

;



Trabajo Práctico Final

March 29, 2024

Introducción a la Investigación Operativa y Optimización

Integrante	LU	Correo electrónico
Calloni, Sol Anabella		
Duran, Valentina Camila		
Montserrat, Luz		



Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

<http://www.exactas.uba.ar>

1 Introducción

El trabajo a continuación consiste en una implementación del siguiente paper *Guillermo Durán, Rodrigo Wolf-Yadlin. (2011, Mayo-Junio) A Mathematical Programming Approach to Applicant Selection for a Degree Program Based on Affirmative Action. Interfaces*. El mismo describe el uso de modelos de programación lineal matemática en el procedimiento de selección de solicitantes de un programa de maestría en gestión de la globalización de la Universidad de Chile para las tres primeras clases (2007, 2008 y 2009), sujetos a criterios de equidad en cuanto a género, origen regional y situación socioeconómica. En nuestro caso, nos enfocamos en imitar lo propuesto en el año 2008, tanto los tres modelos presentados, los porcentajes y el algoritmo de selección, generando además una propuesta alternativa.

2 Datos utilizados

Para conseguir los datos que necesitamos para el trabajo, generamos tres instancias de 100 estudiantes con sus puntajes que van del 1 al 100 y sus características: mujer u hombre, clase alta o no clase alta, de Santiago o del interior. Además, para la nueva propuesta generamos una nueva característica donde aclaramos la edad del participante, con edades que van desde los 22 a los 65 años. La forma en la que generamos estos datos fue por medio de un script de python utilizando los comandos `random.choice` y `random.randint`. Luego, todo esto lo pasamos a un dataset. Hicimos esto tres veces, de manera de tener 3 datasets con tres instancias de candidatos distintas. En la siguiente tabla mostramos cuánta gente de cada grupo tenemos en cada dataset.

	Primer Dataset	Segundo Dataset	Tercer Dataset
Mujeres	49	54	46
No Clase Alta	49	48	53
No de Santiago	47	44	46
Mujeres y No Clase Alta	17	27	23
Mujeres y No de Santiago	27	19	19
No Clase Alta y No de Santiago	20	21	28
Mujeres, No Clase Alta y No de Santiago	9	10	11

Table 1: Resumen de Resultados por Dataset y Categoría

3 Modelos utilizados

3.1 Notación de los conjuntos

- N es el número de personas a admitir. En nuestro caso, son 50.
- K es el conjunto de solicitantes preseleccionados.
- M es el conjunto de todas las solicitantes mujeres.
- R es el conjunto de todos los solicitantes de regiones no pertenecientes a Santiago.
- Q es el conjunto de solicitantes pertenecientes al cuantil de ingresos más bajo.

3.2 Variables y parámetros

- p_i = es la puntuación del solicitante i .

Sin pérdida de generalidad, asumimos que las puntuaciones están ordenadas de mayor a menor.

- $x_i = \begin{cases} 1 & \text{si el solicitante } i \text{ es seleccionado} \\ 0 & \text{si no} \end{cases}$

- y = la posición relativa mayor o igual a los solicitantes seleccionados. Lo utilizaremos para el tercer modelo.

3.3 Funciones objetivo

Utilizamos tres modelos diferentes para poder resolver este problema.

1. Buscamos maximizar la suma de las puntuaciones asignadas a los solicitantes seleccionados

$$FO : \max \sum_{i \in K} x_i \cdot p_i \quad (1)$$

2. Buscamos minimizar la suma de sus clasificaciones

$$FO : \min \sum_{i \in K} x_i \cdot i \quad (2)$$

3. Buscamos minimizar la clasificación del último candidato seleccionado.

$$FO : \min(y + 0.000002) \cdot \sum_{i \in K} x_i \cdot i \quad (3)$$

3.4 Restricciones

Como aclaramos anteriormente, tomamos la decisión de utilizar las restricciones de porcentajes dadas en 2008, las cuales eran, por lo menos un 30% de mujeres, un 55% gente no de Santiago y un 70% pertenecientes a los cuatro cuantiles de ingresos más bajos.

1. El número total de solicitantes a seleccionar está predeterminado por los organizadores del programa:

$$\sum_{i \in K} x_i = N \quad (4)$$

2. Al menos un porcentaje del 30 de los solicitantes seleccionados deben ser mujeres:

$$\sum_{i \in M} x_i \geq \frac{30}{100} \cdot N \quad (5)$$

3. Al menos un porcentaje del 55 de los solicitantes seleccionados deben ser de regiones no pertenecientes a Santiago:

$$\sum_{i \in R} x_i \geq \frac{55}{100} \cdot N \quad (6)$$

4. Al menos un porcentaje del 70 de los solicitantes seleccionados deben pertenecer a los cuatro cuantiles de ingresos más bajos:

$$\sum_{i \in Q} x_i \geq \frac{70}{100} \cdot N \quad (7)$$

5. Para el tercer modelo agregamos la siguiente restricción:

$$i \cdot x_i \geq y \forall i \quad (8)$$

4 Procedimiento del Algoritmo de Selección

El proceso de elección de los 50 candidatos y armado de lista de espera consistió en la implementación en python del Algoritmo de Selección propuesto por el paper. Para cada uno de los tres dataset, lo primero que hicimos fue ejecutar el primer paso del algoritmo. Este consistió en correr tres veces cada modelo para obtener tres corridas distintas de cada uno, una primera ejecución donde obtuvimos la mejor solución, la segunda mejor solución agregando una restricción a los modelos para hacer que la mejor solución sea inviable; derivamos la tercera mejor solución eliminando de manera similar la segunda mejor. Luego, verificamos qué candidatos aparecían en las primeras corridas de cada uno de los tres modelos simultáneamente, estos fueron admitidos inmediatamente en el programa.

