Análisis de Correspondencias Simple entre las variables:

Licenciatura vs Escuela de procedencia

Tabla de Información Cruzada

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Preparatoria | CCH | Particular | Otros | Total.Renglón |
| A | 45 | 20 | 4 | 0 | 69 |
| B | 39 | 28 | 15 | 4 | 86 |
| CC | 20 | 17 | 5 | 1 | 43 |
| CT | 11 | 6 | 4 | 1 | 22 |
| F | 26 | 19 | 10 | 11 | 66 |
| M | 12 | 10 | 9 | 8 | 39 |
| Total.Columna | 153 | 100 | 47 | 25 | 325 |

*Análisis de Perfiles Renglón vía método gráfico*



En la gráfica de correspondencias de perfiles renglón podemos observar similitud entre los perfiles de Biología y Ciencias de la Tierra.

Nótese que mediante esta gráfica podemos ver que la incidencia de Actuarios que provienen de Preparatoria UNAM es mucho mayor que para las otras carreras, mientras que la cantidad de alumnos que provienen de CCH es muy parecida para las distintas carreras.

En cuanto a las escuelas particulares, los matemáticos son los más y los actuarios son los menos.

*Biología y Ciencias de la Tierra*



Los perfiles son muy similares, aunque difieran considerablemente en la cantidad de alumnos provenientes de CCH. Esta similitud puede deberse a que ambas carreras comparten un origen parecido, es decir, las dos carreras provienen de las ciencias naturales, a diferencia de las otras carreras de la facultad que tienen sus bases principalmente en las matemáticas.

El barplot muestra mucha similitud entre las dos carreras en términos de la escuela de procedencia, sin embargo en Preparatoria y CCH podría considerarse significativa la diferencia entre las variables de carrera.

*Análisis de Perfiles Columna vía método gráfico*

**

Respecto a los perfiles columna, los más parecidos son los perfiles que representan Preparatoria y CCH, aunque se tiene una diferencia bastante considerable con respecto a la carrera de Actuaría.

Por otro lado, las escuelas Particulares y Otros difieren bastante en todas las carreras excepto en Ciencias de la Tierra, de hecho, en esa carrera se tienen proporciones muy parecidas respecto a la escuela de procedencia, es decir, en la gráfica de perfiles las líneas que representan a las escuelas de procedencia difieren poco.

*Preparatoria y CCH*

**

Se puede observar que los perfiles difieren poco, excepto en Actuaría donde parece existir una gran discrepancia entre alumnos provenientes de Preparatoria y de CCH, en efecto, hay una mayor proporción de sujetos provenientes de Preparatoria.

Sigue siendo muy visual la gran diferencia que existe entre los alumnos de Actuaría que provienen de Preparatoria y de CCH.

