



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales
Departamento de Matemática

Tesis de Licenciatura

Asignación de votantes a centros de votación utilizando técnicas de Programación Matemática para minimizar el tiempo total invertido

Mauro Giormenti Moravec

Director: Nicolás Stier-Moses
Co-Director: Guillermo Durán

Septiembre de 2014; Buenos Aires, Argentina

ASIGNACIÓN DE VOTANTES A CENTROS DE VOTACIÓN UTILIZANDO TÉCNICAS DE PROGRAMACIÓN MATEMÁTICA PARA MINIMIZAR EL TIEMPO TOTAL INVERTIDO

En la República Argentina, a la hora de asignar votantes a los lugares en que emitirán su voto, el criterio utilizado es repartir a las personas dentro de un área geográfica por orden alfabético. En este trabajo se buscan asignaciones más eficientes, en el sentido que el tiempo total invertido por las personas sea el menor posible.

Se dan modelos iniciales sencillos que sólo tienen en cuenta la distancia caminada entre los domicilios y los centros de votación, y posteriormente se construye sobre éstos otros modelos más sofisticados que tienen en cuenta el tiempo de espera en la cola. Para ello se estudió la dinámica real de las filas de votación con datos recolectados durante las elecciones nacionales de 2013 y se las analizó con herramientas proporcionadas por la Teoría de Colas.

Se utilizan datos reales del padrón electoral para implementar y resolver los modelos propuestos con herramientas de Programación Lineal. Los resultados obtenidos son contrastados con los del padrón actual, y se comprueba que hay mucho por ganar al aplicar los métodos de asignación propuestos. Se discute qué modelos vale la pena implementar, se examinan posibles impedimentos para una aplicación a gran escala de los métodos aquí propuestos, y se realizan recomendaciones.

Palabras claves: Elecciones, Padrón, Problema de Asignación, Programación Lineal, Teoría de Colas, Geocodificación.

ASSIGNMENT OF VOTERS TO THEIR VOTING CENTER USING MATHEMATICAL PROGRAMMING TOOLS TO MINIMIZE THE TOTAL TIME INVESTED

In the Argentine Republic, the criteria used for assigning voters to their voting center consists of dividing all citizens in a certain geographic area by alphabetical order. In this thesis we look for more efficient assignments, in the sense that the total time invested by the voters be the least possible.

We give simple initial models that only account for the distance traveled between the voter's residence and the voting centers, and then build up from them more sophisticated models that also consider the time spent while waiting in the voting queue. In order to do that, the real dynamic of the voting queues was studied during the national elections of 2013, using tools provided by Queuing Theory for the Analysis.

Real data from the Electoral Roll is used for implementing and solving the proposed models with Linear Programming tools. The obtained results are compared with the current situation, and it is observed that there is much to gain from applying the proposed assigning methods. Which of the methods are worthy of applying is discussed, possible obstacles for a large-scale implementation are analyzed, and recommendations are made.

Keywords: Elections, Electoral Roll, Assignment problem, Linear Programming, Queuing Theory, Geocoding.

A mis padres, a quienes les debo todo.

Índice general

1..	Introducción	6
2..	Preliminares: Marco teórico	9
2.1.	Problemas de asignación	9
2.2.	Problemas de equilibrio	10
2.3.	Teoría de colas	11
3..	Modelos	13
3.1.	Un ejemplo ilustrativo	13
3.2.	Modelo básico	13
3.2.1.	Formalización	14
3.2.2.	Ejemplo	14
3.3.	Modelo básico con una restricción extra	15
3.3.1.	Ejemplo	15
3.4.	Modelo con capacidad de escuelas variable	16
3.4.1.	Formalización	17
3.4.2.	Ejemplo	18
3.5.	Modelo con mesas intercambiables entre escuelas	19
3.5.1.	Formalización	19
3.5.2.	Ejemplo	21
3.6.	Modelo con mesas intercambiables entre escuelas y capacidad variable	21
3.6.1.	Formalización	22
3.6.2.	Ejemplo	23
4..	Implementación	25
4.1.	Modelado de las colas de votación	25
4.1.1.	Recopilación de datos durante las elecciones nacionales de 2013	25
4.1.2.	Procesamiento y análisis de los datos obtenidos	26
4.1.3.	Simulación de las colas	31
4.2.	Procesamiento del padrón	35
4.3.	Geocodificación, GIS, Circuitos electorales, Matriz de distancias	36
4.3.1.	Geocodificación de las direcciones procesadas	36
4.3.2.	Agrupamiento en esquinas y cálculo de la matriz de distancias	37
4.3.3.	Partición en circuitos electorales	39
4.4.	Implementación de los modelos	40
5..	Resultados obtenidos	42
6..	Conclusiones	54

1. INTRODUCCIÓN

Periódicamente, en la República Argentina se realizan elecciones obligatorias (con alcance Nacional, Provincial o Municipal). Dado que éstas son presenciales, antes de la elección, funcionarios del Poder Judicial de la Nación designan el lugar al que cada persona habilitada para participar de los comicios se dirigirá para emitir su voto. A lo largo de este trabajo nos referiremos a estos lugares como *Escuelas*, pues la mayoría de los centros de votación lo son.

Este trabajo estudia el proceso a través del cual las personas habilitadas para votar (listadas en el *Registro electoral*) son asignadas a sus escuelas. En otras palabras, analizamos el modo en que se construye el *Padrón electoral*.

La construcción del Padrón está regida por el Código Nacional Electoral [6]. No nos adentraremos en muchos detalles, pero sí es necesario tomar nota de algunos detalles precisados por la ley:

- **Subdivisiones geográficas.** Cada provincia, más la Ciudad Autónoma de Buenos Aires (CABA), conforma un Distrito. Éstos, a su vez, están particionados en Secciones; en el caso de la CABA, cada comuna¹ conforma una Sección diferente. Finalmente, las Secciones se subdividen en Circuitos. La normativa establece que cada persona debe votar en el Circuito Electoral en que está domiciliada.
- **Mesas de votación.** El Padrón de cada Circuito se distribuye en mesas de votación repartidas entre las escuelas. Cada mesa puede recibir un máximo de 350 votantes². Desconocemos el criterio por el cual se decide cuántas mesas poner en cada escuela; sin embargo, es sensato suponer que hay una cierta proporcionalidad entre la cantidad de mesas y el tamaño físico de la escuela.
- **Asignación de votantes a mesas.** Una vez distribuidas y numeradas las mesas, el Padrón de cada Circuito Electoral es ordenado alfabéticamente y asignado en ese orden a las mesas (salvo que una disposición específica de alguna localidad fije algún otro criterio).

Como notará el lector, tras la división en Circuitos no se sigue ningún otro criterio geográfico. Al hacerse la distribución de votantes por orden alfabético, la asignación a escuelas resulta prácticamente aleatoria.

El propósito de esta tesis es estudiar estrategias alternativas a la designación de votantes por orden alfabético, en busca de una asignación más inteligente, que disminuya el tiempo que le consume a los votantes la participación en los comicios.

Para ello se cuenta con datos de un padrón electoral real, en el que se especifican los domicilios de todos los habitantes de la Comuna 11 de la Ciudad de Buenos Aires, así como las escuelas a las que éstos son asignados a votar oficialmente. La disponibilidad de datos

¹ Las comunas son subdivisiones geográficas y administrativas en las que está particionada la Ciudad de Buenos Aires. Las mismas están numeradas del 1 al 15, y sus poblaciones varían de los 150.000 a los 230.000 habitantes.

² Esto es así desde una reforma a la ley hecha en 2009. Antes de eso, la capacidad máxima era de 450 personas.

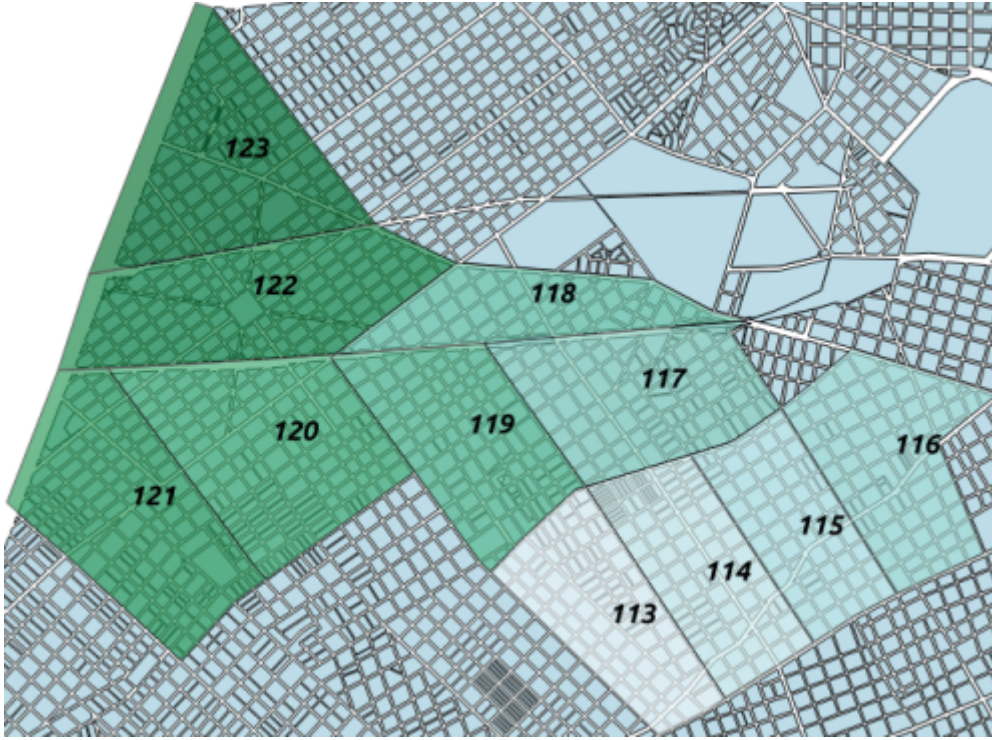


Fig. 1.1: Partición de Circuitos Electorales que componen a la Comuna 11.

reales nos permite analizar rigurosamente el desempeño de los métodos aquí propuestos. La Figura 1.1 muestra la geometría tanto de la Comuna 11 como de sus Circuitos Electorales.

Para comparar la efectividad de nuestras soluciones entre sí y con la asignación oficial en un principio se tendrá en cuenta el tiempo requerido para viajar desde el domicilio hasta la escuela en la que se vota.

Para los modelos siguientes se analizará también la dinámica en las colas de votación, incorporando el tiempo de espera en éstas a los modelos, y estudiando cuánto se puede ganar si se suavizan otras restricciones legales, como la cantidad predeterminada de votantes por mesa, o la cantidad de mesas situadas en cada escuela.

Es interesante notar que, a la hora de revisar en la literatura los problemas relacionados con elecciones, no es frecuente encontrar trabajos como el que aquí se presenta. Las cuestiones electorales son abordadas fundamentalmente desde las Ciencias Sociales, y se estudian problemas de análisis y diseño de sistemas electorales (electoral engineering) o de participación electoral (voter turnout). Lo más similar al tema aquí estudiado (por la importancia de las cuestiones geográficas) es el estudio de las prácticas a partir de las cuales se pueden delimitar los distritos electorales, y cómo la definición de dichos límites puede luego influir en los resultados electorales. Esta delimitación tendenciosa es conocida como gerrymandering.

El problema que se aborda en esta tesis fue previamente tratado en una Tesis de Licenciatura desarrollada por Jeremías Lenzi en la Universidad Nacional de Córdoba, y co-dirigida por Nicolás Stier-Moses y Guillermo Durán [5]. En la misma se desarrolla un modelo lineal buscando minimizar el tiempo promedio invertido por los votantes, teniendo en cuenta una restricción de equidad, a saber, personas que viven cerca no sean enviadas a lugares muy distantes entre sí, para que ninguno pueda sentirse muy desfavorecido. En

este trabajo se propondrá una restricción alternativa (donde cada votante no se compara con sus vecinos, sino respecto de su lugar de votación en elecciones previas). Por otro lado, en este trabajo se profundiza en el estudio de las colas de votación, que en el trabajo previo no fue un aspecto no desarrollado, asumiendo en vez de ello algunas simplificaciones. Finalmente, en este trabajo se parte del modelo base para proponer otros más complejos, en los que se estudia la posibilidad de variar más parámetros y ver en función de ello cómo varía la solución. Algunas de estas nuevas variables a estudiar estaban propuestas como trabajo futuro dentro de la tesis de Lenzi.

El Capítulo 2 brinda un marco teórico a este trabajo, e introduce conceptos y resultados utilizados más adelante. El Capítulo 3 presenta los modelos propuestos para obtener una asignación de votantes a escuelas más eficiente. El Capítulo 4 detalla el proceso realizado para implementar los modelos detallados en el capítulo anterior. En el Capítulo 5 se exhiben y analizan los resultados obtenidos. Finalmente, en el Capítulo 6 se muestran las conclusiones y recomendaciones que se pueden realizar a partir del trabajo aquí presentado.

2. PRELIMINARES: MARCO TEÓRICO

2.1. Problemas de asignación

Dado que el núcleo del problema a atacar consiste en asignar votantes a escuelas, conviene estudiar lo que en la literatura se conoce como Problemas de Asignación. En él se tiene un conjunto de agentes A y otro conjunto de tareas T , ambos con la misma cardinalidad. Asignar al agente $i \in A$ a la tarea $j \in T$ conlleva un costo $C_{i,j}$. Diremos que la función $X : A \times T \longrightarrow \{0; 1\}$, que notaremos X_{ij} , es una asignación si cumple que:

- $\sum_j X_{ij} = 1 \quad \forall i \in A$: Asigna todos los agentes a alguna tarea.
- $\sum_i X_{ij} = 1 \quad \forall j \in T$: Todas las tareas tienen algún agente asignado a ella.

En tal caso, interpretaremos que el agente $i \in A$ es asignado a la tarea $j \in T$ si y sólo si $X_{ij} = 1$.

Resolver el Problema de Asignación implica encontrar una asignación que minimice la función objetivo:

$$\sum_{ij} C_{ij} X_{ij}$$

Si bien esta formulación es un tanto restrictiva (se deben tener la misma cantidad de agentes que de tareas, y la asignación debe ser uno-a-uno), se puede recurrir a trucos sencillos para adaptar el problema a un espectro más amplio de problemas.

Por ejemplo, si se quiere tener varios agentes del mismo tipo, basta con “duplicar” el agente tantas veces como se desee, asignando a cada tarea el mismo costo C_{ij} que tenía el agente original. Análogamente, si lo que se quiere es poder asignar varios agentes a una misma tarea, se usa la misma estrategia, esta vez duplicando la tarea en cuestión.

Esta modificación es de interés, pues podríamos querer modelar como problema de asignación al problema inicial de esta tesis; a saber, el de asignar votantes (que pueden radicar varios en un mismo domicilio) a escuelas (que deben recibir a cientos o miles de votantes), minimizando la cantidad total que éstos deben caminar. Es decir que aquí el costo C_{ij} sería la distancia entre el domicilio i y la escuela j .

Es claro que los problemas de asignación son un caso particular de Problemas Lineales. La razón por la que se podría desear distinguirlos y formularlo como un problema particular es que al tener mayores restricciones, existen algoritmos que, aprovechando sus particularidades, pueden encontrar soluciones con mucha mayor eficiencia que si se usara un algoritmo general de Programación Lineal, como SIMPLEX. Ejemplos de algoritmos que resuelven el Problema de Asignación son el Algoritmo Húngaro, el de Caminos más Cortos Sucesivos o el de Relajación. Nos abstenemos de explicarlos, dado que no serán implementados, pero puede encontrarse más información en [1].

El inconveniente del Problema de Asignación es, justamente, su simpleza. Como ya se adelantó en la sección anterior, y se verá en profundidad en el Capítulo 3, será de nuestro interés sofisticar los modelos para que no sólo optimicen en función de la distancia

entre domicilios y escuelas, sino que también tengan en cuenta los tiempos de espera, y en función de eso cambien las capacidades de las escuelas o muevan mesas de votación de un lugar a otro. En esas circunstancias el modelo provisto por el Problema de Asignación resulta insuficiente para manejar las diversas variables que intervienen, y por lo tanto se hace necesario recurrir a formulaciones lineales más generales para atacar el problema.

Afortunadamente, el tamaño del problema a resolver está acotado, debido a la ya mencionada restricción de que cada votante deba ser asignado a una escuela dentro de su Circuito Electoral. Gracias a esto, incluso en los modelos más complejos el problema puede ser resuelto sin mayores inconvenientes utilizando solvers de Programación Lineal Entera. Como se detallará mejor en la Sección 4.4, en este trabajo se utilizó como solver el GNU Linear Programming Kit (GLPK), que permite resolver problemas Lineales con variables Reales, Enteras o Mixtas.

2.2. Problemas de equilibrio

A la hora de resolver un problema, el enfoque no tiene por qué ser siempre la búsqueda de un óptimo global (en nuestro caso, que la suma del tiempo invertido por todos los votantes sea la mínima posible). La objeción que se puede hacer a tales enfoques es que, en la búsqueda de dicho óptimo, algunas personas se ven mucho más perjudicadas que otras en pos del "bien común".

Por lo tanto, algunas veces en vez de buscar que la solución sea *óptima*, se busca que sea *justa*; es decir, que ninguna persona se vea más perjudicada que otra. De hecho, si en vez de estar asignando nosotros a los votantes se dejase que ellos eligieran dónde votar, la distribución final probablemente tendería a un equilibrio de este tipo, en el que nadie puede mejorar su situación cambiando su lugar de votación, pues todos invierten el mismo tiempo. Es lo que se conoce como un *Equilibrio de Nash* en Teoría de Juegos, o *Equilibrio de Wardrop* en Teoría del Transporte [2] [3].

La ventaja de estar asignando nosotros a los votantes, en vez de dejar evolucionar al sistema por sí solo, es que a veces hay soluciones justas que no son equilibrios. Por lo tanto, estamos en condiciones de elegir la mejor de entre todas las soluciones justas. Sin embargo, estas suelen ser peores que los óptimos globales.

En el caso de la asignación de votantes a escuelas, no tiene demasiado sentido tratar de igualar el tiempo total invertido en la votación (tiempo de viaje más tiempo de espera), pues quienes tienen su domicilio cerca de una escuela están en una situación de ventaja respecto de quienes no, y sería difícil (si no imposible) remediarlo.