Para el segundo paso del algoritmo calculamos las nuevas puntuaciones para cada solicitante no seleccionado, esto lo hicimos utilizando las tres mejores soluciones de cada modelo. Si existían soluciones únicas mejores, segundas mejores y terceras mejores, asignamos a los solicitantes en cada una los coeficientes 1, 0.6 y 0.3, respectivamente. Luego, sumamos estos valores en las tres soluciones y modelos para cada solicitante. Finalmente, multiplicamos este valor por el total de puntos del solicitante para determinar una nueva puntuación y ranking, ambos almacenados en un nuevo dataset como nuevas columnas.

Para el siguiente paso, tomamos la lista que ya teníamos de admitidos hasta el momento para determinar cuántos más de cada categoría se necesitaban para cumplir con los mínimos porcentuales requeridos. A continuación, se ejecutó el Modelo 2 utilizando las nuevas puntuaciones con restricciones que aseguraban, primero, que seleccione al menos el número de candidatos requeridos para cumplir con estos mínimos, y segundo, que aquellos seleccionados igualen el número faltante en la primera selección para satisfacer N, el total del programa. El algoritmo verifica si la solución óptima es única. Si lo es, se seleccionan a los solicitantes en la solución, completando así la lista de admisiones. En caso contrario, se debería pasar al paso 4, el cual consiste en calcular la suma de las puntuaciones de cada solución encontrada en el Paso 3 (es decir, se aplica el Modelo 1), y el grupo con el mayor total de puntos completa la lista de solicitantes admitidos. Si dos o más soluciones proporcionan la misma puntuación, se presentan todas las alternativas a los organizadores del programa para una decisión final. Sin embargo, en nuestro caso con los tres conjuntos de datos obtuvimos una solución óptima única en el tercer paso, con lo cual, no fue necesaria la implementación del Paso 4.

Para finalizar, pasamos al Paso 5 donde armamos la lista de espera. Si el número de solicitantes que aparecen en cualquiera de las nueve ejecuciones pero no son admitidos es mayor a 20, los mejores puntuadores (después de la ponderación) de este grupo se colocan en la lista de espera. Si el número es menor a 20, todos estos solicitantes se colocan en la lista de espera y los solicitantes adicionales necesarios para completarla son elegidos entre los mejores puntuados (antes de la ponderación) entre aquellos que no fueron seleccionados en las mejores soluciones de cualquier modelo.

5 Primeros resultados

5.1 Primer conjunto de datos

Para el primer conjunto de datos, en la primera etapa conseguimos 43 personas que aparecían en las soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes pedidos debíamos elegir 7 estudiantes. Veamos qué nos faltaban con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -4.0. Teníamos 4 mujeres más de las requeridas, por lo que para la siguiente corrida del Modelo 2 no hizo falta restricción para las mujeres.
- Cuantil bajo faltantes: 4.0
- No santiago faltantes: 2.5
- Personas faltantes: 7. Para la nueva corrida cambiamos el valor de N a 7.

Para asegurarnos de satisfacer la cantidad de gente que no es de Santiago y de gente pertenecientes a los cuatro cuantiles de ingresos más bajo modificamos los valores de estas restricciones en el segundo modelo para que elija la cantidad necesaria:

$$\sum_{i \in Q} x_i \geq \frac{4 \cdot 100}{7} = 57.14 \quad (9)$$

$$\sum_{i \in R} x_i \geq \frac{2.5 \cdot 100}{7} = 35.71 \quad (10)$$

Luego, corrimos el Modelo 2 completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dio un valor con función objetivo de 34 y la siguiente fue 35. Obteniendo la siguiente lista final.

Table 2: Lista definitiva de admitidos

Ranking	Puntaje	Género	Clase	Región
1	99	M	SI	SI

Continúa en la siguiente página

Table 2 – Continuación de la tabla

Ranking	Puntaje	Género	Clase	Región
2	94	M	SI	NO
3	93	H	NO	NO
4	92	M	SI	SI
5	91	H	SI	NO
7	90	M	SI	NO
8	90	H	NO	SI
9	89	M	SI	NO
10	88	M	SI	NO
11	88	M	NO	SI
12	87	M	SI	NO
13	87	M	NO	SI
14	87	H	NO	SI
15	86	H	NO	SI
16	86	M	SI	NO
17	84	M	SI	NO
18	84	H	NO	NO
19	81	H	NO	SI
20	80	H	NO	SI
21	79	H	NO	NO
24	77	M	NO	NO
25	77	H	SI	NO
27	75	H	NO	SI
28	75	H	NO	NO
29	73	H	NO	SI
30	70	M	NO	NO
31	67	H	NO	SI
33	65	M	NO	SI
34	64	H	NO	SI
35	62	M	NO	SI
36	62	M	SI	NO
37	62	M	NO	NO
39	60	H	NO	SI
40	58	H	NO	NO
41	58	M	SI	NO
42	53	M	SI	NO
43	53	H	NO	SI
44	52	H	SI	NO
45	52	H	NO	NO
48	51	M	NO	SI
49	51	H	NO	NO
50	50	M	NO	NO
54	48	H	NO	NO
58	43	H	NO	NO
59	43	H	NO	SI
60	41	H	NO	SI
62	38	H	NO	SI
63	38	H	NO	SI
70	33	M	NO	NO
85	23	M	NO	NO

Los porcentajes obtenidos de mujeres fue 46%, de personas que pertenecen a cuantiles bajos fue 70%, y de aquellos que no viven en Santiago fue 56%. Por lo tanto, podemos ver que nuestra lista final cumple con los porcentajes requeridos.

Por otro lado, conseguimos la siguiente lista de espera.

