



UNIVERSIDAD DE CHILE  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

MODELO DE OPTIMIZACIÓN PARA LOCALIZACIÓN DE BASES DE SAMU

TESIS PARA OPTAR AL GRADO DE MAGÍSTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

**CÉSAR IGNACIO VELIS ROMÁN**

PROFESOR GUÍA:  
FERNANDO ORDÓÑEZ PIZARRO

MIEMBROS DE LA COMISIÓN:  
MARCELO OLIVARES ACUÑA  
SEBASTIÁN MAYANZ SALAS

SANTIAGO DE CHILE  
2020

## RESUMEN

El Servicio de Atención Médica de Urgencia (SAMU) es una institución chilena dependiente del Ministerio de Salud, presente tanto en la Región Metropolitana como en el resto de las regiones. Su principal labor es gestionar todas las llamadas realizadas al teléfono de emergencia 131. Durante fines del año 2018, cuando se realiza el presente estudio, el SAMU de la RM contaba con 24 bases, en las cuales se encontraban repartidas 36 ambulancias. De forma regular el SAMU planifica la construcción de nuevas bases para sus ambulancias, pero en la actualidad no se utiliza ninguna heurística o modelo matemático para tomar esta decisión. La ubicación de las bases del SAMU construidas debería reflejar la distribución espacial y responder al eficientemente al comportamiento de la demanda por los servicios de ambulancias.

El objetivo principal de esta tesis es proponer y desarrollar un modelo de optimización que permita localizar de mejor forma las bases de ambulancias del SAMU y maximizar la cobertura de demanda de este servicio de vital importancia. El principal aporte de este modelo radica en que presenta la demanda esperada de los despachos de urgencia en un cuadrante como función de la distancia y el tipo de base más cercana a este. Este modelo refleja el comportamiento observado de los despachos históricos, donde existe una mayor demanda cuando el tiempo de respuesta es menor, lo que está muy correlacionado con la cercanía de la base. Tener una demanda esperada que depende de la localización de las bases de ambulancias es muy novedoso con respecto a los modelos de localización clásicos que presentan una demanda fija dada que puede ser atendida o no. El modelo propuesto en este trabajo, al tomar en cuenta estas variables, supone una innovación en la toma de decisiones estratégicas del SAMU, generando un impacto directo en la calidad de vida de las personas.

Además, se genera un problema de máxima cobertura, utilizando como variables de decisión la ubicación de las distintas bases del SAMU. Este es un modelo de optimización Mixto de gran tamaño y muy complejo de solucionar, en el que se presenta el trade-off entre resolver el problema para un número más agregado de cuadrantes y reducir su complejidad, sacrificando la precisión de la localización o resolverlo para un número mayor de cuadrantes, pero aumentando la complejidad y tiempos de resolución del problema.

Los resultados obtenidos, muestran que utilizando una heurística que permite mover bases ya existentes en un set de opciones, aumenta hasta en un 4% la cobertura de despachos de ambulancias. Además, al resolver el problema de optimización, la cobertura de la demanda aumenta en casi un 18%. Finalmente, se proponen consideraciones que podrían hacer que este modelo que sea aún más preciso.

*“Somos lo que hacemos para cambiar  
lo que somos”*

## AGRADECIMIENTOS

En primer lugar, quiero agradecer el apoyo de mi familia, en especial a mis padres, Rosa y César por su apoyo y amor incondicional. A mi madre por su preocupación, su compañía y por creer siempre en mí, incluso cuando yo no lo hice; a mi padre, por sus buenos consejos, por entregarme las herramientas y compartir el cariño por las matemáticas conmigo. A mi hermana Sofía, mi alter ego, por enseñarme a ir más despacio y disfrutar más la vida.

También quisiera agradecer a mi profesor guía Fernando, por su tiempo, por su paciencia, por creer en mí y darme grandes oportunidades para aprender y crecer tanto profesional como personalmente. Por otro lado, quiero dar gracias al profesor Marcelo por su contante participación y comentarios en el presente trabajo, junto con por permitirme trabajar y aprender mucho de él. Además, agradecer a Sebastián quien siempre tuvo la mejor de las voluntades para el desarrollo de este trabajo, respondiendo dudas y enseñándome sobre un área que era desconocida para mí en un comienzo. Por otro lado, quiero agradecer a Linda Valdés, por su preocupación y por su paciencia para responder todas mis preguntas, que no fueron pocas, antes y durante mi estadía en el MGO. Ya podrás descansar de mí Linda.

Quiero aprovechar de dar las gracias a todos los amigos y compañeros a quienes conocí durante mi estadía en la universidad por todas las risas, el apoyo y las cervezas compartidas junto a una buena conversación; a mis amigos Valentina, Valeria, Pablo, Angi, Xavi, Mati, Tomás, Fari, Robin, Daniel, Ignacio, Manuel, Cristo, Simón, Ximena, Esteban, Seba y Fabián; ha sido un gran gusto compartir todos estos años y experiencias. La universidad no hubiera sido lo mismo sin ustedes.

A mis amigos de la vida, Rodrigo, Francisco, Diego, Pato, Brian y Mattias, les agradezco por todos los años de amistad, los viajes vividos, por siempre sacarme una sonrisa y demostrarme su apoyo. Aprovecho de agradecer a Daniela por su cariño y ánimo durante el camino.

Por último, quiero agradecer a Fernanda, quien ha sido muy importante para mí durante esta recta final que suele ser estresante y de mucho trabajo solitario, por su cariño, sus consejos, por sacarme risas en tiempos de pandemia, donde no abundan. Además, por darse el tiempo y la lata de mirar mi tesis para sugerirme mejoras. Infinitas gracias.

# TABLA DE CONTENIDO

<i>RESUMEN</i> .....	<i>i</i>
<i>AGRADECIMIENTOS</i> .....	<i>iii</i>
<i>ÍNDICE DE TABLAS</i> .....	<i>vi</i>
<i>ÍNDICE DE FIGURAS</i> .....	<i>vii</i>
<b><i>CAPÍTULO 1: INTRODUCCIÓN</i></b> .....	<b><i>1</i></b>
1.1 MOTIVACIÓN .....	1
1.2 DEFINICIÓN DEL PROBLEMA .....	3
1.3 OBJETIVOS .....	5
1.3.1 OBJETIVO GENERAL .....	5
1.3.2 OBJETIVOS ESPECÍFICOS .....	5
1.4 HIPÓTESIS .....	5
1.5 ALCANCES .....	6
1.6 ESTRUCTURA TESIS .....	7
<b><i>CAPÍTULO 2: MARCO CONCEPTUAL</i></b> .....	<b><i>8</i></b>
2.1 MODELO DE CONTEO .....	8
2.1.1 REGRESIÓN DE POISSON .....	8
2.1.2 REGRESIÓN BINOMIAL NEGATIVA .....	9
2.1.3 REGRESIÓN POISSON CERO-INFLADO .....	10
2.2 MODELOS DE LOCALIZACIÓN DE BASES .....	11
2.2.1 LOCATION SET COVERING PROBLEM (LSCP) .....	13
2.2.2 MAXIMAL COVERING LOCATION PROBLEM (MCLP) .....	14
2.2.3 MODELO TEAM (Tandem Equipment Allocation Model) .....	15
2.3 TRABAJOS RELACIONADOS .....	15
<b><i>CAPÍTULO 3: MODELO PREDICTIVO DE DESPACHOS</i></b> .....	<b><i>16</i></b>
3.1 CONSTRUCCIÓN DATA SET PARA MODELO .....	16
3.1.1 DATA DESPACHOS Y LLAMADAS AL SAMU .....	16
3.1.2 DATA PRECENSO 2016 .....	18
3.1.3 DATA BASES SAMU .....	18
3.1.4 GEORREFERENCIACIÓN DESPACHOS .....	19
3.1.5 ASIGNACIÓN DESPACHOS A CUADRANTES .....	19
3.1.6 CONSTRUCCIÓN DATA SET FINAL .....	20
3.2 CONSTRUCCIÓN MODELO PREDICTIVO .....	21
3.2.1 ANALISIS EXPLORATORIO DATA .....	22
3.2.2 NIVELES DE AGREGACIÓN DATA .....	23

3.2.3 PROPUESTAS DE MODELO PREDICTIVO.....	24
3.2.4 ELECCIÓN Y RESULTADOS DEL MODELO.....	26
3.2.5 PODER PREDICTIVO DEL MODELO.....	30
3.2.6 ESTIMACIÓN DE PARÁMETROS PARA DISTINTOS NIVELES DE AGREGACIÓN.....	31
<b><i>CAPÍTULO 4: MODELOS DE OPTIMIZACIÓN.....</i></b>	<b>33</b>
4.1 DEFINICIÓN DE PROBLEMA.....	33
4.2 COMPLEJIDAD DEL PROBLEMA.....	34
4.3 FORMULACIÓN MATEMÁTICA (PPL).....	34
4.4 HEURÍSTICA PARA LA RESOLUCIÓN DE PROBLEMA.....	38
4.4.1 LINEALIZACIÓN DE LA FUNCIÓN OBJETIVO.....	38
4.4.2 HEURÍSTICA PARA RESOLUCIÓN DE PROBLEMAS DE GRAN TAMAÑO.....	39
<b><i>CAPÍTULO 5: RESULTADOS PARA PROBLEMA DE LOCALIZACIÓN DE BASES DEL SAMU EN SANTIAGO.....</i></b>	<b>42</b>
5.1 APLICACIONES MODELO PREDICTIVO.....	42
5.1.1 ALGORITMO Y RESULTADOS DE SIMULACIÓN SIMPLE (BASE FIJADA).....	42
5.1.2 ALGORITMO Y RESULTADOS DE ELECCIÓN DE LOCALIZACIÓN DE BASES (SET DE OPCIONES).....	45
5.2 RESULTADOS MODELO DE OPTIMIZACIÓN.....	47
5.2.1 ANÁLISIS RESULTADOS RESOLUCIÓN DEL PROBLEMA.....	47
5.2.2 TIEMPOS DE RESOLUCIÓN DEL PROBLEMA.....	50
5.2.3 SENSIBILIDAD HEURÍSTICA.....	52
<b><i>6. CONCLUSIONES Y TRABAJOS FUTUROS.....</i></b>	<b>53</b>
<b><i>BIBLIOGRAFÍA.....</i></b>	<b>55</b>
<b><i>Anexos.....</i></b>	<b>58</b>
A. UBICACIÓN Y CARACTERÍSTICAS BASE DEL SAMU.....	58
B. ILUSTRACIÓN EJEMPLO MODELO LSCP.....	58
C. ILUSTRACIÓN EJEMPLO OPTIMIZACIÓN BASES TIJUANA.....	59
D. EJEMPLO RESULTADO MODELO ZIP EN SOFTWARE R.....	59
E. MAPA DE CALOR DESPACHOS DATA ORIGINAL.....	60
F: MAPA DE CENTROS DE 600 CUADRANTES.....	60
G. MAPA DE CALOR SIMULACIÓN PARA BASE EN LO BARNECHEA.....	61
H. MAPA DE CALOR CASO BASE.....	61
I.GEOLOCALIZACIÓN BASES CON DOS AMBULANCIAS.....	62
J. LOCALIZACIÓN BASES CON UNA AMBULANCIA.....	62

# ÍNDICE DE TABLAS

TABLA 1: DISTRIBUCIÓN DE LLAMADAS SAMU. ....	1
TABLA 2: MATRIZ DE CORRELACIÓN. ....	22
TABLA 3: MEDIANA DE DISTANCIA BASE MÁS CERCANA Y SUPERFICIE SEGÚN NIVEL DE AGREGACIÓN.....	24
TABLA 4: RESULTADOS MODELOS. ....	29
TABLA 5: COEFICIENTES MODELO ZIP PARA DISTINTOS NIVELES DE AGREGACIÓN. ....	32
TABLA 6 : ZONAS EVALUADAS CON NUEVAS BASES.....	43
TABLA 7: RESULTADOS SIMULACIONES.....	43
TABLA 8: RESULTADOS MODELO DE OPTIMIZACIÓN.....	48
TABLA 9: TIEMPOS DE RESOLUCIÓN PROBLEMA.....	51
TABLA 10: ANÁLISIS SENSIBILIDAD HEURÍSTICA.....	52

# ÍNDICE DE FIGURAS

FIGURA 1:UBICACIÓN BASES SAMU RM. ....	2
FIGURA 2:ESTRUCTURA DATA ORIGINAL DESPACHOS SAMU. ....	17
FIGURA 3ESTRUCTURA DATA LLAMADAS SAMU. ....	17
FIGURA 4: DATA PRECENSO 2016.....	18
FIGURA 5:ESTRUCTURA DATA BASES SAMU. ....	19
FIGURA 6: ESTRUCTURA DATA FINAL. ....	21
FIGURA 7:DISTRIBUCIÓN DESPACHOS. ....	25
FIGURA 8:SIMULACIÓN DISTRIBUCIÓN DESPACHOS. ....	25
FIGURA 9:RESULTADOS MODELO PREDICTIVO. ....	31
FIGURA 10:RESULTADOS MAPA DE CALOR PARA BASE EN EL BOSQUE.....	44
FIGURA 11:SET DE OPCIONES LOCALIZACIÓN. ....	45
FIGURA 13: RESULTADOS SEGUNDA ITERACIÓN. ....	46
FIGURA 12: RESULTADOS PRIMERA ITERACIÓN. ....	46
FIGURA 14:RESOLUCIÓN GRÁFICA MODELO. ....	50



# CAPÍTULO 1: INTRODUCCIÓN

## 1.1 MOTIVACIÓN

El Servicio de Atención Médica de Urgencia Metropolitano, de ahora en adelante SAMU, existe hace 24 años y atiende a las 52 comunas pertenecientes a la Región Metropolitana, incluyendo las que se encuentran en la periferia [1]. La Región Metropolitana de Santiago es la más pequeña de las 16 regiones del país, cuenta con una superficie de 15.402 kilómetros cuadrados. A pesar de esto, es la región más poblada, con cerca de 7 millones de habitantes [2].

El SAMU en el período comprendido entre abril de 2017 y marzo de 2018 recibió un promedio mensual de 69.343 llamadas, de las cuales, en promedio, 51.799 fueron contestadas y de las que surgieron en 5530 despachos mensuales promedio. En la Tabla 1, se puede apreciar la distribución de las llamadas en el período comprendido en los distintos meses en este rango de tiempo. Durante diciembre de 2017 y marzo de 2018 se puede notar que no se encuentra el registro de los pacientes trasladados, información con la que no contaba SAMU en el momento del estudio.

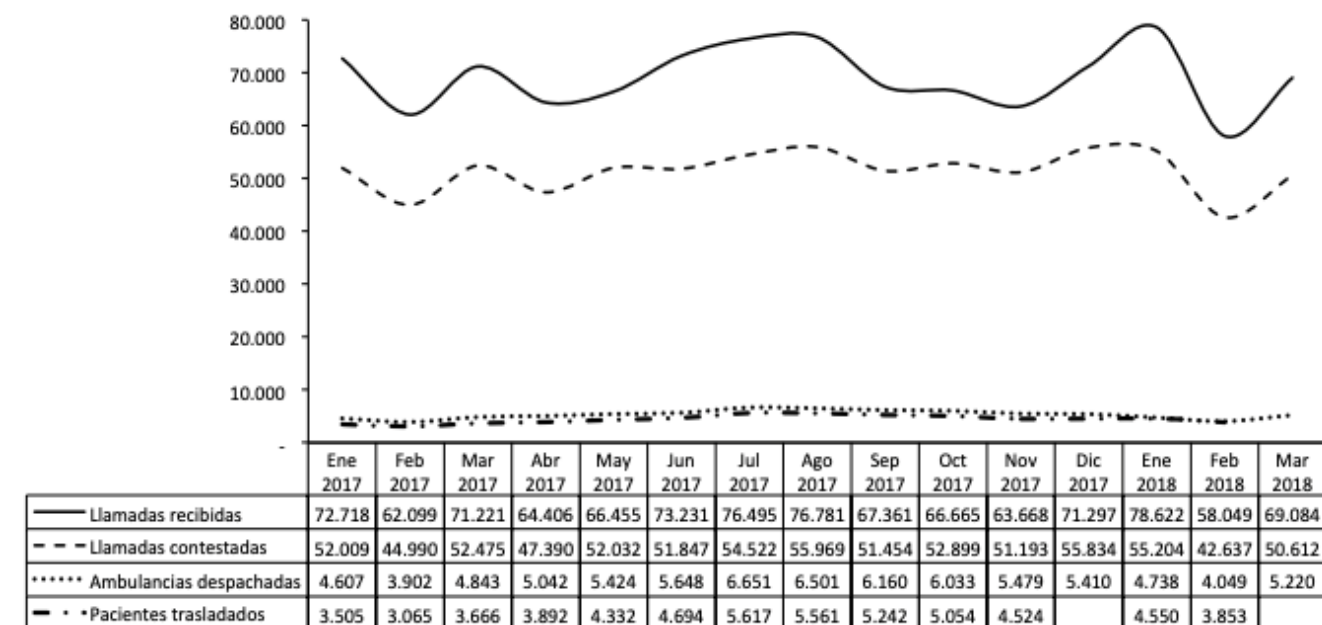


TABLA 1: DISTRIBUCIÓN DE LLAMADAS SAMU.

El SAMU para atender las distintas emergencias, a principios de 2018 cuando se empieza a realizar el presente trabajo, contaba con 24 bases, entre las cuales se distribuían 36 ambulancias (21 ambulancias básicas, 14 avanzadas y 1 ambulancia medicalizada). Las ambulancias avanzadas cuentan con el equipamiento y personal necesario para poder afrontar emergencias que involucran riesgo vital. Mientras que las ambulancias básicas tienen la función de brindar atención de

emergencia a pacientes que no presentan riesgo vital, o en caso de presentarlo, entregan atención básica inicial a la espera de una ambulancia avanzada preste la atención necesaria o trasladan al paciente al centro hospitalario más cercano. En la Figura 1 se puede apreciar un mapa de calor de la Región Metropolitana donde se muestra la distribución de las 24 bases junto con la distribución de los despachos correspondiente a los meses de septiembre y octubre de 2017. Cabe destacar, que las bases se encuentran ubicadas en centros hospitalarios y consultorios. Las zonas más rojas son las que presentan una mayor concentración de despachos, mientras que los círculos o puntos representan la ubicación de las bases. En Anexo A se encuentra la ubicación georreferenciada de las bases junto con el número y el tipo de ambulancias que posee cada una de ellas.

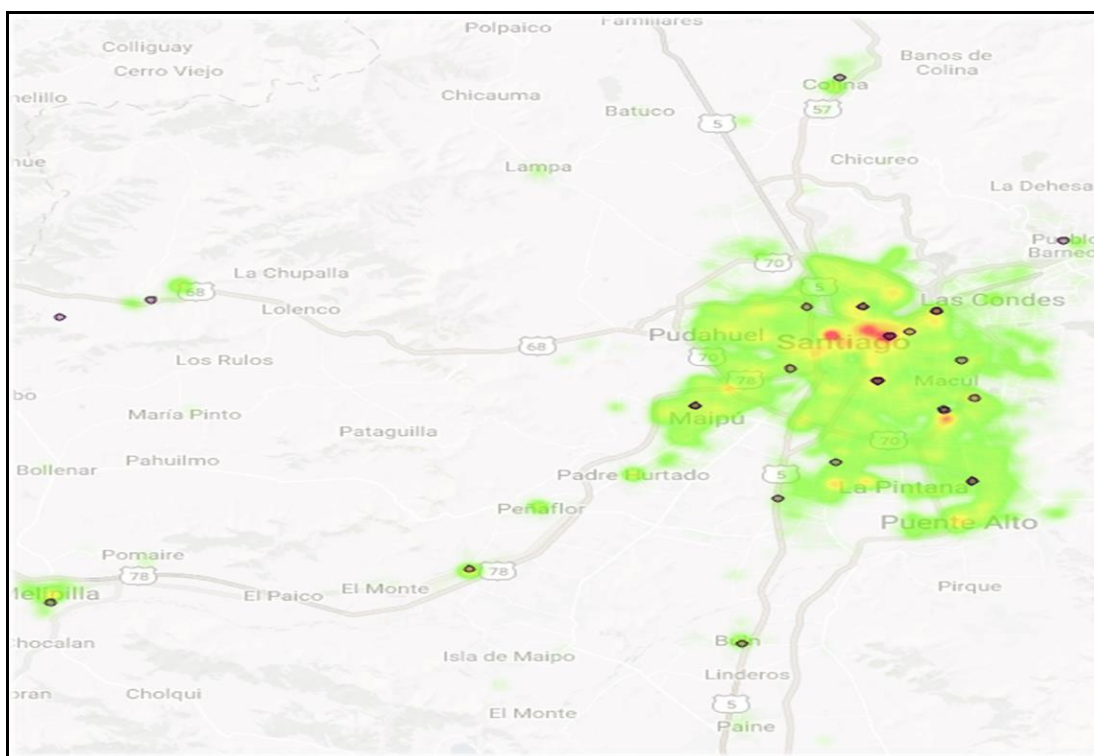


FIGURA 1:UBICACIÓN BASES SAMU RM.

El procedimiento en la actualidad que sigue el SAMU para brindar una atención eficiente es el siguiente:

1. **Recepción de llamados:** En Chile no existe un número centralizado y único de emergencias, las solicitudes de ambulancias que se efectúan al número telefónico 131 son recibidas por el Centro Regulador del SAMU Metropolitano. Esta etapa comprende desde que se recibe la llamada hasta que esta es contestada.
2. **Despacho de Ambulancias:** Una vez que la llamada es contestada, son los trabajadores del Centro Regulador quienes evalúan dicha solicitud. Si ellos consideran que es pertinente, se

comunican radialmente con las ambulancias y son despachadas al lugar de las emergencias. Esta etapa comprende desde que la llamada es contestada hasta que la ambulancia es despachada hacia el lugar del incidente.

3. **Atención de Pacientes:** Una vez que la ambulancia llega al lugar del incidente, se realiza el rescate del paciente y éste es atendido, en caso de ser necesario, por un equipo profesional que incluye paramédicos, técnicos y conductores. Posteriormente, en la mayoría de los casos, el paciente es trasladado hasta un centro hospitalario o asistencial más cercano, lugar donde se le brindan los cuidados y la atención necesaria. Esta etapa comprende desde que la ambulancia es despachada al lugar del accidente hasta que el paciente es trasladado al centro hospitalario.

Un problema actual del SAMU es que la localización de las bases es decidida solo mediante la experiencia práctica de la institución. Para determinar el lugar donde ubicar una nueva base se mira el mapa de distribución de bases actuales y en base a la experiencia de la institución y el set de lugares disponibles, generalmente centros del sistema de salud, se escoge la ubicación, lo que posee escaso nivel de análisis previo. Para establecer el lugar no se utiliza ninguna heurística ni modelo matemático, por lo que podría resultar no del todo eficiente la elección. Un modelo matemático podría dar un mejor sustento a la institución para tomar estas decisiones estratégicas, que poseen gran relevancia para la población. La principal motivación del presente trabajo es aprovechar este gran espacio de mejora y maximizar la cobertura de la demanda.