Sin embargo, lo que sí se podría llegar a nivelar es el tiempo de espera. Aunque de hecho, en ese aspecto el sistema tal y como está ya es *justo*. Como se verá más adelante, el tiempo de espera está directamente relacionado con la cantidad de votantes asignados a una mesa de votación (cuantas más personas, mayor probabilidad de que dos lleguen en horarios similares, y por lo tanto una deba esperar a que la otra termine de emitir su voto).

Algunos de los modelos que se propondrán en el Capítulo 3 mantendrán esta noción de justicia; serán aquellos que no modifiquen el tamaño de las mesas. Sin embargo, otros modelos descartarán esta noción, en pos de mejorar aún más el óptimo global. En el Capítulo 5 se analizará cuánto hay para ganar a cambio de ello.

2.3. Teoría de colas

La Teoría de Colas estudia el comportamiento de sistemas en los que alguien llega a un lugar y debe esperar un cierto tiempo hasta ser atendido. De este modo se pueden modelar desde filas de espera ante las cajas de un banco (o una mesa de votación), hasta redes de semáforos o flujos de paquetes de información a través de Internet.

Se suele llamar *cliente* a quien llega al sistema para ser atendido (por más que lo que se esté modelando sea el envío de un paquete a través de una red). Por otro lado, se llama *servidor* a quién atiende a los clientes (aunque el "servidor", en nuestro modelo, sea una persona atendiendo al público a través de una ventanilla).

Al modelar una cola, hay que especificar diferentes aspectos de la misma, que en conjunto determinarán su dinámica.

En primer lugar, es necesario describir el proceso de llegada de clientes al sistema. Para eso se suele dar la distribución de los tiempos interarribos, es decir, del tiempo que transcurre entre dos llegadas consecutivas de clientes. Notamos $A(t)$ a su función de distribución acumulada. Es común asumir que los tiempos interarribos son independientes y están idénticamente distribuidos.

Asimismo, notamos $B(x)$ a la función de distribución acumulada del tiempo que tarda un servidor en atender un cliente.

Un sistema puede tener más de un servidor atendiendo simultáneamente a los clientes que aguardan en una misma cola. En ese caso, puede darse que distintos servidores posean tiempos de servicio diferentes entre sí. Dado que esto no se dará en el presente trabajo, no ahondaremos en esta posibilidad.

Algunos sistemas tienen una *capacidad de almacenaje* limitada. Esto quiere decir que existe una cota K tal que si un cliente llega cuando la cola alcanzó ese tamaño, el nuevo cliente es automáticamente rechazado, negándosele la posibilidad de ingresar en la cola. Es común asumir que la capacidad de almacenaje es ilimitada.

Finalmente, se debe establecer la *disciplina* de la cola, que determina el orden en que los clientes son atendidos. Hay tres criterios principales: se atiende al que está hace más tiempo (First-Come-First-Serve), al que llegó más recientemente (Last-Come-First-Serve), o a un cliente al azar entre todos los de la cola.

Se suele describir a las colas con la notación $X/Y/n/K$, donde X e Y denotan la distribución de los tiempos interarribos y de servicio, respectivamente; n la cantidad de servidores, y K la capacidad de almacenaje. En los casos en que $K = \infty$, se nota directamente $X/Y/n$. En los lugares de X e Y se ponen diferentes letras para indicar el tipo de distribución con la que se está trabajando.

- M : Indica que la distribución es Markoviana (sin memoria).
- D : Indica que la distribución es determinística.
- G : La distribución es General. Es decir que no se asume nada respecto a la misma; puede ser cualquier cosa.

Por supuesto que los modelos se pueden sofisticar de muchas maneras. Se pueden establecer distintos tipos de cliente, que llegarán cada uno según su propia distribución; se pueden poner varias colas en paralelo y permitir que los clientes pasen de una a otra; se puede permitir que los clientes se cansen de esperar y se vayan; que los clientes se colen o paguen sobornos para avanzar más rápidamente en la fila; que la población de personas

que puedan ir a la cola no sea infinita, sino que esté acotada; etc. Sin embargo, a los efectos de este trabajo será suficiente con manejar los modelos ya introducidos, pues en los capítulos siguientes se trabajará con colas de la forma M/M/1.

Dada una sistema de colas, interesa medir su desempeño y eficiencia. Hay muchas cosas para observar, pero en particular nos interesarán la cantidad de la gente que hay en la cola y, por encima de todo, el tiempo de espera que tienen que afrontar los clientes antes de ser atendidos por un servidor.

Por ejemplo, sean

- $\alpha(t)$: Cantidad de llegadas en el intervalo $(0,t)$.
- $\delta(t)$: Cantidad de salidas en el intervalo $(0,t)$.
- $N(t) := \alpha(t) - \delta(t)$: Cantidad de clientes en el sistema a tiempo t .
- $\gamma(t) := \int N(t)dt$: Tiempo total que todos los clientes pasaron en el sistema.

A partir de éstas, se pueden calcular los siguientes índices:

- $\lambda_t := \alpha(t)/t$: Tasa de llegada promedio en el intervalo $(0,t)$.
- $T_t := \gamma(t)/\alpha(t)$: Tiempo promedio pasado en el sistema por cliente, en el intervalo $(0,t)$.
- $\bar{N}_t := \gamma(t)/t = \lambda_t \cdot T_t$: Cantidad promedio de clientes en el sistema, en el intervalo $(0,t)$.

A menudo no interesa conocer el comportamiento del sistema en todo momento, sino que basta con saber el comportamiento general del mismo. Es decir, cómo se comporta el sistema cuando se lo deja evolucionar el tiempo suficiente, llegando a lo que se denomina *Estado Estacionario*.

Si el sistema tiene estado estacionario, la relación $\bar{N}_t = \lambda_t \cdot T_t$ lleva al *Resultado de Little*.

Proposición 2.3.1 (Resultado de Little). *Supongamos que existen los límites $\lambda := \lim_{t \rightarrow +\infty} \lambda_t$, $T := \lim_{t \rightarrow +\infty} T_t$.*

Entonces también existe $\bar{N} := \lim_{t \rightarrow +\infty} \bar{N}_t$, y vale que $\bar{N} = \lambda \cdot T$.

En estado estacionario, llamamos $\bar{x}$ al tiempo promedio que se tarda en atender a los clientes desde que entran hasta que salen de un servidor. Así como λ es la tasa de llegada de nuevos clientes al sistema, el parámetro $\mu := 1/\bar{x}$ es la tasa de salida.

A partir de ellos, dado un sistema con un solo servidor, se define $\rho := \lambda/\mu$. Notemos que para que el flujo de trabajo en un sistema sea estable, es necesario pedir que $\rho < 1$; pues lo contrario implicaría que están entrando más clientes de los que los servidores pueden procesar, y por lo tanto el tamaño de la cola tenderá a hacerse arbitrariamente grande.

Para las colas M/M/1, la distribución de la longitud de las colas $\bar{N}$ resulta ser geométrica de parámetro $1-\rho$.

3. MODELOS

En este capítulo se introducirán diversas propuestas de modelos que describan el problema introducido en el Capítulo 1, así como formulaciones de Programación Lineal Entera que encuentren soluciones óptimas (según diferentes criterios).

Este capítulo se limita a formular las distintas propuestas de modelos. En el Capítulo 4 se detallará el proceso a través del cual estos modelos fueron implementados, mientras que en el Capítulo 5 se exhibirán los resultados obtenidos.

3.1. Un ejemplo ilustrativo

Con el objetivo de ilustrar más fácilmente los puntos centrales de cómo operan los modelos propuestos, se ha construido un ejemplo "de juguete". A lo largo de este capítulo se irá mostrando cómo actuarían los diversos modelos al buscar la mejor distribución de votantes para el ejemplo.

En la Figura 3.1 se muestra lo que podría ser una asignación arbitraria de votantes para nuestro ejemplo. En el mismo consideramos un universo de 4 manzanas en las cuales hay 2 escuelas, cada una con dos mesas que admiten a 3 personas cada una. Asimismo, hay un total de 12 votantes a los cuales asignar, que están agrupados en 3 esquinas diferentes.

Se ve a simple vista que la asignación dista de ser óptima, pero justamente ese suele ser el caso con la forma en que se asignan votantes actualmente. Si el lector hace las cuentas, podrá constatar que con esta asignación, la suma de las distancias caminadas por los votantes es de 86 cuadras (entre ida y vuelta). En las secciones siguientes se verá de qué manera es mejorada esta distribución.

3.2. Modelo básico

Como primera aproximación al problema, lo natural es formular un modelo que en vez de asignar a las personas por orden alfabético (aleatoriamente), lo haga siguiendo un modelo de cercanía geográfica.

Dada la cantidad de votantes presentes en cada domicilio y la capacidad total de cada escuela (en este punto no nos interesa distinguir entre distintas mesas de votación dentro de una misma escuela), y conociendo la distancia que separa a cada domicilio de cada escuela, se busca minimizar la distancia total recorrida por todos los votantes.

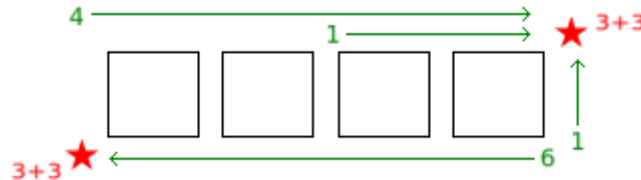


Fig. 3.1: Ejemplo inventado de una asignación arbitraria de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas). Cada número rojo indica una mesa de votación y su capacidad.

Nótese que la distancia total recorrida por una persona es el doble de la distancia que separa su domicilio de su escuela, pues la misma debe ser recorrida para ir y para volver. La solución no se ve afectada en este modelo, pero eso será un factor importante en modelos siguientes, que no miren sólo la distancia recorrida. Por lo tanto, se incluirá este factor en la función objetivo.

3.2.1. Formalización

Conjuntos

- D : Conjunto de Domicilios (agrupados en esquinas) desde los que se acude a votar.
- E : Conjunto de Escuelas en las que se puede votar.

Parámetros

- $vot_i \in \mathbb{Z}$, $i \in D$: Cantidad de votantes en el domicilio i .
- $cap_j \in \mathbb{Z}$, $j \in E$: Capacidad máxima de la escuela j .
- $dist_{ij} \in \mathbb{R}$, $i \in D$, $j \in E$: Distancia entre el domicilio i y la escuela j .

Variables de decisión

- $X_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas a la escuela j .

Función objetivo

$$\min \sum_{i,j} 2 \, dist_{ij} X_{ij}$$

Restricciones

1. Todos los votantes son asignados a alguna escuela.

$$\sum_j X_{ij} = vot_i \quad \forall i \in D$$

2. No se excede la capacidad de ninguna escuela.

$$\sum_i X_{ij} \leq cap_j \quad \forall j \in E$$

En todos los casos las sumatorias se realizan sobre la totalidad de los conjuntos en los que están definidos los parámetros y variables en cuestión.

3.2.2. Ejemplo

La Figura 3.2 muestra cómo actuaría la implementación del Modelo Básico sobre la disposición dada en el ejemplo introducido en la Figura 3.1.

Dado que se busca que la cantidad caminada en total sea lo más pequeña posible, se ve que esta asignación resulta mucho más eficiente. Mientras que en la asignación original la distancia recorrida llegaba a 86 cuadras, ahora se recorren tan solo 34 (entre ida y vuelta).

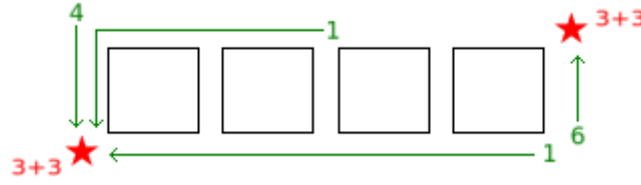


Fig. 3.2: Asignación de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas) resultante de aplicar el Modelo Básico. Cada número rojo indica una mesa de votación y su capacidad.

3.3. Modelo básico con una restricción extra

Pensando en la posibilidad de la implementación real del modelo anteriormente descrito, se decidió estudiar su comportamiento al agregar una restricción extra. Esta restricción tiene en cuenta la asignación actual de votantes (la que figura en el padrón), y procura que en la nueva asignación que se proponga, ninguna persona deba caminar una distancia mayor a la que debe recorrer actualmente. La idea de esto es que ningún ciudadano se sienta individualmente perjudicado por la adopción de este nuevo sistema (a pesar de que sepamos que el óptimo global se verá perjudicado por agregar una nueva restricción). En el Capítulo 5 se analiza la pérdida incurrida por agregar esta restricción.

Dado que este modelo es exactamente igual al detallado en la sección 3.2, salvo que se agregan un parámetro y una familia de restricciones nueva, omitimos repetir la especificación entera del modelo, formalizando tan solo el parámetro y las restricciones nuevas:

Parámetros

- $oficial_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas oficialmente a la escuela j .

Restricciones

1. Para todos los domicilios, se asegura que en ningún caso se mande a nadie a una escuela más lejana que la que tiene asignada actualmente.

$$\sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} X_{ik} \leq \sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} oficial_{ik} \quad \forall j \in E, \quad \forall i \in D$$

3.3.1. Ejemplo

La Figura 3.3 muestra la asignación resultante de aplicar el Modelo Básico con Restricción Extra, teniendo en cuenta que la asignación original era la de la Figura 3.1.

Nótese que como en la asignación original el votante que está solo había sido asignado a una escuela a 2 cuerdas de su domicilio, la restricción extra implica que luego no pueda ser enviado a la otra escuela, que le queda a 3 cuerdas de distancia. Por consiguiente, mientras que en el ejemplo correspondiente a la Figura 3.2 la distancia total recorrida era de 17 cuerdas, al agregar la restricción extra la solución óptima implica un total de 38 cuerdas (entre ida y vuelta).

Era obvio que agregar una nueva restricción implicaría que la solución pudiera empeorar. En el Capítulo 5 se estudiará con datos reales cuánto se ve afectada realmente la solución al agregar esta restricción.

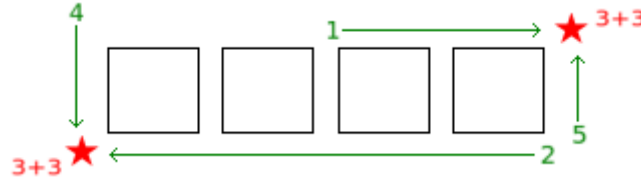


Fig. 3.3: Asignación de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas) resultante de aplicar el Modelo Básico con la restricción extra. Cada número rojo indica una mesa de votación y su capacidad.

3.4. Modelo con capacidad de escuelas variable

En las secciones anteriores se propuso un modelo que buscaba minimizar la distancia recorrida en total por los votantes. Sin embargo, nuestro objetivo no es minimizar las distancias recorridas, sino el tiempo invertido en total por los votantes que participen de los comicios. Por ende, no sólo es de interés minimizar la distancia (y por lo tanto el tiempo) que las personas tardan en ir desde sus domicilios hasta las escuelas, sino también el tiempo de espera que tendrán que aguardar en la cola del cuarto oscuro.

El tiempo promedio que una persona espere en la cola de votación dependerá de la cantidad de personas asignadas a la mesa. Si cada mesa admitiera una sola persona, es seguro que los votantes no tendrían tiempo de espera; mientras que si la capacidad fuese de mil personas, es muy probable que se formen colas sumamente largas. En la Sección 4.1 nos ocuparemos de estimar correctamente el tiempo de espera promedio para cada tamaño de mesa. Por ahora, se asumirá que éste puede ser estimado y usado por nuestro modelo en un parámetro al que denominamos $espera_k$.

Teniendo ese parámetro, se permitirá al modelo variar la cantidad de personas asignadas por mesa. Recordemos que según la legislación actual, el número de personas por mesa debe ser de 350. Nos interesa ver cuánto se podría ganar si hubiera mayor flexibilidad en este aspecto.

Nuestras variables de decisión ya no serán sólo las X_{ij} que determinan cuántas personas del domicilio i son asignadas a la escuela j . Ahora habrá que decidir también qué capacidad se permitirá tener a las mesas de la escuela j . No nos interesa distinguir entre las distintas mesas de una misma escuela, por lo que éstas tendrán la misma capacidad (pudiendo variar en a lo sumo una persona, en el caso de que la cantidad de personas asignadas a la escuela no sea divisible por la cantidad de mesas). Finalmente se introduce otra variable, que determina exactamente cuál es la capacidad de la escuela j .

De este modo, lo que se querrá minimizar es la suma entre el tiempo dedicado a viajar ida y vuelta de domicilio a escuela, y el tiempo de espera promedio.

La idea detrás de permitir aumentar o disminuir el tamaño de las mesas es que en zonas de mayor demanda, tal vez sea buen negocio permitir que las colas de votación sean más largas, dado que eso se compensará con el hecho de que las personas deben caminar menos hasta su centro de votación. Por otro lado, las colas de votación de escuelas más remotas serán más cortas, lo que compensará a los votantes que tengan que acudir desde más lejos.

Se asume que todos los votantes viajan caminando a una velocidad de $4km/h$. Esto es equivalente a $(1/15)km/minuto$. Por ende, el tiempo necesario para recorrer una distancia de $dist_{ij}km$ ida y vuelta es de $2 * 15 * dist_{ij} minutos$. Es por ello que en la función objetivo

aparece un coeficiente de 30 multiplicando a la distancia recorrida.

Por otro lado, si en una escuela las mesas de votación tienen entre $k - 1$ y k personas, el tiempo que esperarán en total las personas que asistan a esa escuela será de $espera_k * cantidad_de_personas_en_la_escuela$. Si bien hay una variable que es la cantidad exacta de personas en la escuela j , no podemos utilizarla porque la ecuación quedaría no lineal. En vez de ello, se estima la capacidad de la escuela como el producto entre la cantidad de mesas (que es un parámetro) y la cantidad de personas por mesa (que es un índice).

Finalmente, se decidió mantener la restricción extra introducida en la Sección 3.3, que fuerza a que ningún votante se vea obligado a ir a una escuela más lejana de la que le tocaba originalmente. La razón de esto es que, como se verá en el Capítulo 5, la solución no se modifica casi nada al agregar dicha restricción.

3.4.1. Formalización

Conjuntos

- D : Conjunto de Domicilios (agrupados en esquinas) desde los que se acude a votar.
- E : Conjunto de Escuelas en las que se puede votar.
- C : Conjunto con las Capacidades permitidas a las mesas de votación.

