Ejercicio 1. Descubrimiento de subgrupos

Autores: Miruna Andreea Gheata,
Miguel Rodríguez Gómez
Primer ejercicio de la asignatura de Técnicas Avanzadas de Minería de Datos del Máster Universitario en Sistemas Inteligentes (MUSI)
Universitat de les Illes Balears
07122 Palma, Islas Baleares, España
miruna.gheata1@estudiant.uib.cat
miguel.rodz.gomz@gmail.com

I. Introducción

En este documento se presenta el trabajo realizado para el primer ejercicio de la asignatura de Técnicas Avanzadas de Minería de Datos del Máster Universitario en Sistemas Inteligentes. A continuación se detalla el contenido de cada apartado.

En el Capítulo 2 se definen los datos empleados para la identificación de subgrupos, así como también las variables objetivo y variables atributo elegidas. A continuación se encontrará el análisis realizado para los subgrupos encontrados para cada variable objetivo.

Por último, el Capítulo 3 contiene un anexo que muestra el código empleado.

II. Descubrimiento de subgrupos

El **descubrimiento de subgrupos** tiene como objetivo identificar subconjuntos dentro de un *dataset* que presenten un comportamiento interesante respecto a un criterio definido, formalizado por una función de calidad [1]. Se pueden identificar 4 elementos importantes: la **variable objetivo**, las variables explicativas conocidas como **atributos**, la **función de calidad** y la **estrategia de búsqueda** mediante la cuál se descubren los subgrupos.

La **variable objetivo** es el criterio que se debe seguir para identificar los subgrupos. Se corresponde con el valor de una de las características del *dataset* para el cual se desea construir el conjunto de reglas que definan el subgrupo. Por ejemplo, si se tiene como característica objetivo `esRojo = 1`, donde el valor 1 indica si el elemento es rojo, el subconjunto encontrado será aquel donde mayor número de elementos rojos existan.

Los **atributos** son aquellas características del *dataset* que se emplean para construir las reglas. La variable objetivo no se puede encontrar dentro de los atributos. Se pueden seleccionar todas las características o elegir un subconjunto de ellas.

La **función de calidad** sirve para determinar el interés del subconjunto encontrado.

Por último, la **estrategia de búsqueda** determina la manera en la que se evalúan los elementos del *dataset* para poder construir los subgrupos de mayor interés para el criterio definido.

II-A. *Elementos empleados*

II-A1. US Census Data: El dataset empleado. El Censo de Estados Unidos es un censo que se realiza cada diez años y que recoge datos de la población de Estados Unidos, incluyendo también inmigrantes ilegales y extranjeros visitantes de larga estancia. El objetivo principal es el de recalcular la asignación de votos electorales de cada estado, basado exclusivamente en la población del mismo. Los datos completos del censo no son de dominio público hasta 72 años después del mismo, aunque se puede disponer de datos agregados y de algunas muestras [2].

Del *dataset* original con 2.458.285 entradas **se ha seleccionado aleatoriamente un 10 %** (245.828) de los datos para poder realizar el estudio de descubrimiento de subgrupos. Se ha decidido emplear este subconjunto de los datos originales debido a la falta de recursos para tratar de forma óptima con todo el *dataset* o incluso con un mayor porcentaje del mismo, ya que en distintas ejecuciones se probó con el 25 % de los datos y requería de una gran cantidad de recursos (tiempo y memoria).

II-A2. Variables objetivo: Tras analizar el *dataset*, se ha decidido emplear como variables objetivo las siguientes:

1. `iLooking = 1`: Personas en búsqueda activa de empleo.
2. `dHours = 5`: Personas que trabajan más de 50 horas semanales.
3. `iSchool = 1`: Personas sin estudios secundarios.
4. `dAge = 7`: Personas mayores de 65 años.
5. `dTravTime = 6`: Personas que tardan más de 1h en llegar al trabajo.
6. `iYearsch = 17`: Personas que tienen un doctorado.
7. `iMilitary = 4`: Personas que no están en un cargo militar.

8. `dAncstry1 = 11`: Personas de origen americano.

II-A3. Funciones de calidad: Se han empleado las funciones de calidad de Adjusted Residuals `ares` y Piatetsky-Shapiro `ps`. El resto de funciones se han descartado debido al elevado tiempo de ejecución.

II-A4. Estrategias de búsqueda: Se han empleado las estrategias de búsqueda BSD `bsd` y SD-Map `sdmap`. El resto de estrategias se han descartado debido al elevado tiempo de ejecución y la cantidad de memoria necesaria (se lanzaba el error de `java.lang.OutOfMemoryError: Java heap space`)

II-B. Consideraciones del análisis

A lo largo de las distintas ejecuciones se han detectado comportamientos similares y, para evitar la redundancia en las explicaciones, se muestran aquí las más importantes:

1. El valor `dIncomeX = 0` aparece en diferentes reglas; según la definición oficial, este valor indica o bien que las personas evaluadas son menores de 16 años o bien que se ha dejado sin contestar [3]. Se ha supuesto que esta variable con este valor hace referencia al segundo caso, es decir, el ingreso proviene de la fuente que se ha marcado pero la persona no ha especificado sus ingresos exactos para ese año.
2. Como bien se ha mencionado, el análisis se ha realizado con un 10 % de los datos originales. Se ha supuesto que esta cantidad de datos es suficiente para poder obtener resultados fiables y cercanos a la realidad (ya que se han seleccionado de forma aleatoria).
3. Se han realizado 3 ejecuciones diferentes para cada variable objetivo, modificando las funciones de calidad y el método de descubrimiento de subgrupos. En concreto, se han utilizado estas combinaciones de parametros:
   - Función de calidad de Piatetsky-Shapiro y el método de SD-Map.
   - Función de calidad de *Adjusted Residuals* y el método de SD-Map.
   - Función de calidad de Piatetsky-Shapiro y el método de BSD.

   Se han realizado estas ejecuciones para observar si los subgrupos obtenidos cambian considerablemente dependiendo de los parámetros empleados.
4. En algunos casos, en vez de realizar las 3 ejecuciones con las diferentes configuraciones, se ha optado por modificar los atributos empleados y obtener reglas más específicas. En estos casos se ha empleado como función de calidad `ps` y método `sdmap`.

II-C. Personas en búsqueda activa de empleo

Se ha analizado **el sector poblacional que buscaba activamente empleo en Estados Unidos en 1990**. Para ello se ha seleccionado la siguiente variable objetivo `iLooking = 1` y los atributos especificados en la Figura 1, donde se puede observar los resultados obtenidos, que nos indican datos de interés.

	quality	p	size	chi2	description
1	2619.71	0.21	14811	16305.94	iYearwrk=2
2	2619.71	0.21	14811	16305.94	iYearwrk=2, iWork89=1
3	2503.25	0.21	14377	15309.15	iYearwrk=2, dHispanic=0
4	2503.25	0.21	14377	15309.15	iYearwrk=2, dHispanic=0, iWork89=1
5	2343.28	0.21	13479	14253.44	iYearwrk=2, iImmigr=0
6	2343.28	0.21	13479	14253.44	iYearwrk=2, iCitizen=0
7	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0
8	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0
9	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iImmigr=0
10	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iImmigr=0, iWork89=1
11	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iWork89=1
12	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iImmigr=0
13	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iImmigr=0, iWork89=1
14	2343.28	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iWork89=1
15	2343.28	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iImmigr=0
16	2343.28	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iImmigr=0, iWork89=1
17	2343.28	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iWork89=1
18	2343.28	0.21	13479	14253.44	iYearwrk=2, iImmigr=0, iWork89=1
19	2338.67	0.05	98816	3060.74	iClass=1
20	2298.06	0.20	13294	13888.27	iYearwrk=2, dPOB=0, iImmigr=0, dHispanic=0

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	127.69	0.21	14811	16305.94	iYearwrk=2
2	127.69	0.21	14811	16305.94	iYearwrk=2, iWork89=1
3	126.13	0.24	10625	15908.07	iYearwrk=2, iClass=1
4	126.13	0.24	10625	15908.07	iYearwrk=2, iClass=1, iWork89=1
5	123.73	0.21	14377	15309.15	iYearwrk=2, dHispanic=0
6	123.73	0.21	14377	15309.15	iYearwrk=2, dHispanic=0, iWork89=1
7	122.25	0.24	10290	14944.97	iYearwrk=2, iClass=1, dHispanic=0
8	122.25	0.24	10290	14944.97	iYearwrk=2, iClass=1, dHispanic=0, iWork89=1
9	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0
10	119.39	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iImmigr=0
11	119.39	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iImmigr=0, iWork89=1
12	119.39	0.21	13479	14253.44	iYearwrk=2, iCitizen=0, iWork89=1
13	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0
14	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iImmigr=0
15	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iImmigr=0, iWork89=1
16	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iCitizen=0, iWork89=1
17	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iImmigr=0
18	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iImmigr=0, iWork89=1
19	119.39	0.21	13479	14253.44	iYearwrk=2, dPOB=0, iWork89=1
20	119.39	0.21	13479	14253.44	iYearwrk=2, iImmigr=0, iWork89=1