## 1.2 DEFINICIÓN DEL PROBLEMA

Como fue planteado en la motivación, el problema que se busca resolver en el SAMU es la localización y distribución óptima de sus bases, para lograr un mejor servicio a la ciudadanía, con una mayor cobertura a las distintas emergencias. En la actualidad, SAMU tiene sus bases ubicadas en centros o instituciones de salud (hospitales, consultorios, etc), entonces decide donde ubicarlas teniendo en cuenta estas limitadas opciones y analizando que zonas de la RM está más desabastecida de bases, lo que resulta tener una baja justificación desde el punto de vista de maximizar la atención de demanda y de reducir los tiempos de respuesta de esta. Por lo que un modelo matemático adecuado permitiría a la institución contar con mayores fundamentos para poder ubicar las bases de manera más eficiente.

Hay que tener en cuenta que la institución cuenta con recursos limitados, un número acotado de ambulancias, por lo que es importante distribuir de la manera más eficiente posible los recursos con los que cuenta. Además, se busca proponer una relación entre la potencial demanda de cada área con la distancia a la que se encuentra la base más cercana a esta. A medida que la base se encuentra más cerca de un área o cuadrante, el número de despachos esperados aumenta como se

analizará en el Capítulo 3. En el problema que se resuelve en este trabajo se tiene en cuenta esta relación y es por eso que no se asume la demanda por cuadrante como un parámetro o una constante determinística, sino que es más bien como una variable.

De esta manera, el problema cuenta con los siguientes elementos:

1. **Cuadrantes:** Son los nodos donde pueden ocurrir emergencias en la red. Estos nodos representan manzanas, los cuales están determinados por polígonos cuyos vértices son puntos georreferenciados. Estos cuadrantes en un principio son obtenidos del PRECENSO 2016.
2. **Despachos:** Los despachos son las situaciones de emergencia donde el SAMU envía una ambulancia a los distintos cuadrantes definidos en el punto anterior. Cada uno de los cuadrantes tendrá asignado un número de despachos, el cual corresponde al número de emergencias que se presentaron dentro del cuadrante durante los meses de septiembre y octubre de 2017.
3. **Características demográficas:** Cada cuadrante presenta características demográficas como cantidad de viviendas, superficie o densidad poblacional asociada a cada uno. Esta información es extraída del PRECENSO 2016.
4. **Bases del SAMU:** Como ya se comentó, son las 24 bases que están distribuidas en toda la Región Metropolitana, 12 de ellas tienen solo una ambulancia y las otras 12 poseen dos ambulancias en su interior.

Para resolver el problema definido, el trabajo se divide en dos etapas:

1. **Modelo de Predicción:** Se utiliza un modelo econométrico llamado Regresión de Poisson, el cual asume que los datos de despachos presentan una distribución de Poisson. Es un modelo de conteo, el cual permite saber cómo influyen las distintas variables, tanto demográficas como las distancias desde cada cuadrante a la base más cerca al número de despachos esperados.
2. **Modelos de Localización de Bases:** Una vez obtenida la función que permite predecir el valor esperado de despachos para cada cuadrante dadas sus características demográficas y las distancias de este a la base más cercana, se propone un modelo de optimización, que permite escoger la localización de las bases y lograr el objetivo que es la maximización de la cobertura de la demanda. Para este modelo se propone una heurística que permite simplificar la resolución de un modelo de gran escala como este.

## 1.3 OBJETIVOS

### 1.3.1 OBJETIVO GENERAL

Desarrollar un modelo de optimización que permita localizar de manera eficiente las 24 bases que posee el SAMU, con el fin de brindar un servicio de mejor calidad y maximizar la cobertura de emergencias.

### 1.3.2 OBJETIVOS ESPECÍFICOS

- Obtener y procesar de datos de despachos realizados por el SAMU.
- Obtener y procesar la data de cuadrantes con sus respectivas características demográficas.
- Obtener distancias desde cada cuadrante a la base más cercana.
- Desarrollar un modelo econométrico que permita predecir el número de despachos esperados y la relación matemática de este con las distintas variables explicativas.
- Agregar cuadrantes y número de despachos para reducir el tamaño del problema.
- Generar heurísticas para decidir dónde construir una base de un pool de opciones.
- Generar un modelo de optimización (PPL) que permita relocalizar todas las bases ya existentes y maximizar la función del número de despachos esperados encontrada por el modelo econométrico.

## 1.4 HIPÓTESIS

Dada la forma en que actualmente se decide la localización de las bases del SAMU, se plantea que el sistema que se utiliza para escoger la ubicación es ineficiente y no permite brindar un buen servicio ni una buena cobertura a las emergencias que surgen en la Región Metropolitana. Por lo mismo, en el presente trabajo se considera que una heurística basada en un modelo predictivo en primera instancia y luego un modelo de optimización combinados podría mejorar la situación, generando una repartición más eficiente de los recursos limitados de SAMU, obteniendo mejores resultados, maximizando la cobertura de demanda de despachos en la Región Metropolitana.

Además, se tiene la hipótesis que es posible generar el modelo que predice el número de llamadas en función de datos demográficas y tiempos de respuesta. Este modelo es desarrollado en profundidad durante el trabajo.

Finalmente, la hipótesis que se tiene en cuenta es que la distribución de las 36 ambulancias en 24 bases puede no ser la óptima. Una distribución de las ambulancias en más bases podría generar una mayor cobertura de las emergencias, pero también podría implicar un mayor costo económico para el SAMU.

## 1.5 ALCANCES

En el presente trabajo se desarrolla un modelo de optimización que permite localizar las bases del SAMU perteneciente a la Región Metropolitana. Para esto, en primer lugar, se utiliza un modelo predictivo que permite relacionar los despachos esperados para cada cuadrante en los cuales se divide la región con distintas variables, algunas demográficas pertenecientes a las características propias de cada cuadrante, las cuales son extraídas desde el PRECENSO 2016 y otras que son calculadas, como la distancia entre cada cuadrante y la base más cercana. Posteriormente, se realizan modelos de optimización, dónde se busca encontrar la localización óptima de las bases maximizando la función de demanda esperada, que es extraída desde el modelo econométrico mencionado anteriormente.

En el presente trabajo se propone una metodología innovadora con respecto a la literatura de este tipo de problemática, ya que la demanda de despachos en cada cuadrante depende de la distancia a la base más cercana, por lo que no es una constante, como en la mayoría de la bibliografía revisada con anterioridad.

Otro alcance importante es que las ambulancias básicas y avanzadas son consideradas de la misma forma con el fin de poder facilitar el estudio. Entonces, cada emergencia, independiente de su gravedad es considerada de la misma forma. Por lo tanto no hay distinción entre los dos tipos de ambulancia.

Además, en la construcción del modelo predictivo se excluyeron comunas que presentan comportamientos anómalos y que podrían generar bastante ruido en el modelo, se excluyeron algunas comunas del sector oriente de la capital por poseer una muy baja tasa de llamadas y Santiago Centro, puesto que poseía cuadrantes con un número elevado de accidentes, lo que es debido al gran movimiento y flujo que presenta la comuna, lo que difícilmente podría ser captado por algún modelo.

Dentro de las limitaciones que se encuentran en el presente trabajo, destaca que no se pudo contar con variables sociodemográficas asociadas a cada uno de los cuadrantes, estas podrían explicar de mejor forma la utilización de ambulancias en casos de emergencia, haciendo el modelo predictivo aún más preciso, ya que el PRECENSO 2016 no contaba con esta data a nivel desagregado o de cuadrantes. Otra importante limitación es la cantidad de datos de despachos con las que se contó, los datos utilizados corresponden a una ventana de tiempo de dos meses de despachos en la Región Metropolitana.

## 1.6 ESTRUCTURA TESIS

El resto de la presente tesis cuenta con la siguiente estructura: en el capítulo 2 se estudia el marco conceptual en el que se desarrolla el trabajo presentado y se comenta sobre los diferentes modelos existentes utilizados en la localización de bases en sistemas de emergencia. Además, se revisa literatura sobre el modelo econométrico que fue utilizado para predecir los despachos para cada cuadrante, que es un modelo de conteo, llamado Modelo de Poisson. En el capítulo 3, se propone la construcción y la elección del modelo econométrico utilizado, así como la estructura y procesamiento de datos que permitieron generar los resultados de dicho modelo. En el capítulo 4, se plantea la definición de diferentes modelos de optimización y heurísticas propuestas que permiten localizar las bases maximizando la función obtenida en el modelo propuesto en el capítulo anterior. El capítulo 5, se centra en el estudio de los resultados obtenidos utilizando las heurísticas y modelos de optimización presentados en el capítulo 4. Por último, el capítulo 6, se plantean en las conclusiones del trabajo y propuestas de trabajos futuros para resolver de mejor manera aún la problemática del SAMU.

## CAPÍTULO 2: MARCO CONCEPTUAL

### 2.1 MODELO DE CONTEO

Como se mencionó anteriormente, la primera parte de la tesis se enfoca en encontrar un modelo econométrico que permita predecir el número de despachos que ocurren dentro de los cuadrantes, el cual es un número entero. Es por lo anterior, que se utiliza un modelo de conteo. A continuación, se presentan los modelos de conteo más utilizados.

#### 2.1.1 REGRESIÓN DE POISSON

Un modelo predictivo para problemas de conteo es el de Regresión de Poisson. Este modelo ha sido estudiado bastante en estadística general, a continuación, se explica más en detalle la construcción de este modelo junto a sus características.

La Regresión de Poisson es un modelo lineal generalizado, en el cual la variable dependiente es discreta y sigue una distribución de Poisson de tasa  $\lambda$  y se busca entender el comportamiento de esta mediante un conjunto de variables explicativas [4]. La Regresión de Poisson se puede utilizar para predecir tanto variables categóricas como numéricas.

El modelo toma una variable dependiente que toma valores en  $\{0, 1, 2, \dots\}$  y estudia su relación con otras variables explicativas  $X$  mediante una regresión. Se pretende construir un modelo para  $\lambda(x) = E(Y|X = x)$ , es decir, la esperanza de  $Y$  condicionada a cada valor que adopten las variables explicativas. En el modelo la variable dependiente  $Y$  se representa como una tasa de Poisson igual a la exponencial de la combinación de las variables explicativas. La fórmula matemática que permite describir el modelo se puede notar en la ecuación 2.1. En ella  $a$  representa el error en la estimación y es una constante [7].

$$Y \sim \text{Poisson}(\exp(\beta'X + a)) \quad (2.1)$$

$$\log(E(Y|X)) = \beta'X + a \quad (2.2)$$

El Modelo de Poisson obtiene el valor de los estimadores  $\beta$  maximizando la función de log-verosimilitud, la cual según [12] viene dada por la ecuación 2.3. En la ecuación  $L(\cdot)$  representa la función de log verosimilitud para el estimador. En esta fórmula el índice  $i$  representa cada uno de los datos de la muestra. En [12] además, se señala que por lo general los paquetes estadísticos de los diversos softwares al correr este modelo ofrecen estimadores  $\beta$  que maximizan la función, pero a pesar de esto, es necesario para obtener inferencias válidas, verificar que se cumpla el supuesto que la media y la varianza condicional para este modelo son iguales (distribución de Poisson). Aunque se ha demostrado que a pesar de que este supuesto no se cumpla (como sucede en la mayoría de los casos), el coeficiente del estimador  $\beta$  aún es válido.



$$L(\beta) = \sum_{i=1}^n \{y_i x_i' \beta - \exp(x_i' \beta) - \log(y_i!)\} \quad (2.3)$$

Una de las principales limitaciones de este modelo según [10] es que no se hace cargo de la heterogeneidad no observada y en particular para problemas relacionados al área de salud puede ser una mala elección, debido a que estos problemas generalmente poseen muchos ceros, es decir, la variable que se busca explicar muchas veces posee un valor 0, lo que resulta no consistente con el Modelo de Poisson.

### 2.1.2 REGRESIÓN BINOMIAL NEGATIVA

No hacerse cargo de la heterogeneidad no observada puede introducir, en primer lugar, sobre dispersión y, en segundo lugar, un número excesivo de ceros [8]. Un modelo matemático de conteo que se hace cargo de la heterogeneidad no observada es la Regresión Binomial Negativa, el cual se describe con mayor detalle a continuación.

Los datos observados se distribuyen siguiendo una distribución de Poisson, pero se presupone un elemento de heterogeneidad no observada que sigue una distribución gamma, que refleja el hecho que la media no se midió de manera perfecta [11]. La distribución binomial negativa se puede apreciar en la ecuación 2.4 donde el parámetro  $\alpha \geq 0$  corresponde a un parámetro de dispersión, mientras que  $\Gamma(\cdot)$  es la función gamma. El parámetro de dispersión es el que define la relación entre la media y la varianza condicional. Por ejemplo, si  $\alpha = 0$ , entonces la media y la varianza son iguales y se obtiene el modelo de Poisson, ya que en esta distribución se cumple que la media es igual a la varianza [12].

$$f(y_i|x_i) = \frac{\Gamma(y_i + \alpha^{-1})}{\Gamma(y_i + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^{y_i} \quad (2.4)$$

La función de log-verosimilitud para este modelo viene dada por la ecuación 2.5, en este caso, a diferencia de la función log-verosimilitud correspondiente al Modelo de Poisson se optimiza dicha función en base a dos parámetros, los cuales son  $\alpha$  y  $\beta$ , trabajo que realizan paquetes estadísticos de los diversos softwares para obtener estos estimadores.

$$L(\beta, \alpha) = \sum_{i=1}^n \left\{ \sum_{j=0}^{y_i-1} \log(\alpha^{-1}) - \log(y_i!) - (y_i + \alpha^{-1}) \log(1 + \alpha \times \exp(x_i' \beta)) + y_i \log(\alpha + y_i x_i' \beta) \right\} \quad (2.5)$$

Un ejemplo de la utilización de este modelo se puede encontrar en [13] donde se utiliza para hacer un análisis tecnológico. En este trabajo los resultados son comparados con los que arroja el Modelo de Poisson con ceros inflados, que es el modelo que se presenta a continuación.

### 2.1.3 REGRESIÓN POISSON CERO-INFLADO

El modelo Binomial Negativo surge y es desarrollado para modelar e incluir la heterogeneidad no observada, de la cual no se hace cargo el Modelo de Poisson, pero dicha heterogeneidad, generalmente, se produce debido a la cantidad excesiva de ceros que se presentan. Estos casos son comunes cuando se analizan datos relativos al uso de servicios de salud, como es el caso de esta tesis [12].

De esta manera, es posible que el mecanismo aleatorio genere la repetición excesiva de un valor, que como se da en este trabajo, puede ser el número cero. Esto implica que dicho valor posee una mayor probabilidad de ocurrencia. El caso específico donde el valor repetido es cero, se da mucho en el contexto de uso de servicios de salud donde se presenta mucho esta situación. También sucede particularmente en este trabajo se observará con más detalle en los próximos capítulos. Es posible que los ceros tengan dos posibles causas. En primer lugar, puede ser que ese evento no suceda nunca para ese individuo o zona, o también puede suceder que el evento no haya ocurrido en la ventana de tiempo considerada en la data que se tiene.

Dado lo anterior, no es posible asumir que ambas instancias, los ceros y no ceros, son generadas por un mismo proceso aleatorio. Es debido a esto, que para datos de conteo cero-inflados, le confiere mayor peso a la probabilidad que la variable de conteo sea igual a 0, al incorporar una probabilidad  $p$  para este segmento y una probabilidad  $1-p$  para el segmento de los valores positivos. En resumen, el modelo de Regresión de Poisson Cero-Inflado es una mezcla entre un Modelo Poisson estándar (el presentado en 2.1.1) y una distribución con función de masa centrada en cero.

De manera más formal y según el trabajo de Lambert [14], el modelo de Regresión Poisson puede especificarse como se muestra en las ecuaciones (2.6), (2.7) y (2.8). En primer lugar, la expresión  $\varphi_i = \Pr(y_i = 0)$  corresponde a la probabilidad de que la variable dependiente o de conteo tenga valor 0, su formulación matemática se puede apreciar en 2.6. Mientras que, la expresión  $\mu_i = \mu(x_i, \beta)$  corresponde al valor esperado cuando la variable de conteo asume un valor positivo, se puede apreciar su formulación en 2.8. Para este modelo, la expresión  $\varphi_i$  no puede tomar valores negativos, es por esta razón que Lambert propuso una parametrización para esta, basándose en la función logística y la ubicó en una regresión, en otras palabras, la parametrizó usando una Logit o Regresión Logística.

$$\varphi_i = \frac{\exp(z_i\gamma)}{1 + \exp(z_i\gamma)} \quad (2.6)$$

$$y_i = 0 \text{ con probabilidad } \varphi_i \quad (2.7)$$

$$y_i \sim Pr(\mu_i) \text{ con probabilidad } (1 - \phi_i) \quad (2.8)$$

El interés del modelo radica en encontrar los valores de los parámetros  $\beta, \gamma$  que maximizan la función de log-verosimilitud. En la función se define una variable para denotar el valor de  $y_i$ , esta toma el valor 1 si  $y_i = 0$  y 0 en caso de que tome cualquier otro valor. La función se puede apreciar en la ecuación 2.9.

$$L(\beta, \gamma) = \sum_{i=1}^n 1(y_i = 0) \log(\exp(z_i \gamma) + \exp(-\exp(x_i \beta))) + 1 - 1(y_i = 0)(y_i x_i \beta - \exp(x_i \beta) - 1 - \exp(z_i \gamma)) \quad (2.9)$$

Para observar un ejemplo práctico sobre el uso de este modelo se puede revisar el trabajo [13], en el que se realiza un análisis tecnológico cuya data posee una cantidad excesiva de ceros. En dicho trabajo, el modelo es comparado en todas las dimensiones posibles con el modelo presentado anteriormente, la Regresión Binomial Negativa.

Dados los resultados mostrados en los trabajos señalados en el párrafo anterior, en donde la regresión resulta ser mejor en varias aristas, y dada las características propias del problema, se escoge la Regresión de Poisson con Ceros-Inflados. Para utilizar este modelo y correr distintas instancias se utilizó el software R. Para entender cómo funciona correctamente el paquete estadístico que fue utilizado en la primera parte de este trabajo, se utilizó como referencia lo realizado en [15], donde no solo se explica el funcionamiento del paquete, sino que se presenta el uso práctico de la misma.

## 2.2 MODELOS DE LOCALIZACIÓN DE BASES

El problema descrito en este trabajo pertenece al tipo de problemas de localización de bases para sistemas de emergencias médicas o EMS (*Emergency Medical Services*). El desarrollo de estos ha sido ampliamente estudiado por diferentes disciplinas como las matemáticas, salud y gestión de operaciones, entre muchas otras. Es un problema de vital importancia para la organización y bienestar de la sociedad, por lo que ha sido un tema de constante estudio y resulta fácil encontrar material sobre este tipo de problemas. El buen funcionamiento de estos servicios es crítico para aumentar la probabilidad de supervivencia de los pacientes a atender.

En la investigación [16] se enumeran las decisiones que deben tomar las personas que están a cargo de gestionar servicios del tipo EMS, entre ellas destacan las siguientes:

- La localización de los puntos desde los cuales despachan los vehículos para la atención de emergencias.

- La regla de asignación para el despacho de vehículos cuando se presenta una emergencia o un incidente que amerite un despacho de vehículos.
- La relocalización de los recursos cuando una zona se encuentra desatendida por efecto del despacho de las unidades que allí operan.

La decisión de estos elementos influye de manera dramática en el tiempo de respuesta del servicio para llegar al lugar de la emergencia o el punto de demanda.

En este trabajo se busca resolver el problema de decisión de la localización de los puntos de donde parten los vehículos, es decir, la ubicación de las bases pertenecientes al servicio de emergencia, teniendo en cuenta que se tienen recursos limitados. Adicionalmente, se debe considerar que el proceso de relocalización de las bases de los móviles se hace más complejo de lo que parece, debido a que la demanda se ve influenciada proporcionalmente con la distancia a la que se encuentran las bases de las áreas de emergencias o cuadrantes y a los cambios demográficos de los mismos, como se puede apreciar en [17].

Tradicionalmente, este tipo de problemas poseen dos decisiones importantes, la primera consiste en escoger la ubicación para localizar la base o el depot, mientras que la segunda consiste en determinar el número de ambulancias que pertenecerán a cada depot. Otra decisión que incluye, generalmente, este tipo de problema es la asignación de la base a cierta área o cuadrante, dado que las bases y/o ambulancias tienen ciertas restricciones de distancia o capacidad a la hora de atender. Para tomar estas decisiones se debe tener en cuenta los potenciales lugares donde podrían ubicarse las bases y la demanda a cubrir de cada una de las zonas geográficas.

Estos modelos se dividen en tres tipos, según se señala en [18], los cuales se enumeran a continuación:

1. **Covering Models:** Este tipo de modelos se caracteriza por maximizar la cobertura para emergencias sin una distancia estándar predefinida.
2. **P-Median Models:** Este tipo de modelos se caracteriza por minimizar el promedio total de servicio para todos los puntos de demanda.
3. **P-Center Models:** El principal objetivo de este tipo de problemas es minimizar la distancia máxima para todos los puntos de demanda.

Dentro de los 3 grupos descritos anteriormente, el grupo utilizado y prevalente en el presente trabajo es el de los modelos de cobertura (*Covering Models*). A continuación, se describen los modelos más famosos y utilizados de este grupo.

### 2.2.1 LOCATION SET COVERING PROBLEM (LSCP)

En [16] se menciona que el problema LSCP es de los primeros modelos propuestos de grafos para localizar vehículos de emergencia. Su principal objetivo consiste en minimizar el número de vehículos requeridos para poder cubrir la demanda en su totalidad. Dentro de sus restricciones se encuentra que cada punto de demanda tiene asignado al menos un vehículo o estación y que toda estación tiene a lo más un vehículo en ella. Este modelo es claramente determinístico, es decir, la demanda es una constante dada antes de la resolución del problema.

Este modelo matemático fue propuesto en 1971 y se puede apreciar con mayor detalle en [19]. El PPL está escrito de la siguiente manera:

$$\min \sum_{i \in W} x_j \quad (1)$$

$$\text{subject to } \sum_{i \in W_i} x_j \geq 1 \quad i \in V \quad (2)$$

$$x_j \in \{0,1\} \quad j \in W \quad (3)$$

En el PPL, la variable  $x_j$  representa si una base o vehículo fue localizado en  $j$ . Es por esto, que en la función objetivo (1) se busca minimizar la sumatoria de estos. La restricción (2) corresponde a que cada elemento que pertenezca al conjunto de las demandas  $V$  debe tener al menos una base o vehículo asignada. Mientras que, la restricción (3) corresponde a la de naturaleza de variables, donde  $x_j$  puede tomar valor 1 o 0 dependiendo si la base o móvil fue asignado a la locación  $j$ . El conjunto  $W$  corresponde a los sitios que son candidatos a tener una ambulancia en su interior.

Una aplicación de este modelo se puede encontrar en el trabajo [20], donde en un caso práctico se alcanza una reducción de 200 a 145 vehículos en la mejor solución obtenida. Además, en Anexo B se presenta una ilustración obtenida de [18] que permite entender de mejor forma el funcionamiento de este modelo.