Parámetros

- $vot_i \in \mathbb{Z}$, $i \in D$: Cantidad de votantes en el domicilio i .
- $dist_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Distancia entre el domicilio i y la escuela j .
- $oficial_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas oficialmente a la escuela j .
- $espera_k \in \mathbb{R}$, $k \in C$: Tiempo de espera promedio en una mesa con k personas asignadas.
- $mesas_j \in \mathbb{Z}$, $j \in E$: Cantidad de mesas de votación disponibles en la escuela j .

Variables de decisión

- $X_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas a la escuela j .
- $capmesas_{jk} \in \{0; 1\}$, $j \in E$, $k \in C$: Variable binaria. Vale 1 si y solo si las mesas de la escuela j tienen asignadas k ó $k - 1$ personas.
- $cap_j \in \mathbb{Z}$, $j \in E$: Capacidad máxima de la escuela j .

Función objetivo

$$\min \sum_{i,j} 30 \, dist_{ij} \, X_{ij} + \sum_{j,k} espera_k \, k \, mesas_j \, capmesas_{jk}$$

Restricciones

1. Todos los votantes son asignados a alguna escuela.

$$\sum_j X_{ij} = vot_i \quad \forall i \in D$$

2. La capacidad de cada escuela es exactamente la cantidad de votantes que asisten a ella.

$$\sum_i X_{ij} = cap_j \quad \forall j \in E$$

3. La capacidad de las escuelas está condicionada por la cantidad de mesas y el tamaño que se les asigna.

$$\sum_k capmesas_{jk} (k-1) mesas_j \leq cap_j \leq \sum_k capmesas_{jk} k mesas_j \quad \forall j \in E$$

4. Las mesas de cada escuela sólo pueden adoptar un rango de tamaños $[k-1, k]$.

$$\sum_k capmesas_{jk} = 1 \quad \forall j \in E$$

5. Para todos los domicilios, se asegura que en ningún caso se mande a nadie a una escuela más lejana que la que tiene asignada actualmente.

$$\sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} X_{ik} \leq \sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} oficial_{ik} \quad \forall j \in E, \quad \forall i \in D$$

En todos los casos en que no se especifica lo contrario, las sumatorias se realizan sobre la totalidad de los conjuntos en los que están definidos los parámetros y variables en cuestión.

3.4.2. Ejemplo

Para saber cómo actuaría la implementación de este modelo sobre la disposición dada en el ejemplo introducido en la Figura 3.1, es necesario conocer cómo se modifican los tiempos de espera en función de los tamaños de la cola, y cuánto tiempo se demora en recorrer una cuadra. Como la idea de estos ejemplos es mantener las cosas lo más sencillo posible, supondremos que se tarda un minuto en recorrer cada cuadra, y que el tiempo de espera en cada mesa es igual (en minutos) a la cantidad de personas asignadas a ella.

Dado que en los ejemplos anteriores había 3 mesas de 3 personas, el tiempo de espera para todas las personas era de 3 minutos. Como había 12 votantes, el tiempo de espera total habría sido de 36 minutos. Por ende, el tiempo total para la Asignación Oficial (Figura 3.1), el Modelo Básico (Figura 3.2) y el Modelo Básico con Restricción Extra (Figura 3.3) habría sido de 122, 70 y 74 minutos, respectivamente.

La Figura 3.4 muestra el resultado de aplicar este modelo. Si el lector hace las cuentas, observará que el tiempo necesario para caminar ida y vuelta es de 26 minutos. Para calcular el tiempo de espera observemos que 4 personas están asignadas a mesas de capacidad 2 (2 minutos de espera), mientras que 8 van a mesas de capacidad 4 (4 minutos de espera). Entonces, en total, el tiempo de espera es de 40 minutos. Luego, el tiempo demorado en total habrá sido de 66 minutos, lo cual mejora todas las soluciones obtenidas hasta el momento (inclusive la del Modelo Básico, que no tenía en cuenta la restricción extra, que sí es satisfecha por este nuevo modelo).

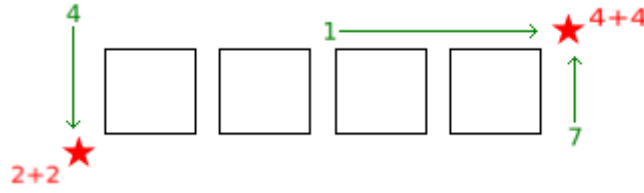


Fig. 3.4: Asignación de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas) resultante de aplicar el Modelo con capacidad de escuelas variable. Cada número rojo indica una mesa de votación y su capacidad.

3.5. Modelo con mesas intercambiables entre escuelas

En el modelo anterior, la idea detrás de permitir aumentar la capacidad de unas escuelas y de disminuir la de otras era que aumentar la capacidad de escuelas situadas en zonas de mayor demanda permitiría que más votantes tengan que caminar menos; mientras que disminuir la de escuelas en zonas de baja demanda haría que fueran menos los que tienen que caminar mucho. Sin embargo, hay que tener en cuenta que el efecto colateral de aumentar el tamaño de una mesa es que habrá un incremento en los tiempos de espera para todas las personas asignadas a dicha mesa.

En esta sección se presenta un modelo que busca aprovechar ese posible beneficio al aumentar y disminuir las capacidades de las escuelas, pero sin modificar los tiempos de espera. Lo que se hará será redistribuir las mesas de votación entre las distintas escuelas en función de dónde es realmente necesaria esa capacidad.

Una ventaja que tiene esta aproximación respecto de la del modelo anterior es que, hasta donde sabemos, no se estaría incumpliendo ninguna legislación (como la que fija el tamaño máximo de las mesas en 350 personas). Por otro lado, para poder aplicar este modelo sí sería necesario conocer los límites físicos de los distintos centros de votación, pues existe la posibilidad de que queramos colocar muchas mesas en una escuela demasiado chica para albergarlas. Sin embargo, dado que no poseemos datos sobre la capacidad de las escuelas, implementaremos un modelo sin poner límites en este aspecto. Por ende, hay que tener en cuenta que la solución que se obtenga no será necesariamente un óptimo alcanzable, sino más bien una cota superior a la mejora que se podría conseguir.

Dado que este modelo no modifica la capacidad de las mesas, el tiempo de espera será el mismo para todas las personas. Por lo tanto, la función a minimizar será la misma que en el modelo básico, es decir, la que tan sólo modifica la distancia total caminada.

Se agrega un nuevo parámetro que contiene la cantidad de mesas que hay originalmente en cada escuela, y una variable que decide para cada escuela cuántas mesas se agregan o se quitan.

Finalmente, se vuelve a incorporar la restricción introducida en la Sección 3.3, que fuerza a que ningún votante se vea obligado a ir a una escuela más lejana de la que le tocaba originalmente.

3.5.1. Formalización

Conjuntos

- D : Conjunto de Domicilios (agrupados en esquinas) desde los que se acude a votar.
- E : Conjunto de Escuelas en las que se puede votar.

Parámetros

- $vot_i \in \mathbb{Z}$, $i \in D$: Cantidad de votantes en el domicilio i .
- $cap_j \in \mathbb{Z}$, $j \in E$: Capacidad máxima oficial de la escuela j .
- $dist_{ij} \in \mathbb{R}$, $i \in D$, $j \in E$: Distancia entre el domicilio i y la escuela j .
- $oficial_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas oficialmente a la escuela j .
- $mesas_j \in \mathbb{Z}$, $j \in E$: Cantidad oficial de mesas de votación disponibles en la escuela j .

Variables de decisión

- $X_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas a la escuela j .
- $\Delta mesas_j \in \mathbb{Z}$, $j \in E$: Cantidad de mesas que se agregan (si la variable es positiva) o se quitan (si es negativa) de la escuela j .

Función objetivo

$$\min \sum_{i,j} 2 \, dist_{ij} X_{ij}$$

Restricciones

1. Todos los votantes son asignados a alguna escuela.

$$\sum_j X_{ij} = vot_i \quad \forall i \in D$$

2. No se excede la capacidad de ninguna escuela. Notar que se tiene en cuenta la capacidad final, es decir, después de tener en cuenta los movimientos de mesas de una escuela a otra.

$$\sum_i X_{ij} \leq cap_j + 350 \, \Delta mesas_j \quad \forall j \in E$$

3. Para todos los domicilios, se asegura que en ningún caso se mande a nadie a una escuela más lejana que la que tiene asignada actualmente.

$$\sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} X_{ik} \leq \sum_{\substack{k/ \\ dist_{ik} \geq dist_{ij}}} oficial_{ik} \quad \forall j \in E, \quad \forall i \in D$$

4. Los cambios de mesas entre escuelas tienen que compensarse entre sí.

$$\sum_j \Delta mesas_j = 0$$

5. Las escuelas deben tener por lo menos una mesa.

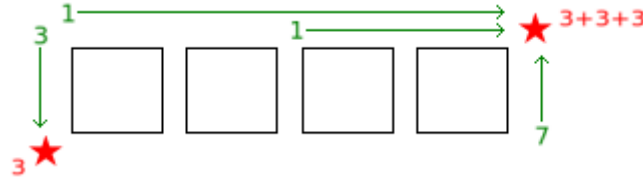


Fig. 3.5: Asignación de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas) resultante de aplicar el Modelo con mesas intercambiables entre escuelas. Cada número rojo indica una mesa de votación y su capacidad.

$$\Delta mesas_j \geq -mesas_j + 1 \quad \forall j \in E$$

En todos los casos en que no se especifica lo contrario, las sumatorias se realizan sobre la totalidad de los conjuntos en los que están definidos los parámetros y variables en cuestión.

Restricción opcional Como se ha mencionado previamente, a la hora de implementar este modelo se optó por no restringir la cantidad de cambios de mesas permitidos entre escuelas. Sin embargo, esta restricción puede agregarse fácilmente, del siguiente modo.

1. Se permite a lo sumo M movimientos de mesas por escuela.

$$|\Delta mesas_j| \leq M \quad \forall j \in E, \text{ para algún } M \in \mathbb{Z} \text{ fijo.}$$

Por otro lado, es razonable pensar que diversas escuelas tengan distintos tamaños, por lo que mientras algunas podrían recibir varias mesas más, otras tienen una capacidad menor. En ese caso, habría que introducir un nuevo parámetro $M_j \in E$ que determine el una cota para la cantidad de mesas extra que puede albergar la escuela j . No es necesario incluir una cota inferior, dado que el modelo ya exigía que ninguna escuela quedase vacía.

Es este escenario, se puede introducir la restricción siguiente.

1. La escuela j admite a lo sumo M_j mesas más de las que ya tiene.

$$\Delta mesas_j \leq M_j \quad \forall j \in E.$$

3.5.2. Ejemplo

Volviendo al ejemplo introducido en la Figura 3.1, la Figura 3.5 muestra cómo sería la solución arrojada por el modelo presentado en esta sección.

Se ve que la distancia de ida y vuelta recorrida en total es de 32 cuadras. Por otro lado, como se vuelve nuevamente al caso de 4 mesas de 3 personas, el tiempo de espera vuelve a ser de 36 minutos. Luego, el tiempo total transcurrido será de 68 minutos; algo ligeramente peor al modelo 2, pero aún superior al modelo básico.

3.6. Modelo con mesas intercambiables entre escuelas y capacidad variable

Finalmente, presentamos un modelo que mezcla las características de los dos modelos anteriores. Se permitirá la reubicación de mesas entre escuelas, y a su vez la capacidad de dichas mesas podrá ser modificada.

Para decidir sobre la ubicación de mesas y capacidades se usará una variable binaria similar a la utilizada anteriormente, sólo que esta vez $capmesas_{jkl}$ decidirá no sólo si en la escuela j la capacidad de las mesas es de $k - 1$ ó k votantes, sino además si se ponen l mesas. Más allá de ese detalle, las restricciones y función objetivo son análogas a las utilizadas anteriormente.

3.6.1. Formalización

Conjuntos

- D : Conjunto de Domicilios (agrupados en esquinas) desde los que se acude a votar.
- E : Conjunto de Escuelas en las que se puede votar.
- C : Conjunto con las Capacidades permitidas a las mesas de votación.
- M : Conjunto con la cantidad de Mesas posibles que puede tener una escuela.

Parámetros

- $vot_i \in \mathbb{Z}$, $i \in D$: Cantidad de votantes en el domicilio i .
- $dist_{ij} \in \mathbb{R}$, $i \in D$, $j \in E$: Distancia entre el domicilio i y la escuela j .
- $oficial_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas oficialmente a la escuela j .
- $espera_k \in \mathbb{R}$, $k \in C$: Tiempo de espera promedio en una mesa con k personas asignadas.
- $mesas_j \in \mathbb{Z}$, $j \in E$: Cantidad oficial de mesas de votación disponibles en la escuela j .

Variables de decisión

- $X_{ij} \in \mathbb{Z}$, $i \in D$, $j \in E$: Cantidad de personas del domicilio i asignadas a la escuela j .
- $capmesas_{jkl} \in \{0; 1\}$, $j \in E$, $k \in C$, $l \in M$: Variable binaria. Vale 1 si y solo si en la escuela j hay l mesas, y éstas tienen asignadas k ó $k - 1$ personas.
- $cap_j \in \mathbb{Z}$, $j \in E$: Capacidad máxima de la escuela j .

Función objetivo

$$\min \sum_{i,j} 30 \, dist_{ij} \, X_{ij} + \sum_{j,k,l} espera_k \, k \, l \, capmesas_{jkl}$$

Restricciones

1. Todos los votantes son asignados a alguna escuela.

$$\sum_j X_{ij} = vot_i \quad \forall i \in D$$

2. La capacidad de cada escuela es exactamente la cantidad de votantes que asisten a ella.

$$\sum_i X_{ij} = \text{cap}_j \quad \forall j \in E$$

3. La capacidad de las escuelas está condicionada por la cantidad de mesas y el tamaño que se les asigna.

$$\sum_{k,l} \text{capmesas}_{jkl} (k-1) l \leq \text{cap}_j \leq \sum_{k,l} \text{capmesas}_{jkl} k l \quad \forall j \in E$$

4. La cantidad de mesas de cada escuela sólo puede ser una, y el rango de tamaños $[k-1, k]$ tiene que ser el mismo para todas.

$$\sum_{k,l} \text{capmesas}_{jkl} = 1 \quad \forall j \in E$$

5. Para todos los domicilios, se asegura que en ningún caso se mande a nadie a una escuela más lejana que la que tiene asignada actualmente.

$$\sum_{\substack{k/ \\ \text{dist}_{ik} \geq \text{dist}_{ij}}} X_{ik} \leq \sum_{\substack{k/ \\ \text{dist}_{ik} \geq \text{dist}_{ij}}} \text{oficial}_{ik} \quad \forall j \in E, \quad \forall i \in D$$

6. La cantidad total de mesas luego de la reasignación debe ser la misma que la que había originalmente.

$$\sum_{j,k,l} \text{capmesas}_{jkl} l = \sum_{j,k,l} \text{capmesas}_{jkl} \text{mesas}_j$$

En todos los casos en que no se especifica lo contrario, las sumatorias se realizan sobre la totalidad de los conjuntos en los que están definidos los parámetros y variables en cuestión.

3.6.2. Ejemplo

La Figura 3.6 muestra cómo sería la solución arrojada por el modelo presentado en esta sección al aplicarse al ejemplo introducido en la Figura 3.1. Se ve que en este caso lo óptimo resulta tanto de mover una mesa como de modificar las capacidades.

Ahora la distancia total recorrida es de 26 cuadras, mientras que el tiempo de espera es de 38 minutos. Por ende, trabajando aún bajo el supuesto de que caminar una cuadra consume un minuto, el tiempo total invertido sería de 64 minutos, lo que constituye una solución mejor que todas las obtenidas hasta este momento.

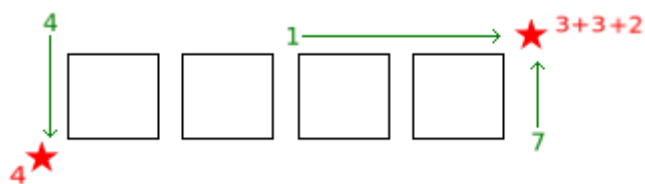


Fig. 3.6: Asignación de 12 votantes (marcados como números verdes) a 2 escuelas (marcadas con estrellas rojas) resultante de aplicar el Modelo con mesas intercambiables entre escuelas y capacidad variable. Cada número rojo indica una mesa de votación y su capacidad.

4. IMPLEMENTACIÓN

4.1. Modelado de las colas de votación

Las colas de votación fueron modeladas como filas M/M/1, siguiendo la notación introducida en la Sección 2.3. En este esquema, el tiempo de interarribo representa el tiempo que transcurre entre que llega una persona y la siguiente al lugar de votación. Por otro lado, el tiempo de servicio es el tiempo que demora una persona entre que la autoridad de mesa toma sus datos y comprueba su identidad, entra al cuarto oscuro para buscar su boleta, y sale para depositar su voto en la urna. A menudo conviene no tener en cuenta la toma de datos como tiempo de servidor, pues las autoridades de mesa suelen adelantar este trámite mientras el cuarto oscuro está aún ocupado por un votante anterior. En la práctica, la toma de datos sólo consume tiempo si a la llegada de un votante el servidor estaba ocioso (o, muy ocasionalmente, si alguna persona pasó muy poco tiempo dentro del cuarto oscuro).

Es natural considerar procesos markovianos tanto a los tiempos interarribos como al tiempo de servicio, pues no hay motivo para suponer que la historia sobre los tiempos de llegada o los tiempos de votación pasados condicionen de alguna manera los futuros.

4.1.1. Recopilación de datos durante las elecciones nacionales de 2013

En 2013 hubo elecciones nacionales en dos oportunidades: las elecciones primarias del 11 de agosto y las generales del 27 de octubre. Ambas ocasiones fueron aprovechadas para recopilar datos sobre el comportamiento de las colas de votación a lo largo del día.

La metodología consistió en acudir a una misma escuela en tres distintos momentos del día, durante entre una y dos horas en cada caso, y observar el comportamiento de dos mesas simultáneamente.

Nuestro principal interés era poder estimar los parámetros que describían tanto la llegada de personas a la mesa de votación, como el tiempo que demoraban las personas en votar. Para ello se contaba con una grilla en la que se anotaba el tiempo en que alguien nuevo llegaba a la cola, o en que una persona entraba o salía del cuarto oscuro. Con estos datos más tarde se pudo reconstruir el comportamiento de la cola (incluyendo la longitud de la cola en cada momento).