*Gráfica de Mosaico*

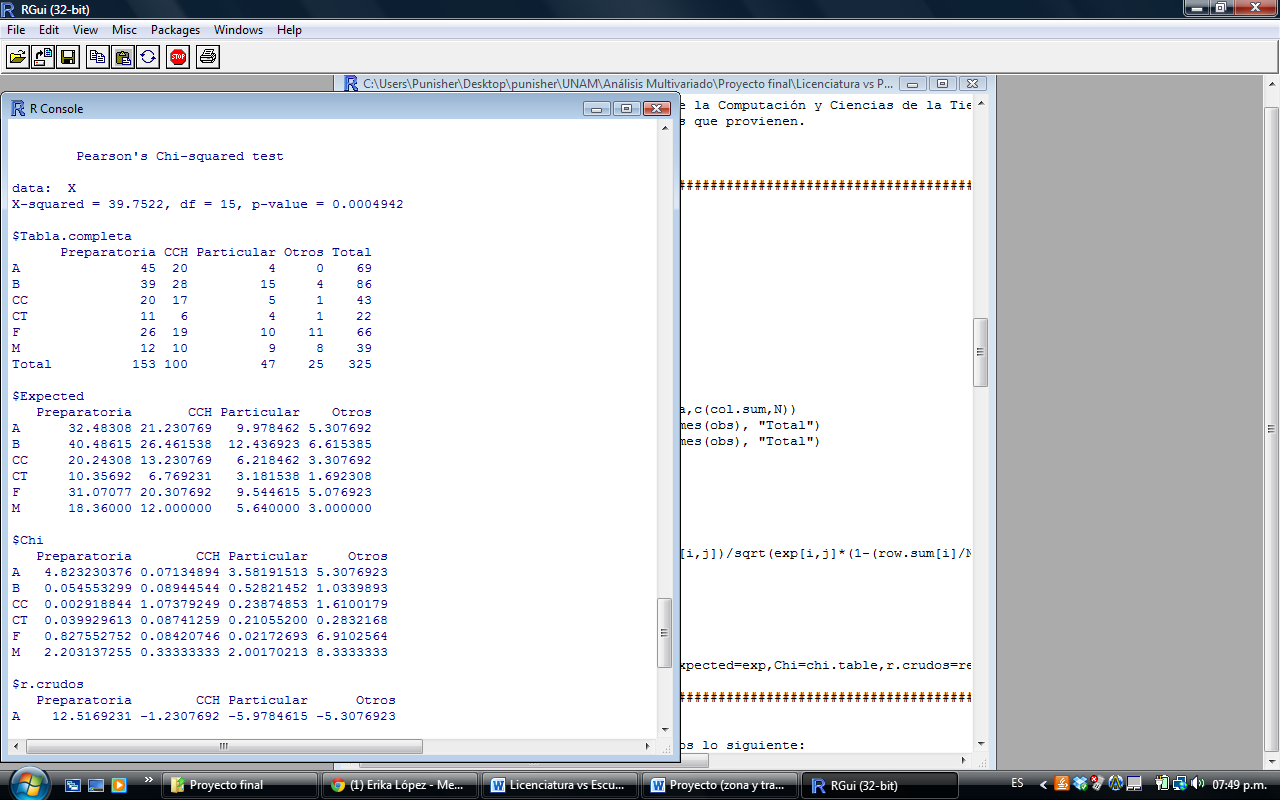


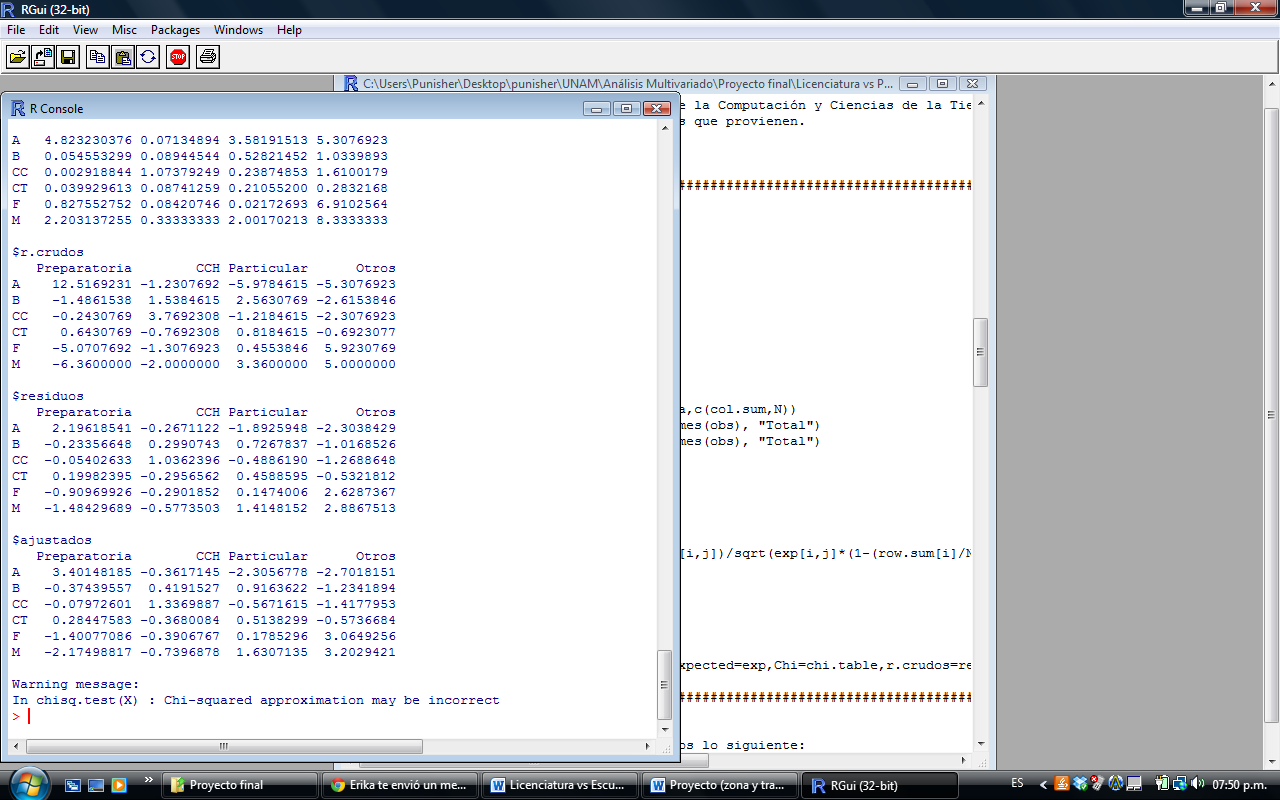
En el gráfico de Mosaico podemos observar que la asociación entre la carrera de actuaría y Preparatoria es muy fuerte, de hecho es una asociación positiva (color azul y tamaño del rectángulo). Esto podría significar que los estudiantes de la carrera de actuaría se distinguen por provenir de Preparatoria UNAM.

Por otro lado, las carreras de Física y Matemáticas presentan una considerable asociación con Otros, es decir, los sujetos provenientes de "otras" escuelas se distinguen por estudiar estas carreras.

Las otras carreras (Biología, Ciencias de la Computación y Ciencias de la Tierra) no muestran asociación alguna con las escuelas de las que provienen.