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	2619.71	0.21	14811	16305.94	iYearwrk=2
2	2619.71	0.21	14811	16305.94	iWork89=1, iYearwrk=2
3	2503.25	0.21	14377	15309.15	dHispanic=0, iYearwrk=2
4	2503.25	0.21	14377	15309.15	dHispanic=0, iWork89=1, iYearwrk=2
5	2343.28	0.21	13479	14253.44	dPOB=0, iImmigr=0, iYearwrk=2
6	2343.28	0.21	13479	14253.44	iCitizen=0, iImmigr=0, iYearwrk=2
7	2343.28	0.21	13479	14253.44	dPOB=0, iYearwrk=2
8	2343.28	0.21	13479	14253.44	dPOB=0, iImmigr=0, iWork89=1, iYearwrk=2
9	2343.28	0.21	13479	14253.44	iImmigr=0, iWork89=1, iYearwrk=2
10	2343.28	0.21	13479	14253.44	iCitizen=0, dPOB=0, iImmigr=0, iYearwrk=2
11	2343.28	0.21	13479	14253.44	iCitizen=0, dPOB=0, iWork89=1, iYearwrk=2
12	2343.28	0.21	13479	14253.44	iCitizen=0, iImmigr=0, iWork89=1, iYearwrk=2
13	2343.28	0.21	13479	14253.44	iCitizen=0, iWork89=1, iYearwrk=2
14	2343.28	0.21	13479	14253.44	dPOB=0, iWork89=1, iYearwrk=2
15	2343.28	0.21	13479	14253.44	iImmigr=0, iYearwrk=2
16	2343.28	0.21	13479	14253.44	iCitizen=0, iYearwrk=2
17	2343.28	0.21	13479	14253.44	iCitizen=0, dPOB=0, iYearwrk=2
18	2343.28	0.21	13479	14253.44	iCitizen=0, dPOB=0, iImmigr=0, iWork89=1, iYearwrk=2
19	2338.67	0.05	98816	3060.74	iClass=1
20	2298.06	0.20	13294	13888.27	dHispanic=0, iImmigr=0, iYearwrk=2

(c) qf = `ps`, method = `bsd`.

Figura 1: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población que están en búsqueda activa de empleo.

En primer lugar la regla más importante para la búsqueda activa de empleo es `iYearwrk = 2`, es decir, personas que cuyo último empleo fue en 1989. De esta regla se podría concluir que la economía estadounidense es sana y dinámica, ya que gran parte de quienes buscan trabajo lo hacen desde solamente el año anterior, indicando que no hay población que lleva años buscando sin encontrar.

Sin embargo, el 13 de Octubre de 1989 hubo una caída del mercado que pudo propiciar un aumento notable de desempleados en ese mismo año (este evento pasó a conocerse como *Black Friday* [4]). Debería entonces compararse con datos años anteriores para poder sacar conclusiones firmes sobre la salud de la economía. Por otro lado, `dPOB = 0`, `iCitizen = 0`, `iImmigr = 0`, `dHispanic = 0` son valores que indican que la siguiente regla mas importante es ser ciudadano estadounidense no migrante. Esto no significa necesariamente que los nacidos en estadounidenses estén desempleados en mayor medida, sino que, al ser el porcentaje de migrantes tan solo del 7.9 % [5] tenemos un dataset desbalanceado y la función de calidad empleada (Piatetsky-Shapiro) no tiene esto en cuenta.

II-D. Personas que trabajan más de 50 horas semanales

Esta sección se centra en las **personas que trabajan más de 50 horas semanales**. Para ello se ha seleccionado como variable objetivo `dHours = 5`. En la Figura 2 se muestran los resultados obtenidos.

	quality	p	size	chi2	description
1	10821.88	0.18	114039	24395.52	iYearwrk=1, dIncome5=0, dIncome6=0
2	10819.83	0.18	115483	24348.17	iYearwrk=1, dIncome5=0
3	10796.53	0.18	118618	24184.34	iYearwrk=1, dIncome6=0
4	10792.32	0.18	120192	24147.81	iYearwrk=1
5	10622.63	0.18	114915	23482.36	iYearwrk=1, dIncome7=0, dIncome6=0
6	10619.09	0.18	116423	23432.68	iYearwrk=1, dIncome7=0
7	10614.10	0.18	111487	23548.91	iYearwrk=1, dIncome5=0, dIncome7=0, dIncome6=0
8	10612.52	0.18	112879	23495.05	iYearwrk=1, dIncome5=0, dIncome7=0
9	10578.40	0.25	65489	29663.27	iYearwrk=1, iFertil=0
10	10552.36	0.25	64919	29682.78	iYearwrk=1, iFertil=0, dIncome6=0
11	10540.27	0.18	106456	23442.02	iYearwrk=1, dPoverty=2, dIncome5=0
12	10525.50	0.18	108863	23261.26	iYearwrk=1, dIncome8=0, dIncome5=0, dIncome6=0
13	10522.92	0.18	110127	23197.08	iYearwrk=1, dIncome8=0, dIncome5=0
14	10520.04	0.25	62967	30091.10	iYearwrk=1, iFertil=0, dIncome5=0
15	10515.95	0.18	110895	23136.86	iYearwrk=1, dPoverty=2
16	10514.20	0.19	105572	23373.30	iYearwrk=1, dPoverty=2, dIncome5=0, dIncome6=0
17	10506.67	0.18	113180	23019.55	iYearwrk=1, dIncome8=0, dIncome6=0
18	10502.12	0.18	114560	22961.47	iYearwrk=1, dIncome8=0
19	10494.76	0.25	62458	30106.92	iYearwrk=1, iFertil=0, dIncome5=0, dIncome6=0
20	10492.13	0.18	109915	23069.98	iYearwrk=1, dPoverty=2, dIncome6=0

(a) Ejecución 1.

	quality	p	size	chi2	description
1	5428.97	0.19	52351	9109.99	iClass=1, iFertil=0
2	5353.41	0.19	50588	9084.12	iClass=1, iFertil=0, dHispanic=0
3	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, iCitizen=0
4	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, iImmigr=0
5	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, dPOB=0
6	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, iImmigr=0, iCitizen=0
7	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, dPOB=0, iCitizen=0
8	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, dPOB=0, iCitizen=0, iImmigr=0
9	5038.54	0.19	46710	8545.28	iClass=1, iFertil=0, dPOB=0, iImmigr=0
10	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, iCitizen=0, dHispanic=0
11	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, iImmigr=0, iCitizen=0, dHispanic=0
12	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, iImmigr=0, dHispanic=0
13	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, dPOB=0, iCitizen=0, iImmigr=0, dHispanic=0
14	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, dPOB=0, iCitizen=0, dHispanic=0
15	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, dPOB=0, iImmigr=0, dHispanic=0
16	5022.68	0.19	46138	8572.21	iClass=1, iFertil=0, dPOB=0, dHispanic=0
17	4445.99	0.22	34237	8542.43	dOccup=1
18	4369.92	0.22	33551	8394.13	dOccup=1, dHispanic=0
19	4277.55	0.13	98816	3943.20	iClass=1
20	4245.62	0.13	95579	3929.57	iClass=1, dHispanic=0

(b) Ejecución 2.

Figura 2: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población trabaja más de 50 horas semanales.

En el primer caso (Figura 2a) puede observarse que la regla más importante es `iYearwrk = 1` pero, dado que esto indica que aquellas personas empleadas en 1990, no nos da información importante. Las siguientes más comunes son `dIncome5 = 0`, `dIncome6 = 0`, haciendo referencia a quienes reciben prestaciones sociales por ser pensionistas, tener discapacidades o situación social precaria. Por otro lado, la regla `iFertil = 0` indica hombres o mujeres menores de edad sin hijos (dado que estamos considerando trabajadores, se trata de hombres), mientras que `dPoverty = 2` indica gente por encima del umbral de pobreza.

En el segundo caso no se tuvieron en cuenta los atributos `iYearwrk`, `dIncome` y `dPoverty`, para observar reglas más específicas. En la Figura 2b observamos cómo `iClass = 1` es la regla más importante (trabajadores en el sector privado), seguida de `iFertil = 0` y después de `iImmigr, dHispanic, dPOB = 0`. Las últimas reglas dan importancia a `dOccup = 1`, indicando profesiones de investigación, médicos y sector financiero.

Usando ambos resultados concluimos que el subgrupo que trabaja (remuneradamente) más de 50 horas son hombres empleados en el sector privado (especialmente investigadores, médicos, finanzas), personas con necesidad de prestaciones sociales y no migrantes. Esto último puede deberse, como se comentó en la sección anterior, a que se trata de un dataset desbalanceado, no a que dicho sector poblacional se vea menos afectado.

II-E. Personas sin estudios secundarios

Ahora se analizará el perfil que tienen **aquellas personas que no atienden a la escuela**, utilizando como variable objetivo `iSchool = 1`. En la Figura 3 se muestran los resultados obtenidos.

Los resultados de las Figuras 3a y 3c son los siguientes: la regla más importante es `iDisabl2 = 2`, que designa personas sin ningún tipo de discapacidad que les impida trabajar. Esto quiere decir que la mayoría de las personas que no tienen estudios secundarios no tienen discapacidad (el 85 %, mientras que el 15 % sí que la tiene). Sin embargo, observando los datos que indican discapacidad dentro del *dataset* empleado, `iDisabl2 = 0, 1, 2`, encontramos que ésta tiene una prevalencia del 6,77 % (mucho menor que el 15 % de discapacitados sin estudios), mientras que no tenerla del 92,23 %, mayor que el 85 % sin estudios). Esto significa que la probabilidad de que alguien con discapacidad no tenga estudios secundarios es más del doble que la media poblacional (2.21 veces más alta). Fijándonos en otros atributos, las siguientes reglas que encontramos son `dIncome3/6/8 = 0`, refiriéndose respectivamente a quienes son autónomos granjeros o personas que reciben prestaciones sociales (por discapacidad, ser familia dependiente, desempleados, veteranos, etc.).

