





R E V I S T A

# INGENIERIA DE SISTEMAS

---

Volumen XXXII

Septiembre 2018

- Problemas de Cutting Stock y Pricing para productos perecibles 5  
*Alejandro Cataldo, Pablo A. Rey, Carlos Riesco, Antoine Sauré.*
- Análisis de redes sociales para mejorar la identificación de patrones de robo de vehículos 29  
*Alejandro Vásquez, Rodrigo Joannon, Richard Weber.*
- Efectos de la contaminación del aire en atenciones de urgencia 55  
*Valeria Scapini, Carla Carrasco, Cinthya Vergara.*
- Asignación de árbitros para las ligas profesionales de básquet de Argentina mediante Investigación de Operaciones. 75  
*Guillermo Durán, Mario Guajardo, Facundo Gutiérrez.*

Publicada por el  
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL  
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

R E V I S T A  
**INGENIERIA DE SISTEMAS**

ISSN 0716 - 1174

---

EDITOR

**Guillermo Durán**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

EDITOR ASOCIADO

**Richard Weber**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

AYUDANTE DE EDICIÓN

**Cinthya Vergara**

*Departamento de Ingeniería Industrial  
Universidad de Chile*

COMITÉ EDITORIAL

**René Caldentey**

*University of Chicago, Estados Unidos*

**Héctor Cancela**

*Universidad de la República, Uruguay*

**Rafael Epstein**

*Universidad de Chile, Chile*

**Luis Llanos**

*CMPC Celulosa, Chile*

**Javier Marengo**

*Universidad Nacional de  
General Sarmiento, Argentina*

**Juan de Dios Ortúzar**

*P. Universidad Católica, Chile*

**Víctor Parada**

*Universidad de Santiago, Chile*

**Oscar Porto**

*GAPSO, Brasil*

**Lorena Pradenas**

*Universidad de Concepción, Chile*

**Nicolás Stier**

*Facebook Core Data Science, Estados Unidos*

Financiado parcialmente por el Instituto Sistemas Complejos de Ingeniería.

---

Las opiniones y afirmaciones expuestas representan los puntos de vista de sus autores y no necesariamente coinciden con las del Departamento de Ingeniería Industrial de la Universidad de Chile.

---

Los artículos sólo pueden ser reproducidos previa autorización del Editor y de los autores.

---

Representante legal: Fernando Ordóñez  
Correo electrónico: ris@dii.uchile.cl  
Diagramación: Cinthya Vergara

Dirección: Av. Beauchef 851, Santiago, Chile.  
Web URL: www.dii.uchile.cl/ris  
Diseño Portada: Gabriella Fabbri

---

---

## Carta Editorial Volumen XXXII

---

Nos es muy grato presentar este nuevo número de la Revista de Ingeniería de Sistemas (RIS) dedicado a temas de frontera en Investigación de Operaciones, Gestión y Tecnología. Queremos agradecer al Instituto Sistemas Complejos de Ingeniería (ISCI) por su colaboración para hacer posible esta publicación.

Este número contiene artículos de académicos y estudiantes de nuestro Departamento de Ingeniería Industrial, de investigadores del ISCI, y de otros académicos y estudiantes de universidades nacionales y del extranjero.

Nuestro objetivo a través de esta publicación es contribuir a la generación y difusión de las tecnologías modernas de gestión y administración. La revista pretende destacar la importancia de generar conocimiento en estas áreas, orientado tanto a problemáticas nacionales como a la realidad de países de características similares de la región.

Estamos seguros de que los artículos publicados en esta oportunidad muestran formas de trabajo innovadoras que serán de gran utilidad e inspiración para todos los lectores, ya sean académicos o profesionales, por lo que esperamos que esta iniciativa tenga la recepción que creemos se merece.

Guillermo Durán  
*Editor*

Richard Weber  
*Editor Asociado*

---

# Llamado a Presentar Trabajos

---

La Revista Ingeniería de Sistemas (RIS) busca constituir un canal de divulgación de los avances en las áreas de Gestión de Operaciones, Tecnologías de Información e Investigación Operativa, que incluya los mundos académico y empresarial. Son particularmente apropiados artículos orientados a la práctica de estas disciplinas, que estimulen su uso o den cuenta de aplicaciones innovadoras de ellas, especialmente en América Latina.