### 2.2.2 MAXIMAL COVERING LOCATION PROBLEM (MCLP)

Mientras el modelo LSCP busca calcular el mínimo requerido de vehículos para operar mientras se satisface la demanda, el problema de localización de cobertura máxima busca maximizar la cobertura alcanzada para cada uno de los puntos de demanda, teniendo como parámetro el número de vehículos disponibles, que en la literatura es designada con la letra  $p$ . Además, este modelo mejora la formulación al ponderar o asignar un peso a la demanda relativa de cada zona, de modo que se tiene en cuenta, que ciertos cuadrantes o zonas podrían poseer un mayor número de incidentes o de emergencias.

En este modelo ya no es estrictamente necesario cubrir el 100% de la demanda, ya que para lograr esto dependerá del número de recursos  $p$  que se tengan, a diferencia del modelo anterior que poseía una cantidad ilimitada de recursos. Una de las formas de alcanzar el 100% es hacer crecer  $p$  lo suficiente, hasta que existan los recursos necesarios para poder cumplir con toda la demanda [16].

El modelo fue descrito en 1974 por Church y ReValle. La formulación matemática de este problema es la siguiente:

$$\min \sum_{i \in W} d_i y_i \quad (1)$$

$$\text{subject to } \sum_{i \in W_i} x_j \geq y_i \quad i \in V \quad (2)$$

$$\sum_{i \in W_i} x_j = p \quad (3)$$

$$x_j, y_i \in \{0,1\}, \quad j \in W, i \in V \quad (4)$$

En la función objetivo (1) se maximiza la demanda cubierta. La restricción (2), en este caso, garantiza que el punto de demanda  $i$  está cubierto solo si una o más bases están asignadas dentro de la distancia estándar. La restricción (3) especifica que el número de instalaciones que deben ser utilizadas son igual a  $p$ , que es el número de recursos con los que se cuenta. La restricción (4) corresponde a la naturaleza de las variables, las cuales deben ser binarias.

Para ver aplicaciones de estos modelos se pueden revisar los trabajos realizados por Eaton, en los cuales se utiliza el modelo para resolver este tipo de problema en Valle del Cauca en Colombia [21] y para Texas en USA [22]. Además, para ver un compilado de trabajos donde se utilizó el modelo se recomienda revisar la publicación [23], donde se encuentran desarrollados distintos problemas donde se utilizó MCLP de manera exitosa. Un último ejemplo donde se utiliza

este modelo de forma eficiente para distribuir de manera eficiente ambulancias aéreas (helicópteros) en este caso es en [3], donde se tiene como objetivo principal brindar cobertura al 90% de la población en menos de 45 minutos, obteniendo importantes resultados.

### 2.2.3 MODELO TEAM (TANDEM EQUIPMENT ALLOCATION MODEL)

Uno de los principales inconvenientes de los modelos anteriormente descritos es que solo manejan una categoría de vehículos y de tipo de incidentes. En muchos servicios, dependiendo del incidente, es enviado un tipo diferente de vehículo, e incluso en ocasiones, son requeridos los dos tipos de vehículos en el lugar del accidente.

El *Tandem Equipment Allocation Model* (TEAM) se basa en el MCLP, con la diferencia de que existe un conjunto de variables adicionales para el segundo tipo de vehículo, por lo tanto, un punto de demanda no se considera cubierto si no hay disponible un vehículo de cada categoría para esa zona. En [24] se describe la aplicación del modelo TEAM para una red de transporte de desechos radiactivos.

## 2.3 TRABAJOS RELACIONADOS

Como se ha apreciado en los puntos anteriores, la bibliografía estudiada con respecto a este tipo de problemas de optimización es abundante. Muchos de los modelos anteriormente descritos se han utilizado para resolver problemas de localización para servicios de emergencias. Este conocimiento se ha utilizado para resolver problemas de localización de servicios de emergencias en varios países, como Austria [27] y Bélgica [28].

Otro ejemplo de esto se da en el trabajo presentado en [26], donde se optimiza la localización de las bases de las ambulancias para Tijuana en México. El problema cuenta con 11 ambulancias ubicadas en 8 bases que quieren ser localizadas. En esta investigación se seleccionan 961 lugares como potencial localización para una base y consiguen las coordenadas utilizando Google places API, como se hará más adelante en el presente trabajo también. En este problema se contaba con 10.512 llamadas las cuales correspondían a 8 meses de emergencias. Para resolver el problema se utiliza una variación del MCLP. Se localizan dos tipos de base y el resultado gráfico de la resolución de este problema se puede apreciar de mejor forma en el Anexo C.

Una aplicación similar y que se tomó como ejemplo para realizar la primera parte de esta tesis, donde se utiliza una Regresión de Poisson con Ceros-Inflados se da en [28]. En este estudio realizado en Suecia se busca predecir el número de llamados de emergencias para ciertas zonas geográficas, utilizando como variables explicativas la población de cada zona y las características como la edad de la población e ingresos de cada zona.

## **CAPÍTULO 3: MODELO PREDICTIVO DE DESPACHOS**

### **3.1 CONSTRUCCIÓN DATA SET PARA MODELO**

Para confeccionar y generar un modelo econométrico predictivo es fundamental la cantidad y calidad de la data para entrenar correctamente el modelo, que tenga un ajuste correcto y un poder predictivo real.

En este punto se presenta el origen de la data utilizada, sus características, limitaciones y el tratamiento y procedimiento que se le dio para que estuviera apta para ser utilizada en la generación del modelo predictivo, que como se dijo anteriormente, busca predecir el número de despachos de emergencias realizados por ambulancias del SAMU asociados a cada cuadrante, utilizando variables explicativas y características sociodemográficas de cada cuadrante.

#### **3.1.1 DATA DESPACHOS Y LLAMADAS AL SAMU**

La data utilizada donde se encuentran los despachos fue proporcionada por el SAMU, la que incluye los despachos de ambulancias realizados entre los meses de septiembre y octubre. El archivo cuenta con 11.800 despachos los cuales fueron recolectados manualmente por trabajadores del Centro Regulador (perteneciente al SAMU), ya que para ese entonces no se contaba con un sistema que pudiera guardar y recolectar esta información de manera automática.

En la Figura 2, se puede apreciar la estructura de la data. La información que se encuentra en la data y que fue utilizada en este trabajo, es la base desde donde partió cada ambulancia hacia el lugar de la emergencia o incidente. Además, se tiene la ambulancia señalada con una ID que atendió la emergencia. Se puede apreciar la dirección, la cual no se encuentra con coordenadas, solo se encuentra la dirección escrita con palabras. Finalmente, se encuentra descrita brevemente la atención que recibió el paciente y el centro hospitalario al cual fue trasladado, en caso de haber sido necesario.



Fecha	Tipo	Clave	QTC	Base	Móvil	Origen	comur	Destino	Actividad	H-Pedido
9-1-2017	B	101	9	56	85	COMPAÑÍA 2259 DEPTO 42	33		RECHAZA TRASLADO	00:15
9-1-2017	B	101	10	13	116	CAMINO RINCONADA 1775	17		RECHAZA TRASLADO	00:34
9-1-2017	B	801	12	23	39	INDEPENDENCIA / HIPODROMO CHILE	7	H SAN JOSE		00:48
9-1-2017	B	309	22	48	112	ALCALDE JORGE MONCKEBERG	19		POR OTROS MEDIOS	01:43
9-1-2017	B	101	24	46	94	LAS ALGAS 4926	19		RECHAZA TRASLADO	01:58
9-1-2017	B	101	25	23	39	PUCON 4972	1		POR OTROS MEDIOS	02:00
9-1-2017	B	309	29	56	113	GENERAL BULNES 41	33		SIN PACIENTE	02:29
9-1-2017	B	107	31	53	127	CALLE CERO 575	40	H DE CURACAVI		02:47
9-1-2017	B	101	32	56	85	BRAVO LUCO 1417	24	POSTA 3		02:52
9-1-2017	B	101	33	68	70	LA ALBORADA 25 / SANTA ROSA PARADERO 26	13		RECHAZA TRASLADO	02:49
9-1-2017	B	101	34	11	28	MANUEL DE AMAT 3171	33		RECHAZA TRASLADO	02:59
9-1-2017	A	607	18	55	80	SAR RENCA		HUAP		01:26
9-1-2017	B	309	37	56	113	SAN PABLO 6600	14			03:23
9-1-2017	B	101	39	13	116	EL ARENAL 1739			POR OTROS MEDIOS	03:40
9-1-2017	B	101	20	23	21	MAIPO 514	28	H SAN JOSE		01:37
9-1-2017	B	309	42	56	113	LOS ARRAYANES 6669	14			04:18
9-1-2017	B	101	43	48	112	SALVADOR 100	20		SIN PACIENTE	04:38
9-1-2017	B	101	45	13	116	RIO MAGDALENA 650	3		RECHAZA TRASLADO	06:25
9-1-2017	B	101	30	23	21	SAN LUIS 1546	7	H SAN JOSE		02:39
9-1-2017	B	101	47	35	117	CARMEN 2341	30	H BARROS LUCO		06:35
9-1-2017	B	350	51	56	85	VIRIPUGI 1632	2	H SOTERO DEL RIO		07:12
9-1-2017	B	101	61	76	114	CALETA ARRAYAN 996	21	H SOTERO DEL RIO		07:48
9-1-2017	B	101	36	23	39	MAR AZUL 2486	1	H SAN JOSE		03:05
9-1-2017	B	309	41	68	70	SAN MIGUEL 1411	12	H LA FLORIDA		04:06

FIGURA 2:ESTRUCTURA DATA ORIGINAL DESPACHOS SAMU.

La data correspondiente a las llamadas de emergencia que fueron recibidas por el SAMU también fue proporcionada por la institución. Este archivo cuenta con el número de llamados que recibió el Centro Regulador de forma muy agregada a nivel comuna. Esta información, a diferencia de la de despachos, no cuenta con la la ubicación exacta de cada llamada, solo es identificable la comuna donde se realizó. Esta data era capturada automáticamente por el sistema para ese entonces. En la Figura 3 se puede apreciar la base de datos, la cual permite entender la naturaleza y el contenido de mejor manera.

COMUNA	sept-16	oct-16	nov-16	dic-16	ene-17	feb-17	mar-17	abr-17	may-17	jun-17	jul-17	ago-17	sept-17	oct-17
ALHUE	19	14	17	21	25	16	24	24	30	28	13	37	23	21
CERRILLOS	180	158	180	169	219	162	177	183	198	232	244	242	216	234
CERRO NAVIA	361	353	312	371	426	325	377	337	334	442	506	556	490	498
COLINA	153	164	156	148	163	148	137	131	171	175	147	201	277	163
CONCHALI	276	318	259	274	317	254	287	223	333	357	407	447	386	408
CURACAVI	84	82	104	89	93	92	73	81	116	76	82	87	73	86
EL BOSQUE	219	224	236	257	177	185	214	215	252	257	381	322	282	293
ESTACION CENTRAL	580	582	496	609	554	516	559	540	597	680	693	775	709	671
HUECHURABA	182	218	207	217	273	174	211	159	209	232	231	250	251	254
INDEPENDENCIA	345	347	378	302	395	322	401	349	405	444	514	498	520	459
LA CISTERNA	378	359	326	355	341	237	319	288	343	381	407	460	484	464
LA FLORIDA	952	866	812	819	774	697	846	786	788	882	923	1064	978	931
LA GRANJA	308	270	195	247	281	222	203	218	242	239	298	305	309	310
LA PINTANA	439	385	335	280	419	300	307	316	409	449	469	480	525	436
LA REINA	153	121	112	147	134	92	147	134	151	134	165	165	172	183
LAMPA	75	56	78	76	96	85	95	63	101	64	89	124	109	113
LAS CONDES	289	322	350	345	352	267	318	295	332	324	369	411	323	407
LO BARNECHEA	217	249	224	251	257	225	209	246	255	246	264	299	332	292
LO ESPEJO	187	180	150	176	213	159	208	170	157	157	183	199	210	217
LO PRADO	165	181	184	210	188	115	190	193	151	214	237	242	219	191
MACUL	226	212	198	203	201	166	227	245	274	292	287	335	313	360
MAIPU	931	875	874	913	970	827	876	895	868	1019	1233	1175	1140	1175
MARIA PINTO	8	5	5	5	1	9	5	8	7	1	4	6	6	12

FIGURA 3:ESTRUCTURA DATA LLAMADAS SAMU.

### 3.1.2 DATA PRECENSO 2016

Para poder generar los cuadrantes, es decir la división de la Región Metropolitana en diversas manzanas, se utilizó data proveniente del PRECENSO 2016. En ella se encuentra la región dividida en 54.000 cuadrantes, los cuales están delimitados por polígonos de hasta doce vértices. Cada uno de los vértices del polígono se encuentra georreferenciado con coordenadas. Dentro de las columnas o variables presentes en la data se encuentran la comuna a la que pertenece cada cuadrante, algunas características sociodemográficas como la cantidad de viviendas presentes en cada uno de ellos, la densidad de población y la superficie en metros cuadrados de cada uno. Para apreciar mejor la estructura de la data se puede observar la Figura 4.

urbano	manzent	total_viv	superficie	dens_viv	geom
CERRILLOS (I	1.3102E+13	45	0.8662	51.9511;3400	339999.672
CERRILLOS (I	1.3102E+13	40	0.5238	76.365;3400	340033.860
CERRILLOS (I	1.3102E+13	35	0.4671	74.9304;3400	340067.829
CERRILLOS (I	1.3102E+13	32	0.3905	81.9462;3400	340100.361
CERRILLOS (I	1.3102E+13	24	0.3358	71.4711;3400	340134.291
CERRILLOS (I	1.3102E+13	19	0.2721	69.8273;3400	340167.566
CERRILLOS (I	1.3102E+13	11	0.224	49.1071;3400	340201.364
CERRILLOS (I	1.3102E+13	30	0.5666	52.9474;3399	339982.010
CERRILLOS (I	1.3102E+13	32	0.3899	82.0723;3400	340015.631
CERRILLOS (I	1.3102E+13	32	0.3845	83.225;3400	340049.384
CERRILLOS (I	1.3102E+13	36	0.5236	68.7548;3400	340228.817
CERRILLOS (I	1.3102E+13	32	0.3775	84.7682;3400	340082.566
CERRILLOS (I	1.3102E+13	32	0.3784	84.5666;3400	340115.886
CERRILLOS (I	1.3102E+13	32	0.3829	83.5727;3400	340149.504
CERRILLOS (I	1.3102E+13	29	0.418	69.378;3402	340258.159
CERRILLOS (I	1.3102E+13	32	0.4093	78.1823;3400	340182.482
CERRILLOS (I	1.3102E+13	26	0.347	74.928;3403	340291.750
CERRILLOS (I	1.3102E+13	18	0.256	70.3125;3400	340323.505
CERRILLOS (I	1.3102E+13	61	0.8941	68.225;3401	340174.863
CERRILLOS (I	1.3102E+13	0	0.383	0;340427.83	340427.550
CERRILLOS (I	1.3102E+13	4	2.1405	1.8687;3432	343243.006
CERRILLOS (I	1.3102E+13	31	0.6875	45.0909;3399	339956.548
CERRILLOS (I	1.3102E+13	32	0.4968	64.4122;3399	339996.301
CERRILLOS (I	1.3102E+13	28	0.4588	61.0288;3400	340034.839
CERRILLOS (I	1.3102E+13	66	0.9962	66.2518;3400	340486.186
CERRILLOS (I	1.3102E+13	28	0.4313	64.92;34007	340073.551
CERRILLOS (I	1.3102E+13	24	0.4014	59.7907;3400	340112.998
CERRILLOS (I	1.3102E+13	25	0.36	69.4444;3400	340147.948
CERRILLOS (I	1.3102E+13	19	0.336	56.5476;3400	340194.633
CERRILLOS (I	1.3102E+13	19	0.2968	64.0162;3400	340232.351
CERRILLOS (I	1.3102E+13	16	0.2714	58.9536;3400	340271.371
CERRILLOS (I	1.3102E+13	7	5.9728	1.172;34266	342667.281

FIGURA 4: DATA PRECENSO 2016.

### 3.1.3 DATA BASES SAMU

Esta data contiene información de las bases con las que contaba el SAMU en noviembre de 2017, y tiene campos importantes como la ID que identifica a cada una de las bases, la comuna en la que se encuentra, la dirección en la que se ubica, el número de ambulancias básicas y avanzadas que contiene cada una. Además, se encuentran georreferenciada las direcciones, en coordenadas UTMN y UTME. Esta última información es necesaria para calcular las distancias entre cada cuadrante y la base más cercana a él. Se puede apreciar la estructura de la data en la Figura 5.

ID	NOM_BASE	NOM_COM	DIRECCION	REFERENCIA	AMB. AVANZADA	AMB. MEDICALIZADA	COORD_X	COORD_Y
1	BASE SAMU 11	SANTIAGO	PORTUGAL 125	HOSPITAL DE URGENCIA ASISTENCIA PUBLICA	0	0	347743.89	6298373.43
2	BASE SAMU 12	ESTACION CENTRAL	LLANCAHUE 840	SAPU CHUCHUNCO	0	0	341448.82	6294866.3
3	BASE SAMU 13	MAIPU	CAMINO A RINCONADA 1001	CRS MAIPU	0	0	335444.19	6290880.19
4	BASE SAMU 14	MAIPU	CAMINO A RINCONADA 1001	CRS MAIPU	2	0	335444.19	6290880.19
5	BASE SAMU 15	MELIPILLA	ORTUZAR 1201	HOSPITAL SAN JOSE DE MELIPILLA	1	0	294619.69	6299361.08
6	BASE SAMU 20	SANTIAGO	PORTUGAL 125	HOSPITAL DE URGENCIA ASISTENCIA PUBLICA	1	1	347743.89	6298373.43
7	BASE SAMU 22	INDEPENDENCIA	DOMINGO SANTA MARIA 1207	SIN REFERENCIA	2	0	346009.55	6301456.87
8	BASE SAMU 23	INDEPENDENCIA	DOMINGO SANTA MARIA 1207	SIN REFERENCIA	0	0	346009.55	6301456.87
9	BASE SAMU 25	COLINA	CALLE B 95	SIN REFERENCIA	1	0	344123.57	6325507.61
10	BASE SAMU 30	SAN BERNARDO	SAN ALFONSO 1201	CESAFAM DR. RAUL CUEVAS	1	0	340872.5	6281172.27
11	BASE SAMU 33	SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	2	0	347051.87	6293647.53
12	BASE SAMU 35	SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	0	0	347051.87	6293647.53
13	BASE SAMU 40	PROVIDENCIA	VITACURA 115	HOSPITAL FELIX BULNES (EX MILITAR)	1	0	350715.38	6301061.86
14	BASE SAMU 44	ÑUÑO A	JUAN MOYA MORALES 1188	SAPU CENTRO URGENCIA ÑUÑO A	1	0	352389.66	6295879.32
15	BASE SAMU 45	PEÑALOLEN	LAS TORRES 5100	HOSPITAL LUIS TISNE	0	0	353290.18	6291982.75
16	BASE SAMU 46	PROVIDENCIA	VITACURA 115	HOSPITAL FELIX BULNES (EX MILITAR)	0	0	350715.38	6301061.86
17	BASE SAMU 48	PROVIDENCIA	RANCAGUA 600	HOSPITAL SALVADOR	0	0	349009.71	6298813.52
18	BASE SAMU 53	CURACAVI	AMBROSIO O'HIGGINS 504	HOSPITAL DE CURACAVI	0	0	300414.7	6301303.4
19	BASE SAMU 55	QUINTA NORMAL	DR LORENZO URCARAY 4364	HOSPITAL FELIX BULNES	2	0	342400.55	6301359.05
20	BASE SAMU 56	SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	0	0	347051.87	6293647.53
21	BASE SAMU 66	PUENTE ALTO	EL PEÑON 39	HOSPITAL SOTERO DEL RIO	1	0	353264.15	6283247.39
22	BASE SAMU 68	LA FLORIDA	FROILAN ROA 6426	CESFAM BELLAVISTA	0	0	351302.29	6290756.76
23	BASE SAMU 76	PUENTE ALTO	EL PEÑON 39	HOSPITAL SOTERO DEL RIO	0	0	353264.15	6283247.39
24	BASE SAMU 90	LA FLORIDA	FROILAN ROA 752	HOSPITAL DE LA FLORIDA	1	0	351349.59	6290598.24

FIGURA 5: ESTRUCTURA DATA BASES SAMU.

### 3.1.4 GEORREFERENCIACIÓN DESPACHOS

Como las direcciones están escritas con palabras y no de manera estructurada, es necesario procesar los datos para corregir errores y normalizar esta columna para que posteriormente permita georreferenciar las direcciones, y así obtener las coordenadas en formato x e y para luego calcular las distancias entre cada cuadrante y la base más cercana.

Para realizar este proceso se utilizó Python para normalizar las direcciones y darles una estructura coherente. Luego se georreferenciaron los 11.800 despachos utilizando la plataforma de la empresa SimpliRoute [29], la cual trabaja sobre la API de Google Maps. Del total de direcciones se pudo rescatar cerca del 70% de ellas, lo cual corresponde a 8.200 despachos. El resto de las direcciones poseían calles inexistentes o direcciones inexactas por lo cual resultó imposible geolocalizarlas de manera correcta.

### 3.1.5 ASIGNACIÓN DESPACHOS A CUADRANTES

Una vez explicados los datos utilizados en el problema junto con sus características y origen, se realizó un proceso de asignación, en el cual se buscaba generar un conteo de cuántos despachos de ambulancias (los que lograron ser geolocalizados) correspondían a cada uno de los cuadrantes generados. Para realizar esto se utilizó la siguiente estrategia: en primer lugar, se consideraron los lados de cada polígono que delimitan cada uno de los cuadrantes como restricciones lineales, por lo que cada vértice fue considerado como la intersección de dos restricciones. Luego, se consideró que un despacho pertenecía a un cuadrante si cumplía todas las restricciones conformadas por las aristas del polígono. Mediante un método iterativo realizado en Python se hizo el conteo de los despachos por cuadrante.

Como sólo existían 8.200 despachos correctamente georreferenciados, estos fueron repartidos entre los 54.000 cuadrantes que traía delimitados el PRECENSO 2016, resultando que menos del 12% de los cuadrantes cuenta con algún despacho de ambulancias del SAMU realizado en su interior. Este número resulta ser más pequeño que el 15% esperado (8.200/54.000) debido que hay algunas zonas, como la de Santiago Centro, donde hay una alta concentración de despachos. Dado esto, se plantea la idea de agregar cuadrantes para que este porcentaje sea mayor y la data sea más consistente a la hora de generar modelo predictivo como el que se muestra en el Capítulo 4.