Table 3: Lista de espera

Ranking	Puntaje	Género	Clase	Región
6	90	M	SI	SI
47	52	M	SI	NO
65	37	M	NO	SI
66	36	H	NO	SI
90	14	M	NO	NO
52	48	M	SI	NO
55	48	M	SI	NO
56	48	H	SI	NO
91	8	H	NO	NO
57	44	M	SI	NO
61	41	H	SI	NO
62	38	H	NO	SI
63	38	H	NO	SI
64	38	M	SI	SI
65	37	M	NO	SI
66	36	H	NO	SI
67	35	H	SI	SI
68	34	H	NO	SI
69	34	H	SI	NO
70	33	M	NO	NO

En nuestro caso para el primer conjunto de datos, estábamos en la situación de tener menos de 20 solicitantes que aparecen en cualquiera de las nueve ejecuciones. Por lo tanto, colocamos a todos estos y elegimos a los adicionales entre los mejores puntuados (antes de la ponderación) de aquellos que no fueron seleccionados en las mejores soluciones.

5.2 Segundo conjunto de datos

Para el segundo conjunto de datos, en la primera etapa conseguimos 42 personas que aparecían en las soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes pedidos debíamos elegir 8 estudiantes. Veamos qué nos faltaba con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -9.0. Tenemos 9 mujeres más de las requeridas, por lo que para la siguiente corrida del modelo 2 no hace falta restricción para las mujeres.
- Cuantil bajo faltantes: 4.0
- No santiago faltantes: 3.5
- Personas faltantes: 8

Para asegurarnos de satisfacer la cantidad de gente que no es de Santiago y de gente perteneciente a los cuatro cuantiles de ingresos más bajo modificamos los valores de estas restricciones en el segundo modelo para que elija la cantidad necesaria:

$$\sum_{i \in Q} x_i \geq \frac{4 \cdot 100}{8} = 50 \quad (11)$$

$$\sum_{i \in R} x_i \geq \frac{3.5 \cdot 100}{8} = 43.75 \quad (12)$$

Luego, corrimos el Modelo 2 completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dió un valor con función objetivo de 43 y la siguiente fue 44. Obteniendo la siguiente lista final.

Table 4: Lista definitiva de admitidos

Ranking	Puntaje	Género	Clase	Región
1	99	H	NO	SI
2	98	M	NO	SI
3	98	M	SI	SI
4	97	M	SI	SI
5	96	M	NO	SI
6	95	M	SI	NO
7	95	H	SI	SI
8	94	M	SI	SI
9	93	M	NO	NO
10	92	H	NO	SI
11	91	H	NO	NO
12	89	H	SI	NO
13	88	H	SI	NO
14	85	M	NO	NO
15	85	H	SI	SI
17	84	M	SI	NO
18	84	M	NO	NO
20	83	M	NO	SI
21	81	M	NO	SI
22	80	H	NO	NO
24	75	M	SI	NO
25	75	H	NO	NO
27	74	M	SI	NO
28	74	M	NO	SI
29	72	M	SI	NO
30	72	M	NO	NO
31	72	H	NO	SI
32	70	H	NO	SI
34	69	H	NO	NO
35	68	H	NO	NO
36	66	H	NO	SI
40	63	M	NO	SI
42	62	M	NO	SI
43	61	H	NO	SI
44	61	H	SI	NO
45	60	H	SI	NO
46	59	M	SI	NO
47	59	H	NO	NO
49	56	H	NO	SI
53	54	M	NO	NO
55	51	M	NO	SI
57	49	M	NO	NO
59	47	M	NO	SI
60	47	H	NO	NO
61	45	M	NO	NO
63	44	M	NO	NO
64	44	M	NO	SI
69	41	H	NO	NO
74	36	H	NO	NO
78	32	H	NO	NO

Los porcentajes obtenidos de mujeres fue 54%, de personas que pertenecen a cuantiles bajos fue 70%, y de aquellos que no viven en Santiago fue 56%. Por lo tanto, podemos ver que nuestra lista final cumple con los porcentajes requeridos.

Por otro lado, conseguimos la siguiente lista de espera.

Table 5: Lista de espera

Ranking	Puntaje	Género	Clase	Región
16	85	M	SI	SI
51	56	H	SI	NO
50	56	H	SI	NO
58	48	H	SI	NO
65	43	H	NO	SI
68	43	M	NO	SI
66	43	M	NO	SI
88	19	M	NO	NO
52	55	H	SI	SI
54	51	M	SI	SI
56	50	H	SI	SI
58	48	H	SI	NO
62	44	M	SI	SI
64	44	M	NO	SI
65	43	H	NO	SI
66	43	M	NO	SI
67	43	M	SI	NO
68	43	M	NO	SI
69	41	H	NO	NO
70	40	H	SI	NO

Para el segundo conjunto de datos, estábamos en la situación de tener menos de 20 solicitantes que aparecen en cualquiera de las nueve ejecuciones. Por lo tanto, completamos de la misma forma que el anterior.

5.3 Tercer conjunto de datos

Para este conjunto de datos, en la primera etapa conseguimos 46 personas que aparecían en las soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes debíamos elegir 6 estudiantes. Veamos qué nos faltaba con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -1.0. Tenemos una mujer más de las requeridas, por lo que para la siguiente corrida del Modelo 2 no hace falta restricción para las mujeres.
- Cuantil bajo faltantes: 2.0
- No santiago faltantes: 1.5
- Personas faltantes: 4

Para asegurarnos de satisfacer la cantidad de gente que no es de Santiago y de gente perteneciente a los cuatro cuantiles de ingresos más bajo modificamos los valores de estas restricciones en el segundo modelo para que elija la cantidad necesaria:

$$\sum_{i \in Q} x_i \geq \frac{2 \cdot 100}{6} = 50 \quad (13)$$

$$\sum_{i \in R} x_i \geq \frac{1.5 \cdot 100}{6} = 37.5 \quad (14)$$

Luego, corrimos el Modelo 2 completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dio un valor con función objetivo de 12 y la siguiente fue 14. Obteniendo la siguiente lista final.

