distintas regiones, la más fuerte parece ser entre "kitchen" y ARMENIAN. Lo último se interpreta como que en Armenia las familias suelen distinguirse por contar con cocina a diferencia de los otros pasises, aunque en realidad la asociación también es fuerte con las otras variables de región.

c). Resultados arrojados por la función *ca().*

Principal inertias (eigenvalues):

1 2 3

Value 0.109688 0.027546 0.009578

Percentage 74.71% 18.76% 6.52%

Rows:

toilet kitchen bath electricity water radio tv set

Mass 0.301735 0.174876 0.032683 0.230406 0.086121 0.111137 0.007822

ChiDist 0.141546 0.518184 0.540103 0.072548 0.084422 0.771961 0.547796

Inertia 0.006045 0.046957 0.009534 0.001213 0.000614 0.066229 0.002347

Dim. 1 0.205680 1.477729 -1.178701 -0.069365 -0.011056 -2.189978 1.092084

Dim. 2 0.198861 0.756318 -2.038259 -0.410883 -0.490570 1.529894 -0.604529

refrigerator

Mass 0.055220

ChiDist 0.501239

Inertia 0.013873

Dim. 1 -0.546507

Dim. 2 -2.789384

Columns:

CRISTIAN ARMENIAN JEWISH MOSLEM

Mass 0.272305 0.257590 0.194548 0.275558

ChiDist 0.332074 0.499714 0.436155 0.236800

Inertia 0.030028 0.064324 0.037009 0.015452

Dim. 1 0.742014 -1.477068 1.194365 -0.195738

Dim. 2 -1.282583 -0.453512 0.657548 1.227144

Con la función ca() obtenemos la masa, la distancia Ji-cuadrada, la inercia y las coordenadas de los puntos que representan a las variables en dos dimensiones. De las variables de nivel socioeconómico, podemos notar que la mayor cantidad de masa la obtuvo la variable "electricity". Las distancias Ji-cuadrada más elevadas las proporcionan las variables "radio", "kitchen", "bath", "tv set" y "refrigerator". Esta distancia representa la "lejania" entre los perfiles renglón o columna de su perfil medio, más aún, si los perfiles difieren poco de sus perfiles medios, entonces el valor de su inercia sería bajo y eso implicaría una pobre asociación entre las variables.

Con respecto a las variables de región; CRISTIAN, ARMENIAN y MOSLEM aportan casi el mismo nivel de masa, JEWISH aporta un poco menos. En cuanto a la distancia Ji-cuadrada, las variables ARMENIAN y JEWISH muestran los valores más altos, sin embargo, éstos no difieren mucho de los valores generados por las otras variables de región. La inercia más alta la proporciona ARMENIAN.

> summary(ca(base))

Principal inertias (eigenvalues):

dim value % cum% scree plot

1 0.109688 74.7 74.7 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

2 0.027546 18.8 93.5 \*\*\*\*

3 0.009578 6.5 100.0

-------- -----

Total: 0.146812 100.0

Rows:

name mass qlt inr k=1 cor ctr k=2 cor ctr

1 | tolt | 302 286 41 | 68 232 13 | 33 54 12 |

2 | ktch | 175 951 320 | 489 892 382 | 126 59 100 |

3 | bath | 33 915 65 | -390 522 45 | -338 392 136 |

4 | elct | 230 984 8 | -23 100 1 | -68 884 39 |

5 | watr | 86 932 4 | -4 2 0 | -81 930 21 |

6 | radi | 111 991 451 | -725 883 533 | 254 108 260 |

7 | tvst | 8 469 16 | 362 436 9 | -100 34 3 |

8 | rfrg | 55 983 94 | -181 130 16 | -463 853 430 |

Columns:

name mass qlt inr k=1 cor ctr k=2 cor ctr

1 | CRIS | 272 959 205 | 246 548 150 | -213 411 448 |

2 | ARME | 258 981 438 | -489 958 562 | -75 23 53 |

3 | JEWI | 195 885 252 | 396 823 278 | 109 63 84 |

4 | MOSL | 276 815 105 | -65 75 11 | 204 740 415 |

Con el summary de ésta función podemos ver que tan bien representadas estan las variables en las distintas dimensiones:

De acuerdo a lo anterior la representación en dos dimensiones es bastante buena, ya que en la primera dimensión se acumula el 74.7% de la inercia (que es como la varianza explicada) y con la segunda dimensión se explica el 93.5% de la inercia.