Veamos que ocurre con los residuos





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

Un estadístico de 39.7522, con 15 grados de libertad y p-value=0.0004942.

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 Actuaría

-Los valores de Matemáticas

-Los valores de Física

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

En la matriz de distancias Ji-Cuadrada, encontramos valores altos para la variable "Actuaría" con respecto a las distintas variables de escuela de procedencia, excepto con la variable "CCH". También la variable "Física" presenta un valor muy alto con respecto a la variable "Otros". Finalmente la variable "Matemáticas presenta grandes valores con respecto a las variables "Preparatoria", "Particular" y "Otros".

Recordemos que la distancia Ji-cuadrada representa la "lejanía" entre los perfiles renglón o columna de su perfil medio, entonces, mientras más grandes sean los valores de las variables renglón con respecto a las variables de columna, la asociación es mayor, ya que la Ji-cuadrada es una medida del nivel de asociación global de las variables en estudio.

Analicemos los residuos.

Recordemos que entre más grande sea el residuo más fuerte es la asociación entre las variables.

La variable Actuaría cuenta con residuos crudos altos con respecto a las variables Preparatoria, Particular y Otros, con residuos negativos para las últimas dos variables. Lo anterior se podría interpretar como que los estudiantes de la carrera de actuaría se distinguen por provenir de Preparatoria UNAM (como ya habíamos mencionado anteriormente) y no provienen de escuelas Particulares ni de Otras.

Por otro lado, Ciencias de la Computación cuenta con residuo crudo significativo y positivo respecto a la variable CCH, es decir, los estudiantes de ésta carrera se distinguen por provenir de CCH.

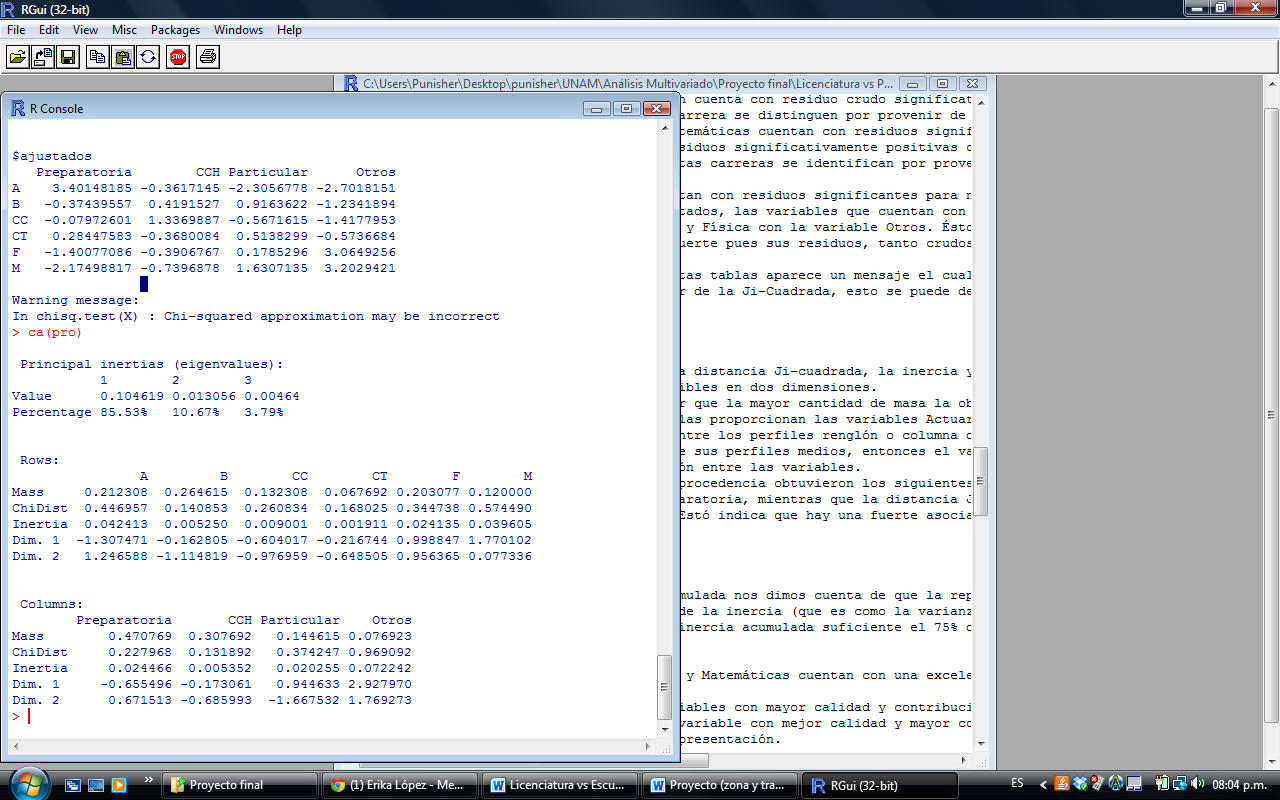
En cuanto a las variables de Física y Matemáticas cuentan con residuos significativamente negativos respecto a la variable Preparatoria y residuos significativamente positivas con respecto a la variable Otros. De aquí que los estudiantes de éstas carreras se identifican por provenir de otras escuelas y no de provenir de Preparatorias UNAM.