También son bienvenidos artículos con análisis del estado del arte en un campo particular y de la forma en que los avances en dicho campo se han utilizado en la práctica.

Se espera que los artículos estén escritos de manera que puedan ser leídos por personas no especialistas en el tema tratado. Se recomienda incluir una lista de lecturas sugeridas para que los lectores no especialistas puedan profundizar en el tema.

---

## Formato del Manuscrito

Los autores deben enviar un archivo en formato PDF del manuscrito que desean someter a referato a:

*Comité Editorial Revista Ingeniería de Sistemas,  
Departamento de Ingeniería Industrial,  
Universidad de Chile.  
Santiago, Chile.  
Email: ris@dii.uchile.cl*

Los manuscritos deben estar formateados para hojas tamaño carta, a doble espacio, márgenes de 2,5 centímetros en todos los lados y su extensión no debe exceder las 20 hojas.

La primera hoja debe contener el título del trabajo, nombre y dirección de los autores (teléfono y correo electrónico del autor de contacto) y un resumen de no más de 150 palabras.

---

## Referencias

Las referencias se deben citar en el cuerpo del texto usando el nombre del autor y el año de publicación, e.g., Morton (1998). Al final del artículo se debe incluir la lista en orden alfabético de las referencias citadas en el texto. Para referencias de revistas científicas el formato es el siguiente: Autor(es), Año de publicación. Título. Nombre completo de la revista , Volumen e.g.:

Kodialam, M. y H. Luss, 1998. Algorithms for Separable Nonlinear Resource Allocation Problems. *Operations Research* , 44(2), 272-284.

Para referencias de libros el formato es el siguiente: autor(es), año de publicación. Título. Editorial, Ciudad; e.g.:

Kleinrock, L., 1975. *Queueing Systems* . John Wiley, New York.

En caso de haber más de una referencia con el mismo autor y año de publicación, se debe usar "a", "b", etc. como sufijo del año de publicación para diferenciarlas.

**Detalles en [www.dii.uchile.cl/ris](http://www.dii.uchile.cl/ris)**

---

---

## PROBLEMAS DE *Cutting Stock* Y *Pricing* PARA PRODUCTOS PERECIBLES

---

ALEJANDRO CATALDO <sup>\*</sup>  
PABLO A. REY <sup>\*\*</sup>  
CARLOS RIESCO <sup>\*</sup>  
ANTOINE SAURÉ <sup>\*\*\*</sup>

### Resumen

Este trabajo presenta una versión integrada de los problemas de *Cutting Stock* y *Pricing* aplicado en la problemática que enfrentan las empresas agroindustriales que disponen diariamente de piezas enteras de insumo que deben ser trozadas para obtener productos finales (problema de *Cutting Stock*). A su vez, dada la naturaleza perecible de sus productos, una correcta fijación de precios, relacionada a las políticas de inventario, resulta relevante para evitar la caducidad de los productos finales (problema de *Pricing*). Para resolver el problema conjunto de trozado y fijación de precios, aquí definido como “*Cutting Stock and Pricing Problem*”, se propone un enfoque de solución basado en la formulación y resolución de dos modelos de programación entera no lineal. El primero considera variables de precio, producción e inventario, siendo un caso particular del problema de *Lot-Sizing*, mientras que el segundo considera variables de producción, inventario y precio, siendo un caso particular de un problema de flujo en redes. Los resultados obtenidos muestran que el primero de estos modelos tiene un mejor desempeño en tiempo de resolución en todas las instancias resueltas. También se observó, en todos los casos, que el inventario de productos finales converge a cero a medida que avanzan los períodos en el horizonte de evaluación.

**Palabras Clave:** *Cutting Stock*, *Pricing*, Optimización, Productor perecibles, Agroindustria.

---

<sup>\*</sup>Departamento de Ingeniería Industrial y de Sistemas Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>\*\*</sup>Departamento de Ingeniería Industrial, Escuela de Ingeniería Universidad Tecnológica Metropolitana, Santiago, Chile.

