

Bases de Datos Sistema de Admisión Escolar 2020

Equipo Sistema de Admisión Escolar

13 de enero de 2021

1. Introducción

El Sistema de Admisión que surge como consecuencia de la Ley de Inclusión tiene la característica de ser centralizado. Todas las postulaciones de los apoderados son realizadas a través de una plataforma web única. Luego, son procesadas utilizando un algoritmo de asignación que le entrega a cada postulante un establecimiento.

El proceso intenta otorgar a cada postulante un establecimiento dentro de la preferencia más alta posible, considerando algunas prioridades especiales y las restricciones de capacidad de los establecimientos. En caso de que haya más postulantes que cupos disponibles, se desempata por medio de loterías, de manera que todos tengan las mismas probabilidades de ser admitidos.¹

2. Bases de Datos

Las bases del Sistema de Admisión Escolar se componen de 6 archivos, cada uno con una versión para la etapa regular, terminada en “1”, y otra para la etapa complementaria, terminada en “2”².

A continuación se describe el contenido de cada uno de estos archivos.

- **A1 y A2:** Datos de la oferta de cursos; los cupos declarados por el establecimiento, las vacantes, y otras características relevantes para la asignación.
- **B1 y B2:** Datos de los postulantes; el nivel al que postula, el género, si es alumno alto rendimiento, condición de prioritario y georeferencia.
- **C1 y C2:** Datos de las postulaciones; rankings declarados por los postulantes, tipo de prioridad de cada postulación, orden en la lotería.
- **D1 y D2:** Datos de resultado; asignación de establecimiento, respuesta de los postulantes para el caso de la etapa regular.
- **E2:** Datos útiles únicamente para replicar el proceso; asignación de establecimiento utilizando la georeferencia con error.
- **F1 y F2:** Datos de relaciones entre postulantes, identifica los que son hermanos, y los que están realizando una postulación en bloque.

¹ Solo los postulantes que tienen prioridades especiales tienen una mayor posibilidad de obtener un cupo, como hermanos de estudiantes ya matriculados o hijos de funcionarios. A una misma prioridad, la probabilidad de quedar es exactamente la misma.

² Desde 2018, uno de los archivos solo existe en su versión para el proceso complementario.

2.1. Información de oferta (A1 y A2)

La unidad de asignación está definida por la tupla de variables **rbd**, código de grado (**cod_grado**), tipo de enseñanza (**cod_ense**), de especialidad (**cod_espe**), de sede (**cod_sede**), de género (**cod_genero**³, que se construye en base a **solo_hombres** y **solo_mujeres**) y de jornada (**cod_jor**). Todas las variables distintas a **rbd** se resumen en una nueva variable, creada para el proceso de admisión, denominada **cod_curso**, por lo que la tupla **rbd-cod_curso** define un “curso” al que se puede postular. No es relevante si el colegio ofrece una o tres salas de clases para cierto **rbd-cod_curso**, para el sistema de admisión cuenta como “un gran curso”, con capacidad declarada por el establecimiento, y representada en la variable **cupos_totales**. Para este cuarto año de implementación, el sistema ofreció 67.340 opciones en el periodo principal⁴, repartidas en 8.064 establecimientos.

Cada **rbd-cod_curso** tiene asociado unos **cupos_totales**, **vacantes**, una latitud y longitud (**lat** y **lon**) y si cobra copago (**con_copago**). Para calcular las vacantes disponibles (**vacantes**), se resta a los **cupos_totales** los estudiantes que continúan en el establecimiento de un año para el otro, es decir los estudiantes que están matriculados en el nivel anterior, descontando los repitentes estimados del nivel anterior y los que postulan para cambiarse. Finalmente se resta los repitentes del nivel. La siguiente fórmula explica el cálculo:

$$Vacantes = cupos_totales - alumnos_continuan_{nivel_anterior} + repitentes_{nivel_anterior} - repitentes_{nivel_actual}$$