Ciencias de la Tierra y Biología no cuentan con residuos significantes para ninguna variable.

Ahora bien, respecto a los residuos ajustados, las variables que cuentan con valores elevados son Actuaría con Preparatoria (nuevamente),y Matemáticas y Física con la variable Otros. Ésto confirma lo anteriormente dicho, la asociación entre estas variables es fuerte pues sus residuos, tanto crudos como ajustados son significativamente grandes.

Es importante aclarar que al calcular estas tablas aparece un mensaje el cual advierte la posible existencia de error al momento de aproximar el valor de la Ji-Cuadrada, esto se puede deber a la cantidad de información utilizada.

ca(pro)



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 Carrera podemos notar que la mayor cantidad de masa la obtuvo Biología.

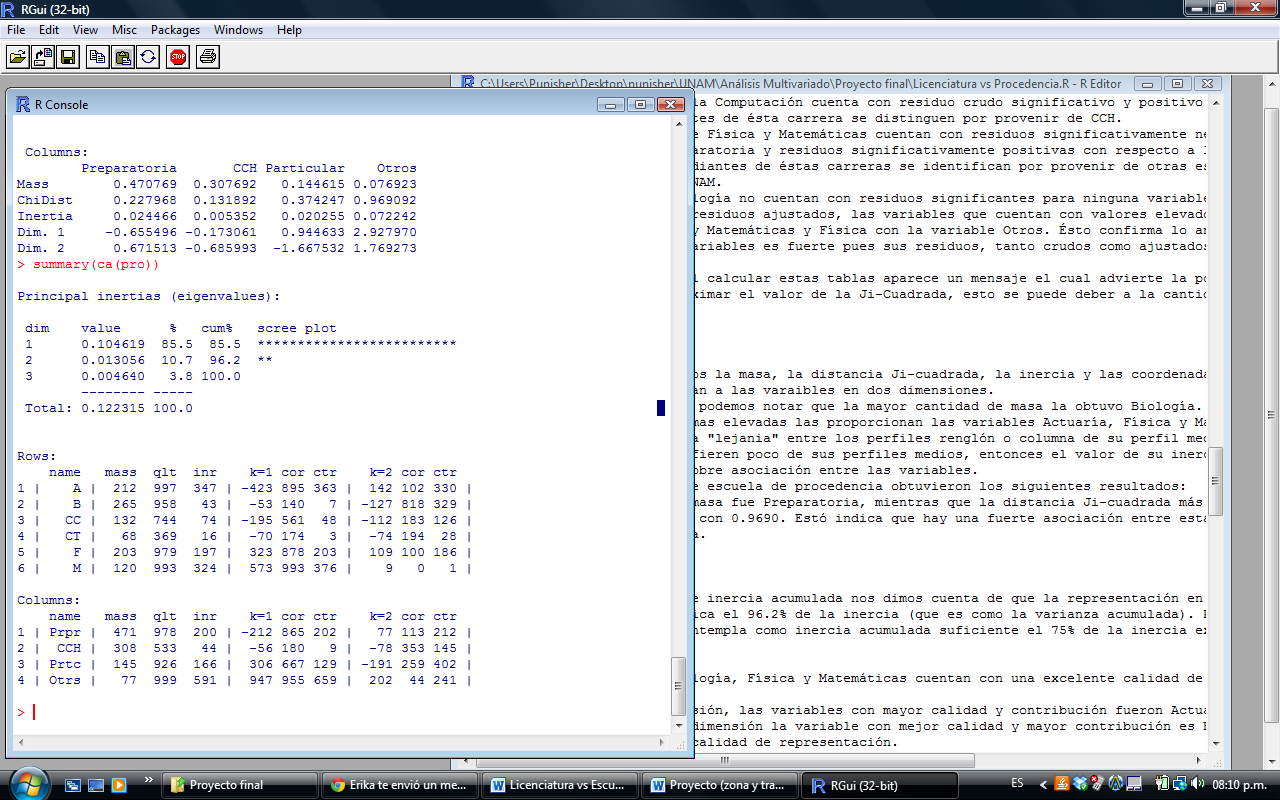
Las distancias Ji-cuadrada más elevadas las proporcionan las variables Actuaría, Física y Matemáticas.

Esta distancia representa la "lejanía" 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.

Ahora bien, las variables de escuela de procedencia obtuvieron los siguientes resultados:

La variable en obtener más masa fue Preparatoria, mientras que la distancia Ji-cuadrada más elevada fue la de la variable Otros con 0.9690. Esto indica que hay una fuerte asociación entre esta variable y alguna variable de carrera.

summary(ca(pro))



Al observar la proporción de inercia acumulada nos dimos cuenta de que la representación en dos dimensiones es bastante buena, ya que explica el 96.2% de la inercia (que es como la varianza acumulada). Recordemos que el criterio empírico contempla como inercia acumulada suficiente el 75% de la inercia explicada.