Renglones.

Para las variables "kitchen", "bath", "electricity", "water", "radio" y "refrigerator" la calidad de representación es muy buena, para las otras variables (toilet y tv set) la calidad de representación es bastante mala.

Ahora bien, en la primera dimensión, las variables con mejor calidad y mayor contribución (en ésta dimensión) son "kitchen" y "radio". Por otro lado, en la segunda dimensión, las variables con mejor calidad son "electricity", "water" y "refrigerator", y las variables con mayor contribución son "refrigerator" y "radio". La variable "bath" cuenta con alta calidad en la dimensión 1, sin embargo no contribuye mucho en ésta dimensión.

Columnas.

Las calidades de representación de todas las variables de región son muy altas. La variable CRISTIAN tiene una similar calidad en la dimensión 1 que en la dimensión 2, pero contribuye más a la dimensión 2. Las variables ARMENIAN y JEWISH tienen una alta calidad y contribución en la dimensión 1, mientras que MOSLEM tiene una alta calidad y contribución en la dimensión 2.

d). Análisis del biplot de perfiles renglón y perfiles columna en 2D y comparación con el biplot en 3D.





En éstas gráficas podemos observar las asociaciones perfiles columna y perfiles renglón en dos dimensiones, por ejemplo "electricity" y "water" se encuentran muy cercanos, esto ya lo habíamos notado con las gráficas iniciales.

Con la primera gráfica podemos visualizar lo que decíamos anteriormente sobre la contribución de las variables a cada dimensión.

Un dato que es importante denotar es que, variables como "toilet" que no se encontraban bien representadas en dos dimensiones, es posible que al momento de ver el biplot en 2D el punto que represente a ésta variable se encuentre a poca distancia de otro, pero esto no significa que sean similares, ya que por su baja calidad de representación puede pasar que en 3 dimensiones ese punto se proyecte a un sitio completamente lejano de la masa de puntos.