### 3.1.6 CONSTRUCCIÓN DATA SET FINAL

Una vez obtenidas todas las bases de datos necesarias, realizados los procesos de georreferenciación de despachos de emergencia y posteriormente la asignación de estos despachos a los distintos cuadrantes generados, se realizó el cruce de las diversas bases para generar una estructura final de data, la que es utilizada para generar el modelo predictivo en el Capítulo 4. Se puede apreciar esta estructura en la Figura 6, donde se observa que la data final cuenta con 54.000 filas, una para cada cuadrante. Dentro de las variables más importantes que contiene este data set final se encuentran:

- **Comuna:** Esta variable señala la comuna a la que pertenece el cuadrante de la fila correspondiente.
- **Total Viviendas:** La variable señala el número total de viviendas que posee cada cuadrante. Esta variable tiene gran variabilidad.
- **Superficie:** Esta variable indica la superficie en kilómetros cuadrados que posee cada cuadrante.
- **Despachos:** Esta variable señala el número de despachos que se dieron dentro de cada uno de los cuadrantes. La variable fue generada según el proceso descrito en el Capítulo 3.1.5. Esta resulta ser la variable que busca ser explicada en el modelo predictivo propuesto en el Capítulo 4 y maximizada posteriormente en el Capítulo 5.
- **Coordenadas UTMN y UTME:** Este sistema de coordenadas utiliza una proyección cilíndrica para representar la tierra sobre un plano. Estas variables en la data señalan la posición del centro correspondiente a cada cuadrante, las que son conseguidas promediando las coordenadas de los vértices que contiene cada uno de los cuadrantes en caso de ser convexos los polígonos. En caso contrario, se utiliza una función en Python para calcular el centro de polígonos no convexos, lo cual se da en menos de un

5% de los 54.000 cuadrantes. Esta variable es muy importante, ya que posteriormente es utilizada para calcular las distancias desde el centro de cada cuadrante a la base de ambulancias más cercanas.

- **Coordenadas Latitud y Longitud:** Esta variable también indica el centro correspondiente a cada uno de los 54.000 cuadrantes. Este tipo de coordenadas resultan ser más útiles cuando se quiere generar mapas, pero a la hora de calcular distancias euclidianas entre dos puntos, resultan ser mucho más adecuada para este cometido las coordenadas UTM.

region	provincia	des_regi	des_prov	des_comu	total_viv	superficie	dens_viv	despachos	utmn	utme	lat	long
13	131	REGION METI SANTIAGO	CERRILLOS		265	3.9776	51.9511	1	340154.91	6294704.8	-70.720293	-33.475215
13	131	REGION METI SANTIAGO	CERRILLOS		32	0.3905	81.9462	0	340102.83	6294767.6	-70.720842	-33.474641
13	131	REGION METI SANTIAGO	CERRILLOS		30	0.5666	52.9474	0	339954.08	6294626.8	-70.722468	-33.475888
13	131	REGION METI SANTIAGO	CERRILLOS		127	1.9587	82.0723	1	339995.64	6294533.7	-70.722037	-33.476734
13	131	REGION METI SANTIAGO	CERRILLOS		575	5.9054	84.7682	3	340091.92	6294450.4	-70.721016	-33.477499
13	131	REGION METI SANTIAGO	CERRILLOS		74	1.0667	78.1823	1	340201.67	6294504.2	-70.719826	-33.477031
13	131	REGION METI SANTIAGO	CERRILLOS		1195	11.3297	0	1	340320.62	6294268.7	-70.718588	-33.479171
13	131	REGION METI SANTIAGO	CERRILLOS		94	5.3692	1.8687	0	343044.81	6294356.8	-70.689264	-33.478728
13	131	REGION METI SANTIAGO	CERRILLOS		44	0.696	69.4444	0	340161.83	6294415.8	-70.72027	-33.477822
13	131	REGION METI SANTIAGO	CERRILLOS		61	1.0179	64.0162	0	340282.27	6294391.8	-70.718979	-33.478056
13	131	REGION METI SANTIAGO	CERRILLOS		8	6.8233	1.172	1	342632.76	6294334.5	-70.693701	-33.478921
13	131	REGION METI SANTIAGO	CERRILLOS		5	2.8497	1.6013	1	343291.42	6294403.1	-70.686603	-33.478399
13	131	REGION METI SANTIAGO	CERRILLOS		149	1.3497	48.4688	0	340496.56	6294260.4	-70.716697	-33.479273
13	131	REGION METI SANTIAGO	CERRILLOS		14	1.8905	12.3066	0	343153.48	6294464	-70.688076	-33.47783
13	131	REGION METI SANTIAGO	CERRILLOS		191	5.8394	17.1512	5	343208	6294140.1	-70.687546	-33.480758
13	131	REGION METI SANTIAGO	CERRILLOS		155	5.9203	49.5868	1	342768.92	6294094.5	-70.692278	-33.481105
13	131	REGION METI SANTIAGO	CERRILLOS		54	0.5375	100.4651	0	339902.24	6294207.2	-70.7231	-33.479664
13	131	REGION METI SANTIAGO	CERRILLOS		56	1.3672	58.5892	0	342760.35	6294125.5	-70.692365	-33.480824
13	131	REGION METI SANTIAGO	CERRILLOS		124	4.5719	0	3	342959.2	6294049.8	-70.690239	-33.481536
13	131	REGION METI SANTIAGO	CERRILLOS		75	6.277	17.2038	0	342601.4	6294090.5	-70.694081	-33.481117
13	131	REGION METI SANTIAGO	CERRILLOS		67	4.2915	15.6123	1	343166.92	6293954.7	-70.688021	-33.482424
13	131	REGION METI SANTIAGO	CERRILLOS		742	6.9518	141.5094	4	339984.11	6294066.2	-70.722245	-33.480947
13	131	REGION METI SANTIAGO	CERRILLOS		60	1.4985	41.8994	0	342091.85	6293975.5	-70.699584	-33.482078
13	131	REGION METI SANTIAGO	CERRILLOS		215	32.131	49.0798	2	341954.86	6293872.9	-70.701076	-33.482983
13	131	REGION METI SANTIAGO	CERRILLOS		207	8.4017	52.0517	3	342801.56	6293817.9	-70.691976	-33.483603
13	131	REGION METI SANTIAGO	CERRILLOS		2	9.252	0.2162	0	339513.77	6293861.6	-70.727342	-33.482721

FIGURA 6: ESTRUCTURA DATA FINAL.

### 3.2 CONSTRUCCIÓN MODELO PREDICTIVO

Una vez definida y procesada la data, se busca encontrar una relación entre la demanda de despachos de ambulancias y diversas variables explicativas. Para cumplir este propósito, en esta etapa se busca construir o generar un modelo predictivo lo suficientemente robusto, para predecir los despachos esperados de acuerdo con ciertas variables explicativas, que son obtenidas desde el análisis exploratorio de las variables. Además, se proponen distintos modelos en búsqueda de uno que presente un mayor nivel predictivo y variables explicativas significativas. Las diversas etapas para la concepción de modelo finalmente escogido se encuentran a continuación.

### 3.2.1 ANÁLISIS EXPLORATORIO DATA

Para proponer variables que deberían ser consideradas por el modelo, se debe primero conocer la relación existente entre las distintas variables y la variable que se busca explicar. Para esto se realiza un estudio de correlación de los despachos efectivos con las siguientes variables:

- Total viviendas
- Distancia a base más cercana
- Número de ambulancias base más cercana
- Distancia a hospital o clínica más cercana

En la Tabla 2, se aprecia una matriz de correlación entre la variable de despachos y las potenciales variables explicativas. Se puede notar que una correlación fuerte y la más alta que se aprecia en la tabla se da entre la cantidad de despachos que fueron realizados y la cantidad de viviendas presentes en ese cuadrante, la cual corresponde a 0.53, esto quiere decir que existe una correlación positiva entre ambas variables y es esperable que cuando aumente en magnitud la cantidad de viviendas, también lo haga la cantidad de despachos. De la tabla, también se puede observar que existe una correlación negativa moderada de -0.26 entre la cantidad de despachos y la distancia a la base más cercana. En palabras simples, si aumenta la variable de distancia, la variable asociada a los despachos debería disminuir. Esto hace sentido ya que se espera que a medida que la base más cercana esté más lejos las personas tienden a llamar menos, ya que los tiempos de cobertura podrían ser demasiado altos o bien la base más cercana no tiene la capacidad de recursos para atender emergencias.

Además, se puede notar en la figura, que existe una correlación positiva moderada de 0.23 entre los despachos y el número de ambulancias que posee la base de ambulancias del SAMU más cercano, por una razón similar a la explicada en el caso de la correlación entre distancia y despacho, que corresponde a que una base con más ambulancias podría dar una mayor cobertura y llegar de mejor manera a las diversas emergencias.

	Despachos	Viviendas	Distancia	Nro Ambulancias
Despachos	1	0.53	-0.26	0.23
Viviendas	0.53	1	-0.1	0.08
Distancia	-0.26	-0.1	1	-0.1
Nro Ambulancias	0.23	0.08	-0.1	1

TABLA 2: MATRIZ DE CORRELACIÓN.

Por otro lado, se generaron dos variables, que correspondían a la distancia más cercana desde cada cuadrante al hospital o clínica más cercana. La georreferenciación de cada centro hospitalario se obtuvo desde la plataforma web del Ministerio de Salud. Al calcular la correlación entre la cantidad de despachos y la distancia al centro hospitalario más cercano resulta ser negativa

como es esperable, pues mientras más cerca esté un centro hospitalario desde el lugar del incidente, es más probable que se pueda transportar a la persona utilizando otro medio de transporte. La fuerza de esta correlación resulta ser bastante baja, solo tiene una magnitud de -0.07.

### 3.2.2 NIVELES DE AGREGACIÓN DATA

Como se mencionó anteriormente, se cuenta con 8.200 despachos georreferenciados correctamente, los que se reparten entre 54.000 cuadrantes, es decir, se deberían esperar solo 0.15 despachos por cada cuadrante en una ventana de tiempo de dos meses. Pero en la práctica menos del 12% de los cuadrantes contiene al menos un despacho en su interior, por lo tanto, cerca del 88% poseen un valor 0, como se puede apreciar en el Marco Conceptual de este trabajo resulta perjudicial para los modelos de conteo [10], incluso para el modelo de Regresión de Poisson con Ceros-Inflados con un nivel tan elevado de valores repetidos. Es debido a lo anterior que se decide agregar cuadrantes a distintos niveles.

Para realizar la agregación de cuadrantes se utilizó Python y el algoritmo se describe a continuación: en primer lugar, se creó una función que permitiera saber si dos polígonos eran colindantes, si esto se cumplía y ambos polígonos o cuadrantes estaban en la misma comuna se agregaban. En segundo lugar, se sumaron algunas variables demográficas para cada cuadrante como el total de viviendas y la superficie en kilómetros cuadrados. Finalmente, se recalculan los centros pertenecientes a cada cuadrante, utilizando la misma estrategia que fue descrita en el Capítulo 3.1.6.

Se hizo esto para varios niveles de agregación, con el fin de probar los modelos predictivos para distintos niveles de agregación y poder elegir el modelo con el mayor poder predictivo. Los distintos niveles de agregación se enumeran a continuación:

- **Nivel 8:** Es el número de cuadrantes original de la data de PRECENSO 2016, es decir, posee 54.000 cuadrantes.
- **Nivel 7:** Este nivel posee 43.000 cuadrantes.
- **Nivel 6:** Este nivel posee 23.300 cuadrantes.
- **Nivel 5:** Este nivel posee 12.500 cuadrantes.
- **Nivel 4:** Este nivel posee 6.300 cuadrantes.
- **Nivel 3:** Este nivel posee 3.900 cuadrantes.
- **Nivel 2:** Este nivel posee 2.000 cuadrantes.
- **Nivel 1.** Este nivel posee 600 cuadrantes.

Para los distintos niveles de agregación, no solo cambia, como es esperable, el número de cuadrantes, sino que además cambian las superficies de cada uno de los estos, las que como se mencionó, se deben sumar al unir dos cuadrantes. Se aprecia que a medida que se van agregando

los cuadrantes, la mediana de la superficie va a aumentando. Esto permite observar de mejor manera en la Tabla 3, de esta figura también se puede notar que a medida que se van agregando los cuadrantes, se mantiene invariante la mediana de la distancia desde cada cuadrante a la base más cercana. Es decir, independiente del nivel de agregación, el 50% de los cuadrantes se encuentra a menos de 3.5 km de la base más cercana.

	Nivel 1 (600)	Nivel 2 (2.000)	Nivel 3 (3.900)	Nivel 4 (6.300)	Nivel 5 (12.500)	Nivel 6 (23.300)	Nivel 7 (43.000)
Mediana distancia(km)	3.5	3.4	3.4	3.5	3.5	3.4	3.5
Mediana superficie(km <sup>2</sup> )	75.8	17.4	8.7	4.6	2.2	1	0.47

TABLA 3:MEDIANA DE DISTANCIA BASE MÁS CERCANA Y SUPERFICIE SEGÚN NIVEL DE AGREGACIÓN.

En Anexo F se observa en un mapa los centros de los cuadrantes para un nivel de agregación de 600 cuadrantes. Para realizar este tipo de mapas se utilizó GMAPS de Google en Python.

### 3.2.3 PROPUESTAS DE MODELO PREDICTIVO

Una vez analizadas las distintas variables que podrían incluirse al modelo, se propusieron diversos modelos predictivos, desde una Regresión Lineal simple hasta modelos de conteo más sofisticados como una Regresión de Poisson con Ceros-Inflados. Finalmente, se descartó una Regresión Lineal simple, porque no mostró un buen nivel predictivo y debido a que un modelo de conteo debería ajustarse mejor a este tipo de problemas, donde se busca predecir la tasa bimensual de despachos de ambulancias. A continuación, se describen los dos modelos que fueron propuestos:

- **Regresión de Poisson:** Es un modelo lineal generalizado en el que la variable explicada posee una distribución de Poisson y el logaritmo de su valor esperado puede ser modelado como una combinación lineal de parámetros desconocidos y variables explicativas. La fórmula matemática de este modelo se presenta en la ecuación (3.1). En ella,  $Y$  representa la variable que busca ser explicada,  $X$  representa el conjunto de las variables explicativas,  $\beta$  representa los parámetros que son obtenidos por el modelo y  $\varepsilon$  representa el error del modelo.

$$\log(E(Y|X)) = \sum_i \beta_i * X_i + \varepsilon \quad (3.1)$$



Una de las limitaciones de este modelo según [10] es que no se hace cargo de la heterogeneidad no observada y en particular para problemas relacionados al área de salud, puede ser una mala elección, debido a que una gran cantidad de ceros puede ser inconsistente para el modelo y pierde mucho poder predictivo.

- **Regresión de Poisson con Ceros-Inflados:** Este modelo es una mezcla entre la Regresión de Poisson y una distribución con función de masa centrada en cero, es decir, asume que la probabilidad que la variable explicativa tome valor cero es mucho mayor que si se asumiera una distribución de Poisson como lo hace el modelo clásico. Este modelo resulta ser muy útil para problemas relacionados al área de salud donde es común encontrar una gran cantidad de ceros. Para revisar la formulación matemática de este modelo ir al Capítulo 2.1.3.

Para ambos modelos descritos, es deseable que la variable a ser explicada siga una distribución de Poisson, aunque no es estrictamente necesario. A pesar de que la variable explicada o dependiente siga una distribución distinta, aún se pueden obtener estimadores consistentes según lo revisado en [12].

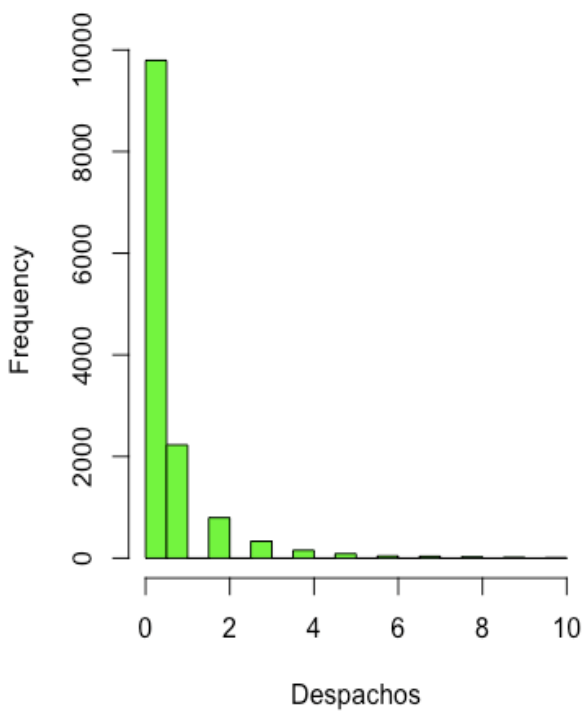


FIGURA 7: DISTRIBUCIÓN DESPACHOS.

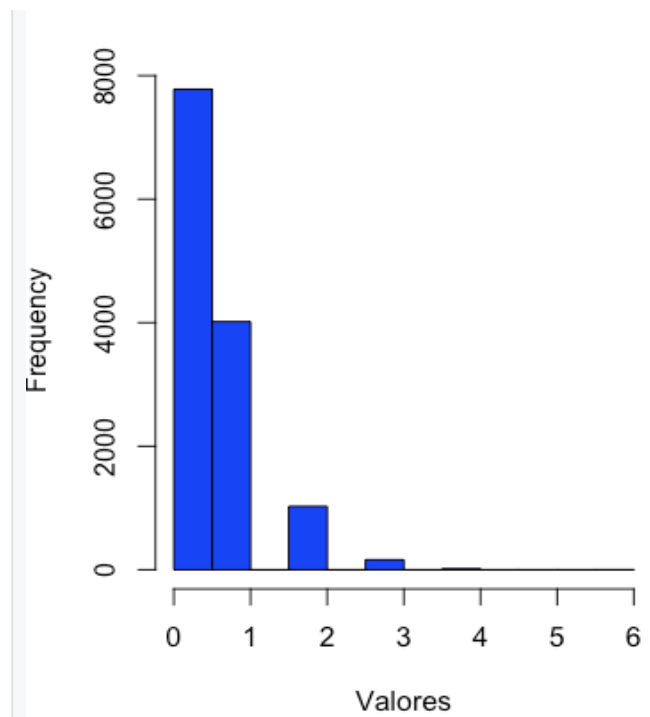


FIGURA 8: SIMULACIÓN DISTRIBUCIÓN DESPACHOS.

La distribución de despachos para el nivel de agregación de 12.500 cuadrantes se puede apreciar en la Figura 7. En este gráfico el número de frecuencia indica el número de cuadrantes que registran la cantidad de despachos indicada en el eje x. De la figura se puede notar que la mayor frecuencia, la moda de la distribución se encuentra centrada en 0 con una frecuencia cercana a 10.000, dicha cantidad de cuadrantes presenta 0 despachos en el período de dos meses estudiado.

En la Figura 8, se muestra la distribución de Poisson esperada asumiendo  $n=12.500$  y  $\lambda=0.67$ , que es el número esperado de despachos por cuadrante. Se puede notar que la distribución posee una menor frecuencia de ceros que la distribución real mostrada en la Figura 7. La moda también es 0, al igual que en la distribución de los datos reales. Si bien ambas distribuciones presentan similitudes a simple vista, la distribución real de los despachos posee una mayor masa de ceros, por lo que resulta conveniente utilizar un modelo de Regresión de Poisson con Ceros Inflados para poder hacerse cargo de esta cantidad extra de ceros en la distribución.

### 3.2.4 ELECCIÓN Y RESULTADOS DEL MODELO

Dado los modelos propuestos en el capítulo anterior, se escogió el Modelo de Regresión de Poisson con Ceros-Inflados pues se ajusta mejor a las características del problema, teniendo en cuenta que el modelo posee casi 2.000 ceros más que los que presentaría una distribución de Poisson normal. Dado esto, utilizar la otra opción, que es el modelo de Regresión de Poisson clásico, generaría un modelo poco consistente de acuerdo con sus estimadores y con bajo poder predictivo. Si bien este modelo puede soportar que la variable explicativa no cumpla una distribución de Poisson, no es capaz de soportar una cantidad excesiva de ceros. Debido a todo lo expuesto anteriormente, es que se determina optar por el modelo ZIP.

Para generar los modelos se tomaron algunas consideraciones con el fin de aumentar el nivel explicativo de este. Las consideraciones se mencionan a continuación:

1. **Generación de dummies para variable distancia:** Debido a que la influencia de la distancia en la cantidad de despachos esperados no es lineal, se generaron 3 dummies para esta variable, con el fin de explicar de mejor forma los despachos. El sentido de esto es que a mayor distancia se espera un número menor de despachos debido a que las personas dejarían de llamar o la base no podría dar cobertura a ese nivel. El efecto de esto no es lineal, es por esto que se crean intervalos utilizando los cuartiles de la distribución de distancias. Estos cuartiles quedan determinados dependiendo del nivel de agregación que se utilice para resolver el problema.

Estas nuevas variables siguen la siguiente lógica:

$$d_1 = \begin{cases} 1 & \text{si distancia} \in [0, \text{primer cuartil distancia}] \\ 0 & \text{si no} \end{cases}$$

$$d_2 = \begin{cases} 1 & \text{si distancia} \in [\text{primer cuartil distancia}, \text{segundo cuartil distancia}] \\ 0 & \text{si no} \end{cases}$$

$$d_3 = \begin{cases} 1 & \text{si distancia} \in [\text{segundo cuartil distancia}, \text{tercer cuartil distancia}] \\ 0 & \text{si no} \end{cases}$$

- 2. Exclusión de Comunas con comportamientos anómalos:** Debido a que hay algunas comunas que tienen características y comportamientos totalmente distintos al resto, deben ser excluidas a la hora de entrenar el modelo. En caso de no hacerlo el modelo podría sesgarse y ser poco consistente. Un ejemplo de esto se da en uno de los cuadrantes de Santiago, que posee 98 despachos, a pesar de que el promedio para los 12.500 cuadrantes es de solo 0,67 despachos por cuadrante. Este caso se debe a que Santiago es la comuna que presenta una mayor población flotante, pero como no hay variables que puedan medir este tipo de comportamiento a nivel desagregado (cuadrante) y para no sesgar el modelo, se escogió quitar estas variables de acuerdo con cierto criterio con las llamadas de emergencia.

El criterio de exclusión es el siguiente: en primer lugar, se calculó la métrica llamadas anuales/persona para cada comuna, En base a esto, se excluyó a las comunas que su promedio estaba sobre o bajo el promedio total al menos dos desviaciones estándar. Bajo este criterio, las comunas excluidas fueron La Reina, Santiago, Lo Barnechea, Las Condes y Vitacura. Las comunas excluidas, salvo Santiago, presentaron muy bajos llamados y son las comunas con mayor ingreso per cápita de la Región Metropolitana (según datos del CENSO 2017). Por lo cual, sería interesante haber podido agregar una variable a nivel cuadrante de ingreso, con lo que se podría haber controlado este tipo de casos, la que no se pudo conseguir. Luego de esto, quedaron 7.420 despachos.