2.2. Información de postulantes (B1 y B2)

Los participantes del proceso están identificados por el **mrnun**, máscara del run generada por el Centro de Estudios de MINEDUC ampliamente utilizada en bases de educación. Cada postulante tiene asociado un nivel al que postula (**cod_nivel**), si es mujer (**es_mujer**), si es clasificado como prioritario, y si pertenece al 20 % superior del ranking de notas de su establecimiento de origen (**alto_rendimiento**). Los archivos de postulantes para el año 2020 se componen de 454.415 postulantes para el periodo principal y 74.111 para el periodo complementario.

Las variables **lat_con_error** y **lon_con_error** corresponden a la latitud y longitud obtenida a partir de la georeferenciación de la dirección declarada por el apoderado. La plataforma de postulación no siempre pudo encontrar una georeferencia precisa de la dirección ingresada por el usuario y en ocasiones eligió la municipalidad de la comuna respectiva como referencia del hogar, debido a que el mapa desplegado no representaba su ubicación. Por lo anterior, se definió una variable que representa la “calidad de la georeferenciación” (**calidad_georef**), para distinguir los casos en que se logró una buena georeferencia o si se imputó la municipalidad.

La variable **calidad_georef** se incluye únicamente en B1. Toma valores enteros desde el 1 al 5. El valor 1 significa que se obtuvo una respuesta única de la *Google Geocoding API*, con calidad “rooftop” o “range_interpolated” en la variable “location_type”. 2 también corresponde a una respuesta única de la *Google Geocoding API*, pero de calidad “geometric_center” o “approximate” en “location_type”. El valor 3 significa que la API entregó múltiples respuestas coherentes. 4 es que los datos ingresados corresponden a las coordenadas de la municipalidad. Y 5 corresponde a cuando el usuario compartió su ubicación, por lo que ésta no fue consultada a *Google Geocoding API*.⁵.

Para proteger la confidencialidad de los postulantes, se modificaron los datos de geolocalización utilizando las recomendaciones del “Geographic Displacement Procedure and Georeferenced Data Release Policy for

³ 1: mujeres, 2: hombres, 3: mixto

⁴ Para el periodo complementario de postulación se añadieron 17 cursos, sin aumentar el número de establecimientos.

⁵ Para mayor información sobre el proceso de construcción de la variable **calidad_georef** contactar a pablo.becerra@mineduc.cl.

the Demographic and Health Survey”⁶. El método de desplazamiento utilizado fue el: “dirección aleatoria, distancia aleatoria”, el que contienen las siguientes etapas:

1. Seleccionar una dirección (ángulo) aleatoria entre 0 y 360 grados.
2. Seleccionar una distancia aleatoria entre 50 y 300 metros.
3. Desplazar el punto a las nuevas coordenadas.
4. Únicamente para B1, chequear que las nuevas coordenadas están en tierra firme, repetir pasos 1,2 y 3 de no cumplirse la condición.

Los números aleatorios en ambos casos provienen de una distribución uniforme con los mínimos y máximos señalados. La distancia mínima de desplazamiento es de 50 m, la media de 175 m y la máxima de 300 m.

2.3. Información de postulaciones (C1 y C2)

Las listas de establecimientos elaboradas por los postulantes del Sistema de Admisión Escolar están representados por el identificador de la persona (**mrnun**), el establecimiento al que postula (**rbd**), el **cod_curso** (que define el nivel, jornada y especialidad), y el orden de preferencia dentro del ranking (**preferencia_postulante**).

Las postulaciones con valor “1” en la variable **agregada_por_continuidad** corresponden a establecimientos que no fueron declarados como preferencia de los postulantes en la plataforma de admisión, sino que se incluyeron artificialmente al final de cada uno de los rankings para efectos del algoritmo de asignación. Estos corresponden a los establecimientos de procedencia de los postulantes, y únicamente se agregan a los postulantes que podrían continuar sus trayectorias educativas sin cambiarse, es decir, los establecimientos ofrecen los niveles a los que postulan y podrían darle continuidad al postulante. Se incluyen al final del ranking porque el postulante es asignado allí en caso de no poder ser asignados a ninguna de sus preferencias.