Utilizando los mismos atributos pero aplicando la función de calidad `ares` (Figura 3b) obtenemos prácticamente los mismos resultados.

	quality	p	size	chi2	description
1	27054.89	0.85	172072	67026.95	iDisabl2=2
2	26424.28	0.87	151965	56889.74	iDisabl2=2, dPoverty=2
3	26424.28	0.87	151965	56889.74	iDisabl2=2, dPoverty=2
4	26403.85	0.85	169518	62632.91	iDisabl2=2, dIncome3=0
5	25975.44	0.85	166999	59565.02	iDisabl2=2, dIncome6=0
6	25864.76	0.87	149416	53970.21	iDisabl2=2, dPoverty=2, dIncome6=0
7	25864.76	0.87	149416	53970.21	iDisabl2=2, dPoverty=2, dIncome6=0
8	25814.10	0.87	149608	53797.14	iDisabl2=2, dPoverty=2, dIncome3=0
9	25814.10	0.87	149608	53797.14	iDisabl2=2, dPoverty=2, dIncome3=0
10	25593.95	0.85	163838	56671.49	iDisabl2=2, dIncome8=0
11	25335.06	0.85	164490	55754.06	iDisabl2=2, dIncome6=0, dIncome3=0
12	25276.11	0.85	163878	55286.13	iDisabl1=2
13	25276.11	0.85	163878	55286.13	iDisabl1=2, iDisabl2=2
14	25264.64	0.87	147102	51078.90	iDisabl2=2, dPoverty=2, dIncome6=0, dIncome3=0
15	25264.64	0.87	147102	51078.90	iDisabl2=2, dPoverty=2, dIncome6=0, dIncome3=0
16	24997.73	0.87	144792	49641.63	iDisabl2=2, dPoverty=2, dIncome8=0
17	24997.73	0.87	144792	49641.63	iDisabl2=2, dPoverty=2, dIncome8=0
18	24976.32	0.85	161417	53207.64	iDisabl2=2, dIncome8=0, dIncome3=0
19	24868.03	0.87	145076	49169.86	iDisabl1=2, dPoverty=2
20	24868.03	0.87	145076	49169.86	iDisabl1=2, dPoverty=2

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	258.90	0.85	172072	67026.95	iDisabl2=2
2	250.27	0.85	169518	62632.91	iDisabl2=2, dIncome3=0
3	248.85	0.86	166750	61928.50	iDisabl2=2, dHispanic=0
4	244.06	0.85	166999	59565.02	iDisabl2=2, dIncome6=0
5	240.81	0.85	164211	57989.50	iDisabl2=2, dHispanic=0, dIncome3=0
6	238.52	0.87	151965	56889.74	iDisabl2=2, dPoverty=2
7	238.52	0.87	151965	56889.74	iDisabl2=2, dPoverty=2
8	238.06	0.85	163838	56671.49	iDisabl2=2, dIncome8=0
9	236.12	0.85	164490	55754.06	iDisabl2=2, dIncome6=0, dIncome3=0
10	235.76	0.85	161979	55583.92	iDisabl2=2, dHispanic=0, dIncome6=0
11	235.13	0.85	163878	55286.13	iDisabl1=2
12	235.13	0.85	163878	55286.13	iDisabl1=2, iDisabl2=2
13	232.31	0.87	149416	53970.21	iDisabl2=2, dPoverty=2, dIncome6=0
14	232.31	0.87	149416	53970.21	iDisabl2=2, dPoverty=2, dIncome6=0
15	232.23	0.87	147771	53930.37	iDisabl2=2, dPoverty=2, dHispanic=0
16	232.23	0.87	147771	53930.37	iDisabl2=2, dPoverty=2, dHispanic=0
17	231.94	0.87	149608	53797.14	iDisabl2=2, dPoverty=2, dIncome3=0
18	231.94	0.87	149608	53797.14	iDisabl2=2, dPoverty=2, dIncome3=0
19	230.67	0.85	161417	53207.64	iDisabl2=2, dIncome8=0, dIncome3=0
20	229.79	0.85	158702	52801.75	iDisabl2=2, dIncome8=0, dHispanic=0

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	27054.89	0.85	172072	67026.95	iDisabl2=2
2	26507.78	0.86	166750	61928.50	dHispanic=0, iDisabl2=2
3	26424.28	0.87	151965	56889.74	dPoverty=2, iDisabl2=2
4	26424.28	0.87	151965	56889.74	dPoverty=2, iDisabl2=2
5	26403.85	0.85	169518	62632.91	dIncome3=0, iDisabl2=2
6	25975.44	0.85	166999	59565.02	dIncome6=0, iDisabl2=2
7	25930.92	0.87	147771	53930.37	dHispanic=0, dPoverty=2, iDisabl2=2
8	25930.92	0.87	147771	53930.37	dHispanic=0, dPoverty=2, iDisabl2=2
9	25864.76	0.87	149416	53970.21	dIncome6=0, dPoverty=2, iDisabl2=2
10	25864.76	0.87	149416	53970.21	dIncome6=0, dPoverty=2, iDisabl2=2
11	25860.29	0.85	164211	57989.50	dIncome3=0, dHispanic=0, iDisabl2=2
12	25814.10	0.87	149608	53797.14	dIncome3=0, dPoverty=2, iDisabl2=2
13	25814.10	0.87	149608	53797.14	dIncome3=0, dPoverty=2, iDisabl2=2
14	25593.95	0.85	163838	56671.49	dIncome8=0, iDisabl2=2
15	25487.08	0.85	161979	55583.92	dHispanic=0, dIncome6=0, iDisabl2=2
16	25393.59	0.87	145351	51313.18	dHispanic=0, dIncome6=0, dPoverty=2, iDisabl2=2
17	25393.59	0.87	145351	51313.18	dHispanic=0, dIncome6=0, dPoverty=2, iDisabl2=2
18	25335.06	0.85	164490	55754.06	dIncome3=0, dIncome6=0, iDisabl2=2
19	25323.38	0.87	145426	51041.60	dIncome3=0, dHispanic=0, dPoverty=2, iDisabl2=2
20	25323.38	0.87	145426	51041.60	dIncome3=0, dHispanic=0, dPoverty=2, iDisabl2=2

(c) qf = `ps`, method = `bsd`.

Figura 3: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población que no tiene estudios secundarios.

Tras examinar las tres ejecuciones se puede concluir que el poco poder adquisitivo (al necesitar ayuda económica) es el factor más determinante a la hora de no poder acceder a estudios secundarios y, además, las personas con discapacidad tienen una probabilidad mucho más alta (2.21 veces) de no acceder a ello.

II-F. Personas mayores de 65 años

En esta sección hemos analizado las características generales de las **personas que tienen más de 65 años**, utilizando como variable objetivo `dAge = 7`. En la Figura 4 se muestran los resultados obtenidos. A diferencia de las otras ejecuciones, se han ido variando los atributos para observar si se puede llegar a obtener información más interesante.

	quality	p	size	chi2	description
1	22610.95	0.79	34506	150901.3	dIncome5=1
2	22352.79	0.79	34013	149264.4	dIncome5=1, dHispanic=0
3	22284.03	0.79	34034	148270.7	dIncome5=1, dIncome3=0
4	22025.86	0.79	33541	146642.9	dIncome5=1, dHispanic=0, dIncome3=0
5	21935.78	0.79	33420	145889.2	dIncome5=1, dIncome2=0
6	21682.04	0.79	32939	144288.3	dIncome5=1, dIncome2=0, dHispanic=0
7	21640.23	0.79	33006	143485.7	dIncome5=1, dIncome2=0, dIncome3=0
8	21525.41	0.79	32594	143483.9	dIncome5=1, dIncome6=0
9	21386.48	0.79	32525	141892.4	dIncome5=1, dIncome2=0, dHispanic=0, dIncome3=0
10	21306.57	0.79	32167	142162.7	dIncome5=1, dIncome6=0, dHispanic=0
11	21276.58	0.79	32395	140915.5	dIncome5=1, dIncome8=0
12	21208.20	0.79	32147	140927.5	dIncome5=1, dIncome6=0, dIncome3=0
13	21025.92	0.79	31921	139348.7	dIncome5=1, dIncome8=0, dHispanic=0
14	20989.35	0.79	31720	139613.2	dIncome5=1, dIncome6=0, dIncome3=0, dHispanic=0
15	20970.74	0.79	31968	138444.9	dIncome5=1, dIncome8=0, dIncome3=0
16	20927.77	0.78	32059	137545.3	dIncome5=1, iImmigr=0
17	20927.77	0.78	32059	137545.3	dIncome5=1, iCitizen=0
18	20927.77	0.78	32059	137545.3	dIncome5=1, iCitizen=0
19	20927.77	0.78	32059	137545.3	dIncome5=1, dPOB=0
20	20927.77	0.78	32059	137545.3	dIncome5=1, dPOB=0, iCitizen=0

(a) Ejecución 1.