- 3. Generación dummy ambulancias:** Al igual que el caso de la distancia desde cada cuadrante, el efecto del número de ambulancias en el número esperado de despachos no es lineal, es decir, el aumento de una ambulancia no implica doble cobertura o aumento del doble en la cantidad de despachos esperados. Es debido a esto, que se genera una variable con la lógica que se muestra a continuación:

$$amb = \begin{cases} 1 & \text{si la base más cercana posee 1 ambulancia} \\ 0 & \text{si posee 2 o más ambulancias} \end{cases}$$

- 4. Generación dummy Centro Hospitalarios:** Se generaron dos dummies que miden la distancia desde cada cuadrante al centro hospitalario más cercano. Esta dummy señala si el

cuadrante se encuentra a una distancia menor a 1.000 metros al centro hospitalario más cercano. Se genera esta variable debido a que se espera que si el lugar del incidente está a menos de un kilómetro de distancia, debería disminuir el número esperado de despachos, debido a que los pacientes podrían llegar por su cuenta o utilizando otro tipo de transporte que no sea ambulancia. La definición de esta dummy es la siguiente:

$$DIST\_Centro = \begin{cases} 1 & \text{si distancia a centro} \leq 1 \text{ km} \\ 0 & \text{si no} \end{cases}$$

Finalmente, utilizando las variables estudiadas y analizadas en el Capítulo 3.2.1 y las consideraciones mencionadas anteriormente, se corrieron dos modelos de Regresión de Poisson con Ceros-Inflados para distintos niveles de agregación. Los resultados de ambos modelos se pueden apreciar en la Tabla 4. M2 es modelo ZIP que incluye los parámetros asociados a las variables de distancia hacia el hospital y el centro de urgencia más cercano, M1 no cuenta con estas dos variables. En los resultados se puede notar que, en los dos modelos, las variables resultan ser estadísticamente significativas. Los modelos fueron corridos en el software R, un ejemplo de estos resultados se pueden ver en el Anexo D. El término que se encuentra arriba en cada recuadro es el valor del coeficiente perteneciente a cada una de las variables, el cual se denota como  $\beta$ . El número que se encuentra entre paréntesis debajo del coeficiente corresponde al error estándar, el que denotaremos con la letra  $Z$  y permite saber que tan confiable es el coeficiente. Para ejemplificar esto, si el intervalo  $[\beta - 1.96 * Z, \beta + 1.96 * Z]$  no contiene el 0 en su interior, se afirma que el coeficiente es estadísticamente significativo con un 95% de confiabilidad.

Los signos de las variables en términos generales son los esperados. El coeficiente asociado a  $\log(\text{viviendas})$  posee un signo positivo, debido a que mientras mayor es la población presente en el cuadrante, se esperaría un mayor número de despachos, por lo que se espera un signo positivo. Lo mismo sucede con las dummies de distancia, mientras menor es la distancia desde el cuadrante a la base más cercana, es esperable una mayor cantidad de despachos y esto se da, dado que las 3 dummies poseen un signo positivo y la dummy que indica una distancia menor presenta una magnitud de coeficiente mayor a las otras dos, lo cual tiene sentido.

Por otro lado, el coeficiente asociado a la dummy número de ambulancias posee un signo esperado, debido a que una base con solo una ambulancia debería tener menos cantidad de despachos esperados con respecto a una base de 2 o más ambulancias. Sin embargo, las variables distancia hospital y distancia centro de urgencia privado, a pesar de ser débilmente significativas poseen signos de coeficientes no esperados. Se esperaría que si la distancia entre el lugar de un incidente (cuadrante) y el centro hospitalario más cercano es menor a un kilómetro debería disminuir la cantidad de despachos esperados, ya que a menor distancia sería más factible llegar hasta el centro asistencial utilizando un medio de transporte alternativo o el paciente podría llegar por su propia cuenta.

Coeficientes/ Nivel Agregación	Nivel 4 (6.300)		Nivel 5 (12.500)		Nivel 6 (23.300 )		Nivel 7 (43.000)	
	M1	M2	M1	M2	M1	M2	M1	M2
<b>Intercepto</b>	-1.01 (0.04)	-0.98 (0.04)	-1.15 (0.04)	-1.11 (0.04)	-1.49 (0.04)	-1.44 (0.04)	-1.87 (0.04)	-1.83 (0.04)
<b>Log(Viviendas)</b>	0.69 (0.13)	0.69 (0.01)	0.55 (0.01)	0.56 (0.01)	0.47 (0.01)	0.47 (0.01)	0.47 (0.01)	0.47 (0.01)
<b>Distancia 1</b>	0.79 (0.04)	0.81 (0.04)	0.81 (0.04)	0.84 (0.04)	0.84 (0.04)	0.86 (0.03)	0.84 (0.04)	0.86 (0.04)
<b>Distancia 2</b>	0.63 (0.04)	0.61 (0.04)	0.63 (0.04)	0.61 (0.04)	0.65 (0.04)	0.63 (0.04)	0.67 (0.05)	0.65 (0.05)
<b>Distancia 3</b>	0.48 (0.04)	0.4 (0.04)	0.46 (0.04)	0.45 (0.04)	0.45 (0.04)	0.45 (0.04)	0.47 (0.04)	0.47 (0.04)
<b>Nro Ambulancias</b>	-0.24 (0.03)	-0.25 (0.02)	-0.24 (0.02)	-0.25 (0.02)	-0.26 (0.03)	-0.27 (0.02)	-0.27 (0.03)	-0.28 (0.03)
<b>Distancia Hospital</b>	0.67 (0.06)		0.6 (0.06)		0.63 (0.06)		0.55 (0.06)	
<b>Distancia Centro Urgencia</b>	0.05 (0.03)		0.1 (0.03)		0.1 (0.03)		0.09 (0.03)	

TABLA 4: RESULTADOS MODELOS.

Una vez analizados los resultados, se elige el modelo M2, pues es el que presenta todas sus variables estadísticamente significativas al 99%. Además, los coeficientes que acompañan a las variables poseen los signos esperados, lo que no se cumple en el modelo M1, ya que las variables dummies relacionadas a las distancias hacia los centros hospitalarios y a los centros de urgencias más cercanos resultan estadísticamente débilmente significativas y no poseen los signos esperados, por lo cual este modelo es finalmente rechazado. Esta es la principal razón por la que se termina escogiendo definitivamente el modelo M2.

Tal como se mencionó anteriormente, el modelo escogido y que será utilizado en los siguientes capítulos es el M2, el cual tiene la fórmula matemática que se muestra en la ecuación (3.2). Las variables  $d1$ ,  $d2$  y  $d3$  corresponden a las variables dummies de distancia, la variable  $nroamb$  corresponde a la dummy de ambulancia, la variable  $\log(viv)$  corresponde al logaritmo del número de viviendas que corresponden a cada cuadrante dividido en 1000 para poder reducir la magnitud de los coeficientes y poder comparar efectos de forma más directa con el resto de los coeficientes. Cabe destacar que, para términos de predicción, esta fórmula se debe desarrollar para cada cuadrante con el fin de obtener la cantidad de despachos esperados para cada uno de ellos. Es por esto que cada variable posee un subíndice  $i$ , que señala el cuadrante al que pertenece.

$$\log(E(desp_i)) = \beta_1 \times \log(viv_i) + \beta_2 \times d1_i + \beta_3 \times d2_i + \beta_4 \times d3_i + \beta_5 \times nroamb + \beta_0 \quad (3.2)$$

### 3.2.5 PODER PREDICTIVO DEL MODELO

Una vez escogido el modelo predictivo a utilizar, se debe comprobar la validez de este, es decir, es necesario verificar que el modelo sea consistente y tenga un poder predictivo adecuado. Para comprobar esto se utilizó el modelo seleccionado anteriormente para predecir los despachos sobre la data actual para ver qué tan bien logra ajustarse. Para poder realizar la predicción se generan las dummies que incluye el modelo para la data real. Posteriormente utilizando la fórmula 3.2 se obtiene el número de despachos esperados para cada uno de los cuadrantes.

Al realizar esta predicción, la suma de los despachos esperados para toda la Región Metropolitana fue de 7.479 despachos, solo 721 despachos menos que la data real, que presenta 8.200 despachos registrados. En la Figura 9, se presenta un mapa de calor generado utilizando Google Maps, donde se mapean las predicciones de despachos realizadas por el modelo escogido. En el mapa, cada uno de los círculos representan las bases actuales del SAMU, que ascienden a 24. Al comparar el mapa de calor generado por el modelo con el correspondiente a la realidad o la data original (Para revisar este mapa ir a Anexos E), se obtienen bastantes semejanzas entre ambos. Se puede notar que la mayor cantidad de despachos es observada en la parte céntrica de Santiago, donde existe una alta densidad de población y, por lo tanto, de viviendas, lo que explica esta concentración en los despachos que logra capturar bien el modelo.

De acuerdo con la suma de la cantidad de despachos esperados que predice el modelo para la Región Metropolitana, junto con el mapa de calor que se genera utilizando esta predicción, el cual parece acercarse bastante a la realidad de la data original de los despachos bimensuales, se puede afirmar que el modelo predictivo presenta un grado de consistencia y un poder predictivo aceptable.

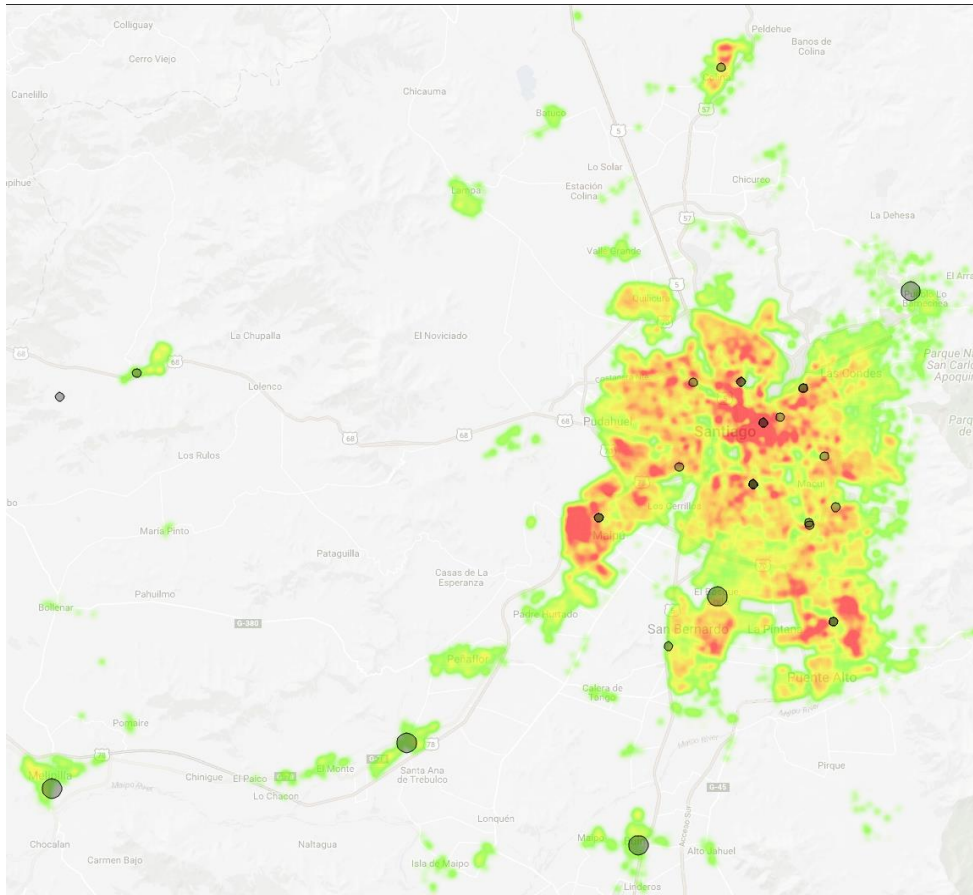


FIGURA 9: RESULTADOS MODELO PREDICTIVO.

### 3.2.6 ESTIMACIÓN DE PARÁMETROS PARA DISTINTOS NIVELES DE AGREGACIÓN

Para generar la heurística para la resolución del problema de optimización en el capítulo 4, primero es necesario entender bien el comportamiento de los coeficientes obtenidos del modelo ZIP (Zero Inflated Poisson) para los distintos niveles de agregación de cuadrantes, para esto se realiza un análisis de sensibilidad donde se corre el modelo seleccionado en distintos niveles de agregación. Los resultados se aprecian en la Tabla 5 donde se observa que a mayor nivel de agregación el intercepto va disminuyendo. También se puede notar que el coeficiente asociado a  $\log(\text{viviendas})$  va aumentando mientras el problema tiene menos cuadrantes. Los coeficientes asociados a los cuartiles de distancia no sufren grandes variaciones de acuerdo con los distintos niveles de agregación y el coeficiente asociado al número de ambulancias presenta una leve disminución mientras menor es el nivel de agregación. La mayor variación en el número de despachos esperados se da debido a la cantidad de viviendas presentes en cada uno de los cuadrantes. Los números denotados entre paréntesis corresponden al error estándar. Se puede notar también que la mayoría de los coeficientes, para casi todos los niveles de agregación resultan tener una significancia estadística al 95%.

Número cuadrantes	600	2.000	3.900	6.300	12.000	23.300	43.000
<b>Intercepto</b>	-0.76 (0.05)	-0.81 (0.04)	-0.85 (0.04)	-0.98 (0.04)	-1.11 (0.04)	-1.44 (0.04)	-1.83 (0.03)
<b>Log(Viviendas)</b>	0.85 (0.02)	0.8 (0.02)	0.72 (0.01)	0.69 (0.01)	0.56 (0.01)	0.47 (0.01)	0.47 (0.01)
<b>Distancia 1 (1er cuartil)</b>	0.8 (0.04)	0.81 (0.04)	0.87 (0.04)	0.81 (0.04)	0.84 (0.04)	0.86 (0.03)	0.86 (0.04)
<b>Distancia 2 (2do cuartil)</b>	0.62 (0.05)	0.64 (0.05)	0.65 (0.05)	0.61 (0.05)	0.61 (0.04)	0.63 (0.04)	0.65 (0.04)
<b>Distancia 3 (3er cuartil)</b>	0.44 (0.05)	0.43 (0.05)	0.5 (0.04)	0.4 (0.04)	0.45 (0.04)	0.45 (0.04)	0.47 (0.04)
<b>Nro Ambulancias</b>	-0.23 (0.02)	-0.24 (0.03)	-0.25 (0.02)	-0.25 (0.02)	-0.25 (0.02)	-0.27 (0.02)	-0.28 (0.03)

TABLA 5: COEFICIENTES MODELO ZIP PARA DISTINTOS NIVELES DE AGREGACIÓN.

El objetivo de realizar este análisis de sensibilidad radica primero en poder entender el cambio y el comportamiento en magnitud de los coeficientes para los distintos niveles de agregación. Luego, en la heurística propuesta siguiente en el siguiente capítulo, se busca resolver el modelo más desagregado posible, puesto que es el que posee un mayor nivel de precisión. Para poder buscar el modelo de más cuadrantes que se puede resolver es necesario conocer los coeficientes de los distintos niveles de agregación.



## CAPÍTULO 4: MODELOS DE OPTIMIZACIÓN

Al generar el modelo predictivo de Poisson con Ceros-Inflados en el capítulo anterior se puede predecir el número de despachos para cada cuadrante, y obtener una función que permite relacionar matemáticamente cómo afecta la distancia desde cada cuadrante a la base más cercana en el número esperado de despachos. Esto abre la posibilidad de mover las bases, de tal forma de maximizar la cantidad de despachos esperados en la Región Metropolitana.

En este capítulo se busca definir y proponer un problema de optimización que permita aprovechar la función obtenida en el capítulo anterior. Además, se plantea una heurística para la resolución del problema a gran tamaño, el cual resulta bastante complejo de resolver.

### 4.1 DEFINICIÓN DE PROBLEMA

El problema de optimización que se presenta consiste en un modelo de localización de bases, donde se busca maximizar la cobertura de la demanda de despachos de ambulancias, utilizando recursos limitados, que en este caso son las 24 bases que contienen las 36 ambulancias. Para efectos del problema se consideran dos tipos de bases, las que poseen una ambulancia y las que poseen dos ambulancias. Existen 12 bases de cada tipo.

En el Capítulo 3, se obtuvo una función matemática que relaciona la esperanza de los despachos con las características demográficas de cada uno de los cuadrantes, la distancia desde cada uno de estos a la base más cercana y el número de ambulancias que tiene la base más cercana. Es decir, la demanda de cada cuadrante no es determinística, sino que depende de las ubicaciones geográficas de las bases. En este caso, la variable de decisión es la ubicación o el cuadrante donde se localiza cada una de las bases y la asignación de cada cuadrante a una de las 24 bases.

Dentro de las restricciones naturales que posee este tipo de problemas se encuentra la capacidad limitada de recursos, es decir de bases, que como ya se dijo, son 12 de cada tipo. También existen restricciones de capacidad de atención de cada base, cada una puede atender a un número limitado de cuadrantes dependiendo de la naturaleza o tipo de la base. Otra restricción natural que aparece consiste en asignar cada cuadrante a una base, que es la que le prestará atención a los incidentes o emergencias que surjan dentro de él. Además, existe una restricción asociada al número máximo de bases que puede haber en cada cuadrante.

## 4.2 COMPLEJIDAD DEL PROBLEMA

Una de las principales dificultades que presenta la resolución del problema radica en el tamaño de éste, pues como se vio en el Capítulo 3.2.2, existen distintos niveles de agregación. Mientras menor sea el nivel de agregación, es decir, mientras más cuadrantes existan, mayor será la matriz de distancia entre los cuadrantes. Este crecimiento es claramente no lineal y vuelve el problema mucho más complejo de lo que ya es. Si bien, el problema se podría resolver utilizando el nivel de agregación más agregado, el de 600 cuadrantes, se pierde bastante precisión a la hora de escoger el lugar donde se ubicará la base, ya que los cuadrantes a este nivel poseen una mayor superficie. Entonces, se genera el trade-off entre la complejidad de resolución de problema y la precisión de la ubicación de las bases para la maximización de cobertura de despachos. Es debido a esto que en el Capítulo 4.4 se buscan heurísticas de resolución para este problema.

Otra gran dificultad para la resolución del problema se encuentra en que como la función matemática que determina los despachos proviene del modelo ZIP, la cantidad de despachos esperados totales es igual a la sumatoria de exponenciales de una combinación lineal de parámetros como se puede apreciar en la ecuación (4.1), lo que claramente resulta ser no lineal y no se puede modelar como la función objetivo de un Problema de Programación Lineal (PPL). El no poder escribir un PPL como tal, hace que el problema se vuelva muy complejo de resolver, ya que se transforma en un problema no lineal.

$$\sum_i^n E(D_i) = \sum_1^n \exp(\beta_1 \times \log(vivi_i) + \beta_2 \times d1_i + \beta_3 \times d2_i + \beta_4 \times d3_i + \beta_5 \times l_i + \beta_0) \quad (4.1)$$

Finalmente, un problema de menor grado que se presenta es que debido a la influencia no lineal de la distancia entre cuadrantes y bases en la cantidad de despachos esperados, la función objetivo presenta dummies de acuerdo a segmentos de distancias. En términos concretos, a la hora de formular matemáticamente el problema se deben incluir variables binarias que puedan captar en qué segmento se encuentra la distancia desde cada cuadrante a la base más cercana. Generar estas variables y darles sentido mediante restricciones representa un desafío importante.

## 4.3 FORMULACIÓN MATEMÁTICA (PPL)

El problema descrito anteriormente, puede ser formulado matemáticamente como un Problema de Programación Lineal Mixto. El problema resulta muy similar a un Maximal Covering Location Problem (MCLP), aunque presenta algunas variaciones para poder adoptar este problema clásico y fuertemente estudiado al problema actual que presenta el SAMU. Una de las principales diferencias entre ese modelo y el que se presenta en este capítulo radica en que en este caso la demanda de emergencia no es estática, si no que varía de acuerdo con el comportamiento de ciertas

variables. Además, como ya se mencionó, la función objetivo no es lineal debido a que las variables se encuentran dentro de una exponencial. A continuación, se muestra el PPL:

### **Función Objetivo (no lineal)**

$$\text{Max } E(D) = \text{Max} \sum_{i=1}^n \exp(\beta_0 + \beta_1 * \log(viv_i) + \beta_2 * t_{i1} + \beta_3 * t_{i2} + \beta_4 * t_{i3} + \beta_5 * l_i)$$

### **Variables**

$$x_j^1 = \begin{cases} 1 & \text{si se ubica una base con una ambulancia en cuadrante } j \\ 0 & \text{si no} \end{cases}$$

$$x_j^2 = \begin{cases} 1 & \text{si se ubica base con dos ambulancias en cuadrante } j \\ 0 & \text{si no} \end{cases}$$

$$y_{ij}^1 = \begin{cases} 1 & \text{si se asigna base tipo 1 ubicada en } j \text{ al cuadrante } i \\ 0 & \text{si no} \end{cases}$$

$$y_{ij}^2 = \begin{cases} 1 & \text{si se asigna base tipo 2 ubicada en } j \text{ al cuadrante } i \\ 0 & \text{si no} \end{cases}$$

$$t_{i1} = \begin{cases} 1 & \text{si se activa segmento 1 para cuadrante } i \\ 0 & \text{si no} \end{cases}$$

$$t_{i2} = \begin{cases} 1 & \text{si se activa segmento 2 para cuadrante } i \\ 0 & \text{si no} \end{cases}$$

$$t_{i3} = \begin{cases} 1 & \text{si se activa segmento 3 para cuadrante } i \\ 0 & \text{si no} \end{cases}$$

$F_i$  = distancia del cuadrante  $i$  a su base asignada

$f_{i1}$  = segmento 1 de distancia de cuadrante  $i$  a ambulancia asignada

$f_{i2}$  = segmento 2 de distancia de cuadrante  $i$  a ambulancia asignada

$f_{i3}$  = segmento 3 de distancia de cuadrante  $i$  a ambulancia asignada

$$l_i = \begin{cases} 1 & \text{si el cuadrante } i \text{ tiene una base de una ambulancia asignada} \\ 0 & \text{si no} \end{cases}$$

## Restricciones

1.  $\sum_j x_j^1 = p^1$  Utilización de bases con 1 ambulancia
2.  $\sum_j x_j^2 = p^2$  Utilización de bases con 2 ambulancias
3.  $x_j^1 + x_j^2 \leq 1 \quad \forall j$  Cada cuadrante puede tener solo un tipo de base
4.  $y_{ij}^1 \leq x_j^1 \quad \forall i, j$  Debe estar activo el recurso 1 para ser asignado
5.  $y_{ij}^2 \leq x_j^2 \quad \forall i, j$  Debe estar activo el recurso 2 para ser asignado
6.  $\sum_j (y_{ij}^1 + y_{ij}^2) = 1 \quad \forall i$  Cada cuadrante debe tener una base asignada
7.  $\sum_j Dist_{ij} * (y_{ij}^1 + y_{ij}^2) = F_i \quad \forall i$  Distancia desde el cuadrante a base asignada
8.  $\sum_j y_{ij}^1 = l_i \quad \forall i$  Si el cuadrante  $i$  tiene una base de una ambulancia asignada
9.  $F_i = f_{i1} + f_{i2} + f_{i3}$  Repartición flujo
10.  $(t_{i2} + t_{i3}) * a_1 \leq f_{i1} \leq a_1$  Flujo segmento 1
11.  $t_{i3} * (a_2 - a_1) \leq f_{i2} \leq (a_2 - a_1) * (t_{i2} + t_{i3})$  Flujo segmento 2
12.  $0 \leq f_{i3} \leq M * t_{i3}$  Flujo segmento 3
13.  $t_{i1} + t_{i2} + t_{i3} = 1$  Solo puede estar en un segmento
14.  $y_{ij}^1 * Dist_{i,j} \leq cotasup1$  Distancia máxima que puede cubrir una base tipo 1
15.  $y_{ij}^2 * Dist_{i,j} \leq cotasup2$  Distancia máxima que puede cubrir una base tipo 2
16.  $y_{ij}^1, x_j^1, y_{ij}^2, x_j^2, t_{i1}, t_{i2}, t_{i3}, l_i \in \{0,1\}$  Naturaleza de las variables
17.  $F_i, f_{i1}, f_{i2}, f_{i3} \in \mathbb{R}$  Naturaleza de las variables

Como se puede apreciar la función objetivo no es lineal, por lo que se propone una solución a este problema en el siguiente capítulo. A pesar de esto, el resto de la estructura del PPL presenta un comportamiento lineal. Dentro de las variables del problema se encuentra la variable  $x_j$ , que indica en qué cuadrante se ubicará una base. Existen dos de estas variables en el problema, ya que hay bases tipo 1 que cuentan con una ambulancia y bases tipo 2 que las que poseen dos ambulancias.