<sup>\*\*\*</sup>Telfer School of Management, University of Ottawa, Ottawa, Canadá.

---

## 1. Introducción

---

En una economía que cada día requiere de productos más competitivos, ya no basta con que sólo los procesos productivos estén a la vanguardia del desarrollo tecnológico. Además, se requiere que los procesos productivos deben ser de alto estándar para así generar una reducción en los costos operacionales. De esta manera, es necesario que la planificación que acompaña a estos procesos asegure el aprovechamiento máximo de las capacidades productivas, la minimización de costos y las mermas asociadas a estos, manteniendo la competitividad de las empresas en el mercado [32].

En el caso de las empresas agroindustriales, éstas deben tener un cuidado especial en su planificación ya que disponen diariamente de insumos que, por la naturaleza del rubro, son perecibles (por ejemplo, vacunos o pollos). Lo anterior impone una limitante sumamente importante a la ventana de tiempo en que estos productos pueden ser procesados y posteriormente vendidos. Por lo tanto, este tipo de empresas debe planificar muy bien sus rutinas de producción y el manejo de la información con que cuentan, con el fin de mejorar su rentabilidad [35]. En algunos casos, los insumos productivos de este tipo de empresas corresponden a piezas enteras que deben ser trozadas de alguna manera específica para así obtener productos finales. A modo de ejemplo, si se troza una pieza de pollo entero de una manera particular, se pueden obtener dos productos finales: una mitad de pollo y dos cuartos de pollo. A este tipo de problema de producción se le conoce como *Cutting Stock Problem* [4] y [17], en él se busca determinar, para cada período, la cantidad de cada uno de los productos finales a producir a través del proceso de corte o trozado de productos iniciales (insumos) y decidir cuánto almacenar de estos productos finales. Los objetivos de este tipo de problemas apuntan a minimizar las pérdidas de materia prima y los costos totales de producción. Los principales modelos y métodos para enfrentar el problema de *Cutting Stock* se discuten en [4] y [17].

Por otro lado, la naturaleza perecible de los productos que producen las empresas agroindustriales, obliga a estas empresas a intentar vender sus productos antes de que alcancen su fecha de expiración (o de caducidad). Se entiende que, pasada esa fecha, el producto pierde la totalidad de su valor llegando, incluso, a producir un costo para las empresas al tener que deshacerse de los productos que se han transformado en desperdicio. De este modo, una política adecuada de fijación dinámica de precios, que actualice periódicamente el precio de los productos finales en función de la cantidad de unidades en

inventario y su fecha de expiración, es una alternativa estudiada para generar mayores ingresos y reducir las mermas por obsolescencia [25].

El estudio combinado de problemas de producción, gestión de inventarios y políticas dinámicas de fijación de precios lleva ya más de dos décadas de desarrollo [10, 11]. Es bajo este contexto que surge la motivación por enfrentar un problema de producción del tipo Cutting Stock, en el cual existe una relación directa entre los insumos y la producción de los productos finales a través del trozado de los primeros, en conjunto con la determinación de políticas dinámicas de precios. Lo anterior, considerando un horizonte de tiempo dividido en períodos y bajo condiciones de perecibilidad de los productos finales. Esto último implica que se puede almacenar inventario de productos finales de un período a otro, pero siempre y cuando no se haya alcanzado su fecha de expiración. Todas estas decisiones se toman buscando maximizar la utilidad total que obtiene la empresa agroindustrial con respecto a una demanda que depende de manera determinística de los precios de los productos finales.

El documento se estructura de la siguiente manera. En la Sección 2 se presenta una revisión bibliográfica sobre el estado del arte sobre problemas de Cutting Stock y Pricing, incluyendo también problemas de Lot Sizing. En la Sección 3 se provee una descripción detallada del problema en estudio. En la Sección 4 se describen modelos construidos para tomar las decisiones de trozado y precio. Además, se detallan los enfoques de solución correspondientes. En la Sección 5 se resumen los principales resultados obtenidos sobre el desempeño de los dos modelos y provee un análisis de sensibilidad para el modelo que mostró un mejor desempeño en su tiempo de resolución. Por último, en la Sección 6 se presentan nuestras principales conclusiones y comentarios finales.