En la elección de octubre, además de estos datos, se observó el género y la edad aparente de cada votante. Para el caso de la edad, se intentó ubicar a las personas en tres franjas etarias: de 16 a 30 años, de 31 a 50, y de más de 50. Esta estimación resultaba sumamente imprecisa, pero fue realizada de todos modos con el objetivo de detectar algún comportamiento muy llamativo en caso que existiera.

En un primer momento se quiso observar más mesas simultáneamente, pero la dificultad de registrar todo lo ocurrido en muchas mesas diferentes durante los momentos de mayor actividad hizo que se decidiera bajar el número de mesas observadas a dos.

A grandes rasgos, se realizaron observaciones entre las 9 y las 11 horas; entre las 12 y las 14; y entre las 16 y las 18. En todos los casos, en cada franja se observaron dos mesas la mitad del tiempo, y otras dos mesas diferentes la otra mitad. Este conjunto de cuatro mesas fue el mismo durante las tres franjas horarias. Al finalizar ambos días, se registraron

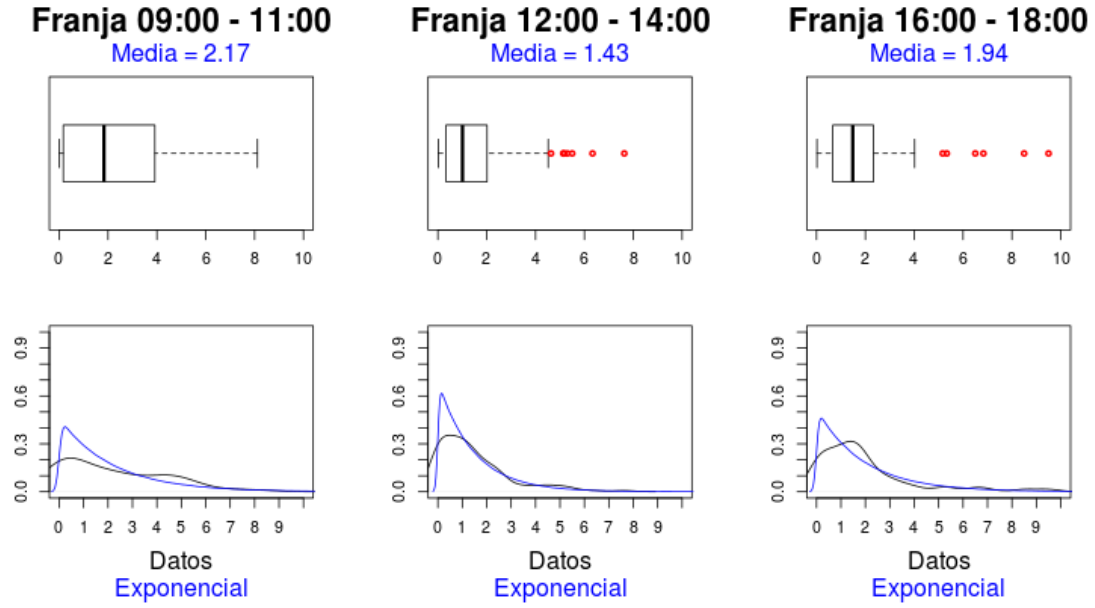


Fig. 4.1: Distribución de los interarrivals de la muestra de Agosto para cada franja horaria, y comparación con una distribución exponencial teórica.

261 llegadas y 258 salidas en las elecciones de agosto, y 188 llegadas y 206 salidas en las de octubre, totalizando 519 y 394 registros con los que realizar el análisis.

4.1.2. Procesamiento y análisis de los datos obtenidos

Se realizó el mismo análisis para ambas muestras (agosto y octubre), con el agregado de unos estudios más en octubre, usando los datos de género y edad. Trataremos esos estudios al final de esta sección.

Tiempos de interarribos

Se separó los datos recolectados ambos meses en sub-muestras agrupadas por franja horaria y mesa observada. En cada sub-muestra se calculó el tiempo de interarribos como la diferencia entre los horarios de llegadas consecutivas. Para facilitar el análisis y la visualización de los datos, se decidió trabajar con muestras más simplificadas, unificando los datos por grandes franjas horarias, a saber: 09:00-11:00, 12:00-14:00 y 16:00-18:00.

Dentro de cada franja se calculó la media entre todos los tiempos de interarribos como forma de estimar la tasa media de llegadas. Este parámetro, al que como se vio en la Sección 2.3 se suele denominar λ , nos dice cuantas personas llegan por minuto en promedio.

Además de ello, es habitual que en colas M/M/n el tiempo de interarribos tenga una distribución exponencial de parámetro λ . Por lo tanto, comparamos la densidad de nuestra muestra con la de una distribución exponencial de parámetro λ , para ver cuán buena era nuestra aproximación. Los resultados obtenidos para las muestras de Agosto y Octubre pueden observarse en las figuras 4.1 y 4.2, respectivamente.

Si bien la densidad de los datos se aproxima razonablemente a los de una exponencial teórica, observamos también que hay diferencias no menores. Una posible explicación sería

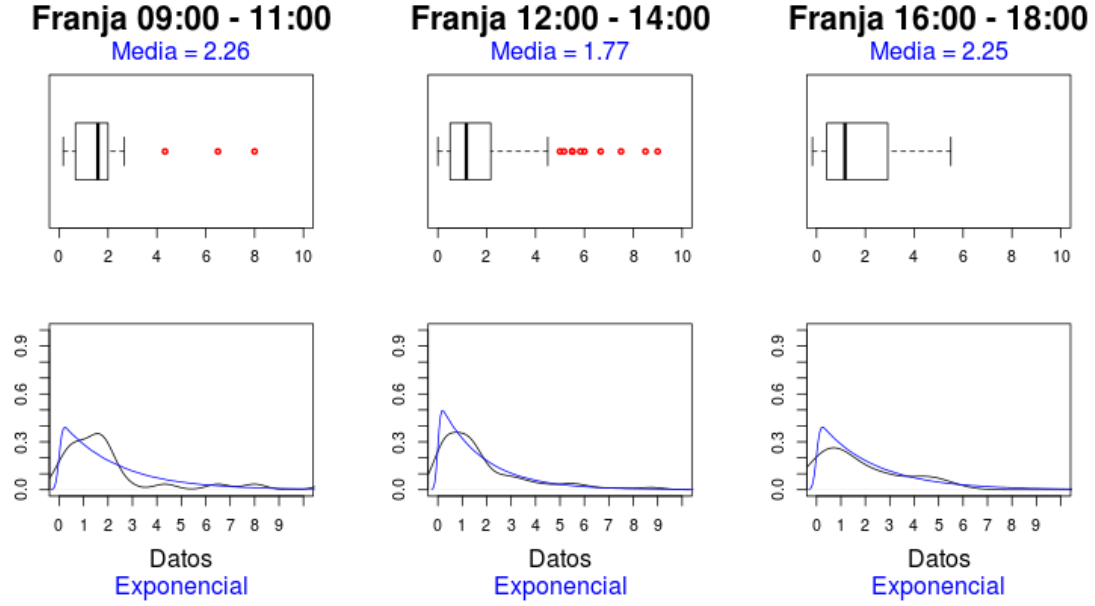


Fig. 4.2: Distribución de los interarribos de la muestra de Octubre para cada franja horaria, y comparación con una distribución exponencial teórica.

que los datos son demasiado pocos para estimar una densidad de modo más preciso. Otra, más problemática (y probable), es que la hipótesis de que el parámetro de llegada de personas es constante dentro de franjas horarias de hasta dos horas es demasiado generosa. Un examen más cercano de cada muestra basta para ver que las tasas de llegada varían dentro de cada sub-muestra (por ejemplo, se observaron muchas más llegadas entre las 12:00 y las 13:30 que entre las 13:30 y las 14:00). A pesar de ello se decidió seguir trabajando con las muestras tal como están: para una mejor caracterización de como varía λ en función del tiempo serían necesarias muchas más muestras sobre el comportamiento de las mesas de votación. Y, a efectos del alcance que se propone este trabajo, esta estimación es suficientemente buena para obtener resultados significativos.

Tiempos de votación

El siguiente paso consistió en buscar la distribución que rige el tiempo de servicio de la cola (es decir, cuánto tiempo demoran las personas en votar). Siguiendo un enfoque similar al de la tasa de interarribos, calculamos la tasa de servicio media para ver si el tiempo de votación seguía una distribución exponencial. Si bien en un comienzo ésta se calculó para cada sub-muestra por separado, un rápido análisis preliminar nos mostró que no había diferencias significativas entre una y otra. Esto tiene sentido, pues es razonable que el tiempo que tarda una persona en votar sea independiente del momento del día en que está votando, o de la mesa en que le tocó votar. Por consiguiente, volvimos a agrupar los tiempos de votación de todas las muestras, y recalculamos el tiempo de votación sobre la muestra unificada. Retomando la notación introducida la Sección 2.3, nos referiremos a este parámetro como $\bar{x}$.

En las figuras 4.3 y 4.4 se puede ver la aproximación de nuestra muestra por la distribución exponencial estimada para las muestras de Agosto y Octubre, respectivamente.

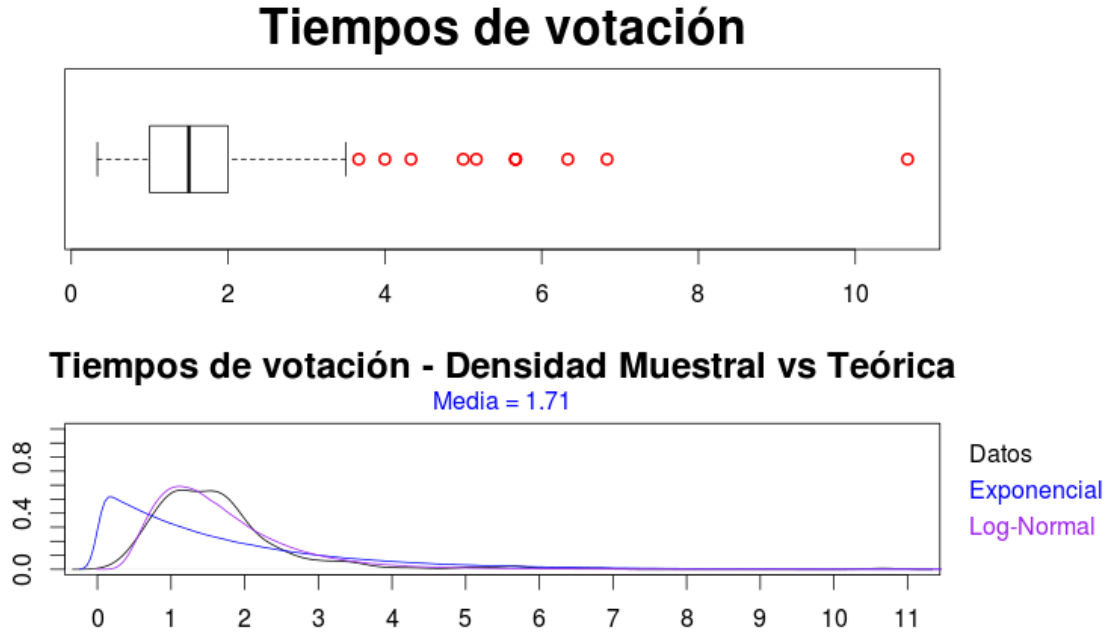


Fig. 4.3: Distribución de los tiempos de votación de la muestra de Agosto, y comparación con distribuciones exponencial y log-normal teóricas.

En ambos casos se ve los tiempos de votación no parecen seguir esta distribución. Considerando que hay un tiempo mínimo necesario para emitir el voto, es razonable que la probabilidad asignada a tiempos cercanos a 0 sea infinitesimal.

El gráfico de la densidad dado que el gráfico de la densidad muestral asemeja al de una log-normal, se decidió probar con esa distribución. Para ello tomamos los logaritmos de los tiempos de votación y calculamos su media y su desviación estándar para estimar los parámetros μ y σ de la log-normal. Como se puede observar en las figuras 4.3 y 4.4, esta aproximación resulta mucho más apropiada.

Tiempos de espera

Otro dato de mucho interés consiste en la estimación de cuánto tiempo de espera deben aguardar los votantes en la cola antes de pasar al cuarto oscuro. Este tiempo, por supuesto, variará a lo largo del día en función de la tasa de llegadas λ . Calculamos esta cantidad de 2 maneras diferentes, a las que llamaremos *empírica* y *teórica*.

Para la empírica se aprovecharon los datos recolectados durante las elecciones: al saber el estado inicial de la cola y las horas exactas a las que llegó y se fue cada persona, resulta trivial reconstruir el comportamiento de la cola a lo largo de todo el tiempo observado. A partir de esta reconstrucción es posible saber exactamente cuánto tiempo estuvo la cola en cada estado. Finalmente, con esta información se puede construir un vector $\vec{p}$ con la probabilidad de encontrar a la cola en cada uno de los estados que tomó.

Para una cola con N personas, se puede estimar el tiempo medio de espera T_N con el Resultado de Little (Proposición 2.3.1), como

$$T_N = N/\lambda.$$

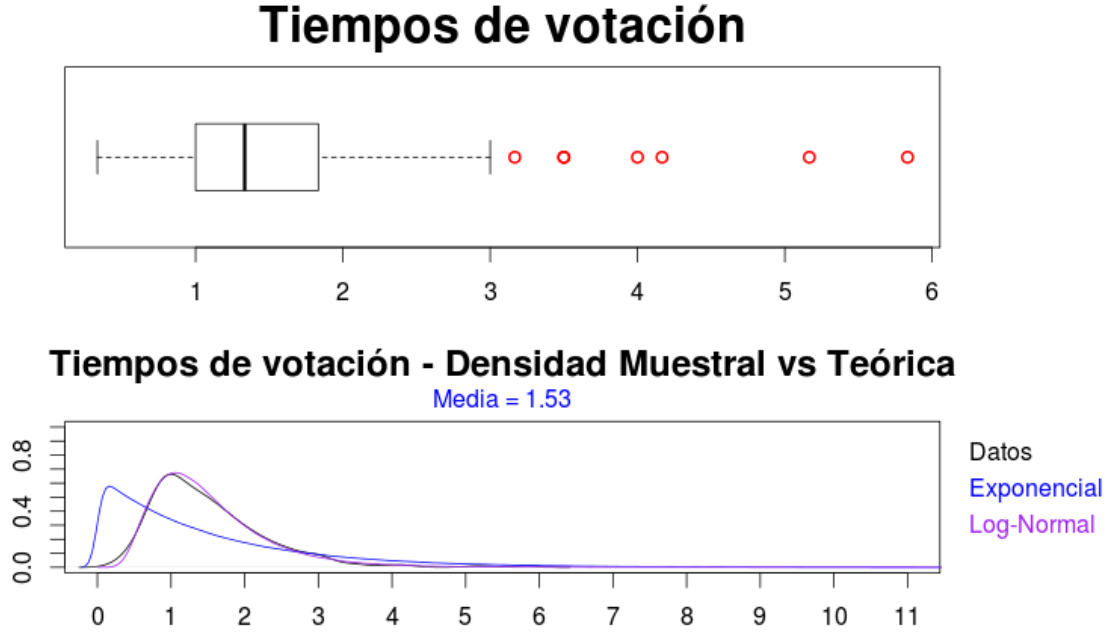


Fig. 4.4: Distribución de los tiempos de votación de la muestra de Octubre, y comparación con distribuciones exponencial y log-normal teóricas.

Finalmente, podemos obtener nuestra primera aproximación para el tiempo de espera T como

$$\mathbb{E}[T] = \vec{T}_N \cdot \vec{p}$$

Llamamos a este tiempo de espera *empírico* dado que se basa en los datos reales sobre el comportamiento de la cola.

El segundo método para estimar los tiempos de espera consiste en usar el conocimiento sobre teoría de colas en estado estacionario introducido en la Sección 2.3. Conocemos el valor de la tasa de servicio μ , que asumimos no varía a lo largo del día, y para cada franja horaria hemos calculado el valor de λ . Esto nos permite calcular, para cada franja, el ratio de servicio $\rho = \lambda/\mu$.

Dado que para las colas M/M/1 la distribución teórica del largo de las colas está dada por una Geométrica de parámetro ρ , podemos calcular el tiempo de espera esperado usando el Resultado de Little:

$$\mathbb{E}[T] = \mathbb{E}[N]/\lambda = \rho/((1 - \rho)\lambda)$$

Los resultados de ambas estimaciones pueden observarse en la Figura 4.5. La razón por la que no hay estimación "teórica" para la franja del mediodía de Agosto es que dado que la tasa de arribos fue tan grande en ese segmento, el ρ resultó ser mayor a 1, y por lo tanto el método anteriormente descrito no se pudo utilizar.

Como puede verse, las estimaciones arrojadas por ambos métodos son muy disimiles entre sí. Al no ser claro cuál de los dos utilizar, se optó por la implementación de un tercer método, que consistió en la simulación del comportamiento de las colas de votación. La totalidad del proceso se explica en detalle en la Subsección 4.1.3.

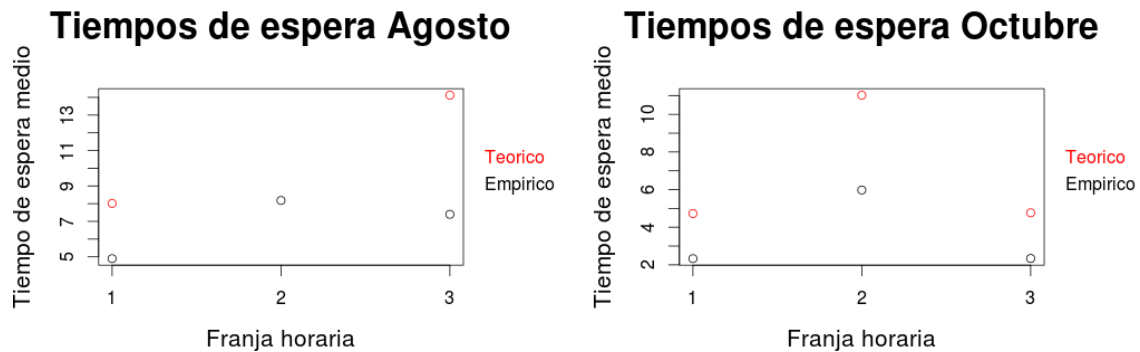


Fig. 4.5: Tiempos de espera en la cola de votación.