Renglones

Las variables Actuaría, Biología, Física y Matemáticas cuentan con una excelente calidad de representación en dos dimensiones.

Respecto a la primera dimensión, las variables con mayor calidad y contribución fueron Actuaría y Matemáticas, mientras que en la segunda dimensión la variable con mejor calidad y mayor contribución es Biología. Las otras variables cuentan con mala calidad de representación.

Columnas

Las variables Preparatoria, Particular y Otros cuentan con calidad de representación muy buena. La razón por la que la variable CCH no tiene buena calidad de representación es porque para todas las carreras mostraba tener similar número de sujetos provenientes de esa escuela.

Las variables Preparatoria y Otros se encuentran muy bien representadas en la primer dimensión, mas aún, la contribución más alta en esta dimensión proviene de la variable Otros. La segunda dimensión no cuenta con variables de escolaridad bien representadas ni que contribuyan en gran medida.



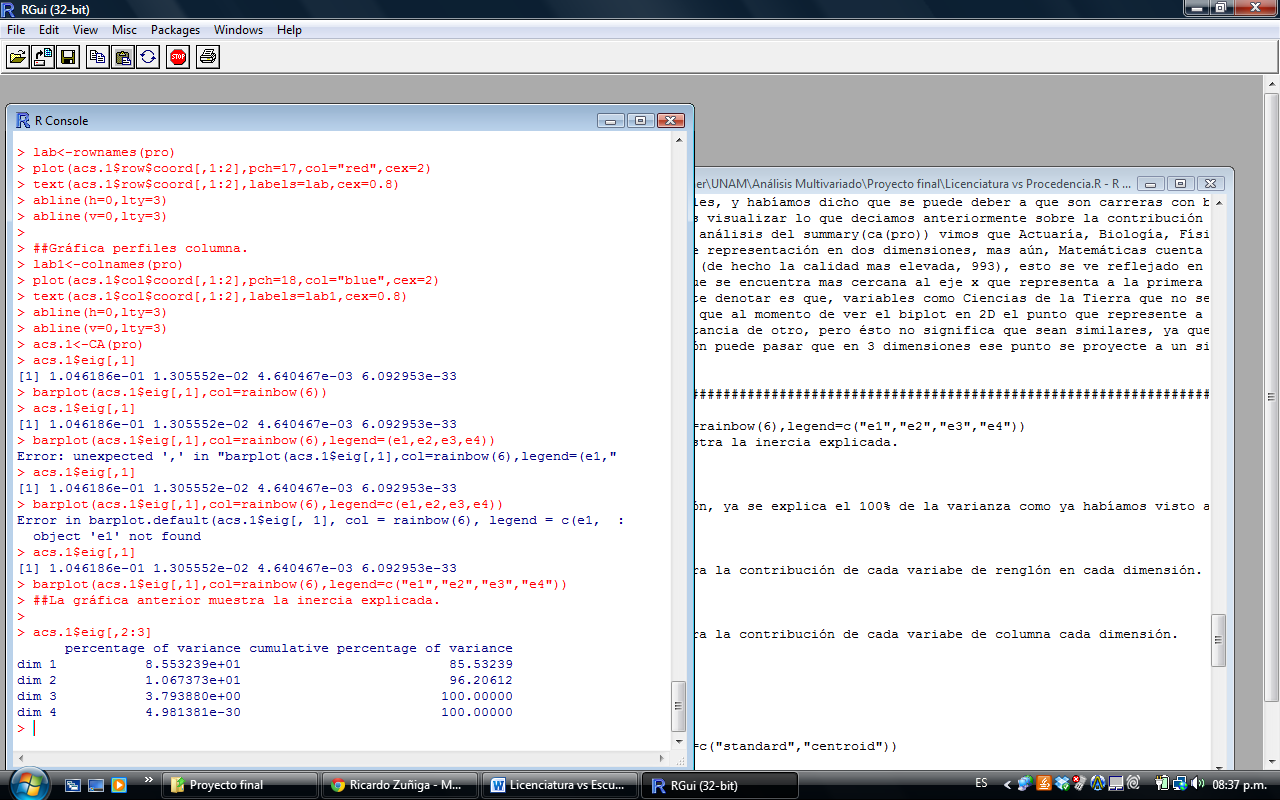


En éstas gráficas podemos observar las asociaciones perfiles columna y perfiles renglón en dos dimensiones, por ejemplo Biología y Ciencias de la Tierra se encuentran muy cercanos, esto ya lo habíamos notado con las gráficas iniciales, y habíamos dicho que se puede deber a que son carreras con bases en ciencias naturales.

Con esta gráfica podemos visualizar lo que decíamos anteriormente sobre la contribución de las variables a cada dimensión. En el análisis del summary(ca(pro)) vimos que Actuaría, Biología, Física y Matemáticas cuentan con una excelente calidad de representación en dos dimensiones, más aún, Matemáticas cuenta con una calidad muy fuerte en la primera dimensión (de hecho la calidad más elevada, 993), esto se ve reflejado en la gráfica de biplot, ya que es la variable que se encuentra más cercana al eje x que representa a la primera dimensión.