---

## 2. Revisión Bibliográfica

---

Los problemas de Cutting Stock y Lot Sizing surgen cuando existe una presión económica para producir artículos anticipadamente con respecto al período en el que serán demandados, y posiblemente vendidos, con el fin de minimizar el desperdicio que se puede generar en su proceso productivo. Sin embargo, los costos de almacenamiento asociados a la producción anticipada de artículos ejercen una presión en sentido opuesto que busca evitar tal anticipación. Cuando existe este trade-off asociado a la producción anticipada de artículos, y se cree que la aplicación de determinados patrones de corte disminuirá el desperdicio resultante en el proceso productivo, a un mayor costo de almacenamiento, entonces se está frente a una problemática que combina

dos problemas de optimización combinatorial bien conocidos en la literatura de planificación de la producción conocidos como problemas de *Cutting Stock* y *Lot Sizing* [23].

Sobre el problema de *Cutting Stock*, Gilmore y Gomory [22] proveen una de las primeras definiciones de este problema, formulando y presentando un método de generación de columnas para su resolución. Desde entonces, su enfoque metodológico ha sido utilizado por muchos investigadores. Por ejemplo, Kallrat et al. [28] y Giannelos et al. [21] describen cómo resolver problemas de *Cutting Stock* en la industria productora de papel. A través de su trabajo, los autores demuestran los beneficios que genera en la industria papelera el uso de modelos matemáticos para apoyar la toma de decisiones de producción. Cheng et al. [12], y más recientemente Cherri et al. [13], proveen una muy buena descripción del estado del arte para problemas de *Cutting Stock*.

Algunos modelos de *Cutting Stock* han sido construidos para enfrentar problemas en la industria de la carne. En particular, Albornoz et al. [3] presentan un modelo aplicado a la industria del ganado. El modelo propuesto por estos autores busca maximizar el beneficio de los productores de carne de cerdo teniendo en cuenta decisiones de producción, inventarios y la perecibilidad de los diferentes productos finales. Mediante los resultados obtenidos, los autores logran validar sus modelos y demostrar que para la industria de la carne es relevante optimizar las decisiones de corte utilizadas en su proceso productivo. En otro trabajo de este tipo, Whitaker y Cammell [40] presentan un nuevo modelo de *Cutting Stock* aplicado a la industria de la carne en Nueva Zelanda. En este trabajo, los autores formulan patrones de corte para distintas partes del animal faenado, lo cual les permite reducir significativamente la dimensionalidad del problema facilitando su resolución.

Con respecto a los problemas de *Lot Sizing*, Amorim et al. [6] nos presentan una revisión exhaustiva de los modelos de producción y reposición de productos perecibles propuestos en la literatura principalmente en la última década. En una primera instancia nos presentan diferentes definiciones de lo que se puede considerar como un producto perecible, para luego mostrar una tabla comparativa de los diferentes modelos de producción y planificación propuestos. Los mismos autores, en su trabajo [5], enfatizan la importancia de incorporar modelos de *Lot Sizing* a la hora de planificar la producción y distribución de productos perecibles, demostrando que permiten lograr una disminución significativa de los costos asociados. Por su parte, Fragkos et al. [20] proponen una descomposición temporal de los problemas de *Lot Sizing* con capacidad limitada y tiempos de setup. Los autores también describen los beneficios de este enfoque de solución, tanto desde el punto de vista de los resultados obtenidos como del uso eficiente de herramientas de resolución modernas.