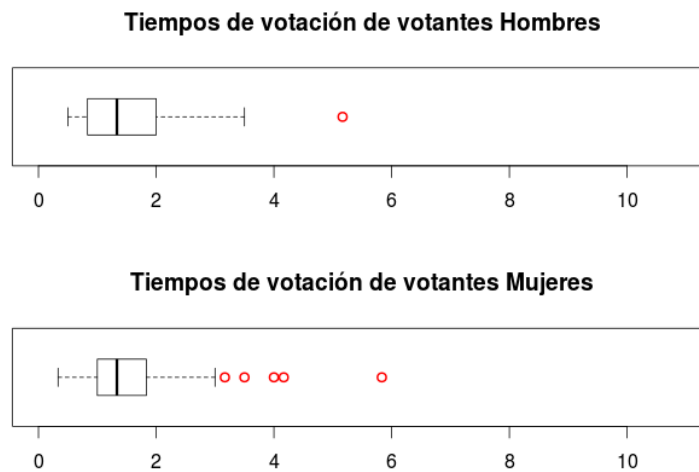


Fig. 4.6: Tiempo que los votantes tardaron en emitir su voto en Octubre, según su género.

Segmentación por edad y género

Como se adelantó en la Sección 4.1.1, durante las elecciones de octubre se observaron el género y la edad aparente de los votantes. Si bien la metodología utilizada resulta extremadamente poco rigurosa, el objetivo era buscar comportamientos que sobresalieran notablemente.

Se analizó los horarios en los que cada grupo acudía a votar, y el tiempo que tardaban en emitir el voto. En las figuras 4.6 y 4.7 se observa la distribución de los tiempos de votación diferenciados por género y edad, respectivamente. En ambos casos puede verse que tanto la mediana como la dispersión son aproximadamente las mismas para todos los casos. Nos sentimos inclinados a pensar entonces que ni el género ni la edad influyen en el tiempo que tardan las personas en votar.

Por otro lado, las figuras 4.8 y 4.9 muestran la hora elegida por los votantes observados para emitir su voto, según género y franja etaria, respectivamente. En la figura 4.8 se ve que no hay mucha diferencia entre el comportamiento de uno y otro género; lo único notable es una diferencia de 10 puntos porcentuales entre los hombres y las mujeres que

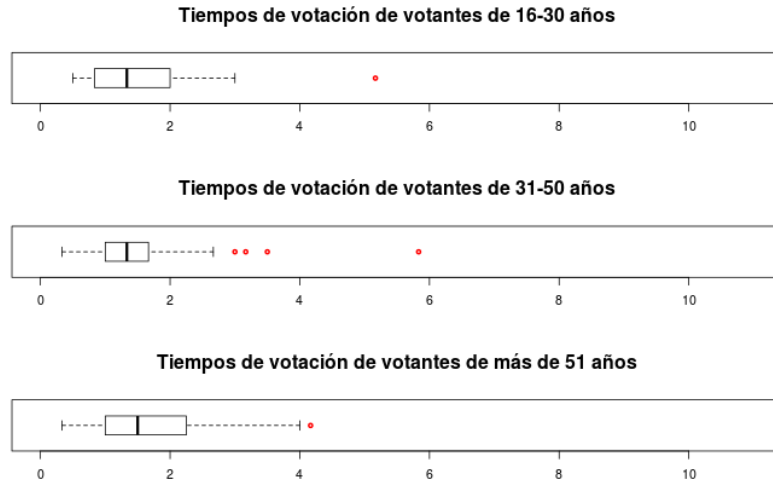


Fig. 4.7: Tiempo que los votantes tardaron en emitir su voto en Octubre, según su franja etaria.

votan a la mañana. En la figura 4.9 sí se ve una diferencia mucho más marcada por parte de los votantes más jóvenes de no acudir durante el horario de la mañana, mientras que votantes de otras franjas etarias están más uniformemente distribuidos.

Si bien la diferencia en el comportamiento de los jóvenes resulta apreciable, se consideró que no ameritaba adecuar los modelos a este hecho, por ejemplo diferenciando las tasas de llegada de votantes por género. Por la tanto, seguimos trabajando con los votantes unificados, y la posibilidad de diferenciar los comportamientos de cada grupo queda planteada para trabajos futuros en los que además se cuente con mejores formas de recolectar esta información.

4.1.3. Simulación de las colas

Para la ejecución de varios de los modelos es necesario conocer el tiempo de espera promedio en función de la cantidad de personas asignadas a una mesa de votación. En vez de usar los métodos utilizados en la subsección *Tiempos de Espera* de la sección 4.1.2, se decidió utilizar los datos recolectados para directamente simular el comportamiento de las colas.

Dado que en la elección de Agosto los tiempos de espera fueron mayores, se decidió usar los parámetros estimados en esa ocasión, para así minimizar la chance de que éstos resultasen subestimados.

Funcionamiento del simulador

El simulador fue programado en Ruby, y su funcionamiento es bastante sencillo. Lo único que es necesario pasar como input son una lista con las tasas de arribo λ que se utilizarán a lo largo del día, y otra lista con los horarios en los que hay un cambio de una franja horaria a la siguiente (es decir, cuando cambia el λ utilizado).

El programa se inicializa en tiempo 0, y en todo momento se fija cuál es el tiempo del siguiente evento a ocurrir. Tan solo hay tres eventos posibles: puede llegar una persona, puede salir de votar una persona, o puede haber un cambio de franja horaria.

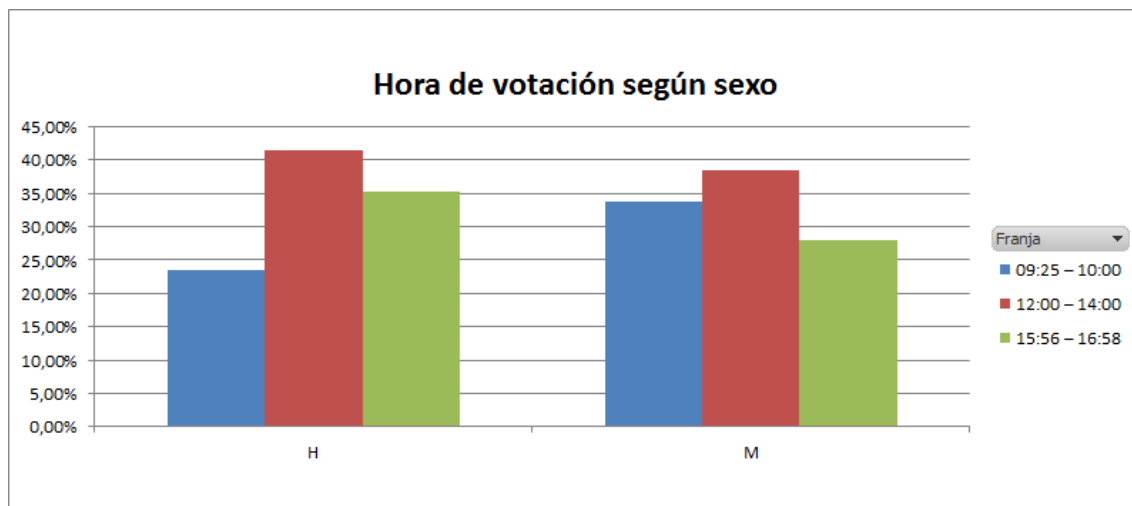


Fig. 4.8: Franja horaria elegida para votar, diferenciada por sexo.

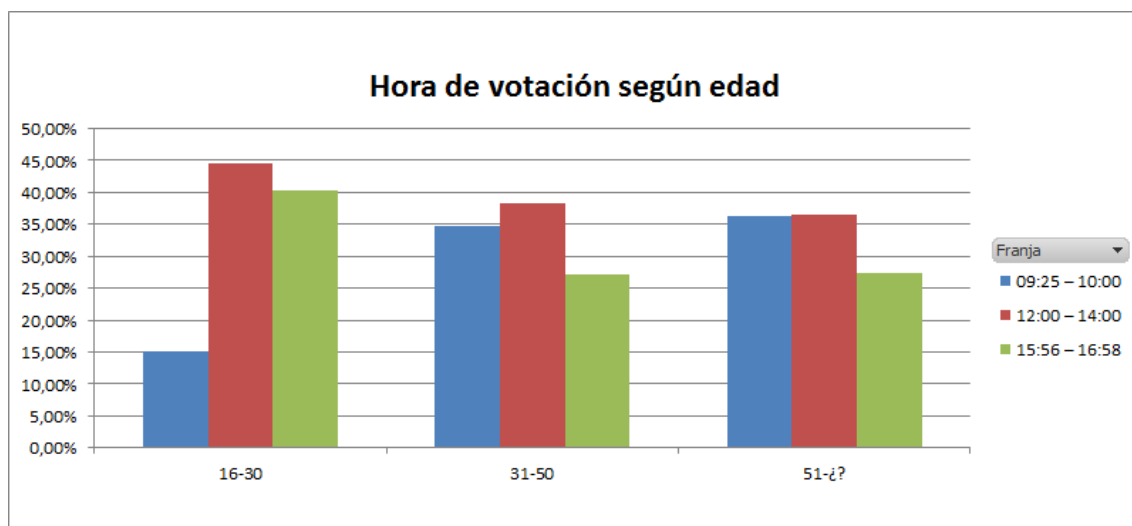


Fig. 4.9: Franja horaria elegida para votar, diferenciada por edad.

Franja Horaria	08-09	09-11	11-12	12-13:30	13:30-15	15-16	16-17	17-18
λ	0.23	0.46	0.46	0.69	0.35	0.51	0.51	0.25

Tab. 4.1: Tasas de llegada de votantes según horario, para una mesa con 350 personas asignadas.

Cuando llega una persona, el simulador suma una persona a la cola (o la manda a votar si ésta está vacía) y calcula un número aleatorio con distribución exponencial usando como parámetro al λ correspondiente a la hora actual. Ese será el tiempo a transcurrir hasta que ocurra la próxima llegada (a menos que la franja horaria cambie en el medio).

Cuando una persona sale de votar (o si llegó una persona cuando la cola estaba vacía), el simulador calcula un número aleatorio con distribución lognormal usando los parámetros calculados para la realización de la figura 4.3; a saber, $\mu = 0,3852106$, $\sigma = 0,5223903$. Este número es el tiempo que pasará hasta que la persona que acaba de entrar a votar salga del cuarto oscuro, y por ende se produzca la próxima salida.

Cuando cambia la franja horaria, se reemplaza el λ actual con el siguiente de la lista, y se actualiza la hora del próximo cambio de franja horaria. Además, se recalcula el horario de la próxima llegada (pues ha cambiado el valor de λ).

Luego, en cada paso, el simulador corrobora cuál es el siguiente evento a ocurrir. Actualiza la hora llevándola al horario de dicha ocurrencia, realiza las acciones correspondientes, e itera. Esto se repite hasta que se hayan simulado 600 minutos (es decir, son las 18hs y no se admiten votantes nuevos), tras lo cual se hace votar a las personas que queden en la cola y termina la simulación.

Al final, el programa cuenta la cantidad de personas que asistieron a votar, cuánto estuvo cada una dentro del sistema, y devuelve el promedio de estas cantidades.

Estimación de las tasas de llegada en los horarios no muestreados

Como se dijo en la subsección anterior, es necesario fijar el valor que tomará λ a lo largo de todo el día. Las mediciones que se realizaron durante las elecciones dejaron franjas horarias desconocidas, por lo que hay que asignar un valor relativamente arbitrario a esos horarios.

Contando con el dato de que la participación durante las elecciones fue de un 77 % del padrón, y que asignadas a las mesas estudiadas habían 350 personas, podemos estimar que a lo largo del día fueron a votar 269 personas.

Contando con ese dato, se partitionaron las 10 horas de votación en ocho franjas horarias, y a cada una se le asignó una tasa de llegada. En algunos casos se refinaron las franjas horarias utilizadas en la Sección 4.1, como sucedió en la del mediodía, donde el comportamiento del sistema cambiaba considerablemente luego de las 13:30. Conjeturamos que las personas que asisten a votar al abrir los comicios y al final del día son pocas, por lo que a esas franjas les asignamos un λ igual a la mitad de sus horarios vecinos. En otros casos, se optó por dejar el mismo valor que alguno de los vecinos inmediatos. Se manipuló un poco los límites de las franjas horarias desconocidas y los valores a asignar para que el valor de votantes esperados para el final del día fuera de 269 personas. En Tabla 4.1 se observa el resultado final (con valores de λ redondeados):

Remarcamos que en caso de una aplicación real de estos métodos, sería necesario realizar este estudio con mayor rigurosidad. Sin embargo, la utilización de estos datos es suficiente para tener una primera aproximación a la hora de pensar en los resultados que se podrían obtener al aplicar los modelos propuestos.

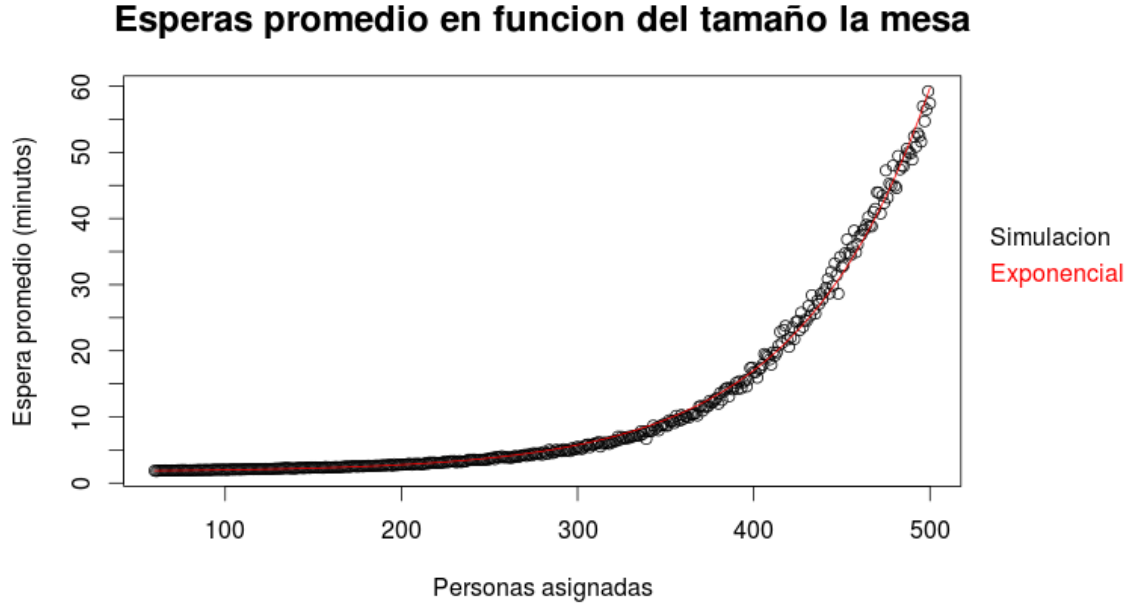


Fig. 4.10: Los círculos marcan el tiempo de espera promedio resultante de 100 simulaciones, para cada cantidad de personas asignadas. La curva es la aproximación de los puntos por una función exponencial.

Cálculo de tasas de llegada según cantidad de personas asignadas a la mesa

Los valores de λ propuestos en la Tabla 4.1 son para el escenario actual, en el que la cantidad de personas asignadas a una mesa es de 350 personas. Para la implementación de los modelos propuestos en el Capítulo 3, sin embargo, es necesario estimar el tiempo de espera promedio a medida que se varía el número de personas asignadas a la mesa.

Para eso se correrá una simulación en la que se variarán las tasas de llegada. Esto es sencillo, si consideramos que los datos de la Tabla 4.1 son representativos de cómo se distribuye la gente a lo largo del día. De este modo, para averiguar la tasa de llegadas λ_k resultante de asignar k personas a una mesa, basta hacer lo siguiente:

$$\lambda_k = \lambda_{350} * (k * 0,77) / 269$$

El factor $0,77$ es necesario para tener en cuenta la cantidad de gente que acudirá realmente a la votación, y no la que está asignada. De este modo, la cantidad de gente esperada para votar a lo largo del día es la que se supone $(k * 0,77)$, y se distribuirá del mismo modo que el dictado por la Tabla 4.1.

Realización de las simulaciones

Finalmente se varió la cantidad de personas asignadas a cada mesa entre 60 y 500. Para cada uno de dichos valores, se efectuaron 100 corridas diferentes del simulador y se tomó el promedio del tiempo de espera medio resultante de cada corrida.

Se tomaron estos valores y se los graficó. El resultado puede observarse en la Figura 4.10. Dado que la correlación es claramente exponencial, se utilizó R para fittear los puntos. En la misma figura se puede observar el resultado.

La formula resultante queda:

$$\text{Tiempo de espera} = e^{-2,5897+0,0133*(\text{Personas asignadas})} + 1.7$$

Dado que deseamos consistencia entre los valores del tiempo de espera para cada número de personas (básicamente, asegurar que si $n < m$, el tiempo de espera asociado a n sea menor que el asociado a m), se usó esa fórmula para recalcular el tiempo correspondiente a cada cantidad de personas, y se lo volcó en un archivo de texto. Es de este archivo del que se nutrirán las implementaciones de los modelos descriptos en el Capítulo 3.

4.2. Procesamiento del padrón

Para la realización de esta tesis, se contaba con datos del padrón correspondiente a la Comuna 11 de la Ciudad Autónoma de Buenos Aires.

Cada fila del padrón representaba a una persona, y contenía 3 campos: su domicilio, la escuela en que vota, y el número de la mesa en que vota.

Lamentablemente, este padrón contaba con bastantes inconvenientes. En primer lugar, estaba desactualizado. Esto no tendría relevancia a los efectos de esta tesis si no fuera porque en 2009 una modificación al Código Nacional Electoral[7] redujo de 450 a 350 el máximo número de personas que podría haber en una mesa (junto a otras modificaciones, como la unificación del padrón femenino con el masculino). Esta situación hace que, si deseamos obtener información realista sobre cómo se desempeñarían nuestros modelos en una situación real, no podamos utilizar como input de nuestros modelos la información sobre las mesas tal cuál figuran en el padrón. Más adelante retomaremos este punto, así como las decisiones que tomamos al respecto.