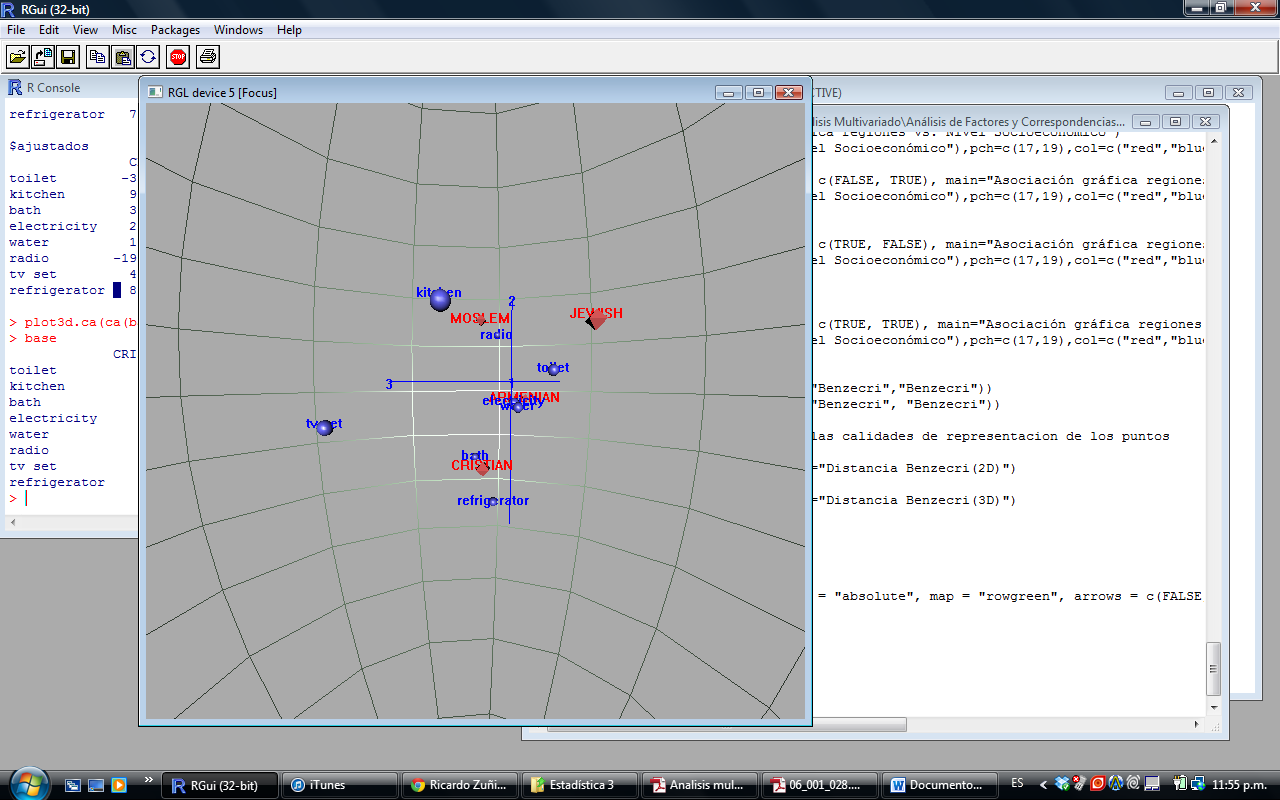






En todos los biplot anteriores se muestra lo mismo, los puntos representados por las distintas variables de región y nivel socioeconómico, sin embargo, en esta última gráfica aparte de mostrar los puntos, también les da un tamaño que hace referencia a su masa. Más aún, a las variables de Región les marca un vector que va del centro a cada punto.

Al comparar el biplot en 2D con el de 3D, nos damos cuenta de que los puntos que representan a las variables renglón “tv set” y “toilet” son colapsados al ser vistos en el biplot en 2D, es decir, al ver el biplot 3D éstos puntos se encuentran alejados de los demás y sobre la tercera dimensión. Esto tiene sentido ya que eran las variables que mostraban una calidad de representación muy baja.



Ahora bien, respecto a las variables columna, JEWISH es una variable que según el summary analizado a partir de la función ca() tiene una calidad de representación muy buena en la primera dimensión, sin embargo al momento de ver el biplot en 3D nos percatamos de que el punto que representa a esa variable se encuentra alejado de la primera dimensión, a comparación con las otras variables columna que se encuentran menos distantes de la primera dimensión.

e).Comparación con la función *anacor()*



Al comparar la gráfica generada por la función anacar() con los resultados antes obtenidos, nos dimos cuenta de que la función anacar() refleja sobre el eje de la primera dimensión los puntos de las variables renglón y disminuye la distancia entre los puntos que representan a las variables columna con el centro.

En general, la gráfica generada por la función anacar() muestra la misma información que los biplots anteriores.

En conclusión, la variable “kitchen” se encuentra fuertemente asociada con las regiones CRISTIAN, ARMENIAN, JEWISH y MOSLEM, ya que basándonos en las distancias Ji-cuadrada y en los residuos crudos y ajustados, podemos darnos cuenta de que la variable “kitchen” tiene fuertes cargas con las variables de región. Habría que ir con un experto de área (en este caso un geógrafo) para saber porque se da ésta fuerte asociación. Lo que significaría esto es que tener cocina está asociado a la región en la que te encuentres. En la región donde la carga fue más fuerte fue en Armenia, aunque fue una carga negativa, esto diría que en Armenia es menos común encontrar hogares con cocina integrada que en las otras regiones. Otra fuerte asociación se dio entre la variable “radio” y Armenia, pero esta vez la carga si fue positiva, entonces en Armenia encontrar hogares donde que cuentan con radio es más común que en las otras regiones.









Al analizar la gráfica de Benzecri nos dimos cuenta de que hay similitudes entre perfiles renglón que no estan bien representadas en dos dimensiones, como: tv set y radio, refrigerator y tv set, tv set y toilet, tv set y water.

Como se puede apreciar, las relaciones entre perfiles que involucran a la variable tv set no se encuentran bien representadas en la segunda dimensión. Lo anterior era un hecho, pues ya habíamos comentado anteriormente que ésta variable no contaba con buena calidad de representación en la segunda dimensión.

En cuanto a las similitudes entre los perfiles columna, estos se encuentran, en general, bien representados.

Para tres dimensiones, tanto los perfiles renglón como los columna ya se encuentran perfectamente representados.