Respecto a los enfoques de solución para problemas de Lot Sizing, Levi et al. [30] presentan un problema que considera múltiples productos y capacidades limitadas de producción e inventario. En su trabajo, los autores demuestran que este problema es NP-completo y proponen un enfoque basado en la solución óptima de la relajación lineal de la formulación del problema. Por su parte, Jiangtao et al. [27] consideran múltiples productos perecibles, con una capacidad de almacenamiento limitada, y describen un enfoque de solución basado en un modelo de cantidad económica de pedido. En el mismo contexto, Luo [31] describe el impacto que tienen las variaciones de precio en este tipo de problemas y Chu et al. [15] resuelven el problema de Lot Sizing para productos perecibles mediante una heurística con resultados que no están lejos del óptimo. En otro estudio, Pahl y Voß [36] concluyen que el aumento de los costos, los retrasos en las entregas, el empeoramiento de la calidad y la insatisfacción de los clientes son efectos negativos que pueden evitarse contabilizando la depreciación de los productos en el proceso productivo. En este sentido, los autores destacan la importancia de incluir la perecibilidad de los productos en las decisiones de planificación. Para mayores detalles respecto al estado del arte en problemas de Lot Sizing sugerimos revisar los trabajos de Bushuev et al. [8] y Karimi et al. [29].

En la literatura también se encuentran diversos estudios sobre el problema combinado de Cutting Stock y Lot Sizing. Por ejemplo, Gramani et al. [24] resuelven el problema combinado considerando en su formulación los costos de producción, inventario y setup, además de los costos por concepto de las mermas de los cortes. Como enfoque de solución, los autores proponen una heurística basada en relajación lagrangiana, utilizando el método del subgradiente para resolver los duales lagrangianos. Por su parte, Wu et al. [41] proponen un método de selección progresiva para resolver este tipo de problemas. Los autores muestran que los resultados obtenidos se acercan a los de la heurística lagrangiana de Gramani et al. y que a través de su enfoque de solución se obtiene un mejor desempeño que aquel obtenido mediante el uso de paquetes comerciales de optimización. En el mismo sentido, Nonås y Thorstenson [24] describen una aplicación de los modelos de Lot Sizing y Cutting Stock en la industria de placas de acero, y presentan diferentes algoritmos de resolución para este tipo de problemas. Melega et al. [34], por su parte, presentan una formulación general para el problema combinado proponiendo una integración a través de los períodos de decisión y entre los diferentes niveles de producción. Para más referencias sobre este problema combinado de Cutting Stock y Lot Sizing, sugerimos revisar el artículo publicado por Melega et al. [33], artículo en el que proveen una clasificación de la literatura existente sobre este tipo de problemas.

Desde la perspectiva de los problemas de Pricing, Chew et al. [14] presentan un modelo que permite determinar los precios óptimos para las diferentes versiones de un producto perecible de tal forma de minimizar el canibalismo entre ellas, además de sugerir la cantidad óptima de orden de nuevos productos. En la misma dirección, Abdollahpour et al. [2] presentan una revisión de diversos modelos de fijación de precios orientados a la industria agroindustrial. En su trabajo, los autores sugieren una estrategia para identificar el modelo adecuado de pricing a utilizar dadas las características de la empresa. Anjos et al. [7] presentan una familia de funciones que permiten determinar una política de pricing. Los autores también demuestran que estas políticas permiten una actualización de los precios dependiendo del comportamiento de los clientes, mejorando así el desempeño de los modelos existentes. En otros trabajos, Cataldo y Ferrer [9] muestran cómo formular y resolver un modelo para apoyar decisiones de precio combinadas con decisiones de composición de productos para empresas de televisión por cable. Por otro lado, Pérez et al. [37] muestran cómo incluir en los modelos de fijación de precios restricciones sobre la disposición a pagar de los consumidores. En dicho artículo, los autores aplican su enfoque a la industria de telecomunicaciones.

Por último, en la literatura se encuentran algunos trabajos que abordan de manera conjunta los problemas de Lot Sizing y Pricing. Por ejemplo, Abad et al. [1] presentan una modificación a los modelos que determinan la cantidad económica de pedido, la cual considera que los productos son perecibles y que se puede satisfacer demanda atrasada. Dye et al. [18] también presentan un modelo de manejo de inventarios del tipo cantidad económica de pedido, el cual permite determinar el precio óptimo de venta permitiendo la opción de crédito inter-compañías. En esta misma dirección, Sana et al. [38] y Feng et al. [19] presentan modelos de manejo de inventario y fijación de precios. Particularmente, Feng et al. consideran productos de repostería que están sujetos a perecibilidad y cuya demanda es sensible a cambios en los precios. Los autores muestran que estos dos factores repercuten en variaciones en la demanda y, por lo tanto, en la utilidad de la empresa. Finalmente, Sana et al. [39] describen un modelo combinado que considera la producción de peces y aves de corral con el fin de aprovechar las mermas de una industria para alimentar a la otra a través de productos como por ejemplo la harina de pescado.