También se encuentra la variable  $y_{ij}$ , la cual indica la asignación de una base ubicada en el cuadrante  $j$  a un cuadrante  $i$ . Existen dos tipos de esta variable debido a la misma razón que la variable  $x_j$ , ambas son variables binarias. Además, se crean las variables binarias  $t_{in}$ , la cual determina en que segmento o cuartil se encuentra la distancia entre el cuadrante y la base asignada a este, funcionan como las dummies de la función determinada para predecir los despachos esperados en el Capítulo 3. Junto a las variables ya mencionadas, se encuentran las variables de flujo, las cuales son números reales continuos, y sirven para determinar el segmento donde se ubica la distancia. Se relacionan con las variables  $t_{in}$  como se verá más adelante. Finalmente, se encuentra la variable  $l_i$  la cual indica si el cuadrante  $i$  tiene asignada una base de tipo 1, es decir, con una ambulancia.

Las restricciones 1 y 2 representan la capacidad de recursos que posee el problema.  $P$  indica el número de bases de cada tipo que se deben utilizar. Como se puede apreciar, el problema está formulado de manera que no permite utilizar menos recursos de los disponibles. La restricción 3 indica que un cuadrante puede tener como máximo una base en su interior, independiente del tipo que ésta sea. Las restricciones 4 y 5 indican que no se puede asignar un recurso si no está localizado en el cuadrante  $j$ . La restricción 6 indica que cada cuadrante debe tener una, y solo una base asignada de cualquier tipo. La número 7 sirve para definir la variable  $l_i$ .

Por otro lado, las restricciones 7 y 9, definen y distribuyen el flujo  $F_i$  que no es más que la distancia desde el cuadrante a la base asignada. Las número 10, 11, 12 y 13 sirven para relacionar las variables entre ellas y para determinar a qué segmento corresponde cada una de las distancias. Finalmente, las número 14 y 15 son restricciones que permiten dar una capacidad máxima de cobertura a cada base localizada. Esta capacidad máxima está dada por una distancia máxima que puede cubrir cada una de las bases, dependiendo al tipo que pertenezca.

## 4.4 HEURÍSTICA PARA LA RESOLUCIÓN DE PROBLEMA

### 4.4.1 LINEALIZACIÓN DE LA FUNCIÓN OBJETIVO

En este capítulo se busca proponer soluciones y heurísticas a los problemas que presenta el modelo, los cuales fueron propuestos en el Capítulo 4.2. Un gran problema que se describe en dicho capítulo es que la función objetivo del modelo matemático no es lineal, lo que hace que el problema con tal estructura se vuelva mucho más complejo de resolver. Para poder lidiar con esto es necesario linealizar esta función, para llegar a ello se realizaron los siguientes pasos.

En primer lugar, como la función objetivo en un principio, es la maximización de la suma de los despachos esperados de cada cuadrante, esta se puede dividir por el número de cuadrantes  $n$  sin afectar el resultado del problema, ya que se divide por una constante positiva. Al hacer eso, la función objetivo es el promedio aritmético de los despachos y queda como se muestra en la ecuación (4.2).

$$\frac{\sum_1^n \exp(\beta_1 \times \log(vivi_i) + \beta_2 \times t1_i + \beta_3 \times t2_i + \beta_4 \times t3_i + \beta_5 \times l_i + \beta_0)}{n} \quad (4.2)$$

Claramente la función presentada en la ecuación (4.2) sigue siendo no lineal con respecto a las variables de decisión. Es por esto que se decide utilizar el promedio geométrico. Según [30], la media geométrica es una buena métrica para utilizarse en modelos de conteo como este, ya que se hace cargo de mejor forma de valores extremos u outliers, los cuales podrían ser números muy grandes o muy pequeños y que ensucian de forma significativa el promedio aritmético simple. Además, señala que esta métrica siempre es igual o menor que la media aritmética. El principal problema que presenta la media geométrica radica en que si uno de los datos es 0 la media también valdrá cero. Afortunadamente en este caso, y debido a que la función posee un intercepto  $\beta_0$ , el número de despachos esperados para cada cuadrante debiera ser distinto de 0. La fórmula matemática de la media geométrica se puede apreciar en la ecuación (4.3). En palabras simples se puede definir como la raíz enésima de la multiplicación de los  $n$  datos.

$$\sqrt[n]{\prod_1^n D_i} \quad (4.3)$$

Desarrollando un poco más esta expresión, expandiendo la variable despachos  $D_i$  queda como se muestra en la ecuación (4.4).

$$\sqrt[n]{\prod_1^n \exp(\beta_1 \times \log(vivi_i) + \beta_2 \times t1_i + \beta_3 \times t2_i + \beta_4 \times t3_i + \beta_5 \times l_i + \beta_0)} \quad (4.4)$$

Al desarrollar la expresión presente en la ecuación (4.4), utilizando propiedades de potencias, queda como se muestra en la ecuación (4.5). Con esta transformación se logra hacer desaparecer la pitatoria para generar una sumatoria como exponente.

$$\sqrt[n]{\exp(\sum_1^n (\beta_1 \times \log(vivi_i) + \beta_2 \times t1_i + \beta_3 \times t2_i + \beta_4 \times t3_i + \beta_5 \times l_i + \beta_0))} \quad (4.5)$$

A pesar de todas las transformaciones realizadas, la función objetivo mostrada en la ecuación (4.5) continúa siendo no lineal, pues presenta variables de decisión del problema de optimización en el exponente. Para solucionar esto, se le aplica logaritmo a esta función, lo que se puede hacer debido a que la función logaritmo es monótona creciente, es decir, a medida que aumenta el argumento del logaritmo, también lo hace la función. En la ecuación (4.6) se puede observar cómo queda la función objetivo una vez que se le aplica el logaritmo a la función, desaparece la exponencial y el radical  $n$  pasa a dividir esta nueva expresión. Una forma de interpretar fácilmente esta expresión es que representa el promedio de los exponentes de la exponencial, que es lo que se busca maximizar.

$$\frac{\sum_1^n (\beta_1 \times \log(vivi_i) + \beta_2 \times t1_i + \beta_3 \times t2_i + \beta_4 \times t3_i + \beta_5 \times l_i + \beta_0)}{n} \quad (4.6)$$

Finalmente, como  $n$  es una constante positiva dividiendo la expresión en la ecuación 4.6, no genera un real aporte al problema a la hora de maximizar la función. Dado esto, es posible sacarlo de la función sin generar ningún cambio en los resultados a obtener. Al hacer esto, la función objetivo queda como se muestra en la ecuación (4.7).

$$\sum_1^n (\beta_1 \times \log(vivi_i) + \beta_2 \times t1_i + \beta_3 \times t2_i + \beta_4 \times t3_i + \beta_5 \times l_i + \beta_0) \quad (4.7)$$

#### 4.4.2 HEURÍSTICA PARA RESOLUCIÓN DE PROBLEMAS DE GRAN TAMAÑO

Como se mencionó en el Capítulo 4.2, una de las principales dificultades que presenta este tipo de problemas a la hora de ser resueltos, radica en su tamaño. Mientras más crece en tamaño, es decir, mientras más cuadrantes o más desagregado está el problema, más compleja se hace su resolución. Esto se debe a que, al hacer crecer el número de cuadrantes, la matriz de distancias entre cada uno de ellos se hace mucho más grande, generando un problema de mucha más complejidad debido a la combinatoriedad.

Resolver un problema de mayor tamaño genera una mayor precisión en los resultados, pues posee cuadrantes con superficies mucho mayores, por lo que permite señalar de mejor forma dónde debería ubicarse cada base del SAMU, y como ya se dijo, resolverlo es mucho más complejo. Entonces queda claro el trade off existente entre la precisión al resolver un problema de mayor tamaño y la dificultad de la resolución de este.

Para poder resolver un problema de mayor tamaño, es necesario generar una heurística que permita dividir el problema en subproblemas que puedan ser resueltos de forma más sencilla para luego poder agrupar estos resultados y formar un cuadro más grande, una solución factible para el problema más grande.

1. **Resolución problema con mayor agregación:** Se propone resolver el problema completo al nivel más desagregado posible, es el que tiene un mayor nivel de precisión posible con respecto a la ubicación de las bases, puesto que a mayor nivel de desagregación los cuadrantes son un número mayor y más pequeños en superficie. El máximo nivel de desagregación que se logró resolver es para 600. Luego, se utilizan los coeficientes asociados a este nivel de agregación obtenidos en el capítulo 3.2.6. Una vez resuelto el problema, se puede obtener una repartición óptima de las bases para este nivel de agregación. Se debe guardar la localización de estas para luego repartirlas en las grillas o subdivisiones que se generan en la siguiente etapa a la hora de resolver los subproblemas con mayor nivel de desagregación, con una mayor cantidad de cuadrantes por zona.
2. **Generación de grillas o subdivisiones:** Posteriormente, se debe dividir el mapa completo de la Región Metropolitana en grillas, con el fin de poder separarlo en subproblemas con una cantidad menor de cuadrantes que el problema más grande resuelto, pero con un nivel de precisión mayor al poseer cuadrantes más pequeños en superficie. El número de grillas que se deben generar está acotado por el problema más grande que se puede resolver (600 cuadrantes en este caso). Entonces para poder resolver el problema general de 3900 cuadrantes, se deben utilizar al menos 7 grillas ( $3.900/600=6.5$ ). Mientras mayor sea el número de grillas generadas habrá una cantidad de cuadrantes menor dentro de cada una y el problema será más simple de resolver, debido a la combinatoriedad del mismo. Sin embargo, se proponen algunas restricciones a la hora de generar y confeccionar estos cuadrantes. Para empezar, la cantidad de cuadrantes deben estar juntos, es decir, no se puede incluir un cuadrante a la grilla si éste no está adyacente a algún cuadrante ya escogido dentro de la grilla. Otra recomendación para no resolver problemas de tanta dificultad como el máximo resuelto (600 cuadrantes en este caso), es balancear la carga de los cuadrantes dentro de lo posible, es decir, intentar que las grillas tengan una cantidad lo más similar posible entre ellas. Para forzar esto, se propone que ninguna grilla tenga más de 1.5 veces los cuadrantes de otra grilla en su interior. Otra restricción que es bueno imponer a la hora de crear las grillas tiene que ver con la cantidad máxima de bases de ambulancias que puede



haber dentro de cada una de las grillas, mientras mayor sea el número de ambulancias dentro de estas mayor será la complejidad del problema, mientras mejor repartidas queden más simple computacionalmente es la resolución de cada uno de los subproblemas. En este caso, se fijó un límite de 4 bases de ambulancias por grilla generada. Otra consideración necesaria para formar las grillas es que los cuadrantes solo pueden pertenecer solo a una, es decir, una vez que es agregado a una zona o grilla no puede pertenecer a otra, esto evita que se considere un cuadrante en dos zonas y que se cuenten dos veces los despachos. Además, a la hora de asignar las bases, se recomienda que, si hay dos bases cercanas, a menos de 7 km, según la localización de la resolución del modelo de 600 cuadrantes, estas queden en una misma grilla para poder evitar lo más posible las complicaciones que se pueden dar en los bordes de cada zona, ya que podrían quedar dos bases de ambulancias muy juntas.

- 3. Fijación de bases según resolución del problema más desagregado:** Finalmente, se deben fijar las bases en cada grilla según la distribución encontrada en la etapa 1 de la heurística y resolver un problema para cada una de las grillas generadas en la etapa 2. Resolver estos problemas debería resultar mucho menos complejo y rápido, ya que en cada uno de estos no se cuenta con las 24 bases iniciales sino, con el número indicado por la asignación de la etapa 1, por lo que combinatorialmente serán problemas más simples de resolver que el resuelto en el paso 1 de la heurística.

La heurística propuesta, si bien es efectiva y simplifica el problema complejo entregando buenos resultados, también presenta algunos potenciales problemas que se recomienda revisar. Por ejemplo, como cada subproblema de cada grilla es resuelto de manera independiente con respecto a los resueltos en otras grillas, es posible que a pesar de obtener el óptimo para cada subproblema no se llegue al óptimo global, es decir, al óptimo resolviendo los problemas de manera simultánea. Puede suceder que haya dos bases muy cercanas, en las fronteras de las grillas, por ejemplo, lo cual claramente no sería una situación óptima. Si bien esto puede pasar, debería ser poco frecuente, pues la resolución del problema busca reducir las distancias entre las bases de ambulancias y los cuadrantes por lo que escoge lugares más céntricos dentro de cada grilla. Pero podría darse si las fronteras poseen una densidad poblacional muy elevada. Además, a la hora de generar las grillas como se explicó, se busca dejar en una misma grilla las bases que quedan cercanas una vez resuelto el primer problema, el del paso 1 de la heurística, por lo que hace más difícil que se de esta situación. A pesar de esto, una vez obtenido el panorama general y resuelto el problema para cada grilla se propone analizar este tipo de casos.

# **CAPÍTULO 5: RESULTADOS PARA PROBLEMA DE LOCALIZACIÓN DE BASES DEL SAMU EN SANTIAGO**

Una vez obtenido un modelo con un aceptable nivel predictivo, haber generado un problema de optimización para poder maximizar la cantidad de despachos esperados para cada uno de los cuadrantes en que se divide la Región Metropolitana y proponer una heurística para la resolución de este modelo matemático, se realizan diversos experimentos utilizando los distintos modelos propuestos hasta acá, donde todos persiguen el mismo objetivo, maximizar la cobertura de despachos del SAMU en la Región Metropolitana. Es por esto, que en el presente capítulo se muestran diversas aplicaciones, junto con las complicaciones y los resultados obtenidos para cada una de ellas.

## **5.1 APLICACIONES MODELO PREDICTIVO**

A continuación, se realizan dos ejemplos prácticos de aplicaciones derivadas del modelo predictivo propuesto en el Capítulo 3 y del modelo de optimización propuesto en el Capítulo 4. Ambos algoritmos se describen a continuación.

### **5.1.1 ALGORITMO Y RESULTADOS DE SIMULACIÓN SIMPLE (BASE FIJADA)**

Una de las posibles aplicaciones que nacen con la obtención del modelo predictivo, es la posibilidad de generar simulación de escenarios. Particularmente en este capítulo, se muestra la simulación al agregar una base fija en una ubicación, volver a calcular las variables de distancia y predecir utilizando los coeficientes obtenidos por los modelos.

El SAMU aportó una lista de lugares donde se evalúa la construcción de nuevas bases, lo que se busca es predecir y entender es cómo se comportaría la demanda de despachos de ambulancias si se agregara una base adicional en los puntos indicados. Se observa la lista de las bases a evaluar entregadas por la organización en la Tabla 6.

Para realizar las simulaciones se tomó el modelo con nivel de agregación de 12.000 cuadrantes. Cada una de las simulaciones, como ya se indicó, se comparan ortogonalmente por separado. Es decir, en cada caso se incluye solamente una de las bases indicadas en la tabla.

Nombre Base	DIRECCION (Latitud,Longitud)	REFERENCIA	AMBULANCIAS
Base Buin	-33.734,-70.739	Hospital de Buin	2
Base Paine	-33.868,-70.747	hospital de Paine	2
Base Melipilla	-33.695,-71.215	Hospital de Melipilla	2
Base Talagante	-33.663,-70.927	Centro Talagante	1
Base El Bosque	-33.563,-70.675	Centro el Bosque	1
Base Lo Barnechea	-33.352,-70.518	Centro lo Barnechea	1

TABLA 6 :ZONAS EVALUADAS CON NUEVAS BASES.

En la Tabla 7 se detallan los resultados de las distintas simulaciones, en ellos se puede ver que la construcción de la base que más impacto generaría en el número de despachos esperados totales sería la que se construiría en la comuna de El Bosque. Su construcción implicaría un aumento de 333.4 despachos en dos meses, lo cual se traduce en un 4.4 % adicional de despachos esperados con respecto al caso base, que es de 7.479 despachos y se obtiene corriendo el modelo con las bases actualmente construidas. También se observa que la construcción de la base que menos impacto tendría según el modelo sería la construida en Paine, la que solo provocaría un aumento de un 0.3% con respecto al caso base.

COMUNA BASE	Despachos	Diferencia	Porcentaje
EL BOSQUE	7813.0	333.4	4.4%
LO BARNECHEA	7605.0	125.3	1.6%
BUIN	7590.3	110.7	1.4%
MELIPILLA	7537.0	57.4	0.7%
TALAGANTE	7513.3	33.7	0.4%
PAINE	7507.5	27.9	0.3%
CASO BASE	7479.6	0.	0%

TABLA 7:RESULTADOS SIMULACIONES.

Para observar algunos de estos resultados, de forma más gráfica y explicativa, se puede ver en la Figura 10 un mapa de calor con la simulación de la construcción de una base en El Bosque. En la imagen, los círculos pequeños corresponden a las bases actuales construidas por el SAMU, mientras que el círculo grande, corresponde al lugar donde se construiría la base simulada. Se puede observar que alrededor del círculo de la nueva base, se torna más rojo el mapa, lo que implica una mayor cantidad de despachos esperados en la zona, debido a la menor distancia que existe entre los cuadrantes presentes en la zona y la base más cercana ahora, que es la construida en El Bosque. Además, es posible notar que la zona donde se construye la base se encuentra sin bases cercanas en un radio considerable.

Para notar de mejor forma la diferencia gráfica al construir la potencial base en la comuna de El Bosque se recomienda mirar el Anexo H, donde se encuentra el mapa de calor correspondiente al caso base, es decir, la simulación del modelo para la data real. Viendo esa imagen, es posible comparar y ver el cambio al construir la base en El Bosque. Se puede observar cómo se torna más roja la zona cercana a la base al realizar la comparación.

Para ver la simulación realizada en la comuna de Lo Barnechea, se puede revisar el Anexo G, donde se aprecia un mapa de calor con las características similares al que se presenta en la Figura 10.

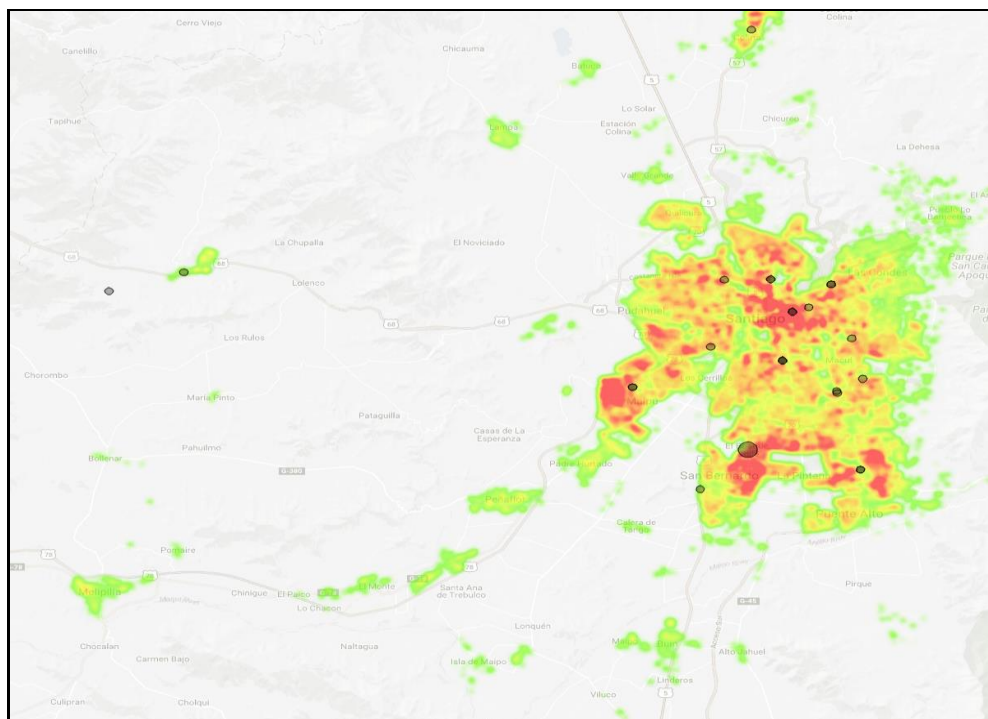


FIGURA 10: RESULTADOS MAPA DE CALOR PARA BASE EN EL BOSQUE.

### 5.1.2 ALGORITMO Y RESULTADOS DE ELECCIÓN DE LOCALIZACIÓN DE BASES (SET DE OPCIONES)

Otra aplicación posible que se puede realizar con el modelo predictivo obtenido, y un poco más compleja a la presentada en la parte anterior, consiste en poder seleccionar el lugar donde se maximizarían los despachos dado un set de opciones donde podría estar la nueva base.

Para ser más específico, esta simulación permite escoger entre potenciales lugares de construcción de una nueva base, el lugar que permita aumentar al máximo los despachos esperados. También se puede utilizar para construir más de una potencial base entre varios puntos de opción. En este caso se utiliza la siguiente heurística propuesta: en primer lugar, la primera base se ubica donde más despachos esperados se dan. Una vez fijada la primera base, se comienza a probar poniendo las otras bases para ver cual, de todas las opciones, en conjunto con la primera base ya fijada, generan una mayor cantidad de despachos globales.

Un experimento realizado utilizando esta simulación consistió en quitar una base construida en la comuna de Quinta Normal y luego agregar dos bases dentro del sector comprendido por las comunas de Quinta Normal, Pudahuel, Cerro Navia, Lo Prado y Renca. Cada una de estas nuevas bases a agregar poseen dos ambulancias. Para poder obtener las potenciales locaciones de esta nueva base, se construye un cuadrilátero de 4 puntos que rodee la zona a estudiar, quedando la zona inscrita a este cuadrilátero. Posteriormente, se rellena el cuadrilátero con 16 puntos, cada uno de estos puntos representa al set de opciones donde se localizarán estas dos bases. En la Figura 11, se aprecian las opciones donde se pueden construir las bases, estas opciones están representadas por los círculos grandes ubicados en la zona nor-poniente de Santiago.