Table 6: Lista definitiva de admitidos

Ranking	Puntaje	Género	Clase	Región
1	100	H	SI	NO
2	100	H	SI	SI
3	99	M	SI	SI
4	98	H	SI	NO
5	97	M	NO	SI
6	96	M	SI	SI
7	94	H	SI	SI
8	94	H	SI	SI
9	93	H	NO	SI
10	93	M	NO	SI
11	92	M	SI	NO
12	91	H	NO	SI
13	90	H	NO	SI
14	88	M	NO	NO
15	88	M	NO	NO
16	88	H	NO	NO
17	87	M	NO	NO
18	86	H	SI	SI
19	85	H	NO	NO
20	85	H	NO	NO
21	85	M	NO	SI
22	84	M	SI	SI
23	83	M	SI	NO
24	83	M	SI	SI
25	81	M	NO	NO
26	79	H	NO	NO
27	79	H	SI	NO
28	78	H	NO	SI
30	75	H	SI	NO
31	75	H	NO	NO
32	74	H	NO	SI
34	73	M	NO	NO
35	73	H	NO	NO
36	73	M	SI	NO
37	71	H	NO	SI
38	71	H	NO	SI
39	70	M	NO	SI
40	70	H	NO	NO
41	68	M	NO	SI
42	68	H	NO	NO
43	66	H	NO	SI
46	64	H	NO	NO
48	63	H	NO	SI
50	61	H	NO	NO
52	61	M	NO	NO
53	60	H	NO	NO
55	54	H	NO	NO
57	53	H	NO	NO
61	45	H	NO	NO
63	42	M	NO	NO

Los porcentajes obtenidos de mujeres fue 38%, de personas que pertenecen a cuantiles bajos fue 70%, y de aquellos que no viven en Santiago fue 56%. Por lo tanto, podemos ver que nuestra lista final cumple con los porcentajes requeridos.

Por otro lado, conseguimos la siguiente lista de espera.

Table 7: Lista de espera

Ranking	Puntaje	Género	Clase	Región
47	64	M	SI	NO
49	63	H	SI	NO
54	56	H	NO	SI
58	52	M	NO	SI
64	40	M	NO	NO
54	56	H	NO	SI
56	53	M	SI	NO
58	52	M	NO	SI
59	52	H	NO	SI
60	46	H	SI	SI
61	45	H	NO	NO
62	44	M	SI	NO
63	42	M	NO	NO
64	40	M	NO	NO
65	39	M	SI	NO
66	35	M	SI	SI
67	35	M	SI	SI
68	35	H	NO	NO
69	34	M	SI	SI
70	34	H	SI	NO

Para el tercer conjunto de datos, nuevamente estabamos en la situación de tener menos de 20 solicitantes que aparecen en cualquiera de las nueve ejecuciones. Por lo tanto, completamos de la misma forma que los anteriores.

5.4 Conclusión de los resultados

Veamos cómo se reparten las distintas categorías e intersecciones entre ellas en los ingresantes finales al programa.

	Primer dataset	Segundo dataset	Tercer dataset
Mujeres	23	27	19
No clase alta	35	35	35
No de Santiago	28	28	28
Mujeres y No de clase alta	7	12	11
Mujeres y No de Santiago	14	10	12
No de Clase alta y No de Santiago	8	8	14
Mujeres, No de clase alta y No de clase alta	2	4	6

Table 8: Cantidad de admitidos de cada categoría

En general, la cantidad de mujeres admitidas supera la meta establecida en todos los conjuntos de datos. Es decir, se seleccionaron más mujeres de las requeridas, lo que sugiere que este criterio fue satisfecho de manera más holgada en comparación con las otras restricciones.

En todos los conjuntos de datos, la cantidad de personas no pertenecientes a la clase alta alcanza la meta de 35 en todas las listas de admitidos. Parecería que esta condición es más restrictiva, ya que pedimos un 70% que es bastante alto, por lo que llega justo a este pedido.

En todos los conjuntos de datos, la cantidad de personas no provenientes de Santiago también alcanza la meta de 28 en todas las listas de admitidos. Al igual que con la restricción de "No Clase Alta", esta restricción se cumple de manera constante y se llega al límite establecido.

Al estudiar la cantidad de gente que se encuentran en más de una categoría, podemos notar que para llegar a los porcentajes solicitados el modelo en general se vio obligado a elegir a distintas personas.

En resumen, el modelo parece estar cumpliendo adecuadamente con las restricciones establecidas. Las variaciones en las cifras pueden atribuirse a la complejidad del problema y la necesidad de equilibrar múltiples restricciones

simultáneamente. La capacidad del modelo para cumplir con los porcentajes pedidos de forma justa en algunas categorías puede deberse a la priorización de ciertos criterios sobre otros durante la optimización.

5.5 Comparación con lo obtenido en el paper

En el paper en el año 2008 la cantidad de admitidos fue 51, a diferencia del nuestro que eran 50 personas. Además, cuando se ejecutó el algoritmo de selección en el paper, las mejores soluciones de los tres modelos coincidieron en 49 (de un posible 51) solicitantes en el Paso 1. Por lo tanto, el algoritmo tuvo que ejecutar el Paso 3 antes de pasar al Paso 5, al igual que nosotras. Sin embargo, nuestra cantidad de coincidencias fue 43 para el primer conjunto de datos, 42 para el segundo y 46 para el tercero.