Los archivos C1 y C2 también contienen el resultado del orden establecido en base a la generación de números aleatorios. Para crear la secuencia de números aleatorios se utilizó una semilla definida por los seis últimos sismos de magnitud 3 o más del 1 de octubre de 2020⁷, y el algoritmo de generación de números aleatorios (α o β) escogido por el establecimiento. Este orden aleatorio es el que define qué estudiantes tienen prioridad para entrar a cada establecimiento, en el caso de que exista más postulantes que vacantes.⁸

El ordenamiento que se utiliza como input del algoritmo de asignación es el definido por la variable **lote-ria_original**. A partir de este número, se ordenan las postulaciones para cada **rbd-cod_curso**, considerando las siguiente prioridades:

- 1° Postulantes que tienen matrícula asegurada en el establecimiento ya que provienen de él. En caso de que estos postulantes no logren cambiarse a través del Sistema, su establecimiento les garantiza la continuidad de estudios (**prioridad_matriculado**= 1).
- 2° Postulantes que tienen un hermano matriculado en el establecimiento (**prioridad_hermano**= 1).
- 3° Los hijos de funcionarios del establecimiento (**prioridad_hijo_funcionario**= 1).

⁶ Burgert, Clara R., Josh Colston, Thea Roy, and Blake Zachary. 2013. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spatial Analysis Reports No. 7. Calverton, Maryland, USA: ICF International. Available at <http://dhsprogram.com/pubs/pdf/SAR7/SAR7.pdf>

⁷ Para el caso del proceso complementario la fecha fue el 30 de noviembre de 2020.

⁸ Para más detalles del proceso de generación de número aleatorios, revisar el documento adjunto a las bases públicas del año 2017 “Implementación de algoritmos de generación de órdenes aleatorios para el Sistema de Admisión Escolar 2017”, elaborado por el Departamento de Ingeniería Industrial de la Universidad de Chile.

4° Los ex-alumnos del establecimiento (**prioridad_exalumno= 1**).

5° Todos los restantes.

Entre las postulaciones con el mismo tipo de prioridad, se ordenan según **loteria_original**, de menor a mayor. Para las cuotas especiales de estudiantes prioritarios, PIE y alta exigencia el orden es levemente diferente. Para mayores detalles revisar la Sección 3.1, y el documento “Implementación algoritmo de asignación SAE 2017”, elaborado por el Departamento de Ingeniería Industrial de la Universidad de Chile.

2.4. Información de resultados (D1 y D2)

El resultado de asignación luego de ejecutar el algoritmo está registrado en la dupla de variables **rbd_admitido** y **cod_curso_admitido**.

La base D1 incluye las variables **respuesta_postulante** y **respuesta_postulante_post_lista_espera**, que registran la decisión del apoderado al revisar el resultado de la etapa regular (o de las listas de espera, según corresponda). Esta variable toma los valores de 1 si acepta la asignación, 2 si acepta la asignación y espera por si corre la lista de espera en alguna de sus otras mayores preferencias, 3 si rechaza la asignación, 5 si no se registró respuesta o 6 si está obligado a esperar lista de espera (pues no tiene asignación)⁹. En caso de que no se registrara respuesta (5), el Sistema asume que la asignación fue aceptada por el apoderado.

Nuevamente para la base de la etapa regular, la variable **rbd_admitido_post_resp** junto con **cod_curso_admitido_post_resp** representan el establecimiento admitido luego de procesar las respuestas y que corran las listas de espera. Estas variables están vacías para los apoderados que rechazan (**respuesta_postulante=3**).

La base D2 no incluye respuesta o asignación posterior a lista de espera, pues en la etapa complementaria no hay proceso de lista de espera.