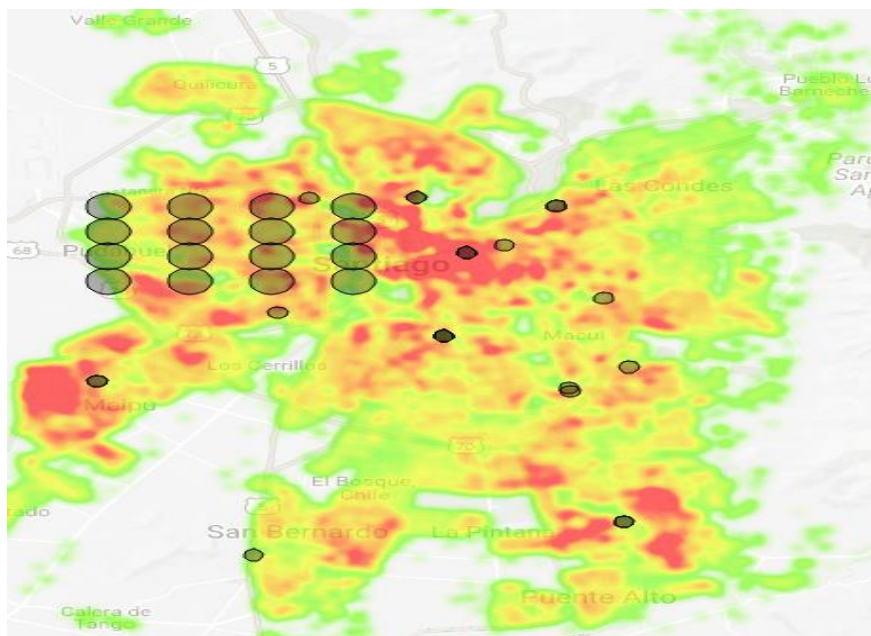


FIGURA 11:SET DE OPCIONES LOCALIZACIÓN.

Como se mencionó, en primer lugar, se fija la base que genera un mayor aumento en los despachos. La primera base en este experimento fue localizada en el punto de coordenadas (-33.60,-70.71). Esta nueva localización genera una cantidad de despachos esperados de 7.648, lo que representa un aumento de un 2.3%. Este punto y cómo afecta la localización de la base en este lugar es apreciable en la Figura 12. En ella se torna roja la zona cercana al punto. Una vez fijada esta primera locación, se procede a buscar la segunda, lo que se realiza probando las 15 opciones restantes en conjunto con la que ya fue fijada. El punto escogido en esta segunda iteración tiene coordenadas (-33.457,-70.737) y genera un aumento, en conjunto con el primer punto fijado, de un 4.0% del número esperado de despachos, lo cual se traduce en un valor esperado de 299.13 despachos adicionales. En la Figura 13 se puede apreciar un mapa de calor que presenta los dos puntos donde son fijadas las bases una vez concluida la simulación. En esta figura se observa cómo se vuelve un poco más roja la zona donde se localizan ambas bases.

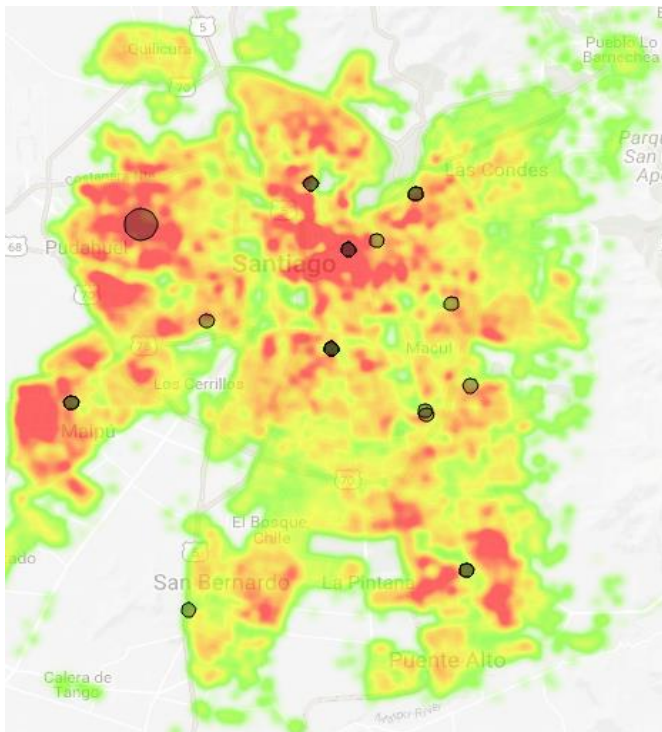


FIGURA 13: RESULTADOS PRIMERA ITERACIÓN.

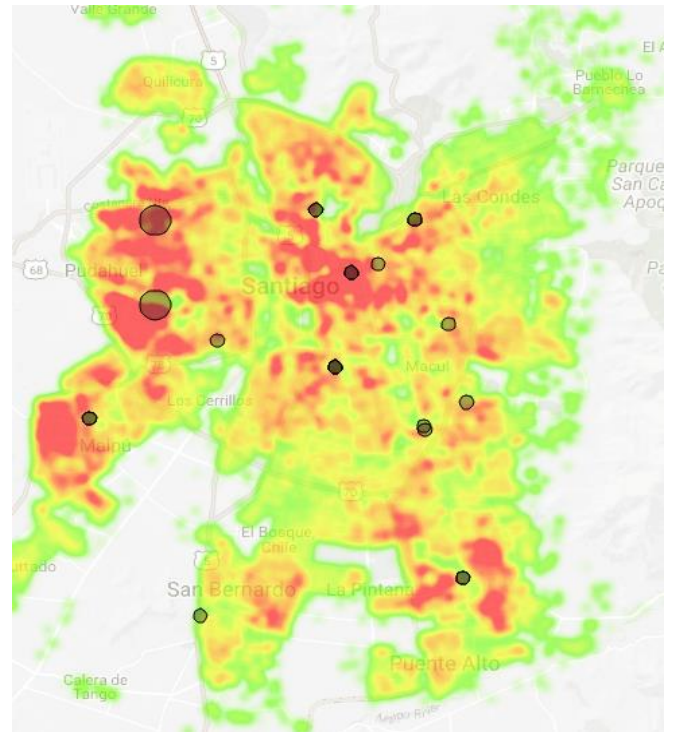


FIGURA 12: RESULTADOS SEGUNDA ITERACIÓN.

Los resultados obtenidos utilizando la heurística mencionada para esta simulación son bastantes alentadores. Sin embargo, hay que tener en cuenta que se puede estar en presencia de un máximo local, ya que combinatorialmente hay muchas más combinaciones de las que fueron probadas. Existen 120 combinaciones posibles y en la heurística tan solo se probaron 15 de estas.

## **5.2 RESULTADOS MODELO DE OPTIMIZACIÓN**

En el Capítulo 4, se definió un modelo de optimización que permite relocalizar todas las bases que posee el SAMU. Sin embargo, es un problema complicado de resolver, por lo que se propone una heurística que permitiera facilitar el desarrollo de este. En este capítulo se busca mostrar los resultados de la heurística propuesta, junto con un análisis de estos.

### **5.2.1 ANÁLISIS RESULTADOS RESOLUCIÓN DEL PROBLEMA**

Como se propuso en la heurística descrita en el Capítulo 4.4.2 se debe, en primer lugar, resolver el problema para el mayor nivel de desagregación posible, es decir, el nivel que posee una mayor cantidad de cuadrantes y por ende, es el más difícil que se pueda resolver. Posteriormente, y tal como se indicó en el Capítulo 4, se divide Santiago en grillas siguiendo ciertas restricciones que se explican en el capítulo ya mencionado. Si bien el problema fue resuelto para distintos números de grilla. En el caso que se muestra a continuación se confeccionaron 10 grillas, en el Capítulo 5.2.2 se menciona porqué fue escogido particularmente este número.

Posteriormente, se incluyeron las bases según se indica para la resolución del problema más gregado, es decir, se asignaron las bases a cada una de las grillas de acuerdo con la ubicación óptima para el problema más agregado. Finalmente, se resolvieron 10 subproblemas para el nivel de agregación de 3.900 cuadrantes. La elección de esta cantidad de grillas es explicada más adelante con mayor nivel de detalle. Cada uno de estos problemas fue mucho más simple de resolver ya que solo se debían repartir como máximo 4 bases en vez de las 24 que se deberían localizar en caso de resolver el problema original con 3900 cuadrantes.

La creación de las zonas es generada como se explica en el capítulo 4.4.2. Cada grilla, como se explica, tiene una serie de restricciones a la hora de ser confeccionada.

Para la resolución del problema se utilizaron los siguientes valores para los parámetros: la distancia máxima que podía cubrir una base de tipo 1 fue considerada como 9 km, mientras que la distancia máxima a la que una base de tipo 2 puede cubrir una emergencia fue considerada como 12 km, debido a que poseen al menos dos ambulancias y una mayor cobertura con respecto a una base de tipo 1.

En la Tabla 8, se encuentran los resultados de la resolución del modelo utilizando la heurística anteriormente descrita. En ella se puede notar que la primera iteración obtuvo un nivel de 8493.3 despachos esperados. Mientras que, al resolver los 10 subproblemas, utilizando el nivel de agregación de 3.900 cuadrantes, se obtuvo una cantidad de despachos totales esperados que asciende a 8.809 despachos. En la tabla se observa la distribución de las distintas bases en las zonas que quedan determinadas por la resolución del problema a nivel agregado.

	<b>Cuadrantes</b>	<b>Número de viviendas</b>	<b>Bases Tipo 1</b>	<b>Bases Tipo 2</b>	<b>Despachos Esperados</b>
<b>Caso Base</b>	12.000	2.223.664	12	12	7479
<b>Nivel Agregado</b>	600	2.223.664	12	12	8493.3
<b>Zona A</b>	484	298.591	1	2	1143.3
<b>Zona B</b>	453	233.560	2	1	964.9
<b>Zona C</b>	389	237.581	2	0	901.6
<b>Zona D</b>	458	228.099	1	2	1022.5
<b>Zona E</b>	392	218.040	2	1	863.8
<b>Zona F</b>	399	208.776	1	1	827.1
<b>Zona G</b>	438	248.107	1	2	943.3
<b>Zona H</b>	303	173.286	0	1	686.5
<b>Zona I</b>	371	190.278	1	1	714.2
<b>Zona J</b>	385	187.346	1	1	742.2
<b>Total Zonas</b>	3.900	2.223.664	12	12	8809.4

TABLA 8: RESULTADOS MODELO DE OPTIMIZACIÓN.

Los resultados obtenidos resultan bastante alentadores y la heurística parece funcionar de manera correcta. Al resolver el primer problema agregado con 600 cuadrantes se presenta un aumento de un 13.3% en el número de despachos esperados con respecto al caso base (el modelo aplicado tal como están distribuida las bases). Mientras que al resolver los 10 subproblemas por separado y luego agruparlos, se genera un aumento en los despachos esperados de un 17.5% con



respecto al caso base. Se puede observar que las bases se reparten, generalmente, de acuerdo con el número de cuadrantes presentes en cada una de las zonas y esto posee una correlación alta con los despachos esperados.

Los resultados mostrados en la Tabla 8 son el óptimo de cada problema considerando un GAP de 5%. En Anexo I y Anexo J se encuentran las tablas con la ubicación geolocalizada con latitud y longitud de cada uno de los tipos de bases.

Además, en esta misma tabla, se aprecia que el número de despachos esperados está muy correlacionado al número de viviendas presentes en cada una de las grillas, al igual que la distribución de las bases en cada uno de estos. Es esperable que mientras mayor sea la población existente en cada una de las grillas, se encuentren un mayor número de ambulancias, y bases preferentemente de tipo 2 y por lo tanto, un mayor número de despachos.

En la Figura 14, se aprecia visualmente cómo se distribuyen las bases al resolver los 10 subproblemas. Los círculos azules corresponden a las bases con dos ambulancias, mientras que los verdes a las que poseen una ambulancia. Se puede observar claramente que las bases con que poseen dos ambulancias se encuentran principalmente en el centro de la Región Metropolitana, lo cual es esperable, debido a la alta densidad poblacional que posee esta zona. Fuera del centro solo se encuentra una base más de este tipo, en Melipilla, debido a la alta densidad poblacional que posee esta comuna también. Se puede notar, además, que las bases de dos ambulancias están más cargadas hacia el sector poniente de la capital, debido a la densidad poblacional que presentan los cuadrantes de esta región. Se revisan las soluciones bordes, es decir, si es que hay dos bases de distintas grillas a menos de 5 kms de distancia, lo que no se da, principalmente debido a la restricción de dejar las bases cercanas de la resolución del problema de 600 cuadrantes en una misma grilla.

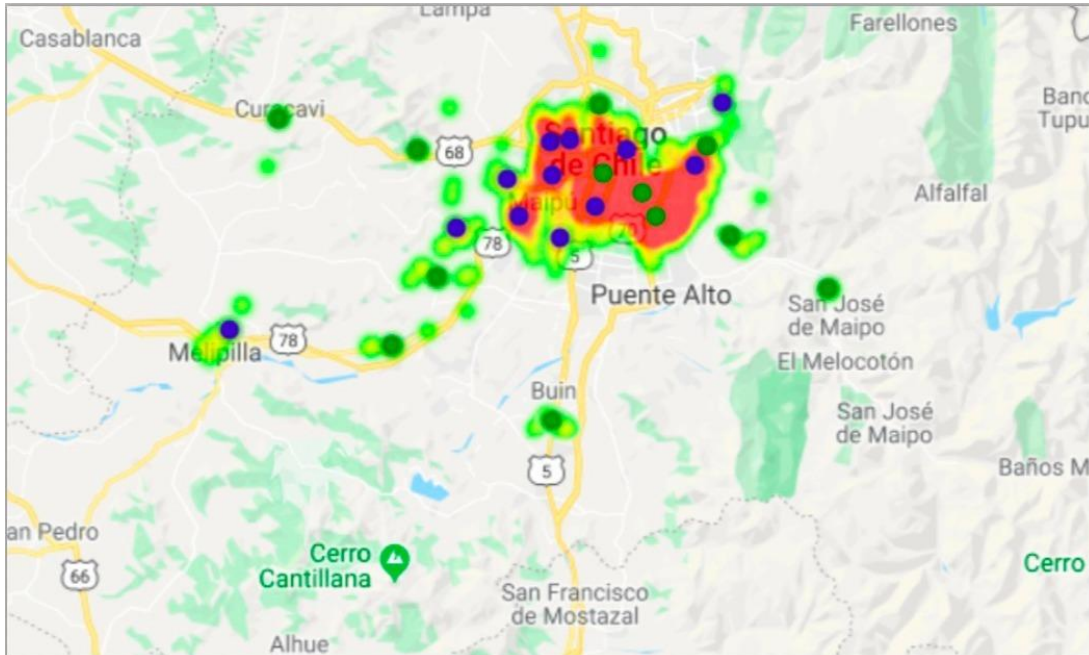


FIGURA 14:RESOLUCIÓN GRÁFICA MODELO.

## 5.2.2 TIEMPOS DE RESOLUCIÓN DEL PROBLEMA

Una vez resueltos exitosamente los problemas de optimización, es necesario analizar los tiempos de resolución para poder discutir sobre la eficiencia y escalabilidad. Cabe destacar que los problemas fueron resueltos utilizando el software de optimización AMPL en un computador con procesador 1.8 Intel Core i5 y una memoria RAM de 8 GB.

En la Tabla 9, se detallan los tiempos de resolución para cada uno los 10 problemas resueltos al momento de ejecutar la heurística propuesta. En ella se puede ver que el problema que más tardó en ser resuelto fue el problema más agregado, el cual poseía 600 cuadrantes tardando un poco más de 2 horas y media en ser resuelto. Mientras que, de los 10 subproblemas resueltos, el que más tardó fue el de la Zona A con un tiempo de 82 minutos, es decir, el modelo más agregado que resuelve el problema para los 600 cuadrantes tardó un 89% más de tiempo que el modelo de la Zona A. En total, la resolución de las 10 zonas tomó un tiempo total de 670 minutos, es decir, un poco más de 11 horas.

Además, se debe destacar que no fue posible la resolución directa de problemas más desagregados. El problema más grande que se pudo resolver sin necesidad de recurrir a la heurística y distribuyendo las 24 bases de ambulancias del SAMU fue de 600 cuadrantes, pero tomó alrededor de 4 horas. Al intentar resolver uno con 650 cuadrantes, no convergió a un óptimo luego de estar durante 8 horas resolviendo el problema. Es fácil notar que la relación entre el número de cuadrantes con los que cuenta el problema, es decir, el tamaño y el tiempo de resolución de este es absolutamente no lineal.

Finalmente, resolver el problema simple más agregado tomó 155 minutos, lo cual es un 23% del primer tiempo mencionado. La resolución de la heurística entera para los 3.900 cuadrantes, considerando la resolución del modelo agregado y de los 10 subproblemas, tomó un tiempo de 825 minutos, lo cual es poco más de 13 horas y media. Sin embargo, los 670 minutos que toma resolver los 10 subproblemas luego de la resolución del problema de nivel agregado (600 cuadrantes) genera casi un 4% más de despachos con respecto a la cantidad de despachos esperados obtenidos. Además, como ya se mencionó permite con mucha mayor exactitud y precisión ubicar las bases, dada la menor superficie que tienen los cuadrantes en el problema de 3900 cuadrantes.

	<b>Tiempos de Resolución (minutos)</b>
<b>Nivel Agregado</b>	155
<b>Zona A</b>	82
<b>Zona B</b>	73
<b>Zona C</b>	63
<b>Zona D</b>	75
<b>Zona E</b>	64
<b>Zona F</b>	61
<b>Zona G</b>	71
<b>Zona H</b>	53
<b>Zona I</b>	61
<b>Zona J</b>	67
<b>Total Zonas</b>	670
<b>Total Final</b>	825

TABLA 9: TIEMPOS DE RESOLUCIÓN PROBLEMA.

### 5.2.3 ANÁLISIS SENSIBILIDAD HEURÍSTICA

Para poder observar cómo influye el número  $n$  de grillas confeccionadas, se resuelve el problema utilizando la heurística variando este parámetro. Los resultados de este experimento se pueden visualizar en la Tabla 10. Como era de esperar, a medida que aumenta el número de grillas se presenta un leve aumento en el número de despachos, puesto que son mejor distribuidas las bases. Se puede apreciar que el aumento en el número de grillas no genera un aumento siempre en los tiempos de resolución, no tiene un comportamiento lineal, y esto se debe a que, si bien una mayor cantidad de cuadrantes permite resolver problemas más simples, cada problema por simple que sea tiene un tiempo de resolución base. De la tabla, también se puede notar que el menor tiempo de resolución de los experimentos corridos se obtiene al realizarlo con 10 grillas con poco más de 11 horas y consigue un 3.7% en el aumento de despachos con respecto a los obtenidos en el problema resuelto de 600 cuadrantes. Los dos niveles más próximos, con 9 y 11 grillas poseen tiempos considerablemente mayores. En el capítulo anterior se escoge este nivel para mostrar en detalle lo expuesto, ya que es el que representa un aumento considerable en el número de despachos esperados en el menor tiempo posible.

Además, hay que tener en cuenta que para poder generar grillas se requiere un trabajo computacional previo, entonces mientras mayor sea el número de grillas escogidas, mayor será este tiempo de procesamiento computacional anterior a la resolución. Este trabajo computacional si bien es mucho menor al tiempo de la resolución de los problemas de optimización, a medida que se aumenta el número  $n$  de grillas se va volviendo cada vez menos despreciable.

Número de grillas	Despachos esperados	Cambio c/r problema 600 cuadrantes	Tiempo de resolución (minutos)
-	8493.3	-	155
7	8745.6	2.9%	1071
9	8782.1	3.4%	735
10	8809.4	3.7%	670
11	8817.2	3.8%	721
15	8856.7	4.2%	812

TABLA 10: ANÁLISIS SENSIBILIDAD HEURÍSTICA.

## 6. CONCLUSIONES Y TRABAJOS FUTUROS

El sistema de predicción y optimización de despachos propuestos en esta tesis significan un avance sustancial para el Servicio de Atención de Urgencias Médicas de Santiago, tanto en términos de eficiencia de recursos como en niveles de cobertura de incidentes de emergencia. Este modelo permite con una metodología innovadora optimizar recursos y maximizar la cobertura de la organización, impactando positivamente en la operación y planificación del SAMU. Los resultados obtenidos resultan ser bastante interesantes: construir un modelo predictivo que permite obtener y estimar los despachos esperados mediante una función para luego generar un modelo de optimización utilizando esta como función objetivo del problema, es un desafío que no se había realizado con anterioridad y genera una muy buena base para realizar nuevos trabajos sobre esta metodología, teniendo en cuenta los buenos resultados que se obtuvieron. El gran aporte del trabajo realizado radica en generar una demanda que es variable y sensible a la ubicación de las bases y el número de ambulancias que poseen estas, a diferencia de los modelos clásicos de localización que la consideran como una constante dada.

Otro aspecto importante que destacar es que el modelo predictivo generado, el cual fue un modelo de conteo llamado Regresión de Poisson con Ceros Inflados, permitió ver cómo afectan ciertas variables demográficas y las distancias desde cada cuadrante a la base de ambulancias más cercana. Al comienzo, se tenía la hipótesis que la distancia desde cada cuadrante a la base más cercana y las características de dicha base, estaban muy relacionadas al número de despachos esperados para el cuadrante. El modelo predictivo permitió validar esta hipótesis, ya que resultó ser estadísticamente significativa al 99%, afirmando que la distancia juega un papel fundamental a la hora de poder dar cobertura a las emergencias que se producen en Santiago.

Una vez validada la hipótesis principal, se logró generar algoritmos sobre el modelo predictivo propuesto. Una primera aplicación es la simulación al agregar una nueva base. Esta aplicación permitió saber cómo se comportaría la demanda al agregar una base más. Con esta simulación se obtuvieron importantes resultados, como que al agregar una base en El Bosque (una zona sin bases cercanas), se espera un aumento de un 4.4% en el número de despachos. Otra aplicación más dinámica, fue la que permite escoger un lugar para localizar una base teniendo en cuenta un set de opciones. Al realizar un experimento en la zona norponiente de Santiago utilizando esta aplicación se obtuvieron alentadores resultados como que, al sacar una base, relocalizarla y agregar otra base en esta zona, se podría incrementar hasta en un 4% los despachos esperados en Santiago.

Luego, se presenta un modelo de optimización más complejo, el cual busca relocalizar las 24 bases presentes actualmente en la Región Metropolitana. La formulación de este modelo es compleja matemáticamente, pero entrega muy buenos resultados. Utilizando una heurística para la resolución de problemas de gran tamaño, con muchos cuadrantes, se obtiene que siguiendo la

localización de bases que arroja el modelo, se podría aumentar en un 17% la cobertura de despachos esperados de emergencia. Este resultado podría ser aún mejor si se analizaran y perfeccionaran los casos de óptimos locales que puede arrojar la heurística, por lo que se sugiere un estudio de estos casos que podrían encontrarse en las fronteras de las grillas.

Si bien los resultados del modelo y la metodología propuesta resultan ser bastantes alentadores, hay muchas cosas que mejorarían los resultados y harían que los modelos fueran mucho más robustos y fuertes. Sería bueno aplicar la metodología propuesta en este trabajo para un mayor horizonte de tiempo. Recordemos que sólo se utilizaron dos meses de data para calibrar el modelo. Una mayor ventana de tiempo permitiría generar un modelo con mayor poder predictivo. Otra cosa que produciría mejoras sustanciales es la inclusión de variables con las que no se pudo contar como, por ejemplo, el número de llamados por cuadrante, esto permitiría geolocalizar el origen de las llamadas, pero a la fecha de esta Tesis el SAMU solo tenía esta información a nivel comuna, dicha variable permitiría explicar mejor los despachos esperados. Otra variable que se propone incluir es alguna que permita medir el nivel socioeconómico de cada cuadrante. Esta variable podría ser el ingreso promedio. También sería aún más preciso y mejor obtener los tiempos de recorrido en vez de las distancias de un cuadrante a otro. Ahora hay aplicaciones como Google Maps que permiten calcular este tipo de métricas por lo que es fuertemente sugerido.