Por otro lado, debido a que cuatro de los solicitantes en lista de espera del paper estaban presentes en una solución ejecutada, el número total presente en cualquiera de las nueve ejecuciones fue de 55. En nuestro caso, hubo 16 en el primer y segundo conjunto de datos y 9 en el tercero que aparecieron en alguna de las ejecuciones de los modelos pero no fueron seleccionados en el Paso 1. Por lo tanto, el total de presentes en cualquiera de las ejecuciones fue de 59 para el primer conjunto, 58 para el segundo y 55 para el último.

Otra cosa a analizar es que en el caso del conjunto de datos del paper, el Modelo 2 generó dos segundas mejores soluciones, y ambos Modelos 1 y 2 ofrecieron la misma solución óptima, que diferió del Modelo 3 en dos solicitantes. Mientras tanto, en nuestro caso, para el Modelo 1 con el primer conjunto de datos la 1° y 2° mejor solución dieron igual valor de función objetivo; para el segundo conjunto ocurrió lo mismo para la 2° y 3° mejor solución. Además, con el Modelo 2 se obtuvo igual valor de la función objetivo para la 2° y 3° mejor solución del primer conjunto de datos. También, para las mejores soluciones del Modelo 1 y Modelo 2 nos dió la misma lista de personas.

Finalmente, si revisamos las tablas anteriores, podemos observar que los porcentajes para las personas que no pertenecen a Santiago y las que pertenecen a los cuantiles bajos no cambiaron en ninguno de los tres conjuntos de datos y son exactamente la cantidad de personas mínimas que requeríamos de dichas categorías. Sin embargo, el porcentaje de mujeres fue diferente en los tres conjuntos y siempre se superó la cantidad mínima que se quería de mujeres. Estos resultados se asemejan a los del año 2008, donde la cantidad seleccionada de personas pertenecientes a los cuantiles más bajos de ingresos y que no eran de Santiago se alcanzó de manera exacta, pero la cantidad de mujeres se superó por una.

6 Nueva propuesta para la selección de admitidos

Como explicamos anteriormente en la sección de datos utilizados, para la nueva propuesta generamos una nueva característica de cada participante donde aclaramos la edad de cada uno, con edades que van desde los 22 a los 65 años. El objetivo principal de esto fue pensar que se quería priorizar a las personas más jóvenes en el programa, debido a que por lo general estas son las que más pueden verse beneficiadas por un proyecto de esta índole. Por lo tanto, agregamos una restricción que exigió que el 35% de las personas elegidas tengan entre 22 y 32 años solo en el Modelo 2. Como lo estamos agregando en un solo modelo y no en los tres, no tenemos una certeza de que en la lista definitiva de admitidos obtengamos efectivamente un 35% de jóvenes, pero darles un poco más de prioridad al armar las listas.

Para correr el algoritmo de selección, como tomamos la decisión de reemplazar el segundo modelo por este nuevo, ya que nuestra idea fue inspirarnos en él agregando la restricción para las edades. Por lo tanto, mantuvimos las variables de decisión, función objetivo y restricciones del Modelo 2. Sin embargo, en el nuevo modelo agregamos el conjunto de personas y la siguiente restricción:

- A es el conjunto de solicitantes con edades entre 22 y 32 años.

$$\sum_{i \in A} x_i \geq \frac{35}{100} \cdot N \quad (15)$$

6.1 Resultados obtenidos y comparación con método original

6.1.1 Primer conjunto de datos

Para el primer conjunto de datos, en la primer etapa conseguimos 42 personas que aparecían en las primeras soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes pedidos debíamos elegir 8 estudiantes. Veamos qué nos faltaba con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -3.0. Teníamos 3 mujeres más de las requeridas, por lo que para la siguiente corrida del Modelo 2 no hizo falta restricción para las mujeres.
- Cuantil bajo faltantes: 4.0
- No santiago faltantes: 3.5
- Jóvenes faltantes: 17.5
- Personas faltantes: 8. Para la nueva corrida cambiamos el valor de N a 8.

Los resultados arrojados nos indicaron que no se eligió ningún joven entre los seleccionados. Debido a esto, tomamos la decisión de volver a correr el modelo sin considerar la elección de los jóvenes, que igual deberían haberse visto beneficiados al asignarles la nueva puntuación. Entonces, corrimos el Modelo 2 original completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dio un valor con función objetivo de 46 y la siguiente 47. Obteniendo la siguiente lista final.

Table 9: Lista definitiva de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
1	99	M	SI	SI	57
2	94	M	SI	NO	51
3	93	H	NO	NO	30
4	92	M	SI	SI	55
5	91	H	SI	NO	26
7	90	M	SI	NO	31
8	90	H	NO	SI	22
9	89	M	SI	NO	25
10	88	M	SI	NO	24
11	88	M	NO	SI	23
12	87	M	SI	NO	41
13	87	M	NO	SI	46
14	87	H	NO	SI	39
15	86	H	NO	SI	24
16	86	M	SI	NO	49
17	84	M	SI	NO	33
18	84	H	NO	NO	43
19	81	H	NO	SI	24
20	80	H	NO	SI	61
21	79	H	NO	NO	59
24	77	M	NO	NO	26
25	77	H	SI	NO	47
27	75	H	NO	SI	43
28	75	H	NO	NO	59
29	73	H	NO	SI	37
30	70	M	NO	NO	56
31	67	H	NO	SI	25
33	65	M	NO	SI	40
34	64	H	NO	SI	51
35	62	M	NO	SI	44
36	62	M	SI	NO	39
37	62	M	NO	NO	55
39	60	H	NO	SI	37
40	58	H	NO	NO	24
41	58	M	SI	NO	40
42	53	M	SI	NO	36
43	53	H	NO	SI	56