Un dato que es importante denotar es que, variables como Ciencias de la Tierra 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 ésto 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.

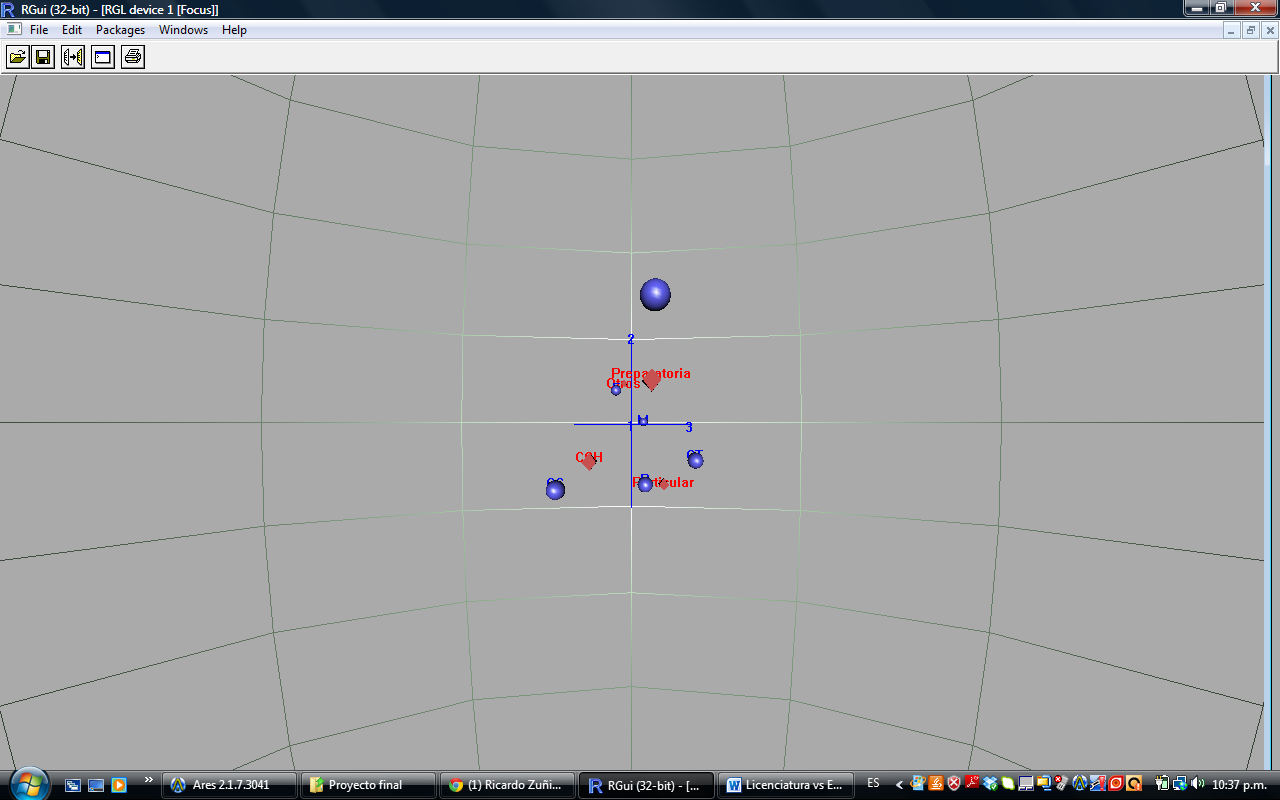
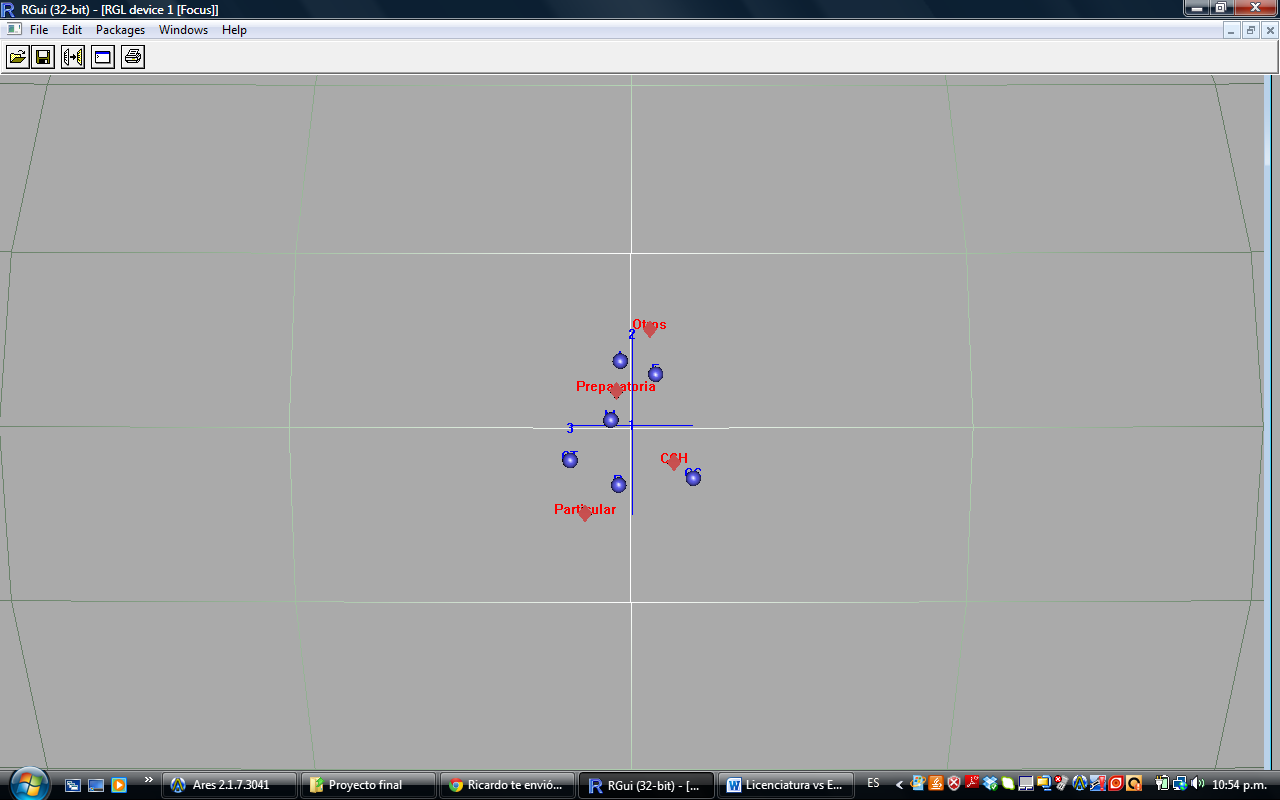
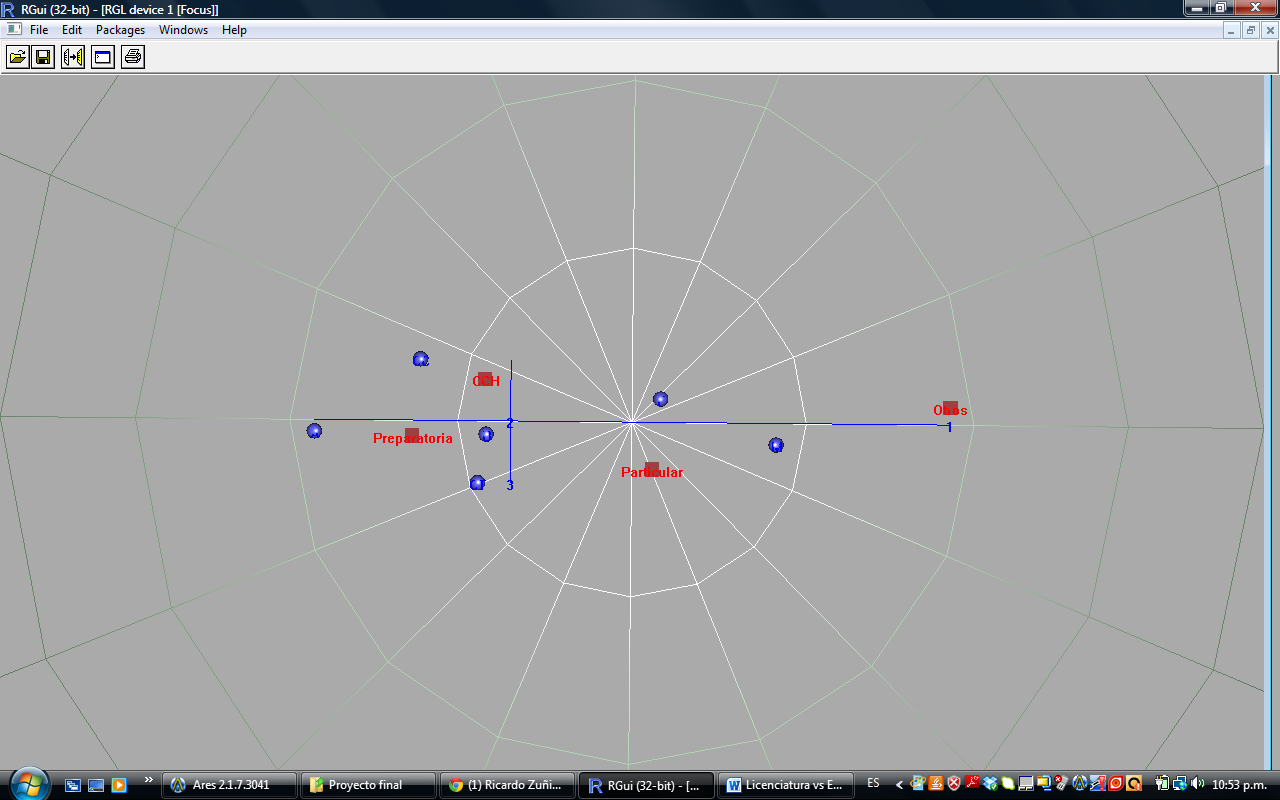




El barplot anterior muestra gráficamente el nivel de inercia explicado para cada dimensión. Como ya habíamos mencionado, la representación en 1 dimensión es bastante buena pues acumula más del 85% de inercia, sin embargo, si consideramos la acumulación hasta la segunda dimensión la representación es muchísimo mejor. Ya para la tercera dimensión se acumula el 100% de la inercia explicada.