Además, se recomienda considerar la resolución del problema no asumiendo las ambulancias básicas y avanzadas como iguales. Se podría, por ejemplo, considerar que ambas ambulancias cubren distintas demandas. También se propone resolver el problema de localización para las ambulancias individualmente y no como bases, esto se podría realizar tomando cada una de las 36 ambulancias como una base. Esto permitiría analizar de mejor forma la repartición de las bases y debería permitir generar mayor cobertura que la obtenida.

Finalmente, este proyecto sienta las bases para el desarrollo de posibles trabajos utilizando la metodología que acá se presenta. Se propone utilizar este tipo de metodologías para otros servicios de emergencia con características similares, como podría ser una base de Bomberos. Además, se podría realizar esta metodología para redistribuir las ambulancias del SAMU y no solamente las bases como se realizó en el presente trabajo. Es decir, se podría tratar cada ambulancia como una base con una ambulancia en ella, es decir, como una base de tipo 1. Esta mejora podría incrementar aún más la cobertura y arrojar resultados importantes. Una vez concluida la optimización de los recursos del SAMU para brindar un mayor nivel de cobertura a las emergencias, surgen diversas dudas y es natural plantearse lo siguiente: ¿Cómo cambiarían los resultados si existiera una demanda con cierta estacionalidad o algunos peaks? ¿Cómo se podría incluir teoría de colas y utilizar este modelo para tomar decisiones operacionales además de estratégicas?

## BIBLIOGRAFÍA

- [1] Servicio de Atención de Urgencias. <http://www.samu.cl/preguntas-frecuentes/>, septiembre 2019.
- [2] Instituto Nacional de Estadística. <https://www.ine.cl/estadisticas/demograficas-y-vitales>, septiembre 2019.
- [3] J. Roislien, P. Van den Berg, T. Lindner, E. Zakariassen, O. Uleberg, K. Aardal, J Van Essen. Comparing Population and Incident Data for optimal Air Ambulance Base Locations in Norway, *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 26, article 42.
- [4] A. Agresti. *An introduction to categorical data analysis*. New York, Wiley, 1996.
- [5] F. Moksony and R. Hegedus. The use of Poisson regression in the sociological study of suicide. *Corvinus journal of sociology and social policy* vol.5, 2014.
- [6] E. Ezzahid. Poisson regression and Zero-inflated Poisson regression: application to private health insurance data, 2012.
- [7] E. García. Aplicación de Modelos de Regresión de Poisson Bivariados a los resultados de los partidos de la Liga Española de Fútbol, Universidad de Vigo, 2014.
- [8] Jones AM. *Applied health economics*. Oxford: Routledge, 2007.
- [9] J.Hardin and J Hilbe . *Generalized linear models and extensions*. 2nd ed. Texas: Stata Press, 2007.
- [10] Cameron AC, Trivedi PK. *Regression analysis of count data*. Cambridge: Cambridge University Press, 1998.
- [11] Hilbe J. *Negative binomial regression*. Cambridge: Cambridge University Press, 2007.
- [12] A. Salinas, B. Manrique and S. Sosa-Rubí. Análisis estadístico para datos de conteo: aplicaciones para el uso de los servicios de salud, *Salud Publica Mex* 2009 vol. 51, no. 5;51:397-406, 2009.
- [13] J. Kim and S. Jun Zero-Inflated Poisson and Negative Binomial Regressions for Technology Analysis, *International Journal of Software Engineering and Its Applications* Vol. 10, No. 12 (2016), pp. 431-448, 2016.

- [14] D. Lambert. Zero-inflated Poisson regression with an application to defects in manufacturing. *Technometrics* 1992; 34:1-14, 1992.
- [15] A. Zeileis, C. Kleiber and S. Jackman. Regression Models for Count Data in R, *Journal of Statistical Software*, Volume 27, Issue 8, 2008.
- [16] O. Parra. Revisión del estado del arte en modelos de localización y relocalización de vehículos para atención de emergencias, 2010.
- [17] S. Sasaki, A. Comber Suzuki and H. Brunsdon. Using genetic algorithms to optimise current and future health planning - the example of ambulance locations. *International Journal of Health Geographics*, 9 (4), 1-10, 2010.
- [18] X. Li, Z. Zhao, X. Zhu and T. Wyatt. Covering models and optimization techniques for emergency response facility location and planning:a review, 2011.
- [19] C. Toregas, R. Swain, C. ReVelle and Bergman L. The location of emergency service facilities, *Oper Res* 19:1363–1373, 1971.
- [20] J. Hong, Y. Yeong. Application Of The Set-Covering Model For The Minimum Cost Transportation Problem, 2008.
- [21] D. Eaton, R. Church, V. Bennett, B. Hamon and L. Lopez. On deployment of health resources in rural Valle del Cauca, Colombia. *TIMS Stud Manage Sci* 17:331–359, 1981.
- [22] D Eaton, M. Daskin, D. Simmons, B. Bulloch, G. Jansma. Determining emergency medical service vehicle deployment in Austin, Texas. *Interfaces* 15:96–108, 1985.
- [23] C. Chung. Recent Applications of the Maximal Covering Location Planning (M.C.L.P.) Model. *The Journal of the Operational Research Society*, 37 (8), 735-746, 1986.
- [24] G. List, M. Turnquist. Routing and Emergency Response Team Siting for High Level Radioactive Waste Shipments. *IEEE Transactions on Engineering Management*, 45 (2), 141-152, 1998.
- [25] A. Murray. Optimising the spatial location of urban fire stations. *Fire safety journal* 62, 64-71, 2013.



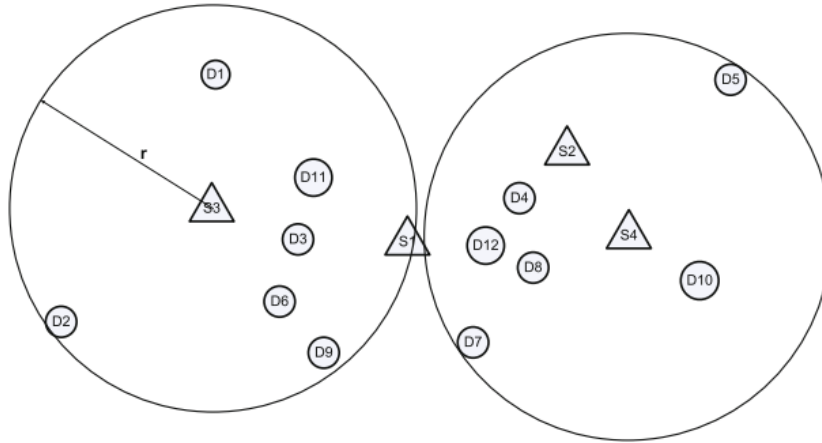
- [26] J. Dibene, Y. Maldonado, C. Vera, M. Oliveira, L. Trujillo and O. Schütze. Optimizing the location of ambulances in Tijuana, Mexico. *Computers in Biology and Medicine* 80, 107-113, 2017.
- [27] K.F. Doerner, W.J. Gutjahr, R.F. Hartl, M. Karall, M. Reimann, Heuristic solution of an extended double-coverage ambulance location problem for Austria, *CEJOR* 13 (2005) 325–340, 2005.
- [28] K. Steins, N. Matinrad and T. Andersson. Forecasting the Demand for Emergency Medical Services, *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
- [29] Documentación API. <https://documentation.simpliroute.com/es/>, octubre 2019.
- [30] K. Roenfeldt, H. Jackson. *Better than Average: Calculating Geometric Means Using SAS*, Foundation for the Advancement of Military Medicine, 2018.

# ANEXOS

## A. UBICACIÓN Y CARACTERÍSTICAS BASE DEL SAMU

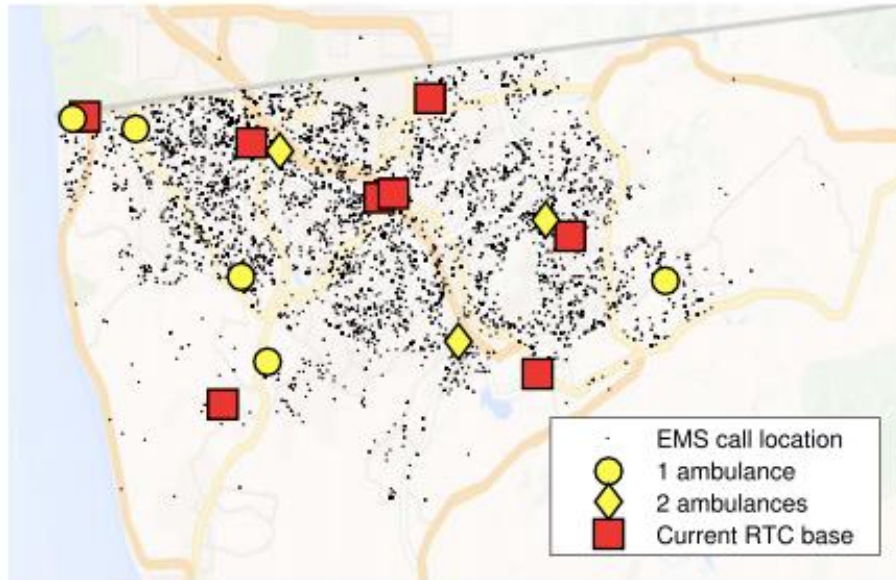
NOM_COM	DIRECCION	REFERENCIA	TIPO_BASE	AMB. BASICAS	AMB. AVANZADA	AMB. MEDICALIZADA	COORD_X	COORD_Y
SANTIAGO	PORTUGAL 125	HOSPITAL DE URGENCIA ASISTENCIA PUBLICA	BASICA	3	0	0	347743.89	6298373.43
ESTACION CENTRAL	LLANCAHUE 840	SAPU CHUCHUNCO	BASICA	1	0	0	341448.82	6294866.3
MAIPU	CAMINO A RINCONADA 1001	CRS MAIPU	BASICA	2	0	0	335444.19	6290880.19
MAIPU	CAMINO A RINCONADA 1001	CRS MAIPU	AVANZADA	0	2	0	335444.19	6290880.19
MELIPILLA	ORTUZAR 1201	HOSPITAL SAN JOSE DE MELIPILLA	AVANZADA	0	1	0	294619.69	6299361.08
SANTIAGO	PORTUGAL 125	HOSPITAL DE URGENCIA ASISTENCIA PUBLICA	AVANZADA	0	1	0	347743.89	6298373.43
INDEPENDENCIA	DOMINGO SANTA MARIA 1207	SIN REFERENCIA	AVANZADA	0	1	1	346009.55	6301456.87
INDEPENDENCIA	DOMINGO SANTA MARIA 1207	SIN REFERENCIA	BASICA	2	0	0	346009.55	6301456.87
COLINA	CALLE B 95	SIN REFERENCIA	AVANZADA	0	1	0	344123.57	6325507.61
SAN BERNARDO	SAN ALFONSO 1201	CESAFAM DR. RAUL CUEVAS	AVANZADA	0	1	0	340872.5	6281172.27
SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	AVANZADA	0	2	0	347051.87	6293647.53
SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	BASICA	2	0	0	347051.87	6293647.53
PROVIDENCIA	VITACURA 115	HOSPITAL FELIX BULNES (EX MILITAR)	AVANZADA	0	1	0	350715.38	6301061.86
ÑUÑO A	JUAN MOYA MORALES 1188	SAPU CENTRO URGENCIA ÑUÑO A	AVANZADA	0	1	0	352389.66	6295879.32
PEÑALOLEN	LAS TORRES 5100	HOSPITAL LUIS TISNE	BASICA	1	0	0	353290.18	6291982.75
PROVIDENCIA	VITACURA 115	HOSPITAL FELIX BULNES (EX MILITAR)	BASICA	1	0	0	350715.38	6301061.86
PROVIDENCIA	RANCAGUA 600	HOSPITAL SALVADOR	BASICA	2	0	0	349009.71	6298813.52
CURACAVI	AMBROSIO O' HIGGINS 504	HOSPITAL DE CURACAVI	BASICA	1	0	0	300414.7	6301303.4
QUINTA NORMAL	DR LORENZO URCARAY 4364	HOSPITAL FELIX BULNES	AVANZADA	0	1	0	342400.55	6301359.05
SAN MIGUEL	ALCALDE PEDRO ALARCON 901	HOSPITAL BARROS LUCO	BASICA	2	0	0	347051.87	6293647.53
PUENTE ALTO	EL PEÑON 39	HOSPITAL SOTERO DEL RIO	AVANZADA	0	1	0	353264.15	6283247.39
LA FLORIDA	FROILAN ROA 6426	CESFAM BELLAVISTA	BASICA	1	0	0	351302.29	6290756.76
PUENTE ALTO	EL PEÑON 39	HOSPITAL SOTERO DEL RIO	BASICA	3	0	0	353264.15	6283247.39
LA FLORIDA	FROILAN ROA 752	HOSPITAL DE LA FLORIDA	AVANZADA	0	1	0	351349.59	6290598.24

## B. ILUSTRACIÓN EJEMPLO MODELO LSCP



**Fig. 1** An illustration of the LSCP.  $D_1$  to  $D_{12}$  are demand points while  $S_1$  to  $S_4$  are EMS facilities. A demand point is covered by a facility if the distance between the demand point and the facility is within  $r$ . A feasible solution is given in which  $S_3$ ,  $S_4$  cover all the demand points

### C. ILUSTRACIÓN EJEMPLO OPTIMIZACIÓN BASES TIJUANA



### D. EJEMPLO RESULTADO MODELO ZIP EN SOFTWARE R

```
Call:
glm(formula = fl$despachos[s1] ~ log(viv) + d1 + d2 + d3 + n1,
     family = poisson(link = log))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3682 -0.9942 -0.7635 -0.0011 23.7603

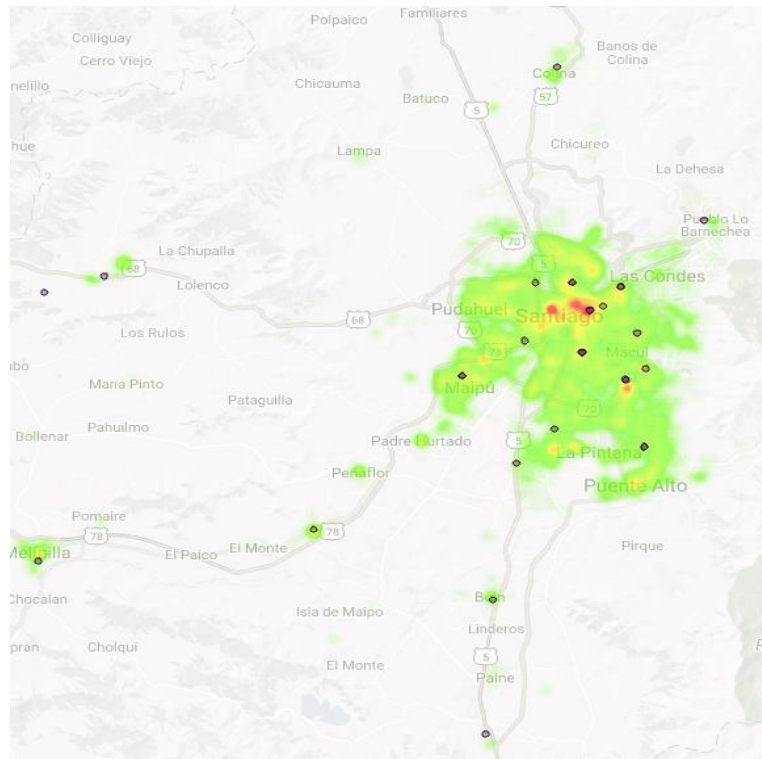
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.11832    0.03616  -30.93 <2e-16 ***
log(viv)     0.55967    0.01290   43.38 <2e-16 ***
d1TRUE       0.83609    0.03868   21.61 <2e-16 ***
d2TRUE       0.61768    0.04037   15.30 <2e-16 ***
d3TRUE       0.45209    0.04222   10.71 <2e-16 ***
n1TRUE      -0.25230    0.02505  -10.07 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

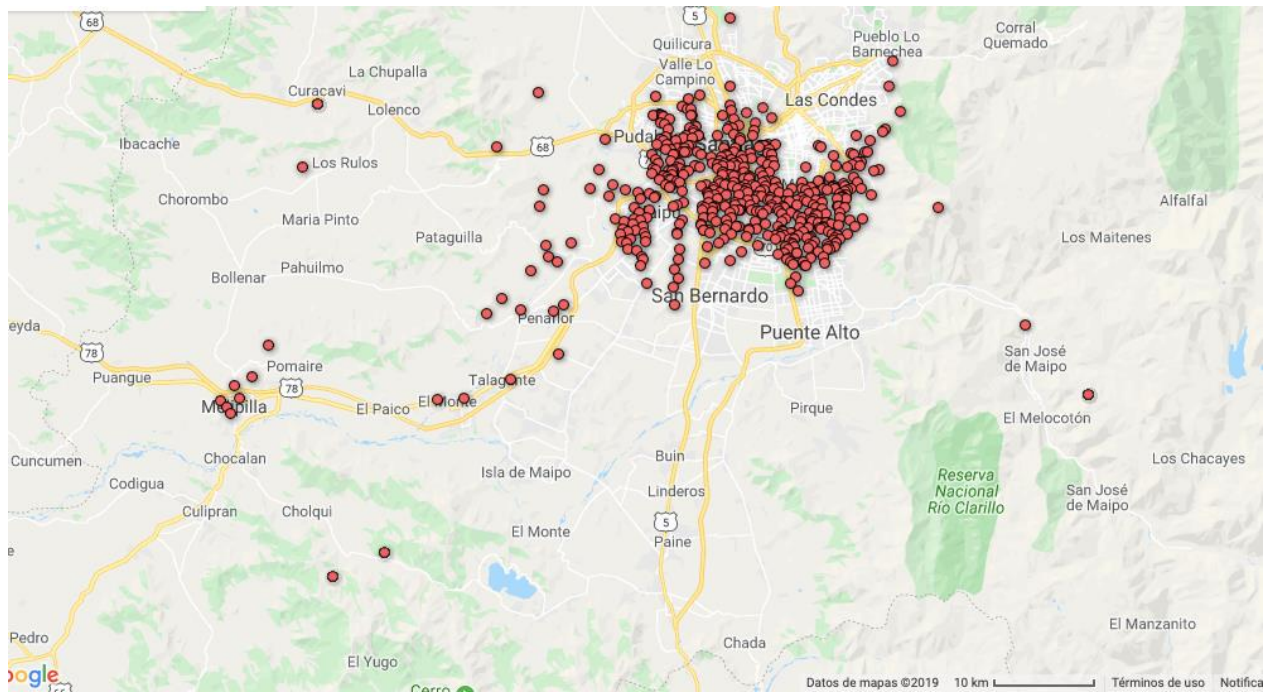
    Null deviance: 21588  on 12027  degrees of freedom
Residual deviance: 18799  on 12022  degrees of freedom
AIC: 26719

Number of Fisher Scoring iterations: 6
```

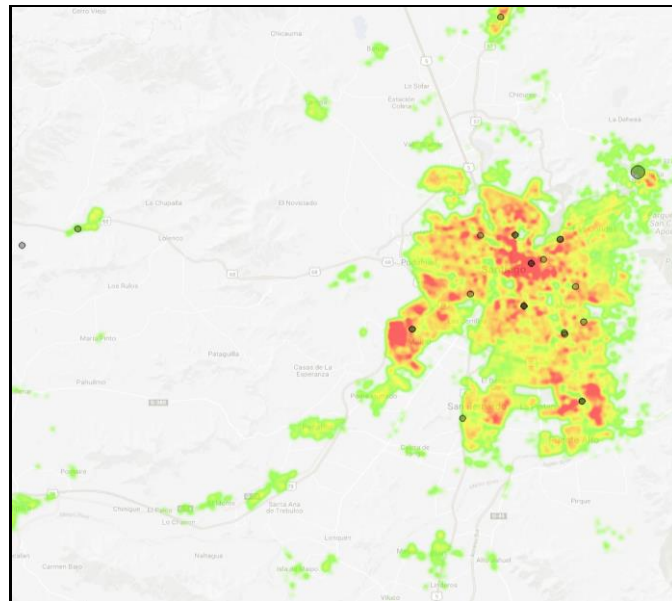
## E. MAPA DE CALOR DESPACHOS DATA ORIGINAL



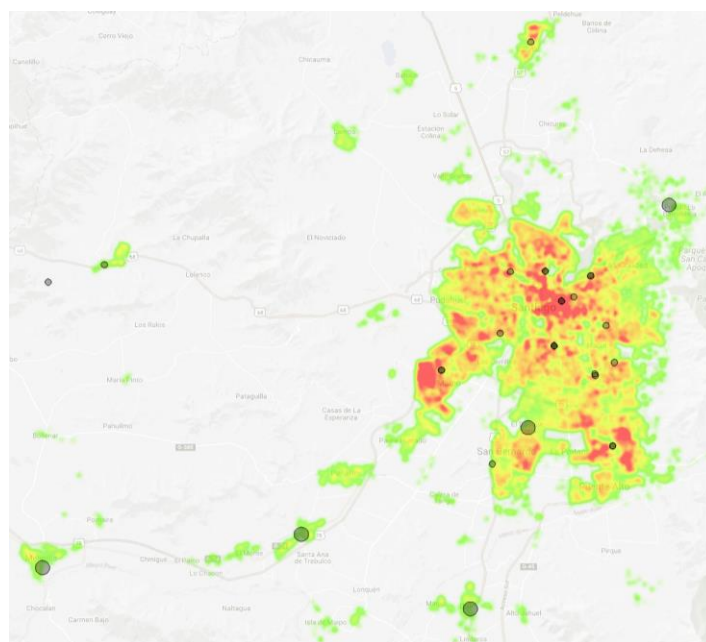
## F: MAPA DE CENTROS DE 600 CUADRANTES



## G. MAPA DE CALOR SIMULACIÓN PARA BASE EN LO BARNECHEA



## H. MAPA DE CALOR CASO BASE



## I.GEOLOCALIZACIÓN BASES CON DOS AMBULANCIAS

Latitud	Longitud
-33.6593	-71.1978
-33.5527	-70.7336
-33.5404	-70.8784
-33.5278	-70.7896
-33.5173	-70.6822
-33.4850	-70.8057
-33.4803	-70.7445
-33.4670	-70.5446
-33.4497	-70.6391
-33.4413	-70.7476
-33.4387	-70.7180
-33.3956	-70.5047

## J. LOCALIZACIÓN BASES CON UNA AMBULANCIA

Latitud	Longitud
-33.3964	-70.6772
-33.4123	-71.1274
-33.4449	-70.5257
-33.4506	-70.9324
-33.4776	-70.6735
-33.5004	-70.6167
-33.5272	-70.5986
-33.5991	-70.9057
-33.6128	-70.3563
-33.5469	-70.4931
-33.6791	-70.9681
-33.7296	-70.7638