Continúa en la siguiente página

Table 9 – Continuación de la tabla

Ranking	Puntaje	Género	Clase	Región	Edades
44	52	H	SI	NO	35
45	52	H	NO	NO	33
48	51	M	NO	SI	37
49	51	H	NO	NO	55
50	50	M	NO	NO	36
54	48	H	NO	NO	45
58	43	H	NO	NO	45
59	43	H	NO	SI	58
60	41	H	NO	SI	28
62	38	H	NO	SI	58
63	38	H	NO	SI	30
70	33	M	NO	NO	43
85	23	M	NO	NO	65

Podemos notar que para este primer conjunto de datos la lista definitiva de admitidos no se vió afectada por el agregado de la condición de edad a los participantes. Veamos entonces si la lista de espera tuvo algún cambio:

Table 10: Lista de espera de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
6	90	M	SI	SI	24
56	48	H	SI	NO	28
47	52	M	SI	NO	57
65	37	M	NO	SI	33
66	36	H	NO	SI	40
68	34	H	NO	SI	30
52	48	M	SI	NO	58
55	48	M	SI	NO	47
90	14	M	NO	NO	40
95	6	H	NO	NO	32
59	43	H	NO	SI	58
61	41	H	SI	NO	52
62	38	H	NO	SI	58
64	38	M	SI	SI	51
65	37	M	NO	SI	33
66	36	H	NO	SI	40
67	35	H	SI	SI	35
68	34	H	NO	SI	30
69	34	H	SI	NO	31
70	33	M	NO	NO	43

Se puede apreciar que la lista de espera sí sufrió modificaciones. Por ejemplo, el participante que estaba rankeado 56, en la lista de espera original de este conjunto de datos estaba en el puesto número 8. Sin embargo, en este caso nos dió que está en el segundo puesto, lo que podemos atribuirlo a que tiene 28 años y por lo tanto pertenece al grupo de los jóvenes. Por otro lado, un participante que no se encontraba en la lista anterior pero sí en la de la nueva propuesta es el 95, quien tiene 32 años. Además, hubieron varias modificaciones en cuanto al orden de los participantes.

6.1.2 Segundo conjunto de datos

Para el segundo conjunto de datos, en la primer etapa conseguimos 37 personas que aparecían en las soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes pedidos debíamos elegir 13 estudiantes. Veamos qué nos faltaba con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -6.0. Teníamos 6 mujeres más de las requeridas, por lo que para la siguiente corrida del Modelo 2 no hizo falta restricción para las mujeres.

- Cuantil bajo faltantes: 5.0
- No santiago faltantes: 6.5
- Jóvenes faltantes: 17.5
- Personas faltantes: 13. Para la nueva corrida cambiamos el valor de N a 13.

Nuevamente, los resultados obtenidos nos indicaron que no se eligió ningún joven entre los seleccionados. Al igual que antes, tomamos la decisión de volver a correr el modelo sin considerar la elección de los jóvenes. Entonces, corrimos el Modelo 2 original completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dió un valor con función objetivo de 104 y la siguiente 106. Obteniendo la siguiente lista final.

Table 11: Lista definitiva de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
1	99	H	NO	SI	36
2	98	M	NO	SI	36
3	98	M	SI	SI	43
4	97	M	SI	SI	52
5	96	M	NO	SI	49
6	95	M	SI	NO	49
7	95	H	SI	SI	34
8	94	M	SI	SI	44
9	93	M	NO	NO	22
10	92	H	NO	SI	38
11	91	H	NO	NO	58
12	89	H	SI	NO	53
13	88	H	SI	NO	27
14	85	M	NO	NO	30
15	85	H	SI	SI	22
17	84	M	SI	NO	36
18	84	M	NO	NO	39
20	83	M	NO	SI	58
21	81	M	NO	SI	46
22	80	H	NO	NO	34
24	75	M	SI	NO	56
25	75	H	NO	NO	56
27	74	M	SI	NO	47
28	74	M	NO	SI	33
29	72	M	SI	NO	26
30	72	M	NO	NO	55
31	72	H	NO	SI	56
32	70	H	NO	SI	33
34	69	H	NO	NO	45
35	68	H	NO	NO	23
36	66	H	NO	SI	33
40	63	M	NO	SI	38
42	62	M	NO	SI	52
43	61	H	NO	SI	26
44	61	H	SI	NO	41
45	60	H	SI	NO	52
47	59	H	NO	NO	60
49	56	H	NO	SI	48
51	56	H	SI	NO	24
53	54	M	NO	NO	56
55	51	M	NO	SI	47

Continúa en la siguiente página

Table 11 – Continuación de la tabla

Ranking	Puntaje	Género	Clase	Región	Edades
57	49	M	NO	NO	46
59	47	M	NO	SI	38
60	47	H	NO	NO	48
61	45	M	NO	NO	37
63	44	M	NO	NO	24
64	44	M	NO	SI	34
69	41	H	NO	NO	45
74	36	H	NO	NO	64
78	32	H	NO	NO	26

Podemos notar que en este caso la lista definitiva sí se vió afectada por el agregado de la condición de edad, pero no hubo un cambio significativo. La única modificación que hubo fue que se cambió el participante 46 con 62 años por el 51 en la nueva propuesta, el cual tiene 24 años. Veamos ahora la lista de espera:

Table 12: Lista de espera de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
46	59	M	SI	NO	62
16	85	M	SI	SI	64
33	69	M	SI	SI	32
38	65	M	SI	SI	24
39	64	M	SI	SI	28
50	56	H	SI	NO	52
58	48	H	SI	NO	41
68	43	M	NO	SI	40
67	43	M	SI	NO	22
66	43	M	NO	SI	36
65	43	H	NO	SI	44
73	38	H	SI	NO	22
77	33	H	SI	NO	31
85	22	H	NO	SI	28
88	19	M	NO	NO	58
99	1	M	NO	SI	24
91	13	M	SI	SI	52
84	25	H	SI	NO	36
86	21	M	NO	SI	61
98	4	M	NO	SI	47

Se puede apreciar que la lista de espera sí sufrió bastantes modificaciones. Primero que nada podemos observar que el participante 46 que era el que nuestro nuevo modelo dejó afuera de la lista definitiva, esta primero en la lista de espera. Además, podemos notar que los participantes 33, 38, 39, 73, 77, 85 y 99 aparecen en la nueva lista de espera pero no aparecían en la original, lo cual tiene sentido porque son todos menores a 33 años. Podemos concluir entonces que en el segundo conjunto de datos lo que más se vió afectado fue la lista de espera, subiendo a participantes con mayor Ranking debido a su edad.