No obstante, el mayor inconveniente que presenta el padrón es que el mismo no está construido como una base de datos consistente. Aparentemente, la información sobre el domicilio de cada persona se ingresa de forma mecanográfica, sin seguir un patrón común para la forma de referirse a cada calle (ni preocuparse por la correctitud ortográfica). Como consecuencia, cada calle puede encontrarse escrita de decenas de formas diferentes¹. Tampoco se separa en campos diferentes el nombre de la calle de numeración, piso o departamento (si los hubiere), lo que acarrea incluso más molestias.

Para poder llevar adelante el trabajo, resultaba imprescindible identificar todas las formas distintas en que se escribía cada calle, y llevarlas a una forma común (que además pudiera ser inequívocamente reconocida por la herramienta que luego utilizaríamos para la geocodificación). Se utilizaron algunas heurísticas que facilitaron el proceso.

En primer lugar, se hizo una lista con todos los nombres de calles que aparecían, y se eliminó de ellos todas aquellas palabras que tuvieran dos o menos caracteres. Esta medida se tomó para disminuir la cantidad de casos a revisar más tarde. De este modo, tanto la calle que aparecía como *Av Joaquín González* y la que lo hacía como *Joaquín V. González* pasan a llamarse *Joaquín González*.

Luego, con la lista ya reducida, se procedió a agrupar distintas formas de escribir calles. Para eso se creó primero una función que, dados dos strings, se fijaba si alguno estaba incluido en el otro; donde la noción de inclusión es que todas las letras del más pequeño aparezcan igual o más veces que las letras que componen al más grande. Luego se corrió un algoritmo que recorra toda la lista de direcciones y las agrupaba con el siguiente criterio: si

¹ Por ejemplo, *Joaquín V. González* es referida como *Joaquin V Gonzalez*, *Av Joaquin V Gonzalez*, *J V Gonzalez*, *J B Gonzalez*, *J Gonzalez*, *V Gonzalez*, *Gonzalez*, *Joaquin V Gonzales*; entre otros

dada una dirección existía un grupo en el que cada uno de sus miembros estuviese incluido en ella, o la incluyera, entonces se agregaba a la dirección al grupo; si no, dicha dirección pasaba a conformar un grupo nuevo (en principio formado por ella sola).

Llegados a este punto, sin embargo, fue necesario comenzar una revisión manual para terminar el trabajo empezado por las heurísticas. Como era esperable, hubo muchos casos de direcciones que hacían referencia a la misma calle pero no fueron agrupadas, en general por culpa de errores de ortografía (por ejemplo, *González* y *Gonzáles* no están incluidos uno en el otro) o por abreviar distintas partes de un nombre (por ejemplo *Paula A. de Sarmiento* y *P. Albarracín de Sarmiento*). Por otro lado, si bien se dieron menos casos, ocurrió que se agruparon calles distintas que casualmente resultaron estar una incluida en la otra, a pesar de no estar relacionadas entre sí. Finalmente, otro problema que requirió aún más revisión manual fue que ciertas calles, al quitar las palabras de dos caracteres o menos, se volvieron indistinguibles entre sí (por ejemplo *P. A. Sarmiento* y *F. P. Sarmiento*, o *J. V. González* y *A. González*). Para estos casos, se debió volver al padrón original y agrupar las calles desde allí directamente.

Una vez que se tuvo a todas las calles agrupadas, para cada grupo se eligió una forma común de denominar a la calle en cuestión. Luego se regresó al padrón original, y se cambiaron los nombres originales de las calles por los del representante de su grupo.

Al final, a pesar de que las heurísticas ahorraron la mayoría del esfuerzo, la revisión manual demandó muchas horas de trabajo. Si bien retomaremos este tema en las Conclusiones, adelantamos que esta necesidad de trabajo manual es la mayor traba que encontramos a la hora de pensar en implementar estos modelos a gran escala.

En los casos en que resultaba imposible determinar correctamente la dirección en el padrón (ya sea porque el nombre la calle era ambiguo o incorrecto, la numeración estaba ausente, o cualquier otro motivo), se decidió eliminar esa fila del padrón.

A la hora de implementar los modelos propuestos en el capítulo anterior, diferenciar cada domicilio habría implicado una complejidad innecesaria. En vez de ello, se agrupó a los domicilios cercanos entre sí, considerando que todas esas personas vivían en la esquina más cercana. De este modo llevamos la cantidad de variables necesarias para correr el modelo a una cantidad fácilmente manejable, trabajando bajo el supuesto que los pocos metros de diferencia extra que debería caminar cada persona no tienen un impacto significativo en la solución obtenida.

Lo que se hizo entonces fue redondear la numeración de cada domicilio para que ésta quedase múltiplo de 100. De este modo la distorsión que se provoca suele ser de a lo sumo 50m (considerando cuadradas de aproximadamente 100m y una numeración distribuida uniformemente).

Con el padrón en estas condiciones pasamos a geocodificar la dirección de cada domicilio y de cada escuela, proceso que explicamos en la siguiente sección.

4.3. Geocodificación, GIS, Circuitos electorales, Matriz de distancias

4.3.1. Geocodificación de las direcciones procesadas

El siguiente paso fue asignar a cada dirección de nuestro padrón sus coordenadas geográficas (latitud y longitud). Este proceso es conocido como Geocodificación.

Para geocodificar los puntos utilizamos el API de codificación geográfica de Google [8]. Su funcionamiento es bastante sencillo: se realiza una solicitud via web a través de una dirección url que contiene los parámetros deseados (nombre de la calle, numeración,

ciudad, país), y el API envía una respuesta en formato XML (o json) que contiene, entre otras cosas, las coordenadas de la dirección solicitada.

Es importante recalcar que el API de Google tiene un límite de 2500 solicitudes diarias. Si se desea hacer más, es necesario usar varias computadoras (o enmascarar el IP de la propia); o comprar la versión comercial del API, que eleva el límite a 100.000 consultas diarias. En nuestro caso esto no fue un problema, pues gracias al proceso de unificación descrito en la sección anterior, nuestras direcciones no superaban las 1600. Pero si se deseara implementar estos modelos a mayor escala, habría que tener en cuenta este problema.

Para geocodificar la totalidad de las direcciones del padrón se implementó en Ruby un programa que recorría nuestra lista de direcciones, a partir de las que creaba la url para hacer la solicitud de información al API (como la estructura de dichas urls es genérica, básicamente era cuestión de concatenar strings). Finalmente se parseaba la respuesta XML recibida en busca de la latitud y la longitud del punto recién geocodificado. Esta tarea resultó muy sencilla al usar la librería *nokogiri* de Ruby. Casi la totalidad de las direcciones pudieron ser codificadas automáticamente de esta manera, gracias al trabajo previo de llevar todas las direcciones a una forma común y correcta. En los pocos casos en que el API falló, una revisión manual fue suficiente para obtener la información deseada. Afortunadamente en esta fase las fallas fueron extremadamente pocas, por lo que el trabajo manual fue casi nulo.

4.3.2. Agrupamiento en esquinas y cálculo de la matriz de distancias

Una ventaja de contar con las direcciones geocodificadas es que luego se las puede visualizar con un sistema de información geográfica (GIS, por sus siglas en inglés). Por ejemplo, la Figura 4.11 muestra la distribución de escuelas de toda la Comuna 11. La herramienta utilizada a lo largo de este trabajo fue el Quantum GIS (QGIS)[9], que se encuentra disponible bajo licencia GNU-GPL. Además de permitir visualizar nuestra información, el uso de un GIS permite utilizar datos relacionados que otras personas o instituciones ya hayan distribuido. Por ejemplo, el gobierno de la Ciudad Autónoma de Buenos Aires tiene publicados extensos repositorios con mucha información geográfica[10]. En particular, nos fue particularmente útil la información sobre las calles de la ciudad. También se utilizaron los datos sobre las manzanas para hacer más fácil la visualización de la información.

Contar con información geográfica sobre las calles nos permite utilizar una función del GIS que, dado un conjunto de puntos sobre el mapa y la geometría de las calles, calcula una matriz con las distancias de caminata entre todos tales puntos. En nuestro caso, esos puntos serían las esquinas de la comuna, cuyas coordenadas calculamos como se detalla a continuación.

Cuando durante la geocodificación calculamos las coordenadas de los domicilios, lo que hicimos en realidad fue solicitar al API las coordenadas de cada calle a alturas terminadas en 00. Éstas, si bien son buenas aproximaciones, no son realmente las coordenadas de las esquinas. Más aún, genera el problema que en las intersecciones de calles paralelas tendremos dos coordenadas distintas aproximando a una misma esquina (una por cada calle). Para evitar este inconveniente y lograr unificar realmente todos los datos cercanos en una misma esquina, utilizamos otra función del GIS. Esta función toma un dataset de líneas (en nuestro caso, las calles de la ciudad) y devuelve otro dataset que contiene los puntos en los que dichas líneas se intersecan. Estos puntos son los que queremos que representen a cada esquina. Finalmente se calculó, para cada punto devuelto por el API, cuál es el punto-esquina más cercano a él (determinado como el que minimizaba la distancia

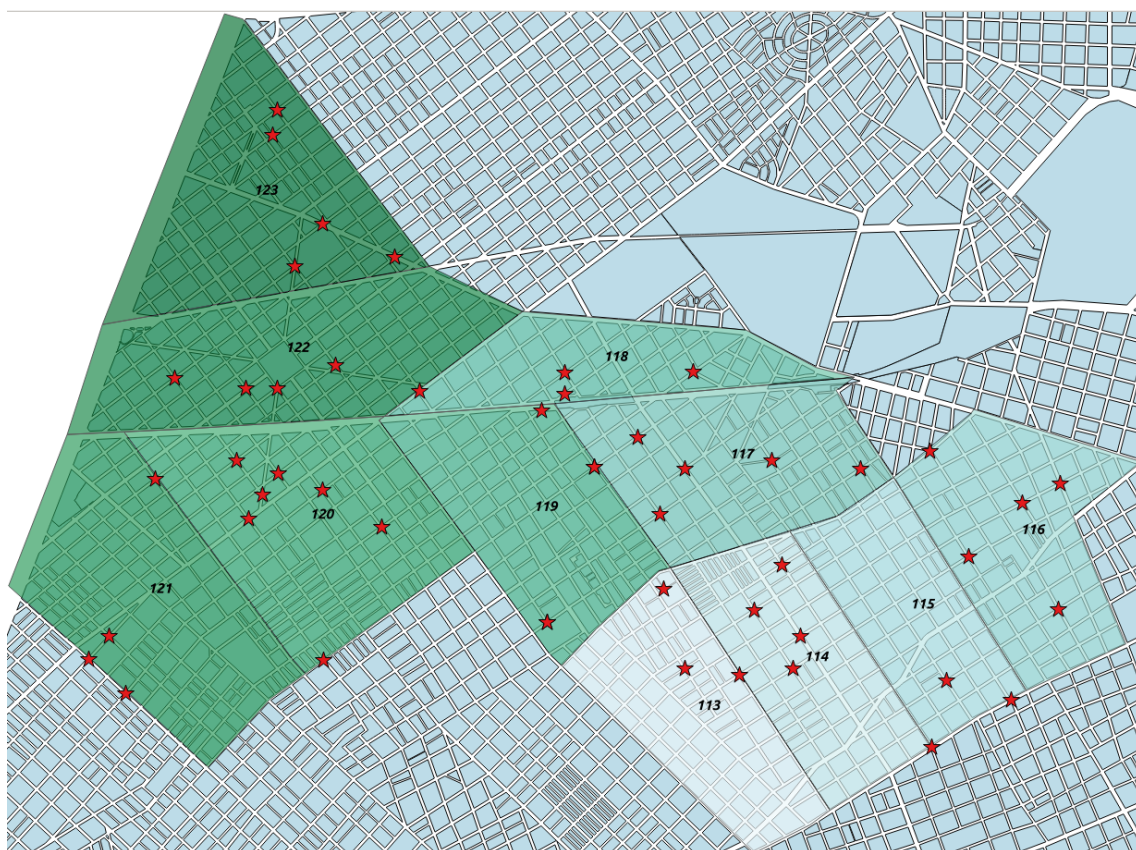


Fig. 4.11: Distribución de escuelas en cada Circuito Electoral de la Comuna 11 de la CABA. Cada estrella representa una escuela diferente.

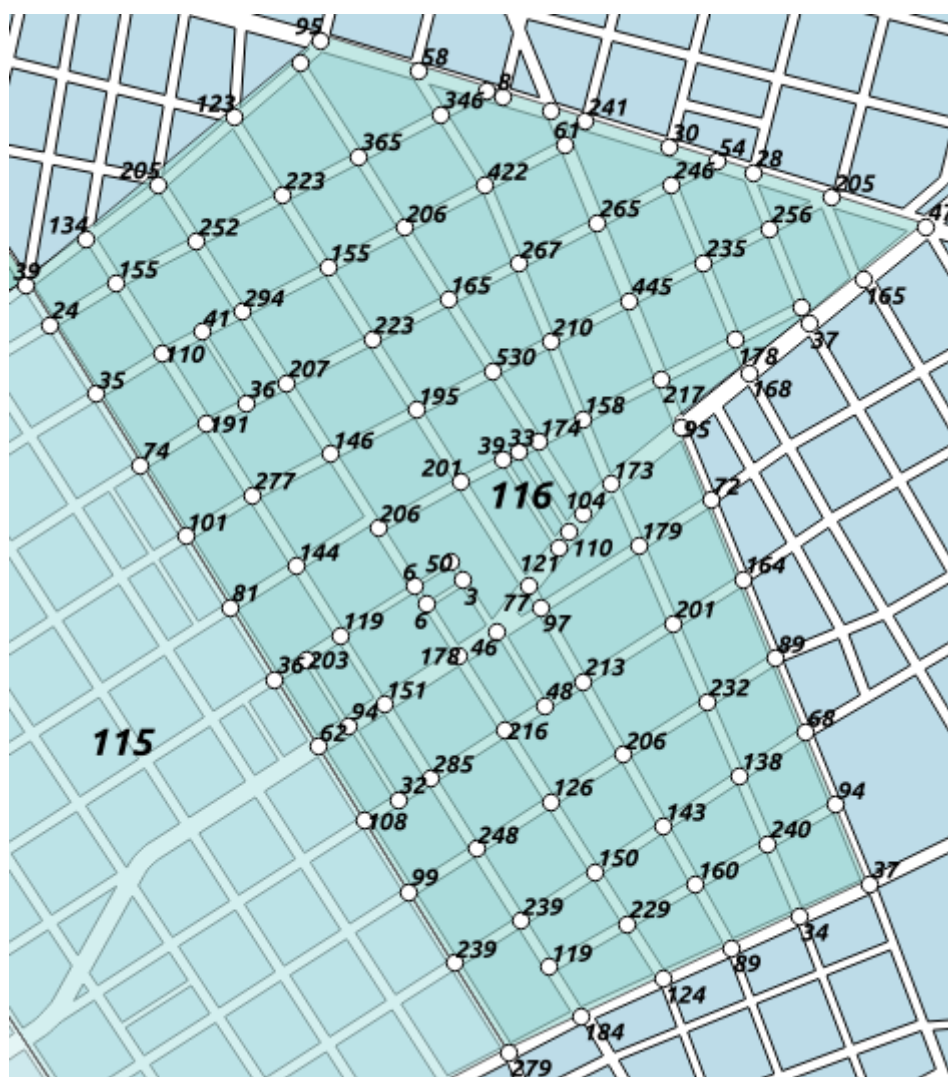


Fig. 4.12: Cantidad de votantes en cada punto-esquina del circuito 116.

euclídea), y se le asignó todos los votantes que había agrupados en el punto-API. A partir de ese momento pasamos a trabajar con los puntos-esquina.

Como último paso, desechamos a todos los puntos-esquina que no tuvieran asignados por lo menos un votante, por no tener ninguna relevancia para nuestros modelos. La Figura 4.12 muestra un ejemplo de cómo quedan agrupados los votantes.

4.3.3. Partición en circuitos electorales

Dado que muchos de nuestros modelos se atañen al Código Nacional Electoral en cuanto a que los ciudadanos deben votar dentro de su propio Circuito Electoral, debimos asignar a cada domicilio su número de circuito (el cual no figuraba entre los datos originales del padrón que nos brindaron). Las Figuras 1.1 y 4.11 muestran un mapa de la Comuna 11 y de los Circuitos Electorales que la componen.

Como los circuitos están claramente delimitados por calles, fue fácil usar el GIS para crear una capa de polígonos cuyas aristas recorrieran las calles que delimitan cada Circuito

Electoral. Luego se podría intersecar dichos polígonos con los puntos de los domicilios y así determinar a qué circuito pertenece cada punto. Sin embargo, este enfoque trae ciertos problemas. El primero es que como nuestra forma de agrupar no distingue entre veredas, en las fronteras entre dos circuitos se entremezclan los domicilios pertenecientes a ambos. Probablemente este problema se solucionaría rehaciendo la etapa de procesamiento del padrón, esta vez distinguiendo entre numeraciones pares e impares al redondear las alturas de los domicilios. Sin embargo, esto no vale la pena, dado el segundo problema que nos encontramos: resulta que el propio diseño de los padrones no respeta completamente la legislación. Al analizar con mayor detenimiento la asignación oficial de votantes a escuelas, nos encontramos con casos en los que una escuela ubicada en el circuito 120 recibe todos sus votantes del 119; u otro caso en que una escuela situada en el 119 recibe gente del 113 (que es adyacente a él). Afortunadamente lo que sí se pudo determinar es que a pesar de que en una misma escuela puedan votar personas de distintos circuitos, en las mesas de votación todos los votantes son homogéneos.

Lo que se hizo entonces fue valernos del GIS para, usando la función de intersección entre polígonos y puntos, asignar a cada *escuela* el circuito en el que se encuentra. Luego corregimos a mano los casos de escuelas que recibían a todos sus votantes de otros circuitos. Hecho esto volvimos al padrón, y a cada domicilio le asignamos el circuito de la escuela a la que está asignado oficialmente. Por último, volvimos a revisar manualmente el caso de la única escuela que recibe votantes de dos circuitos a la vez, determinando a qué circuito pertenecen los domicilios de cada mesa (algo bastante sencillo, considerando que basta determinar el circuito de un representante por mesa).

Tab. 4.2: Cantidad de votantes por Circuito Electoral

Circuito Electoral	Cantidad de Votantes
113	12249
114	15083
115	12150
116	16973
117	17640
118	13261
119	17756
120	18390
121	15010
122	13733
123	13374
Total	165619

4.4. Implementación de los modelos

En las secciones anteriores de este capítulo se detallaron los procesos utilizados para obtener los datos necesarios para correr los modelos explicados en el Capítulo 3. Habiendo geocodificado los domicilios y escuelas, calculado la matriz de distancias y los tiempos de espera asociados a cada capacidad de mesas de votación, ya se estaba en condiciones de implementar los modelos y obtener las soluciones buscadas.