	quality	p	size	chi2	description
1	12096.48	0.71	20975	66774.68	iYearwrk=6
5	11957.17	0.71	20620	66264.15	iYearwrk=6, dHispanic=0
6	11936.50	0.71	20473	66466.04	iYearwrk=6, dHispanic=0, dIndustry=0
7	11936.50	0.71	20473	66466.04	iYearwrk=6, dHispanic=0, dIndustry=0, iClass=0
8	11936.50	0.71	20473	66466.04	iYearwrk=6, dHispanic=0, iClass=0
2	12076.39	0.71	20816	67014.04	iYearwrk=6, dIndustry=0
12	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0
14	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0, iCitizen=0
15	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0, iCitizen=0
16	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0, iImmigr=0
17	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0, iImmigr=0, iCitizen=0
18	11081.57	0.71	19195	60755.59	iYearwrk=6, dPOB=0, iImmigr=0, iCitizen=0
10	11081.57	0.71	19195	60755.59	iYearwrk=6, iCitizen=0
11	11081.57	0.71	19195	60755.59	iYearwrk=6, iCitizen=0
20	11063.72	0.71	19057	60961.43	iYearwrk=6, iCitizen=0, dIndustry=0, iImmigr=0
13	11081.57	0.71	19195	60755.59	iYearwrk=6, iCitizen=0, iImmigr=0
19	11081.57	0.71	19195	60755.59	iYearwrk=6, iCitizen=0, iImmigr=0
3	12076.39	0.71	20816	67014.04	iYearwrk=6, iClass=0
4	12076.39	0.71	20816	67014.04	iYearwrk=6, iClass=0, dIndustry=0
9	11081.57	0.71	19195	60755.59	iYearwrk=6, iImmigr=0

(b) Ejecución 2.

	quality	p	size	chi2	description
1	9812.93	0.23	96005	14408.53	dIndustry=0, dHispanic=0
2	9812.93	0.23	96005	14408.53	iClass=0, dHispanic=0
3	9812.93	0.23	96005	14408.53	iClass=0, dHispanic=0, dIndustry=0
4	9711.95	0.23	99571	13939.84	dIndustry=0
5	9711.95	0.23	99571	13939.84	iClass=0
6	9711.95	0.23	99571	13939.84	iClass=0, dIndustry=0
7	9134.01	0.76	14485	53584.92	iMarital=1
8	8994.60	0.77	14184	52995.38	iMarital=1, dHispanic=0
9	8964.03	0.27	66076	14560.76	dIndustry=0, dAncstry2=1, dHispanic=0
10	8964.03	0.27	66076	14560.76	iClass=0, dAncstry2=1, dHispanic=0
11	8964.03	0.27	66076	14560.76	iClass=0, dAncstry2=1, dHispanic=0, dIndustry=0
12	8903.36	0.26	69153	13964.22	dIndustry=0, dAncstry2=1
13	8903.36	0.26	69153	13964.22	iClass=0, dAncstry2=1
14	8903.36	0.26	69153	13964.22	iClass=0, dAncstry2=1, dIndustry=0
15	8815.85	0.23	90204	11915.72	dIndustry=0, dPOB=0, dHispanic=0, iImmigr=0
16	8815.85	0.23	90204	11915.72	dIndustry=0, iCitizen=0, dHispanic=0
17	8815.85	0.23	90204	11915.72	dIndustry=0, iImmigr=0, dHispanic=0
18	8815.85	0.23	90204	11915.72	dIndustry=0, iImmigr=0, dHispanic=0, iCitizen=0
19	8815.85	0.23	90204	11915.72	dIndustry=0, iImmigr=0, dHispanic=0, iCitizen=0
20	8815.85	0.23	90204	11915.72	dIndustry=0, iCitizen=0, dHispanic=0

(c) Ejecución 3.

Figura 4: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población mayor a 65 años.

En el primer caso (Figura 4a), la regla más importante es `dIncome5 = 1`, conformando el 79 % de las personas de dicha edad. No es de extrañar, ya que dicha categoría incluye a personas que reciben pensión. Otras reglas relevantes (aunque siempre combinadas con la anterior) son `dHispanic, iCitizen, iImmigr = 0`, que indican ciudadanos de EEUU no migrantes. No es de extrañar, pues las personas migrantes son un porcentaje bajo de la población

y, además, es en términos generales joven. También aparecen `dIncome6-7 = 0`, siendo ayudas económicas a personas mayores, discapacitadas o remuneración por antiguo empleo.

En el segundo caso (Figura 4b) eliminamos el atributo `dIncome` y añadimos `dIndustry`. La regla más importante es `iYearwrk = 6`, con un 71 % de personas mayores de 65 años, y que indica personas que trabajaron por última vez en 1979 o antes. Puesto que una persona en EEUU puede jubilarse a partir de los 62 años, es normal encontrar un porcentaje tan alto. La siguiente regla es `dIndustry = 0`, que hace alusión a personas que no trabajan. Después viene `iClass = 0` que, al indicar personas jubiladas que llevan más de 5 años sin trabajar, no añade más información. Otras reglas relevantes son `dHispanic, iCitizen, iImmigr, dPOB = 0`, comentadas anteriormente.

En tercer y último lugar (Figura 4c), eliminamos el atributo `iYearwrk = 6`, encontrando como reglas más importantes `iClass = 0`, `dHispanic = 0` y `dIndustry = 0`. Sin embargo, son relevantes también `iMarital = 1`, que se refiere a personas viudas y afecta al 76 %, y `dAncestry2 = 1`, indicando personas que tienen ascendencia proveniente de Europa occidental.

Podemos concluir que la población mayor de 65 años recibe en su mayoría pensión, no trabaja, es en su mayoría estadounidense, no migrante, y es muy probable que sea viuda.

II-G. Personas que tardan más de 1h en llegar al trabajo

Se han analizado los factores que influyen en el hecho de que **una persona tarde más de 60 minutos en llegar a su trabajo**. La variable objetivo elegida es (`dTravtime = 6`). En la Figura 5 se muestran los resultados obtenidos tras realizar 2 ejecuciones diferentes modificando los atributos empleados.

Los resultados son muy parecidos: las personas que tienen un tiempo de conmutación más elevado no tienen ninguna discapacidad que les impida trabajar (`iDisabl2 = 2`), pertenecen al sector poblacional que se encuentra por encima del umbral de pobreza (`dPoverty = 2`) y, por último, los ingresos que reciben vienen dados por:

- La Seguridad Social (`dIncome = 3`).
- El trabajo en el campo (`dIncome = 5`).
- Asistencia pública (`dIncome = 6`).
- Pensiones de jubilación o de discapacidad (`dIncome = 7`).

.

En los resultados de la Figura 5b, correspondiente a la ejecución con el algoritmo de `ares`, se incluyen también las variables que muestran que los varones son aquellos que tardan mucho más en llegar a trabajar (`iFertil = 0`, `iSex = 0`). Sin embargo, esto no implica que las mujeres tardaban menos en llegar a trabajar. Entran en factor:

	quality	p	size	chi2	description
1	2537.40	0.05	130404	4034.15	iDisabl2=2, dIncome5=0, dPoverty=2
2	2537.40	0.05	130404	4034.15	iDisabl2=2, dIncome5=0, dPoverty=2
3	2521.58	0.05	128381	3977.09	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
4	2521.58	0.05	128381	3977.09	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
5	2504.84	0.05	128446	3924.64	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
6	2504.84	0.05	128446	3924.64	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
7	2492.17	0.05	126156	3879.88	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0
8	2492.17	0.05	126156	3879.88	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0
9	2489.46	0.05	126444	3871.93	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0, dIncome3=0
10	2489.46	0.05	126444	3871.93	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0, dIncome3=0
11	2472.74	0.05	124305	3817.46	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0, dIncome6=0
12	2472.74	0.05	124305	3817.46	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0, dIncome6=0
13	2460.39	0.05	124281	3779.39	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0, dIncome3=0
14	2460.39	0.05	124281	3779.39	iDisabl2=2, dIncome5=0, dPoverty=2, dIncome7=0, dIncome3=0
15	2459.05	0.04	143933	3888.52	iDisabl2=2, dIncome5=0, dIncome6=0
16	2446.44	0.05	125399	3737.74	iDisabl1=2, dIncome5=0, dPoverty=2
17	2446.44	0.05	125399	3737.74	iDisabl1=2, dIncome5=0, dPoverty=2
18	2446.44	0.05	125399	3737.74	iDisabl1=2, dIncome5=0, dPoverty=2, iDisabl2=2
19	2446.44	0.05	125399	3737.74	iDisabl1=2, dIncome5=0, dPoverty=2, iDisabl2=2
20	2443.59	0.05	123676	3727.64	iDisabl1=2, dIncome5=0, dPoverty=2, dIncome6=0, iDisabl2=2

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	65.05	0.06	65092	4232.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2
2	65.05	0.06	65092	4232.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2
3	65.05	0.06	65092	4232.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, iFertil=0
4	65.05	0.06	65092	4232.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, iFertil=0
5	65.05	0.06	65092	4232.10	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2
6	65.05	0.06	65092	4232.10	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2
7	64.79	0.06	64395	4197.43	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
8	64.79	0.06	64395	4197.43	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0, iFertil=0
9	64.79	0.06	64395	4197.43	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
10	64.79	0.06	64395	4197.43	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0, iFertil=0
11	64.79	0.06	64395	4197.43	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
12	64.79	0.06	64395	4197.43	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome6=0
13	64.58	0.06	63459	4171.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
14	64.58	0.06	63459	4171.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0, iFertil=0
15	64.58	0.06	63459	4171.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
16	64.58	0.06	63459	4171.10	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0, iFertil=0
17	64.58	0.06	63459	4171.10	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
18	64.58	0.06	63459	4171.10	iFertil=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0
19	64.33	0.06	62782	4138.08	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0, iFertil=0, dIncome6=0
20	64.33	0.06	62782	4138.08	iSex=0, iDisabl2=2, dIncome5=0, dPoverty=2, dIncome3=0, dIncome6=0