6.1.3 Tercer conjunto de datos

Para el tercer conjunto de datos, en la primer etapa conseguimos 45 personas que aparecían en las soluciones óptimas de los 3 modelos. Por lo tanto, para completar la lista de 50 personas cumpliendo los porcentajes pedidos debíamos elegir 5 estudiantes. Veamos qué nos faltaba con los resultados arrojados en el Paso 2:

- Mujeres faltantes: -1.0. Teníamos 1 mujer más de las requeridas, por lo que para la siguiente corrida del Modelo 2 no hizo falta restricción para las mujeres.
- Cuantil bajo faltantes: 3.0

- No santiago faltantes: 1.5
- Jóvenes faltantes: 17.5
- Personas faltantes: 5. Para la nueva corrida cambiamos el valor de N a 5.

Al igual que en los anteriores dos conjuntos de datos, en el último los resultados obtenidos nos indicaron que no se eligió ningún joven entre los seleccionados. Nuevamente, tomamos la decisión de volver a correr el modelo sin considerar la elección de los jóvenes. Corrimos el Modelo 2 original completando así la lista de admisión ya que obtuvimos una única solución óptima, la cual nos dio un valor con función objetivo de 20 y la siguiente 21. Obteniendo la siguiente lista final.

Table 13: Lista definitiva de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
1	100	H	SI	NO	41
2	100	H	SI	SI	34
3	99	M	SI	SI	64
4	98	H	SI	NO	31
5	97	M	NO	SI	44
6	96	M	SI	SI	25
7	94	H	SI	SI	64
8	94	H	SI	SI	59
9	93	H	NO	SI	55
10	93	M	NO	SI	51
11	92	M	SI	NO	44
12	91	H	NO	SI	29
13	90	H	NO	SI	34
14	88	M	NO	NO	27
15	88	M	NO	NO	43
16	88	H	NO	NO	65
17	87	M	NO	NO	38
18	86	H	SI	SI	23
19	85	H	NO	NO	61
20	85	H	NO	NO	59
21	85	M	NO	SI	33
22	84	M	SI	SI	58
23	83	M	SI	NO	25
25	81	M	NO	NO	65
26	79	H	NO	NO	56
27	79	H	SI	NO	29
28	78	H	NO	SI	64
30	75	H	SI	NO	34
31	75	H	NO	NO	40
32	74	H	NO	SI	64
34	73	M	NO	NO	41
35	73	H	NO	NO	29
36	73	M	SI	NO	28
37	71	H	NO	SI	26
38	71	H	NO	SI	40
39	70	M	NO	SI	26
40	70	H	NO	NO	64
41	68	M	NO	SI	38
42	68	H	NO	NO	60
43	66	H	NO	SI	50
46	64	H	NO	NO	54
47	64	M	SI	NO	52

Continúa en la siguiente página

Table 13 – Continuación de la tabla

Ranking	Puntaje	Género	Clase	Región	Edades
48	63	H	NO	SI	40
50	61	H	NO	NO	31
52	61	M	NO	NO	36
53	60	H	NO	NO	55
54	56	H	NO	SI	28
55	54	H	NO	NO	32
57	53	H	NO	NO	32
61	45	H	NO	NO	47

Se puede apreciar que hubieron dos participantes que fueron modificados respecto de la lista original. En la primer lista definitiva tenemos a los estudiantes que rankeaban en los puestos 24 y 63 de 53 y 33 años respectivamente. Sin embargo, en la lista de la nueva propuesta no se encuentran presentes, y en su lugar tenemos al 47 y 54, el primero de 52 años y el segundo de 28 que estaban en el primer y tercer puesto de la lista de espera originalmente. Veamos ahora las modificaciones de la lista de espera.

Table 14: Lista de espera de nueva propuesta

Ranking	Puntaje	Género	Clase	Región	Edades
24	83	M	SI	SI	53
29	76	H	SI	SI	28
49	63	H	SI	NO	61
56	53	M	SI	NO	26
58	52	M	NO	SI	52
63	42	M	NO	NO	33
72	32	H	NO	SI	29
64	40	M	NO	NO	52
59	52	H	NO	SI	64
60	46	H	SI	SI	48
61	45	H	NO	NO	47
62	44	M	SI	NO	25
63	42	M	NO	NO	33
64	40	M	NO	NO	52
65	39	M	SI	NO	45
66	35	M	SI	SI	22
67	35	M	SI	SI	61
68	35	H	NO	NO	63
69	34	M	SI	SI	53
70	34	H	SI	NO	63

La lista de espera del tercer conjunto de datos sufrió numerosas modificaciones. Para empezar, a los participantes rankeados 24 y 63 que eran los que no habían entrado en la nueva propuesta a diferencia de la original, los podemos encontrar en la nueva lista de espera. Por otro lado, los participantes 29 y 72 que tampoco habían entrado en la lista de espera original, se encuentran en la nueva, lo cual tiene sentido porque ambos pertenecen al grupo de jóvenes. Por último, podemos observar modificaciones en cuanto al orden de algunos participantes.