Se optó por implementar los modelos usando Pyomo, un Lenguaje de Modelado Algebraico que busca tener una sintaxis similar a la de AMPL, pero que al estar escrita en Python es una pieza de Software Libre completamente gratuita (se distribuye bajo licencia BSD). Como solver se usó el GNU Linear Programming Kit (GLPK), otra pieza de Software Libre distribuida bajo licencia GNU GPL.

Cada modelo se implementó por separado (aunque por la superposición de los modelos, se reutilizó bastante código entre uno y otro). Dado que era necesario correr cada modelo con distintos parámetros cada uno correspondiente a un Circuito Electoral diferente, se utilizaron scripts que automatizaran la creación de los archivos de texto que serían usados de input para los diversos modelos. Luego se creó otro script para correr los modelos de todos los Circuitos Electorales automáticamente.

Finalmente, dado que las soluciones eran devueltas en formato *.yaml*, se utilizó la librería de Ruby *yaml* para parsear la solución y generar archivos que recopilasen tanto el valor de la función objetivo, como los valores asignados a cada variable.

5. RESULTADOS OBTENIDOS

En esta sección se exhibirán y analizarán los resultados obtenidos a la hora de implementar los modelos propuestos en el Capítulo 3.

Antes que nada, obsérvese nuevamente la Figura 4.11, que muestra la distribución de las escuelas disponibles a lo largo de la comuna. Es interesante notar que circuitos como el 113 y el 115 concentran sus escuelas en porciones muy acotadas de su territorio. Resulta esperable entonces que los resultados en zonas como estas sean peores que los de otras más homogéneas, como 116 y 118.

En las tablas siguientes, se mostrará para cada Circuito el tiempo promedio que la solución de cada modelo demandará de los votantes desde que salen de su casa hasta que regresan a ella. Algunos modelos sólo minimizan la distancia caminada, sin modificar la capacidad de las mesas. En esos casos, se considera que todos los votantes demoran el tiempo de espera asociado a mesas con capacidad para 350 personas (9.59 minutos por persona promedio). El campo *Distancia Recorrida per capita (minutos)* es el valor de *Distancia Recorrida per capita (km)* multiplicado por 15, que representa el tiempo que se demora en recorrer dicha distancia a una velocidad de 4km/h .

La Tabla 5.1 muestra las distancias y tiempos esperados según la asignación observada en el padrón del que se tomaron los datos. Los datos presentados en esta tabla serán nuestro punto de referencia para evaluar el desempeño de los modelos propuestos. La Figura 5.1 muestra en un ejemplo como la asignación de votantes oficial resulta prácticamente aleatoria, pues en prácticamente todas las esquinas los votantes son enviados a todas las escuelas posibles.

La Tabla 5.2 contiene los resultados obtenidos al implementar el Modelo Básico, descrito en la Sección 3.2. La Tabla 5.3, por otro lado, contiene los resultados de dicho modelo con la restricción extra, explicada en la Subsección 3.3.

Por otro lado, en la Figura 5.2 se contrastan los desempeños de la asignación oficial con los del Modelo Básico (con y sin restricción).

Como primera observación, notamos que los modelos obtienen una mejora significativa en todos los casos. La mínima mejora obtenida por el Modelo Básico es de 5 minutos (Cir-

Tab. 5.1: Tiempos y distancias promedio para la asignación de votantes oficial

Circuito Electoral	Distancia Recorrida per capita (km)	Distancia Recorrida per capita (minutos)	Tiempo de votación per capita (minutos)	Tiempo total per capita (minutos)
113	2.23	33.50	9.59	43.09
114	1.51	22.67	9.59	32.26
115	2.14	32.05	9.59	41.63
116	1.89	28.33	9.59	37.91
117	1.97	29.50	9.59	39.09
118	1.68	25.20	9.59	34.79
119	2.30	34.53	9.59	44.12
120	1.91	28.62	9.59	38.21
121	2.00	29.96	9.59	39.55
122	1.76	26.34	9.59	35.92
123	1.65	24.77	9.59	34.35
Total	1.92	28.73	9.59	38.32

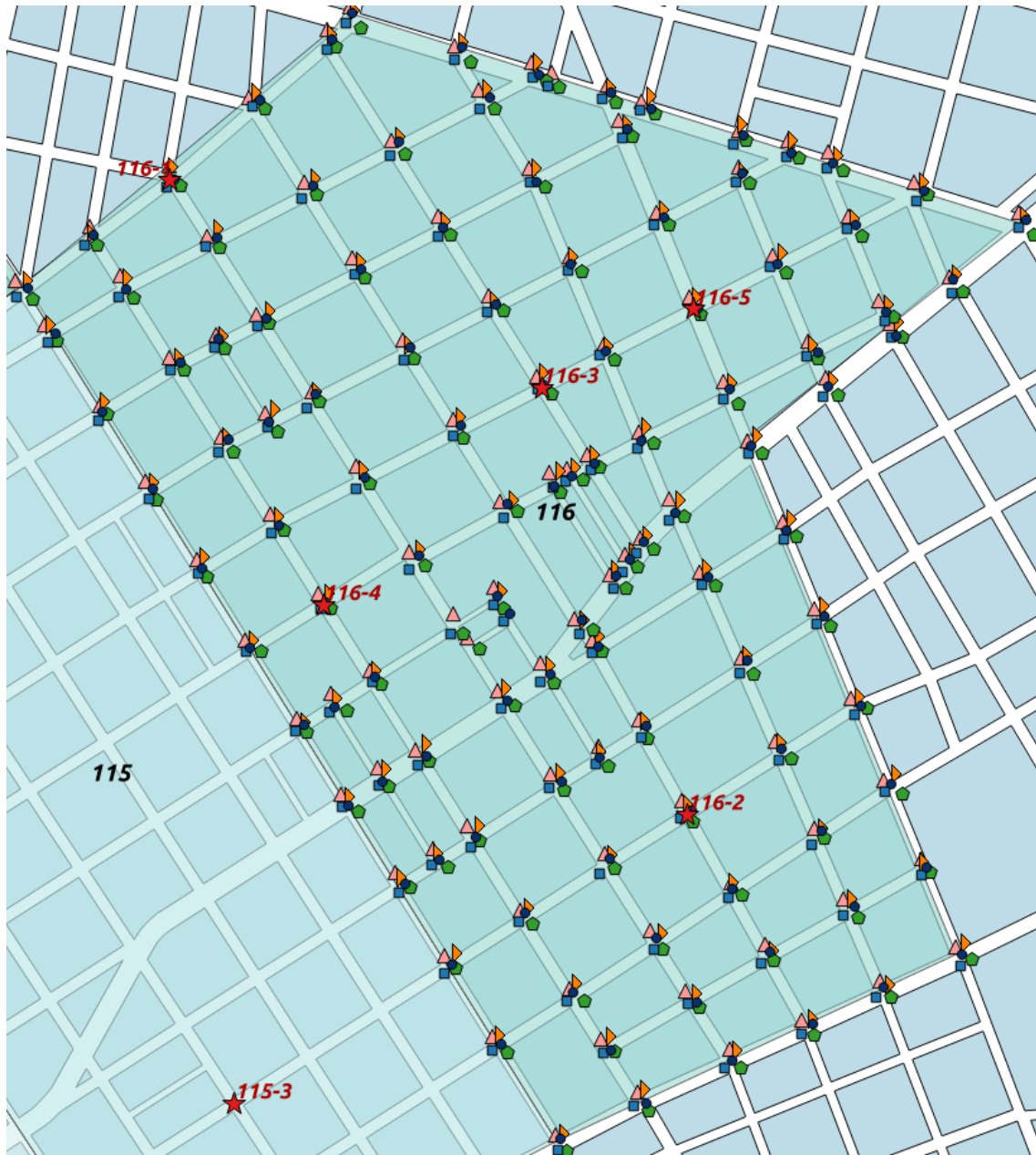


Fig. 5.1: Asignación Oficial de votantes para el Circuito Electoral 116. Las estrellas rojas representan escuelas. Cada uno de los otros símbolos representa votantes enviados a una escuela diferente.

Tab. 5.2: Tiempos y distancias promedio para la asignación de votantes del Modelo Básico

Circuito Electoral	Distancia Recorrida per capita (km)	Distancia Recorrida per capita (minutos)	Tiempo de votación per capita (minutos)	Tiempo total per capita (minutos)
113	1.70	25.52	9.59	35.10
114	1.01	15.09	9.59	24.68
115	1.80	26.97	9.59	36.56
116	0.90	13.53	9.59	23.12
117	0.94	14.16	9.59	23.75
118	0.74	11.09	9.59	20.68
119	1.51	22.59	9.59	32.17
120	1.24	18.55	9.59	28.14
121	1.37	20.49	9.59	30.08
122	1.21	18.14	9.59	27.72
123	0.95	14.28	9.59	23.87
Total	1.20	18.02	9.59	27.61

Tab. 5.3: Tiempos y distancias promedio para la asignación de votantes del Modelo Básico con la restricción extra

Circuito Electoral	Distancia Recorrida per capita (km)	Distancia Recorrida per capita (minutos)	Tiempo de votación per capita (minutos)	Tiempo total per capita (minutos)
113	1.87	28.08	9.59	37.66
114	1.01	15.21	9.59	24.80
115	1.81	27.10	9.59	36.69
116	0.91	13.71	9.59	23.30
117	0.96	14.40	9.59	23.99
118	0.75	11.31	9.59	20.90
119	1.55	23.22	9.59	32.81
120	1.25	18.74	9.59	28.32
121	1.38	20.67	9.59	30.26
122	1.24	18.61	9.59	28.19
123	0.95	14.29	9.59	23.88
Total	1.23	18.43	9.59	28.02

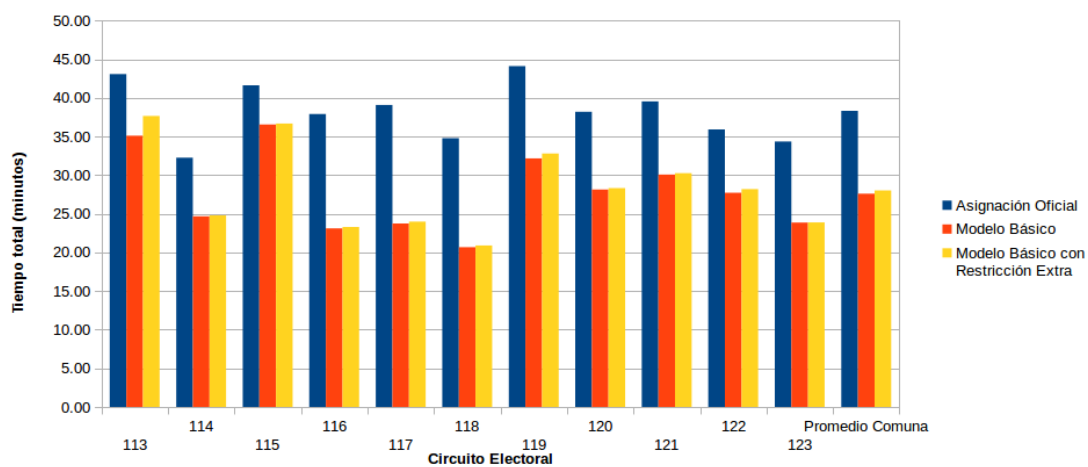


Fig. 5.2: Comparación entre los tiempos totales per cápita obtenidos por la asignación oficial y por el Modelo Básico (con y sin restricción).

cuito 115), mientras que la máxima llega a ser de 15 minutos y veinte segundos (Circuito 117). La reducción de tiempo promedio entre todas las personas de la comuna resulta ser de 10 minutos y 40 segundos.

Pero además, observamos que hay una correlación entre las mejoras obtenidas y la distribución de escuelas dentro de cada circuito electoral. Los casos en los que peores resultados se obtuvo fueron en aquellos lugares en que todas las escuelas (o casi todas) se encuentran muy cercanas entre sí, como es el caso de los Circuitos 113, 114 y, muy especialmente, 115 (ver Figura 4.11). Por otro lado, las comunas donde mejores resultados se obtuvieron (116, 117, 118) poseen una distribución bastante uniforme de escuelas. Es lógico encontrar esta correlación, pues este modelo se basa en encontrar a cada persona el establecimiento que le haga caminar lo menos posible. Si todos los centros de votación están cerca entre sí, entonces todas las personas tendrán que caminar sí o sí hasta esa zona del Circuito, sin importar cuán cerca o lejos se encuentren. Mientras que en casos como el del Circuito 116 se puede hacer que personas que antes caminaban 15 cuadras ahora caminen 2, en el 115 a lo sumo se les podrá ahorrar 3 ó 4 cuadras. En las Figuras 5.3 y 5.4 se puede observar cómo quedan distribuidos los votantes en ambos casos (115 y 116). En dichas figuras se ve claramente como en el Circuito 115 los votantes asignados a todas las escuelas se distribuyen todo a lo largo del Circuito (aunque se agrupan a lo ancho), mientras que en el 116 la distribución es mucho más compacta, con cada escuela recibiendo, en general, votantes de sus zonas circundantes.

Volviendo al análisis de la Figura 5.2, ésta permite también comparar el desempeño entre el Modelo Básico con y sin restricción. Se observa que con la sola excepción del Circuito 113, que empeora 2 minutos al agregar la restricción, en todos los demás casos la diferencia de desempeño entre uno y otro modelo es despreciable. La decisión de incorporar la restricción a todos los modelos que se probaron posteriormente se debe a haber observado que la misma no afectaba demasiado el desempeño de los modelos (mientras que agregaba una propiedad deseable a los resultados obtenidos).

En la Tabla 5.4 se observan los resultados para el Modelo 2, el cual ya tenía en cuenta el tiempo de espera en las filas de votación, y en función de ello variaba la capacidad de las escuelas. Dado que la función objetivo minimizaba directamente la suma del tiempo de espera y el que lleva recorrer ida y vuelta de domicilio a escuela, la tabla exhibe sólo el tiempo total para cada Circuito. Lo mismo sucede en el caso de la Tabla 5.6, que representa al Modelo 4 presentado en la Sección 3.6. La Tabla 5.5, por otro lado, muestra los resultados del modelo con mesas intercambiables entre escuelas (Modelo 3) presentado en la Sección 3.5, que minimiza sólo la distancia. Por ende, en ese caso el Tiempo de votación vuelve a ser el estimado para el Modelo Básico.

Dado que del modelo 2 en adelante se decidió incorporar la restricción extra introducida en la Subsección 3.3, corresponde comparar el desempeño de dichos modelos con el del Modelo Básico con restricción extra. En la Figura 5.5 se muestra dicha comparación.

Lo más notable que muestra la Figura 5.5 es que el comportamiento del Modelo 2 es casi idéntico al del Modelo Básico, mientras que lo mismo sucede entre los modelos 3 y 4. Recuérdesse que la diferencia en ambos casos es que a los modelos 2 y 4 se les permitía variar la cantidad de personas que hay en las mesas de las distintas escuelas. El hecho de que el desempeño en ambos casos sea casi idéntico nos hace pensar que no hay prácticamente ninguna ganancia en otorgar esa flexibilidad al sistema. Al menos no mientras la cantidad de personas promedio por mesa sea de 350.

En la Figura 4.10, se ve que el tiempo de espera extra en que se incurre al sumar una

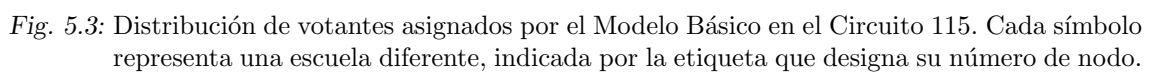


Fig. 5.3: Distribución de votantes asignados por el Modelo Básico en el Circuito 115. Cada símbolo representa una escuela diferente, indicada por la etiqueta que designa su número de nodo.

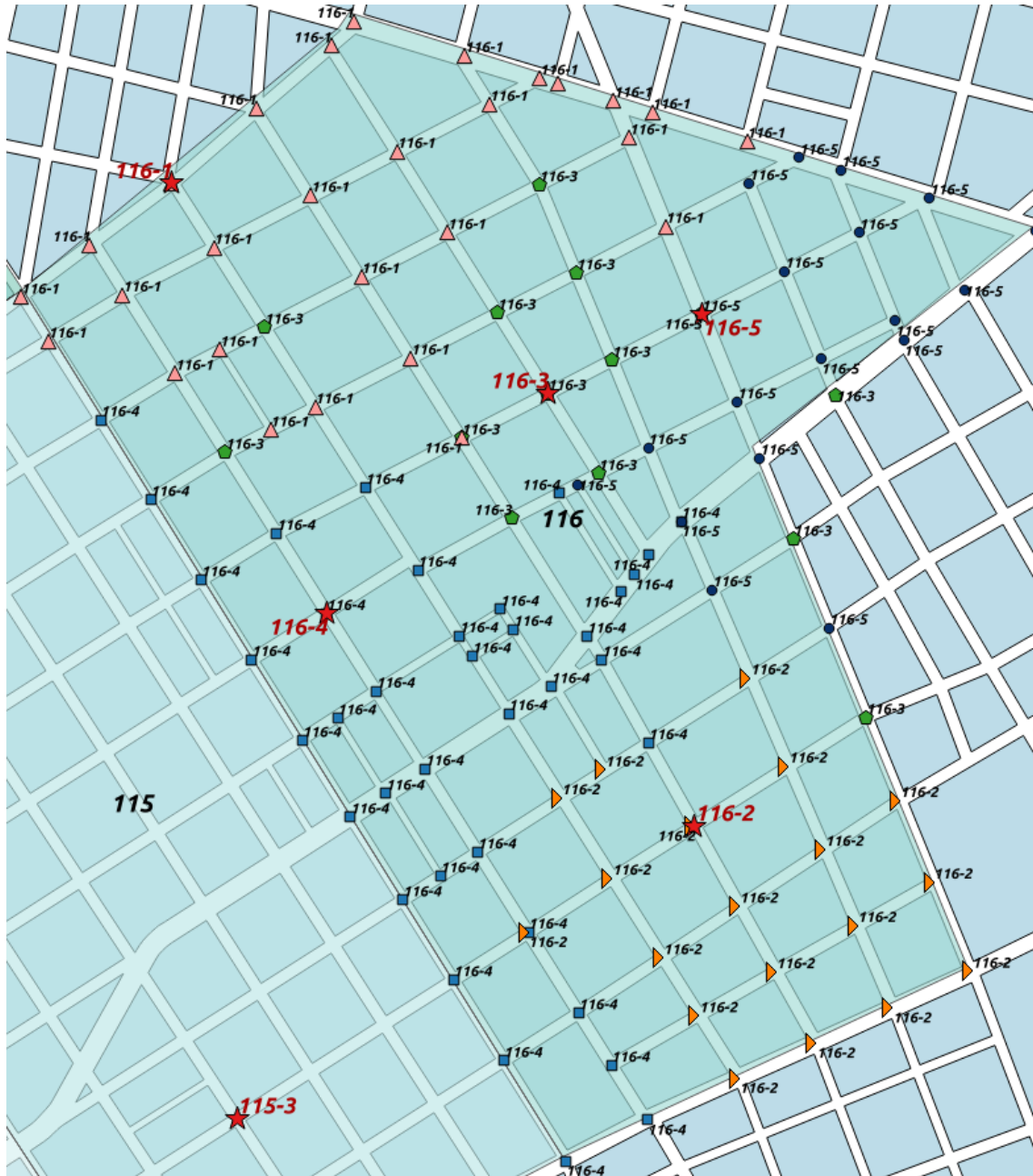


Fig. 5.4: Distribución de votantes asignados por el Modelo Básico en el Circuito 116. Cada símbolo representa una escuela diferente, indicada por la etiqueta que designa su número de nodo.