Una vez terminada la asignación de la ronda complementaria, quedan los postulantes de ambas etapas con matrícula garantizada en el establecimiento asignado, sin por esto estar obligados a matricularse en el establecimiento. En caso de que quieran cambiar de establecimiento, pueden buscar uno que tenga cupos disponibles y matricularse directamente. Esta etapa se denomina “regularización”.

2.5. Información para replicar proceso (E2)

En vista de que las bases B1 y B2 no contienen la georeferencia real, sino una con error, no es posible replicar la asignación de la ronda complementaria con exactitud¹⁰. El resultado de la asignación si se ejecutara el algoritmo utilizando las variables **lat_con_error** y **lon_con_error** en vez de las reales, está almacenado en la variable **rbd_admitido_geo_con_error** y **cod_curso_admitido_geo_con_error** de la Tabla E2. Con este resultado es posible verificar si se ha replicado con precisión el algoritmo de asignación. **rbd_admitido_geo_con_error** difiere de **rbd_admitido** solo en un conjunto parcial de los postulantes y todas estas diferencias se explican por la asignación por cercanía.

2.6. Información de relacion entre postulantes (F1 y F2)

Las relaciones de hermandad y condiciones de postulación en bloque de los postulantes están representadas por pares de identificadores **mrnun_1** y **mrnun_2**. La variable **mismo_nivel** toma el valor 1 cuando

⁹ El valor 4 no existe en la base producto de un cambio de flujo en el proceso del año 2018: en los años anteriores los apoderados podían rechazar la asignación y participar inmediatamente en las listas de espera.

¹⁰ Igual que los años 2018 y 2019, en 2020 no se construye la tabla E1. Con el resto de las tablas es posible replicar con exactitud la asignación de la ronda regular.

ambos estudiantes postulan al mismo nivel, **es_hermano** indica si el par es hermano consanguíneo, y **postula_en_bloque** marca si el apoderado escogió que el Sistema priorizara que ambos postulantes quedaran en un mismo establecimiento, en desmedro de las preferencias declaradas para cada uno.

3. Algoritmo de Asignación

El mecanismo de Aceptación Diferida que utiliza el Sistema de Admisión Escolar es considerado como el de mayor aceptación dentro de la literatura de *Matching* para resolver el problema de emparejamiento. Este algoritmo, propuesto por el economista y matemático David Gale y el filósofo y matemático Lloyd Shapely (1962¹¹), cumple con ciertas condiciones deseables para un mecanismo de asignación, como lo es la *estabilidad* de la asignación y el ser *a prueba de estrategias*, pero no es *Pareto eficiente*. En tanto, Abdulkadiroglu, A., y Sönmez, T. (2003¹²) proponen los pasos que realiza el algoritmo de asignación, donde en primera instancia cada estudiante postula a su primera preferencia, y si los cupos son suficientes cada establecimiento acepta tentativamente a los postulantes, privilegiando a partir de la lista de prioridad. En el caso que existan estudiantes que no hayan sido tentativamente aceptados, postulan a su siguiente preferencia, y si los cupos son suficientes, cada establecimiento aceptará tentativamente a los nuevos estudiantes, sujeto a las prioridades establecidas. El algoritmo se itera hasta que todos los postulantes sean asignados a un establecimiento de preferencia o hasta que no queden vacantes disponibles para asignar.

Este mecanismo es utilizado en otros 12 sistemas centralizados: Ámsterdam, Boston, Chicago, Seúl, Ghana, Antwerp (Bélgica), Finlandia, Hungría, Comunidad Francesa (Bélgica), Nueva York, Inglaterra, y Francia.

¹¹ Gale, D. and L. S. Shapley (1962) College admissions and the stability of marriage, American Mathematical Monthly, 69: 9-15.

¹² Abdulkadiroglu, A., and Sönmez, T. (2003). School choice: A mechanism design approach. American economic review, 93(3), 729-747.