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	2537.40	0.05	130404	4034.15	iDisabl2=2, dIncome5=0, dPoverty=2
2	2537.40	0.05	130404	4034.15	iDisabl2=2, dIncome5=0, dPoverty=2
3	2521.58	0.05	128381	3977.09	iDisabl2=2, dIncome6=0, dIncome5=0, dPoverty=2
4	2521.58	0.05	128381	3977.09	iDisabl2=2, dIncome6=0, dIncome5=0, dPoverty=2
5	2504.84	0.05	128446	3924.64	iDisabl2=2, dIncome3=0, dIncome5=0, dPoverty=2
6	2504.84	0.05	128446	3924.64	iDisabl2=2, dIncome3=0, dIncome5=0, dPoverty=2
7	2492.17	0.05	126156	3879.88	iDisabl2=2, dIncome5=0, dIncome7=0, dPoverty=2
8	2492.17	0.05	126156	3879.88	iDisabl2=2, dIncome5=0, dIncome7=0, dPoverty=2
9	2489.46	0.05	126444	3871.93	iDisabl2=2, dIncome3=0, dIncome6=0, dIncome5=0, dPoverty=2
10	2489.46	0.05	126444	3871.93	iDisabl2=2, dIncome3=0, dIncome6=0, dIncome5=0, dPoverty=2
11	2472.74	0.05	124305	3817.46	iDisabl2=2, dIncome6=0, dIncome5=0, dIncome7=0, dPoverty=2
12	2472.74	0.05	124305	3817.46	iDisabl2=2, dIncome6=0, dIncome5=0, dIncome7=0, dPoverty=2
13	2460.39	0.05	124281	3779.39	iDisabl2=2, dIncome3=0, dIncome5=0, dIncome7=0, dPoverty=2
14	2460.39	0.05	124281	3779.39	iDisabl2=2, dIncome3=0, dIncome5=0, dIncome7=0, dPoverty=2
15	2459.05	0.04	143933	3888.52	iDisabl2=2, dIncome6=0, dIncome5=0
16	2446.44	0.05	125399	3737.74	dIncome5=0, iDisabl1=2, dPoverty=2
17	2446.44	0.05	125399	3737.74	iDisabl2=2, dIncome5=0, iDisabl1=2, dPoverty=2
18	2446.44	0.05	125399	3737.74	dIncome5=0, iDisabl1=2, dPoverty=2
19	2446.44	0.05	125399	3737.74	iDisabl2=2, dIncome5=0, iDisabl1=2, dPoverty=2
20	2443.59	0.05	123676	3727.64	dIncome6=0, dIncome5=0, iDisabl1=2, dPoverty=2

(c) qf = `ps`, method = `bsd`.

Figura 5: Las 20 mejores reglas obtenidas en cada ejecución para el subgrupo de la población que tardan >1hr en llegar al trabajo.

1. La cantidad de mujeres que se encuentran en el subconjunto del *dataset* empleado para el análisis (ya que puede haber muchos más datos sobre hombres que de mujeres).
2. El porcentaje de mujeres trabajadoras en la década de los 90. Tal como se puede observar en la Figura 6, hay aproximadamente un 79 % de hombres y un 59 % de mujeres trabajando.

Lo único que se puede afirmar tras visualizar los resultados es que en el subconjunto encontrado se abarcan más datos relacionados a los hombres trabajadores que a las mujeres.

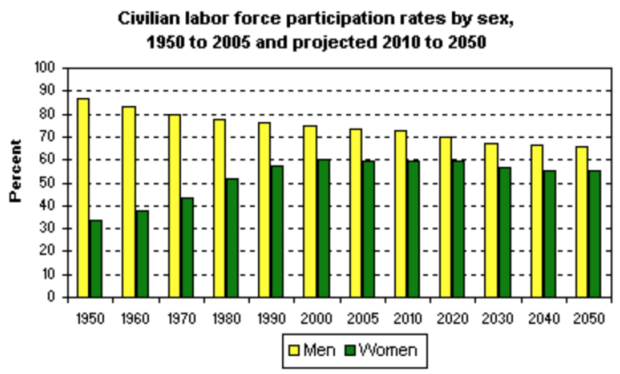

Figura 6: Porcentaje de hombres y de mujeres con trabajo a lo largo de los años. (Fuente: [6]

La precisión de los subgrupos encontrados no es muy elevada (0.05 con `ps + sdmap` y `ps + bsd`, y un 0.06 con `ares + sdmap`).

Para visualizar los datos reales y entender por qué se daban casos de una conmutación superior a una hora para llegar al trabajo se tienen las Figuras 7, 9 y 8.

TABLE 3-1 Commuters (Millions) by Residence Location

	1990		2000	
Worker Residence	No.	%	No.	%
All workers	115.0	100.0	128.3	100.0
In metro area	89.6	77.9	99.1	77.2
In central city	32.2	28.0	34.4	26.8
In suburb	57.4	49.9	64.7	50.4
In nonmetro area	25.4	22.1	29.2	22.8

Note: Data adjusted to reflect 1980 geographic definitions.

Figura 7: Conmutadores (en millones) por su lugar de residencia. (Fuente: [7])

TABLE 3-2 Basic Metropolitan Commuter Flow (Millions)

Origin	Workers	No. of Workers
Central City	Who work in same central city	24.5
	Who work in same suburb	7.5
	Who work outside home metro area	2.3
	Total (rounded) living in a central city	**34.4**
Suburb	Who work in same central city	16.6
	Who work in suburb of same metro area	40.8
	Who work outside home metro area	7.3
	Total (rounded) living in a suburb	**64.7**

Figura 8: Flujo de conmutación del lugar de residencia al lugar de trabajo. (Fuente: [7])

En la Figura 7 se muestra una tabla con los trabajadores que conmutaban agrupados por su lugar de residencia. Se puede observar que hay un 49.9 % con la residencia en las afueras. Combinando esto con los datos de la Figura 8, se

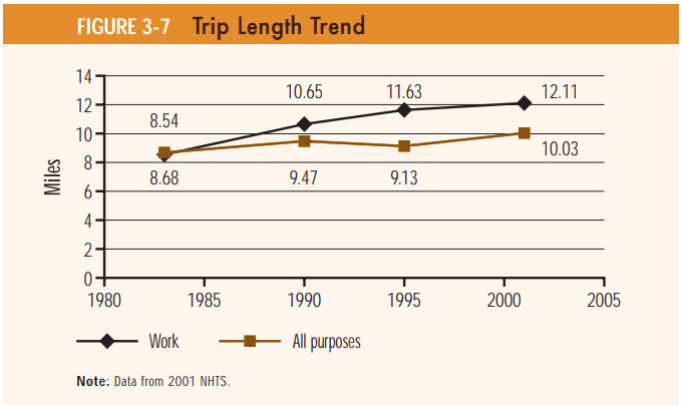

Figura 9: Distancia media recorrida para llegar al trabajo. (Fuente: [7])

tiene que los trabajadores que viven a las afueras suelen ir o bien a la ciudad a trabajar o bien en algún otro destino dentro del mismo área.

Es evidente, por lo tanto, que muchos de estos trabajadores pueden llegar a tardar un tiempo considerable desde su residencia hasta su lugar de trabajo.

En la Figura 9 se muestra que en media los trabajadores viajan unos 17 kilómetros para alcanzar su trabajo. Teniendo en cuenta que muchos de estos trabajadores pueden ser discapacitados (aunque no tengan ninguna discapacidad que les impida trabajar) y no se sabe si tienen vehículo propio o no, se puede suponer que hay ciertas limitaciones de transporte que hacen que el viaje sea más largo.

II-H. Personas que tienen un doctorado

El siguiente análisis se ha realizado entorno al **sector poblacional que posee un doctorado**. La variable objetivo elegida es (`iYearsch = 17`). En la Figura 10 se muestran los resultados obtenidos con las 3 ejecuciones realizadas.

En las Figuras 10a y 10c se obtienen resultados muy similares, mostrando que aquellas personas que tienen un doctorado viven solos (`iRemplpar = 0`), no tienen ninguna discapacidad que les impida trabajar (`iDisabl2 = 2`), tienen ingresos que proceden de la asistencia pública (`iIncome6 = 0`) y tienen una de las siguientes ocupaciones (`dOccup = 1`):

- Legisladores.
- Ocupaciones relacionadas con la gestión.
- Ocupaciones profesionales especializadas (ingenieros, científicos, médicos, profesores, conservadores de arte).
- Científicos sociales y urbanistas.
- Trabajadores sociales, recreativos y religiosos.
- Abogados y jueces.