Tab. 5.4: Tiempos y distancias promedio para la asignación de votantes del Modelo con capacidad de escuelas variable

Circuito Electoral	Tiempo total per capita (minutos)
113	37.80
114	24.94
115	36.31
116	22.68
117	23.37
118	20.65
119	32.42
120	28.67
121	29.22
122	28.33
123	24.12
Total	27.80

Tab. 5.5: Tiempos y distancias promedio para la asignación de votantes del Modelo con mesas intercambiables entre escuelas

Circuito Electoral	Distancia Recorrida per capita (km)	Distancia Recorrida per capita (minutos)	Tiempo de votación per capita (minutos)	Tiempo total per capita (minutos)
113	1.35	20.24	9.59	29.83
114	0.88	13.19	9.59	22.78
115	1.67	25.04	9.59	34.63
116	0.76	11.42	9.59	21.01
117	0.74	11.05	9.59	20.63
118	0.68	10.22	9.59	19.81
119	1.30	19.57	9.59	29.16
120	1.02	15.28	9.59	24.86
121	1.27	19.01	9.59	28.60
122	0.96	14.34	9.59	23.93
123	0.92	13.85	9.59	23.43
Total	1.04	15.53	9.59	25.11

Tab. 5.6: Tiempos y distancias promedio para la asignación de votantes del Modelo con mesas intercambiables y capacidades variables

Circuito Electoral	Tiempo total per capita (minutos)
113	30.05
114	22.83
115	34.32
116	20.97
117	20.45
118	19.70
119	28.97
120	25.26
121	27.75
122	24.16
123	23.49
Total	25.05

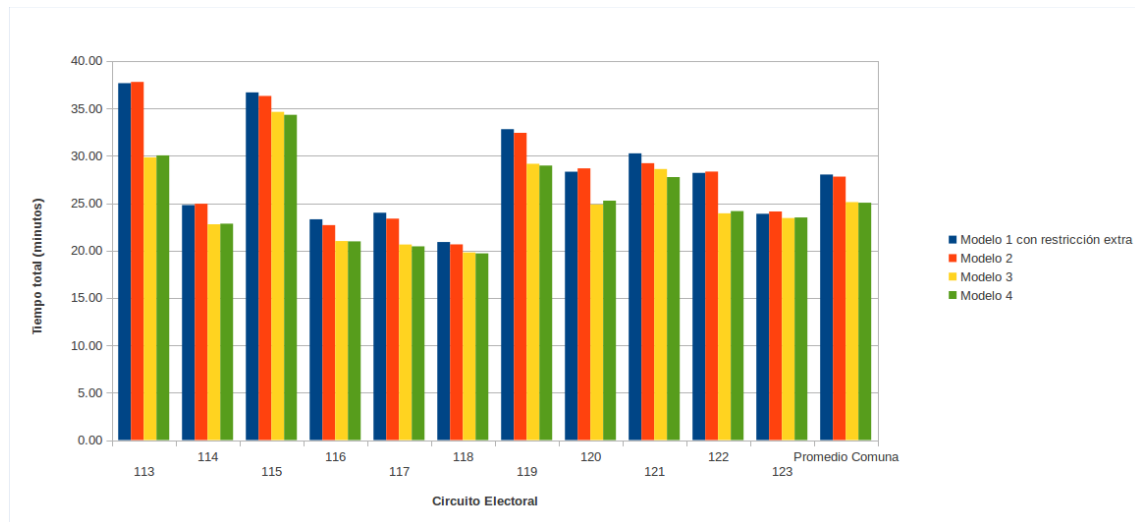


Fig. 5.5: Comparación entre los resultados obtenidos por el Modelo Básico con restricción extra y los Modelos 2, 3 y 4.

persona a una mesa es mayor que el que se gana al quitarla. Al principio esta diferencia es mínima, pero alrededor de las 350 personas la velocidad con que crece el tiempo de espera es demasiado grande como para que el intercambio resulte favorable. Por ejemplo, véase en la Tabla 5.7 que los tamaños asignados a las mesas, salvo contadas excepciones, no se alejan demasiado del tamaño promedio que tenían originalmente dichas mesas.

Por otro lado, sí hay una diferencia apreciable entre los dos primeros y los dos últimos modelos (en promedio, 3 minutos). Esto indica que es posible mejorar el desempeño del sistema si se trasladan mesas de votación de una escuela a otra. No obstante, se debe tener en cuenta que estos modelos significan una cota superior a la ganancia que se podría obtener, pues los modelos funcionan bajo el supuesto de que es posible mover las mesas de votación de un lugar a otro, cuando en realidad es posible que algunas escuelas sean demasiado chicas y no posean espacio físico suficiente para albergar demasiadas mesas. En la Tabla 5.7 se puede observar que las soluciones del modelo 4 a menudo involucran llevar grandes cantidades de mesas a una escuela (a menudo más de dos decenas), mientras que otras se dejan prácticamente desiertas. Para implementar un modelo realista, habría que poseer datos concretos sobre las capacidades concretas de cada escuela, y así poder agregar esas restricciones a los modelos.

En las Figuras 5.6 y 5.7 se muestra la distribución geográfica resultante tras aplicar el Modelo 4 en los Circuitos 115 y 116, respectivamente. A diferencia de lo observado en las Figuras 5.3 y 5.4, estas distribuciones aproximan a las que resultarían de asignar a cada domicilio a su escuela más cercana. Esto es razonable, pues la capacidad de mover irrestrictamente las mesas electorales de un lugar a otro nos permite situarlas donde sean necesarias para que todas las personas caminen lo menos posible.

Tab. 5.7: Cantidad de mesas y tamaño de las mismas para cada escuela en la solución óptima del Modelo 4.

Circuito Electoral	Código Escuela	Capacidad	Cantidad de mesas
113	96339	363	8
113	95065	323	2
113	96198	360	4
113	96239	363	20
114	96353	348	18
114	96390	354	6
114	96210	353	7
114	96208	352	12
115	96562	345	6
115	95941	345	4
115	96592	348	25
116	96570	349	11
116	52911	347	5
116	96698	353	11
116	96621	340	10
116	96582	343	12
117	96290	345	7
117	96310	339	13
117	52671	344	7
117	96235	354	8
117	96405	346	10
117	95129	353	6
118	184727	348	6
118	186082	351	13
118	184809	346	10
118	184811	350	9
119	95129	348	18
119	94208	347	4
119	184824	349	12
119	95065	348	17
120	94480	353	7
120	94682	353	22
120	94666	354	2
120	94676	356	2
120	94599	354	12
120	94486	355	7
121	82483	333	2
121	94523	340	17
121	82482	341	12
121	94565	344	13
122	184458	351	23
122	207213	354	3
122	184473	353	4
122	184353	354	9
123	184330	356	8
123	184202	350	9
123	184539	351	7
123	184541	351	9
123	184778	352	5

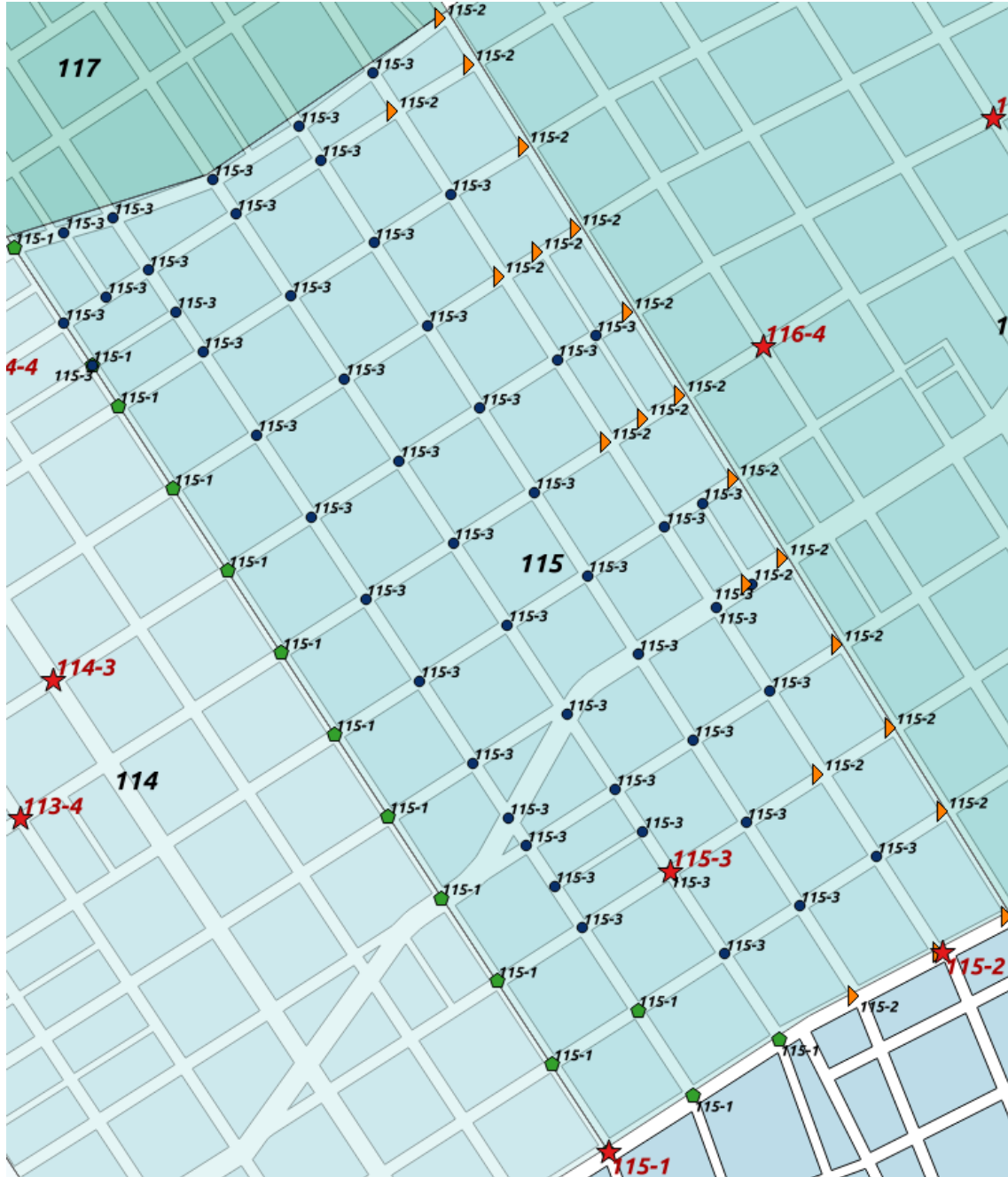
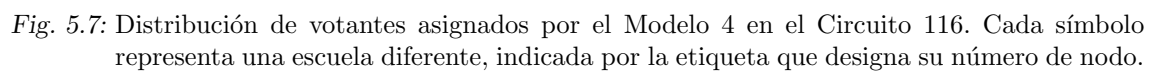


Fig. 5.6: Distribución de votantes asignados por el Modelo 4 en el Circuito 115. Cada símbolo representa una escuela diferente, indicada por la etiqueta que designa su número de nodo.



6. CONCLUSIONES

Como se mencionó en el Capítulo 5, el análisis de los resultados indica que preocuparse por otorgar a distintas escuelas distintos tamaños de mesa no tiene sentido, pues el resultado obtenido es básicamente el mismo. Este hecho sólo debería ser revisado en caso que se decida disminuir notablemente la cantidad de personas promedio que votan por mesa; pues en ese caso la ganancia marginal de hacer que algunas personas voten más cerca de su domicilio podría superar a la pérdida marginal en que se incurre al aumentar la capacidad de una escuela. Asimismo, retomando las nociones introducidas en la Sección 2.2, vemos que sacrificar una solución *justa* en la que todos los votantes tengan (en esperanza) un tiempo de espera idéntico es algo que no parece valer la pena.

Ahora bien, queda pendiente como trabajo futuro mejorar el estudio de la relación entre la cantidad de votantes por mesa y los tiempos de espera, sobretodo si se tiene en cuenta que recientemente (en el año 2009) esta cantidad fue modificada (de 450 a 350 personas por mesa), y que en caso de que esto ocurra nuevamente, tener más información que ayude a la toma de la decisión sería deseable. Para tal fin, recordamos que la estimación de las tasas de arribo de votantes descrita en la Sección 4.1 se puede mejorar bastante, en caso que se desee obtener una simulación más precisa. Esta mejora se obtendría tomando franjas horarias más pequeñas, censando la totalidad de los comicios, y tomando muestras sobre más mesas.

Los resultados obtenidos permiten concluir que sí hay mucho por ganar con el hecho de asignar a las personas a una escuela más cercana a su domicilio. La sola aplicación del Modelo Básico reduce el tiempo invertido promedio en más de un 26 %. Además, se puede incluir sin coste significativo una restricción que hace que nadie empeore su situación respecto de la asignación oficial, y de este modo evitar quejas de la población ante una eventual aplicación de los modelos propuestos.

Sí puede ayudar considerablemente la reubicación de mesas, como muestra el resultado obtenido al implementar el Modelo 3. Lo interesante es que para implementar esto no es necesario modificar ninguna legislación, sino solamente obtener información sobre las capacidades máximas de cada establecimiento.

Al comparar los desempeños de los modelos en distintos Circuitos Electorales, vemos que hay mucho que ganar si se distribuyen las escuelas uniformemente a lo largo de cada Circuito. Tener esto en cuenta requeriría un trabajo previo al de la distribución de votantes, que consiste en escoger convenientemente las escuelas que estarán abiertas para votar. Desconocemos los requerimientos legales a la hora de designar una escuela como centro de votación, y tampoco se cuenta con datos sobre posibles opciones. Por lo tanto, el estudio de este problema (ya sea por separado o tratándolo simultáneamente con el de la distribución de votantes) queda pendiente para un trabajo futuro.

Por lo pronto, es interesante notar que para implementar las sugerencias hechas en este trabajo, sólo es necesario modificar el CNE[6] en un aspecto menor, que es el que dictamina que los votantes de cada Circuito Electoral deben ser asignados por orden alfabético. Esta modificación no conllevaría (creemos) la necesidad de grandes cambios (como sí lo haría una propuesta de modificación de los Circuitos Electorales en sí mismos).

Probablemente el mayor impedimento para la implementación a gran escala de estos modelos es la mala situación en que se encuentran los padrones, en los cuales las calles

son escritas de un modo no homogéneo, y por lo tanto la tarea de agrupar las diversas direcciones en una base de datos funcional puede acarrear mucho trabajo manual. Esto es así al menos para la versión del padrón que se pudo conseguir para la realización de este trabajo (que está desactualizado en al menos 5 años). Esta no es una traba completamente restrictiva, desde luego. Tan solo demandaría trabajo adicional, ya sea manual o implementando a través de herramientas de reconocimiento automatizado de texto.

Finalmente, remarcamos que este trabajo muestra que hay mucho por ganar si se implementan algunas de las sugerencias aquí realizadas. La distribución de votantes por orden alfabético, si bien es sencilla e intuitiva, no posee ninguna virtud práctica en sí misma, y cualquiera de los modelos aquí propuestos mejora notablemente el desempeño total del sistema. Del mismo modo, vale la pena estudiar la posibilidad de una distribución más uniforme de los centros de votación a lo largo de cada distrito. Como se ha mostrado, la implementación de estos modelos es factible, y los impedimentos que se han señalado son superables sin demasiadas complejidades. Creemos entonces que las ideas aquí propuestas podrían ser implementadas próximamente al menos en algún distrito a modo de prueba, con la idea de que en el mediano o largo plazo puedan ser llevadas a escala nacional.

Bibliografía

- [1] Ahuja, R. K., T. L. Magnanti, and J. B. Orlin (1993). *Network Flows: Theory, Algorithms, and Applications*. Prentice-Hall, Englewood Cliffs, NJ.
- [2] Beckmann, M. J., C. B. McGuire, and C. B. Winsten (1956). *Studies in the Economics of Transportation*. Yale University Press, New Haven, CT.
- [3] Wardrop, J. G. (1952). *Some theoretical aspects of road traffic research*. Proceedings of the Institution of Civil Engineers, Part II, Vol. 1 , 325–378.
- [4] Kleinrock, L. (1975). *Queueing Systems, Vol. I: Theory*. Wiley Interscience.
- [5] Lenzi, Jeremías (2013). *Asignación Efciente de Votantes a Escuelas mediante Programación Matemática: El Caso de la Ciudad de Pergamino*.
- [6] Código Nacional Electoral. Ley N° 19.945.
- [7] Ley N° 26.571. Artículo 86.
- [8] API de codificación geográfica de Google. La documentación puede encontrarse en <https://developers.google.com/maps/documentation/geocoding/?hl=es>
- [9] Quantum GIS. Puede encontrarse información del proyecto en <http://www.qgis.org>
- [10] Datos geográficos publicados por el gobierno de la Ciudad de Buenos Aires. <http://data.buenosaires.gob.ar/dataset>