En la gráfica biplot anterior se pude notar la cercanía entre las carreras Ciencias de la Tierra y Biología, y simultáneamente la cercanía de estas carreras con la escuela de procedencia CCH, sin embargo, recordemos que la variable CCH no ayudaba a distinguir las variables de carrera, entonces ¿Por qué gráficamente parece que estas variables se encuentran ampliamente relacionadas? Esta pregunta es fácil de responder ya que la calidad de representación de la variable CCH en la primera y segunda dimensión es muy mala, al igual que sus respectivas contribuciones. Esto nos lleva a pensar que esta variable es colapsada en dos dimensiones de tal suerte que parece estar asociada con estas variables, entonces si vemos la representación en tres dimensiones es posible que esta variable se separe bastante de la masa de puntos. Confirmemos lo anteriormente dicho.. (Gráfica 3D)



Como anteriormente mencionamos, la variable CCH se encuentra alejado de las variables de carrera aunque graficándolo en dos dimensiones parecía que esta variable se encontraba estrechamente relacionada con las variables de Biología, Ciencias de la Tierra y Ciencias de la Computación.

Por otro lado, la variable Física y Matemáticas efectivamente se encuentran cercanas a la variable de escolaridad Otros. Esto nos ayuda a confirmar que estas variables de Carrera se encuentran asociadas a la variable Otros. Cabe destacar que tanto en el biplot en 2D como en el de 3D los puntos que representan a las variables Física y Matemáticas no solo se encuentran cercanos al vector que representa Otros, aparte se encuentran en la misma dirección (vectorialmente hablando), esto es interpretado como que la asociación que tienen con esta variable es positiva, es decir, los estudiantes que estudian estas carreras se distinguen por provenir de otras escuelas. Lo anterior hace sentido ya que muchos de los estudiantes de estas carreras (Física y Matemáticas) provienen de estados de la república mexicana lejanos a la zona metropolitana.

Otra observación que vale la pena mencionar es sobre la asociación de la carrera de Actuaría con la escuela de procedencia Preparatoria. Como se había observado en la gráfica de mosaico, estas variables se encuentran altamente asociadas, y en efecto, positivamente asociadas. Lo anterior es confirmado con las gráficas biplot, ya que como ocurrió con las variables anteriormente descritas, la posición física de la variable Actuaría con respecto a la variable Preparatoria es de cercanía y de dirección similar.

*Gráficas de Benzecri*





Las anteriores gráficas de Benzecri muestran que son pocas las similitudes entre perfiles tanto de renglón como de columna que se encuentran mal representadas en dos dimensiones, por ejemplo, la asociación entre Ciencias de la Computación y Ciencias de la Tierra cuentan con una regular representación en la segunda dimensión, ya que el punto que representa esta asociación se encuentra alejado de la recta de Benzecri. Otro punto que se encuentra fuera de la recta de 45° es el que representa la asociación entre las variables Matemáticas y Ciencias de la Computación, sin embargo, tanto esta diferencia como la de la asociación CC-A son poco considerables ya que no se encuentran extremadamente alejados de la recta. Respecto a los perfiles columna, todos los puntos se encuentran muy cercanos (o sobre) la recta de Benzecri, lo cual habla de una buena representación de estas asociaciones en la segunda dimensión.

En la tercera dimensión todas las asociaciones son perfectas, lo cual tiene mucho sentido ya que habíamos visto que en la tercera dimensión ya se acumulaba el 100% de la inercia explicada.

*Conclusiones*

Con respecto a los perfiles renglón y columna, encontramos que las variables Actuaría y Preparatoria se encuentran estrechamente asociadas, al igual que las variables Física y Matemáticas con otras escuelas. De aquí que los estudiantes de Actuaría y los estudiantes de Física y Matemáticas se distinguen por provenir de Preparatoria y de otras escuelas, respectivamente. Los estudiantes de las otras carreras (Biología, Ciencias de la Tierra y Ciencias de la Computación) no se distinguen por provenir de alguna escuela en particular.

En cuanto a los perfiles columna, CCH y Preparatoria cuentan con similitudes respecto a la carrera que estudian los alumnos que provienen de ellas. Esta relación no muestra un sentido evidente para nosotros, así que requerimos de un experto de área que nos pueda dar una explicación lógica y razonable de este hecho.

Por otro lado, al analizar los perfiles renglón, observamos que Biología y Ciencias de la Tierra son carreras con mucha similitud respecto a las escuelas de las que provienen los alumnos que las estudian, aunque estos alumnos no se distingan por provenir de alguna escuela en específico. Ya habíamos mencionado nuestra suposición sobre este hecho; estas carreras tienen base en las ciencias naturales a diferencia de las otras carreras que tienen bases matemáticas.