Además, en la Figura 10b (correspondiente con la ejecución con la función de calidad de `ares`), se muestra también que

	quality	p	size	chi2	description
1	731.37	0.03	34153	4035.66	dOccup=1, iRemplpar=0
2	730.99	0.03	34237	4023.17	dOccup=1
3	730.16	0.03	33979	4039.57	dOccup=1, iRelat2=0, iRemplpar=0
4	729.78	0.03	34062	4027.16	dOccup=1, iRelat2=0
5	727.56	0.03	33669	4041.91	dOccup=1, iDisabl2=2, iRemplpar=0
6	727.18	0.03	33753	4029.23	dOccup=1, iDisabl2=2
7	726.34	0.03	33497	4045.75	dOccup=1, iDisabl2=2, iRelat2=0, iRemplpar=0
8	725.96	0.03	33580	4033.16	dOccup=1, iDisabl2=2, iRelat2=0
9	725.85	0.03	33826	4007.21	dOccup=1, dIncome6=0, iRemplpar=0
10	725.47	0.03	33910	3994.68	dOccup=1, dIncome6=0
11	724.63	0.03	33654	4010.90	dOccup=1, dIncome6=0, iRelat2=0, iRemplpar=0
12	724.25	0.03	33737	3998.45	dOccup=1, dIncome6=0, iRelat2=0
13	723.95	0.03	32699	4101.88	dOccup=1, dPoverty=2, iRemplpar=0
14	723.95	0.03	32699	4101.88	dOccup=1, dPoverty=2, iRemplpar=0
15	723.60	0.03	32777	4089.62	dOccup=1, dPoverty=2
16	723.60	0.03	32777	4089.62	dOccup=1, dPoverty=2
17	722.71	0.03	32531	4105.72	dOccup=1, dPoverty=2, iRelat2=0, iRemplpar=0
18	722.71	0.03	32531	4105.72	dOccup=1, dPoverty=2, iRelat2=0, iRemplpar=0
19	722.37	0.03	32608	4093.55	dOccup=1, dPoverty=2, iRelat2=0
20	722.37	0.03	32608	4093.55	dOccup=1, dPoverty=2, iRelat2=0

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iDisabl2=2, iRemplpar=0
2	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iDisabl2=2, iRemplpar=0
3	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRemplpar=0
4	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRemplpar=0, iFertil=0
5	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRemplpar=0
6	91.89	0.08	5835	8442.88	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRemplpar=0, iFertil=0
7	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iRemplpar=0
8	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iRemplpar=0
9	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iRemplpar=0
10	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iRemplpar=0, iFertil=0
11	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iRemplpar=0
12	91.79	0.08	5916	8424.59	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iRemplpar=0, iFertil=0
13	91.77	0.08	5802	8422.38	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iDisabl2=2, iRelat2=0, iRemplpar=0
14	91.77	0.08	5802	8422.38	dOccup=1, dIndustry=9, iFertil=0, dPoverty=2, iDisabl2=2, iRelat2=0, iRemplpar=0
15	91.77	0.08	5802	8422.38	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRelat2=0, iRemplpar=0
16	91.77	0.08	5802	8422.38	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iRelat2=0, iRemplpar=0
17	91.71	0.08	5856	8409.90	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2
18	91.71	0.08	5856	8409.90	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iFertil=0
19	91.71	0.08	5856	8409.90	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2
20	91.71	0.08	5856	8409.90	dOccup=1, dIndustry=9, iSex=0, dPoverty=2, iDisabl2=2, iFertil=0

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	731.37	0.03	34153	4035.66	iRemplpar=0, dOccup=1
2	730.99	0.03	34237	4023.17	dOccup=1
3	730.16	0.03	33979	4039.57	iRemplpar=0, iRelat2=0, dOccup=1
4	729.78	0.03	34062	4027.16	iRelat2=0, dOccup=1
5	727.56	0.03	33669	4041.91	iRemplpar=0, iDisabl2=2, dOccup=1
6	727.18	0.03	33753	4029.23	iDisabl2=2, dOccup=1
7	726.34	0.03	33497	4045.75	iRemplpar=0, iRelat2=0, iDisabl2=2, dOccup=1
8	725.96	0.03	33580	4033.16	iRelat2=0, iDisabl2=2, dOccup=1
9	725.85	0.03	33826	4007.21	iRemplpar=0, dIncome6=0, dOccup=1
10	725.47	0.03	33910	3994.68	dIncome6=0, dOccup=1
11	724.63	0.03	33654	4010.90	iRemplpar=0, iRelat2=0, dIncome6=0, dOccup=1
12	724.25	0.03	33737	3998.45	iRelat2=0, dIncome6=0, dOccup=1
13	723.95	0.03	32699	4101.88	iRemplpar=0, dPoverty=2, dOccup=1
14	723.95	0.03	32699	4101.88	iRemplpar=0, dPoverty=2, dOccup=1
15	723.60	0.03	32777	4089.62	dPoverty=2, dOccup=1
16	723.60	0.03	32777	4089.62	dPoverty=2, dOccup=1
17	722.71	0.03	32531	4105.72	iRemplpar=0, iRelat2=0, dPoverty=2, dOccup=1
18	722.71	0.03	32531	4105.72	iRemplpar=0, iRelat2=0, dPoverty=2, dOccup=1
19	722.37	0.03	32608	4093.55	iRelat2=0, dPoverty=2, dOccup=1
20	722.37	0.03	32608	4093.55	iRelat2=0, dPoverty=2, dOccup=1

(c) qf = `ps`, method = `bsd`.

Figura 10: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población con doctorado.

estas personas son hombres (`iFertil = 0`, `iSex = 0`) y que pertenecen a la industria de servicios profesionales (`dIndustry = 9`), donde se encuentran la medicina, la ingeniería, los museos de arte y las galerías, y la investigación (entre muchos otros).

La precisión de los subgrupos no es muy elevada; sin embargo, hay una mejora al emplear la función de calidad de `ares`: se obtiene un 0.08 de precisión, mientras que con las otras dos ejecuciones se ha obtenido una precisión de 0.03. Destacar, además, que la ejecución con `ares` crea un subgrupo mucho más pequeño que las otras ejecuciones (5.853 y 34.153 elementos, respectivamente).

Por lo tanto, los subgrupos realizados contienen hombres que viven solos y que tienen doctorados científicos mayoritariamente (ya sean de ingeniería, medicina u otra rama de la ciencia) o, en todo caso, doctorado en bellas artes.

II-I. Personas que no están en un cargo militar

Los siguientes subgrupos que se pretenden encontrar son aquellos que contengan **el sector poblacional que no están actualmente en un cargo militar**. Para ello se ha seleccionado como variable objetivo `iMilitary = 4`.

En la Figura 11 se muestran los resultados obtenidos para las 3 ejecuciones realizadas. Como en los anteriores casos, los resultados obtenidos son muy similares excepto los de la ejecución con la función de calidad `ares`.

Con una precisión de 0.83, las personas que no están en un cargo militar están actualmente trabajando (`iYearwrk = 1`) y sus ingresos vienen dados por:

- La Seguridad Social (`dIncome = 3`).
- El trabajo en el campo (`dIncome = 5`).
- Asistencia pública (`dIncome = 6`).
- Pensiones de jubilación o de discapacidad (`dIncome = 7`).

.

La información obtenida por estos subgrupos no es muy interesante, ya que no se aprecia ninguna diferencia respecto a las personas que tardan más de una hora en alcanzar su trabajo.

Sin embargo, si se analizan los resultados proporcionados por la función de calidad de `ares` (Figura 11b), se muestra que las personas pertenecientes a este subgrupo son mujeres casadas (`iMarital = 0 + iSex = 1`) pero que no se encuentran con su pareja en este instante (con una precisión de 0.98). Además, también presentan como fuente de ingresos aquellas ofrecidas por la administración de veteranos, pensiones alimentarias e ingresos por ser familia de una persona dentro con un rango militar (`dIncome = 7`).

Por lo tanto, se puede suponer que las personas de este grupo son mayoritariamente mujeres casadas cuyos maridos trabajan o residen en otro lugar y probablemente éstos trabajen en un cargo militar, explicando las ayudas económicas previamente comentadas.

II-J. Población de origen americano

Por último, se realizará un estudio del **sector poblacional procediente de América** (los que se identifican como americanos). En este sector se encuentran todas las personas con raíces americanas (caucásicos, afro-americanos), canadienses y nativas americanas. Para ello se ha seleccionado como variable objetivo `dAncstry1 = 11`.

	quality	p	size	chi2	description
1	22229.68	0.83	116423	34832.25	iYearwrk=1, dIncome7=0
2	22073.03	0.83	114267	34417.62	iYearwrk=1, dIncome7=0, dIncome3=0
3	21856.89	0.83	112879	33805.18	iYearwrk=1, dIncome5=0, dIncome7=0
4	21829.86	0.83	114915	33639.34	iYearwrk=1, dIncome7=0, dIncome6=0
5	21699.58	0.83	110941	33415.29	iYearwrk=1, dIncome5=0, dIncome7=0, dIncome3=0
6	21671.77	0.83	112777	33239.52	iYearwrk=1, dIncome7=0, dIncome3=0, dIncome6=0
7	21584.36	0.82	120192	32763.75	iYearwrk=1
8	21535.18	0.83	111124	32901.32	iYearwrk=1, dIncome8=0, dIncome7=0
9	21487.29	0.83	111487	32736.72	iYearwrk=1, dIncome5=0, dIncome7=0, dIncome6=0
10	21462.94	0.82	117891	32434.45	iYearwrk=1, dIncome3=0
11	21370.83	0.83	109054	32516.39	iYearwrk=1, dIncome8=0, dIncome7=0, dIncome3=0
12	21352.58	0.82	115483	32165.71	iYearwrk=1, dIncome5=0
13	21329.34	0.83	109561	32360.34	iYearwrk=1, dIncome5=0, dIncome7=0, dIncome3=0, dIncome6=0
14	21216.25	0.82	113468	31828.18	iYearwrk=1, dIncome5=0, dIncome3=0
15	21181.52	0.81	118618	31575.29	iYearwrk=1, dIncome6=0
16	21174.05	0.83	109791	31877.84	iYearwrk=1, dIncome8=0, dIncome7=0, dIncome6=0
17	21147.26	0.83	107772	31919.18	iYearwrk=1, dIncome8=0, dIncome5=0, dIncome7=0
18	21058.39	0.82	116348	31260.32	iYearwrk=1, dIncome3=0, dIncome6=0
19	21009.52	0.83	107737	31506.99	iYearwrk=1, dIncome8=0, dIncome7=0, dIncome3=0, dIncome6=0
20	20995.67	0.82	114560	31129.51	iYearwrk=1, dIncome8=0