7 Análisis cambiando restricciones

En el caso del paper, el mejor valor posible para el Modelo 1 es 3325.4, la suma de las puntuaciones de los 51 mejores candidatos sin aplicar ninguna otra restricción. Para el Modelo 2, el valor de la función objetivo, asumiendo que no hay otras restricciones, es 1326; para el Modelo 3, es 51.2662. Para nuestros datos, el mejor valor posible para el Modelo 1 varía según el conjunto de participantes, para el primero es 3674, para el segundo es 3833 y para el último da 4017. Por otro lado, el mejor valor para los Modelo 2 y 3 no varía según el conjunto de datos, sino varía según la cantidad de admitidos, por lo tanto si no consideramos restricciones los valores son 1275 y 50.255 respectivamente.

Dirigiéndonos al proceso de 2008, si se elimina la restricción de origen regional, el algoritmo finaliza al completar el Paso 1 (lo que implica que los tres modelos ofrecen la misma mejor solución), después de seleccionar el mínimo exacto para las admisiones de mujeres y del cuantil de ingresos más bajos, y uno menos de solicitante no perteneciente a Santiago que el mínimo requerido. Si se excluye la restricción del cuantil de ingresos, el algoritmo nuevamente finaliza cuando se ha ejecutado el Paso 1 después de seleccionar los números exactos mínimos de candidatas mujeres y no pertenecientes a Santiago. El porcentaje de candidatos del cuantil de ingresos más bajos es del 64.7 por ciento (tres solicitantes del cuantil de ingresos más bajos menos que el mínimo requerido cuando se incluye esta restricción).

En nuestro caso, si se elimina la restricción de origen regional, el algoritmo finaliza al completar el Paso 1 (lo que implica que los tres modelos ofrecen la misma mejor solución) para los tres conjuntos de datos. Superamos la cantidad de mujeres por bastante más de lo pedido, necesitábamos 15 y conseguimos 22, 30 y 18 para cada dataset. La cantidad de personas para la restricción de cuantiles bajos lo alcanzamos justo con lo pedido. Por otro lado, obtuvimos 22, 21 y 25 personas que no son de Santiago, cuando pedían 28. Si se excluye la restricción del cuantil de ingresos, el algoritmo nuevamente finaliza cuando se ha ejecutado el Paso 1 para los tres dataset. Superamos la cantidad de mujeres por bastante más de lo pedido, necesitábamos 15 y conseguimos 27, 28 y 20 para cada dataset. La cantidad de personas para la restricción de origen regional lo alcanzamos justo con lo pedido. Por otro lado, obtuvimos 26, 25 y 30 personas de los cuantiles más bajos, cuando pedían 35.

8 Conclusiones

En este extenso análisis del proceso de selección de candidatos, exploramos restricciones y modificaciones al modelo propuesto por el paper.

Pudimos observar que las restricciones originales, como el porcentaje de mujeres, cuantiles bajos y no residentes en Santiago, han sido satisfechas en la mayoría de los casos, incluso algunas veces superando los límites mínimos requeridos. Notamos que en particular el único porcentaje que cambió a lo largo de los tres conjuntos de datos fue el de las mujeres, ya que el de cuantiles bajos fue siempre 70% y el de Santiago 56%, cumpliendo de manera casi exacta el porcentaje solicitado. Sin embargo, los porcentajes de mujeres fueron de 46%, 54% y 38% para cada dataset, lo que nos muestra que el requerimiento de mujeres se cumplió con amplia diferencia. Esto podríamos pensar que se debe a que los datos contienen mayoría mujeres, pero como mencionamos en los datos utilizados, los datasets tienen 49, 54 y 53 mujeres respectivamente, es decir, casi la mitad de los participantes son mujeres.

A pesar que en el estudio de 2008 del paper se tomaba una persona más que en el nuestro, la cantidad de coincidencias en las mejores soluciones fue alta en ambos casos: 49 en el primero y entre 42 y 46 en el estudio actual para distintos conjuntos de datos. Es relevante notar que en el primer estudio, hubo 55 solicitantes presentes en alguna ejecución de los modelos, comparado con 59, 58 y 55 en los distintos conjuntos de datos del estudio actual.

Por otro lado, una conclusión a la que podemos llegar viendo nuestra propuesta nueva es que capaz la lista definitiva de admitidos no se vió muy alterada, más que nada en el primer dataset que no sufrió ni una modificación. Esto podemos atribuirlo a que en la lista definitiva lo que más peso tiene es el puntaje y las otras características. Además, es importante notar que la restricción de edad solamente estaba en un modelo y no en los tres. Si hubieramos querido que en la lista definitiva de admitidos se respetase realmente el 35% de jóvenes, podríamos haber agregado la restricción de edad a los tres modelos. No obstante, pudimos observar que las listas de espera de los tres conjuntos de datos se vieron muy afectadas por participantes jóvenes que subieron de puesto o fueron agregados.

Posteriormente, observamos como se modificaban las listas a medida que sacábamos las restricciones de cuantiles bajo y región de los modelos. Pudimos notar que al quitar la del cuantil bajo conseguíamos que se cumpla las otras condiciones pedidas pero no se alcanzaba el porcentaje deseado para este grupo y lo mismo al sacar la restricción de la región. Esto nos da el indicio de que el algoritmo es útil para asegurar una mayor igualdad en cuanto a la elección de los ingresantes cumpliendo con diversas restricciones y ajustandose a cambios en las condiciones del problema.

Finalmente, este análisis detallado ha proporcionado una visión profunda de cómo las restricciones y modificaciones afectan el proceso de selección de integrantes en el programa, destacando la importancia de la adaptabilidad del algoritmo y las consideraciones éticas y morales en la toma de decisiones.