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	191.65	0.98	54519	36730.91	iMarital=0, iSex=1
2	190.97	0.98	54215	36469.19	iMarital=0, iSex=1, dIncome3=0
3	190.76	0.98	54703	36388.25	iYearwrk=1, iSex=1
4	189.93	0.98	54328	36072.93	iYearwrk=1, iSex=1, dIncome3=0
5	189.05	0.98	53370	35738.78	iMarital=0, iSex=1, dIncome6=0
6	188.47	0.98	53699	35522.24	iYearwrk=1, iSex=1, dIncome6=0
7	188.40	0.99	52912	35493.35	iRspouse=1, iSex=1
8	188.40	0.99	52912	35493.35	iRspouse=1, iSex=1, iMarital=0
9	188.39	0.98	53076	35489.31	iMarital=0, iSex=1, dIncome6=0, dIncome3=0
10	188.36	0.99	52900	35478.56	iMarital=0, iSex=1, dIncome8=0
11	188.33	0.98	53565	35468.50	iYearwrk=1, iSex=1, dIncome7=0
12	187.72	0.99	52610	35237.58	iRspouse=1, iSex=1, dIncome3=0
13	187.72	0.99	52610	35237.58	iRspouse=1, iSex=1, dIncome3=0, iMarital=0
14	187.71	0.99	52610	35233.74	iMarital=0, iSex=1, dIncome8=0, dIncome3=0
15	187.66	0.98	53330	35215.71	iYearwrk=1, iSex=1, dIncome6=0, dIncome3=0
16	187.55	0.98	53216	35176.70	iYearwrk=1, iSex=1, dIncome7=0, dIncome3=0
17	186.63	0.83	116423	34832.25	iYearwrk=1, dIncome7=0
18	186.43	0.99	52057	34755.40	iMarital=0, iSex=1, dIncome7=0
19	186.11	0.98	52590	34638.61	iYearwrk=1, iSex=1, dIncome7=0, dIncome6=0
20	186.10	0.99	51915	34634.77	iRspouse=1, iSex=1, dIncome6=0

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	22229.68	0.83	116423	34832.25	dIncome7=0, iYearwrk=1
2	22073.03	0.83	114267	34417.62	dIncome3=0, dIncome7=0, iYearwrk=1
3	21856.89	0.83	112879	33805.18	dIncome7=0, dIncome5=0, iYearwrk=1
4	21829.86	0.83	114915	33639.34	dIncome6=0, dIncome7=0, iYearwrk=1
5	21699.58	0.83	110941	33415.29	dIncome3=0, dIncome7=0, dIncome5=0, iYearwrk=1
6	21671.77	0.83	112777	33239.52	dIncome3=0, dIncome6=0, dIncome7=0, iYearwrk=1
7	21584.36	0.82	120192	32763.75	iYearwrk=1
8	21535.18	0.83	111124	32901.32	dIncome8=0, dIncome7=0, iYearwrk=1
9	21487.29	0.83	111487	32736.72	dIncome6=0, dIncome7=0, dIncome5=0, iYearwrk=1
10	21462.94	0.82	117891	32434.45	dIncome3=0, iYearwrk=1
11	21370.83	0.83	109054	32516.39	dIncome3=0, dIncome8=0, dIncome7=0, iYearwrk=1
12	21352.58	0.82	115483	32165.71	dIncome5=0, iYearwrk=1
13	21329.34	0.83	109561	32360.34	dIncome3=0, dIncome6=0, dIncome7=0, dIncome5=0, iYearwrk=1
14	21216.25	0.82	113468	31828.18	dIncome3=0, dIncome5=0, iYearwrk=1
15	21181.52	0.81	118618	31575.29	dIncome6=0, iYearwrk=1
16	21174.05	0.83	109791	31877.84	dIncome8=0, dIncome6=0, dIncome7=0, iYearwrk=1
17	21147.26	0.83	107772	31919.18	dIncome8=0, dIncome7=0, dIncome5=0, iYearwrk=1
18	21058.39	0.82	116348	31260.32	dIncome3=0, dIncome6=0, iYearwrk=1
19	21009.52	0.83	107737	31506.99	dIncome3=0, dIncome8=0, dIncome6=0, dIncome7=0, iYearwrk=1
20	20995.67	0.82	114560	31129.51	dIncome8=0, iYearwrk=1

(c) qf = `ps`, method = `bsd`.

Figura 11: Las 20 mejores reglas obtenidas en cada ejecución que describen el sector de la población que no están en un cargo militar.

	quality	p	size	chi2	description
1	4263.29	0.33	31245	4286.07	dPoverty=1
2	4019.91	0.39	20783	5462.62	dPoverty=1, iRPOB=10
3	3915.44	0.23	92709	1707.48	dDepart=0, iRPOB=10
4	3850.17	0.35	25149	4223.02	dPoverty=1, dDepart=0
5	3721.05	0.22	155256	1556.80	iRPOB=10
6	3671.88	0.24	79532	1611.75	dRearning=0, iRPOB=10
7	3566.44	0.24	77805	1538.29	dRearning=0, iRPOB=10, dDepart=0
8	3551.88	0.35	22824	3918.85	dPoverty=1, dRearning=0
9	3535.21	0.40	17380	4976.67	dPoverty=1, iRPOB=10, dDepart=0
10	3525.50	0.24	75634	1526.60	iMarital=4, iRPOB=10
11	3496.51	0.35	21725	3970.16	dPoverty=1, dRearning=0, dDepart=0
12	3416.56	0.25	59125	1671.87	iMarital=4, iRPOB=10, dDepart=0
13	3288.31	0.40	15941	4665.11	dPoverty=1, iRPOB=10, dRearning=0
14	3205.27	0.40	15344	4593.00	dPoverty=1, iRPOB=10, dRearning=0, dDepart=0
15	3192.28	0.25	54271	1549.82	iMarital=4, dRearning=0, iRPOB=10
16	3174.95	0.23	75990	1234.89	iMarital=4, dDepart=0
17	3161.77	0.24	70720	1276.31	dIndustry=0, iRPOB=10, dRearning=0
18	3161.77	0.24	70720	1276.31	iClass=0, iRPOB=10, dDepart=0, dIndustry=0, dOccup=0, dRearning=0
19	3161.77	0.24	70720	1276.31	iClass=0, iRPOB=10, dDepart=0, dIndustry=0, dRearning=0
20	3161.77	0.24	70720	1276.31	iClass=0, iRPOB=10, dDepart=0, dIndustry=0, dOccup=0, dRearning=0

(a) qf = `ps`, method = `sdmap`.

	quality	p	size	chi2	description
1	73.91	0.39	20783	5462.62	dPoverty=1, iRPOB=10
2	70.55	0.40	17380	4976.67	dPoverty=1, iRPOB=10, dDepart=0
3	68.30	0.40	15941	4665.11	dPoverty=1, iRPOB=10, dRearning=0
4	67.77	0.40	15344	4593.00	dPoverty=1, iRPOB=10, dRearning=0, dDepart=0
5	65.47	0.33	31245	4286.07	dPoverty=1
6	64.98	0.35	25149	4223.02	dPoverty=1, dDepart=0
7	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10
8	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0
9	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0, dIndustry=0
10	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0, dIndustry=0, dRearning=0
11	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0, dIndustry=0
12	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0, dIndustry=0, dRearning=0
13	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dDepart=0, dRearning=0
14	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dIndustry=0
15	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dIndustry=0, dRearning=0
16	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dIndustry=0
17	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dIndustry=0, dRearning=0
18	63.88	0.40	13627	4081.11	dPoverty=1, dOccup=0, iRPOB=10, dRearning=0
19	63.88	0.40	13627	4081.11	dPoverty=1, dIndustry=0, iRPOB=10
20	63.88	0.40	13627	4081.11	dPoverty=1, dIndustry=0, iRPOB=10, dDepart=0

(b) qf = `ares`, method = `sdmap`.

	quality	p	size	chi2	description
1	4263.29	0.33	31245	4286.07	dPoverty=1
2	4019.91	0.39	20783	5462.62	iRPOB=10, dPoverty=1
3	3915.44	0.23	92709	1707.48	iRPOB=10, dDepart=0
4	3850.17	0.35	25149	4223.02	dDepart=0, dPoverty=1
5	3721.05	0.22	155256	1556.80	iRPOB=10
6	3671.88	0.24	79532	1611.75	iRPOB=10, dRearning=0
7	3566.44	0.24	77805	1538.29	iRPOB=10, dDepart=0, dRearning=0
8	3551.88	0.35	22824	3918.85	dRearning=0, dPoverty=1
9	3535.21	0.40	17380	4976.67	iRPOB=10, dDepart=0, dPoverty=1
10	3525.50	0.24	75634	1526.60	iRPOB=10, iMarital=4
11	3496.51	0.35	21725	3970.16	dDepart=0, dRearning=0, dPoverty=1
12	3416.56	0.25	59125	1671.87	iRPOB=10, dDepart=0, iMarital=4
13	3288.31	0.40	15941	4665.11	iRPOB=10, dRearning=0, dPoverty=1
14	3205.27	0.40	15344	4593.00	iRPOB=10, dDepart=0, dRearning=0, dPoverty=1
15	3192.28	0.25	54271	1549.82	iRPOB=10, dRearning=0, iMarital=4
16	3174.95	0.23	75990	1234.89	dDepart=0, iMarital=4
17	3161.77	0.24	70720	1276.31	iRPOB=10, dDepart=0, dRearning=0, dIndustry=0
18	3161.77	0.24	70720	1276.31	iRPOB=10, dDepart=0, dRearning=0, dIndustry=0, iClass=0
19	3161.77	0.24	70720	1276.31	iRPOB=10, dDepart=0, dRearning=0, dIndustry=0, dOccup=0, iClass=0
20	3161.77	0.24	70720	1276.31	iRPOB=10, dDepart=0, dRearning=0, dIndustry=0, dOccup=0

(c) qf = `ps`, method = `bsd`.

Figura 12: Las 20 mejores reglas obtenidas en cada ejecución que describe el sector de la población que se identifica de origen americano.

En la Figura 12 se recogen los resultados obtenidos de las 3 ejecuciones realizadas. Para empezar, es sorprendente que la regla más importante indique que este subgrupo se encuentra por debajo del umbral de pobreza (`dPoverty = 1`). Además, se muestra que éstas personas no están trabajando (`dDepart = 0`) y que no tienen ningún ingreso (`dRearning = 0`). Otros datos que se indican es que estas personas residen en el mismo lugar de nacimiento (`iRPOB = 0`), que no han estado casados nunca (`iMarital = 4`) y que son personas menores a 16 años o que no han trabajado desde 1985 (`dIndustry = 0`).

Por lo tanto, se puede determinar que en estos subgrupos se han encontrado a los menores de 16 años que, además, se encuentran en una situación económica desfavorable.

III. Anexo: Atributos empleados

III-A. *Personas en búsqueda activa de empleo*

```
DiscoverSubgroups(data_sample, as.target("iLooking", "1"),
                  new("SDTaskConfig",
                      attributes=c("dAge",
```

```
                                        "dAncstry1",
                                        "dAncstry2",
                                        "iCitizen",
                                        "iClass",
                                        "iImmigr",
                                        "dPOB",
                                        "dPoverty",
                                        "iWork89",
                                        "iSex",
                                        "iYearsch",
                                        "iYearwrk",
                                        "iEnglish",
                                        "dHispanic")
    ))
```

III-B. Personas que trabajan más de 50 horas semanales

```
DiscoverSubgroups(data_sample, as.target("dHours", "5"),
                            new("SDTaskConfig",
                                attributes=c("iCitizen",
                                             "iClass",
                                             "dIndustry",
                                             "dHispanic",
                                             "dOccup",
                                             "iImmigr",
                                             "iFertil",
                                             "dPOB",
                                             "iYearsch"))
    )
```

III-C. Personas sin estudios secundarios

```
DiscoverSubgroups(data_sample, as.target("iSchool", "1"),
                            new("SDTaskConfig",
                                attributes=c("dAncstry1",
                                             "dAncstry2",
                                             "iClass",
                                             "iLang1",
                                             "dPoverty",
                                             "iCitizen",
                                             "iDisabl1",
                                             "iDisabl2",
                                             "iEnglish",
                                             "iImmigr",
                                             "dIndustry",
                                             "dOccup",
                                             "dPoverty",
                                             "dPOB",
                                             "iSex",
                                             "dIncome1",
                                             "dIncome2",
                                             "dIncome3",
                                             "dIncome4",
                                             "dIncome5",
                                             "dIncome6",
                                             "dIncome7",
                                             "dIncome8",
                                             "iFertil",
                                             "dAge")))
```

III-D. Personas mayores de 65 años

```
DiscoverSubgroups(data_sample, as.target("dAge", "7"),
                            new("SDTaskConfig",
                                attributes=c("iSex",
                                             "iCitizen",
                                             "dYrsserv",
                                             "dIndustry",
                                             "iYearsch",
                                             "iYearwrk",
                                             "dIncome1",
                                             "dIncome2",
                                             "dIncome3",
                                             "dIncome4",
                                             "dIncome5",
                                             "dIncome6",
                                             "dIncome7",
                                             "dIncome8",
                                             "iFertil",
                                             "dPOB",
                                             "iEnglish",
                                             "iMarital",
                                             "dHispanic",
                                             "iImmigr",
                                             "iCitizen",
                                             "iClass")))


DiscoverSubgroups(data_sample, as.target("dAge", "7"),
                            new("SDTaskConfig",
                                attributes=c("iSex",
                                             "dAncstry1",
                                             "dAncstry2",
                                             "iCitizen",
                                             "dYrsserv",
                                             "dIndustry",
                                             "iYearsch",
                                             "iYearwrk",
                                             "iFertil",
                                             "dPOB",
                                             "iEnglish",
                                             "iMarital",
                                             "dHispanic",
                                             "iImmigr",
                                             "iCitizen",
                                             "iClass")))


DiscoverSubgroups(data_sample, as.target("dAge", "7"),
                            new("SDTaskConfig",
                                attributes=c("iSex",
                                             "dAncstry1",
                                             "dAncstry2",
                                             "iCitizen",
                                             "dYrsserv",
                                             "dIndustry",
                                             "iYearsch",
                                             "iFertil",
                                             "dPOB",
                                             "iEnglish",
                                             "iMarital",
                                             "dHispanic",
                                             "iImmigr",
                                             "iCitizen",
                                             "iClass",
                                             "iFertil")))
```

III-E. Personas que tardan más de 1h en llegar al trabajo

```
DiscoverSubgroups(data_sample, as.target("dTravtime", "6"),
                            new("SDTaskConfig",
                                attributes=c("dAncstry1",
                                             "dAncstry2",
                                             "iClass",
                                             "iLang1",
                                             "dPoverty",
                                             "iCitizen",
                                             "iDisabl1",
                                             "iDisabl2",
                                             "iEnglish",
                                             "iImmigr",
                                             "dIndustry",
                                             "dOccup",
                                             "dPoverty",
                                             "dPOB",
                                             "iSex",
                                             "dIncome1",
                                             "dIncome2",
                                             "dIncome3",
                                             "dIncome4",
                                             "dIncome5",
                                             "dIncome6",
                                             "dIncome7",
                                             "dIncome8",
                                             "iFertil",
                                             "iMilitary")
    ))
```

III-F. Personas que tienen un doctorado

```
DiscoverSubgroups(data_sample, as.target("iYearsch", "17"),
                            new("SDTaskConfig",
                                attributes=c("dAncstry1",
```

```
                                        "dAncstry2",
                                        "iClass",
                                        "dDepart",
                                        "iLang1",
                                        "dPoverty",
                                        "iCitizen",
                                        "iDisabl1",
                                        "iDisabl2",
                                        "iEnglish",
                                        "iImmigr",
                                        "dIndustry",
                                        "dOccup",
                                        "iRemplpar",
                                        "dPoverty",
                                        "dPOB",
                                        "iRvetserv",
                                        "iSex",
                                        "dIncome1",
                                        "dIncome2",
                                        "dIncome3",
                                        "dIncome4",
                                        "dIncome5",
                                        "dIncome6",
                                        "dIncome7",
                                        "dIncome8",
                                        "iFertil",
                                        "iMilitary",
                                        "iRelat1",
                                        "iRelat2")))
```

III-G. *Personas que no están en un cargo militar*

```
DiscoverSubgroups(data_sample, as.target("iMilitary", "4"),
                              new("SDTaskConfig",
                                  attributes=c("dAge",
                                        "iCitizen",
                                        "iClass",
                                        "dDepart",
                                        "iImmigr",
                                        "dIndustry",
                                        "iMarital",
                                        "dIncome1",
                                        "dIncome2",
                                        "dIncome3",
                                        "dIncome4",
                                        "dIncome5",
                                        "dIncome6",
                                        "dIncome7",
                                        "dIncome8",
                                        "dOccup",
                                        "iRPOB",
                                        "dPoverty",
                                        "dRearning",
                                        "iRspouse",
                                        "iTmpabsnt",
                                        "iSex",
                                        "iYearsch",
                                        "iYearwrk"))
    )
```

III-H. *Población de origen americano*

```
DiscoverSubgroups(data_sample, as.target("dAncstry1", "11")
    ,
                              new("SDTaskConfig",
                                  attributes=c("dAge",
                                        "iClass",
                                        "dDepart",
                                        "dIndustry",
                                        "iMarital",
                                        "iMobility",
                                        "dIndustry",
                                        "dOccup",
                                        "iRemplpar",
                                        "iRPOB",
                                        "dPoverty",
                                        "dRearning",
                                        "iRspouse",
                                        "iTmpabsnt",
                                        "iSex",
                                        "iYearsch",
                                        "iFertil",
                                        "iYearwrk"))
    )
```

Referencias

[1] *Subgroup Discovery – Advanced Review*. https://tinyurl.com/wmvj648w. Accessed: 2021-03-28.

[2] *US Census Data (1990) Data Set*. https://tinyurl.com/y332ed2y. Accessed: 2021-03-28.

[3] *1990 Census of Population and Housing Technical Documentation*. https://tinyurl.com/29hdtr24. Accessed: 2021-03-28.

[4] *Friday the 13th mini-crash*. https://tinyurl.com/3bxzwffs. Accessed: 2021-03-28.

[5] *U.S. Immigrant Population and Share over Time, 1850-Present*. https://tinyurl.com/5xr3n7uc. Accessed: 2021-03-28.

[6] *Changes in men's and women's labor force participation rates*. https://tinyurl.com/4vbcch27. Accessed: 2021-03-28.

[7] *Commuting in America III - The Third National Report on Comuting Patterns and Trends*. https://tinyurl.com/3s37y247. Accessed: 2021-03-28.