Basados en nuestra revisión bibliográfica, podemos concluir que el problema integrado de Cutting Stock y Pricing no ha sido abordado en la literatura. En consecuencia, creemos que el desarrollo y resolución de un modelo de este tipo, con consideraciones de perecibilidad de los productos, es un aporte al estado del arte para ambos tipos de problemas. Adicionalmente, este problema

está inspirado en un caso real, como es el de producción y fijación de precios en el faenado y trozado de productos agroindustriales.

---

### 3. Descripción del Problema

---

Como se ha mencionado en las secciones anteriores, en este artículo se considera la problemática que enfrenta una empresa agroindustrial al momento de tomar sus decisiones de producción y fijación de precios, definiendo así un problema combinado de Cutting Stock y Pricing. Específicamente, se considera el caso de una empresa que en cada período de decisión recibe una cantidad conocida de piezas enteras y debe decidir cómo trozarlas, de manera de transformar cada una de estas piezas en productos finales. Para ilustrar lo anterior, si se considera que la pieza entera corresponde a un pollo entero recién faenado, una forma de trozar esta pieza permite producir una unidad de medio pollo, una unidad de cuarto de pollo y dos unidades de un octavo de pollo. Ahora bien, una forma alternativa de trozar esta unidad de pollo entero permite producir cuatro unidades de cuarto de pollo. A estas distintas maneras de trozar una pieza les llamaremos patrones de corte. Para efectos de la descripción y posterior formulación de este problema, asumiremos que existe un patrón de corte que mantiene la pieza entera sin trozar y que, por lo tanto, la pieza entera también corresponde a un producto final.

Para determinar el precio óptimo al que se debe ofrecer cada producto final en el mercado, se deben tener en cuenta las siguientes consideraciones:

- (i) El precio de cada producto final debe respetar ciertas reglas de factibilidad predefinidas
- (ii) El precio de cada producto final debe respetar una relación de consistencia con el precio de los demás productos finales.

Por ejemplo, para el caso (i), se podría considerar un producto que alcanza su demanda máxima en los  $r_{min}$  pesos y su demanda mínima en los  $r_{max}$  pesos. De este modo, una regla de factibilidad de precios debería garantizar que se cumpla con  $r_{min} \leq p \leq r_{max}$ , donde  $p$  es el precio del producto en pesos. Para el caso (ii), por ejemplo, se podría considerar que dado que una unidad de medio pollo (M) está compuesta de dos unidades de cuarto de pollo (C), entonces se debería garantizar que  $p_M \leq 2p_C$ .

Sobre el horizonte de tiempo considerado en la planificación, éste es dividido en unidades de tiempo de igual tamaño, las cuales llamamos períodos. De este modo, las decisiones de producción y fijación de precios se toman en

cada período. También se asume que se conoce la cantidad de unidades en inventario de cada producto final al comienzo del horizonte de planificación ( $t = 0$ ) y el valor residual de cada unidad del producto final al finalizar el horizonte de planificación.

Los productos finales se encuentran relacionados entre sí a través del proceso de producción de los mismos, el cual corresponde a un problema de Cutting Stock. Además, es importante mencionar que la demanda por cada producto final se asume independiente de los otros productos finales, y determinística y lineal respecto al precio del producto final. Adicionalmente, dado que los productos agroindustriales son perecibles, se asume conocida la cantidad de días de vida útil para cada producto final (por ejemplo, en el caso de la industria avícola, la vida útil los productos finales es de no más de 12 días).

Sobre los costos del problema, se considera el costo de producción de los productos finales que está asociado exclusivamente al trozado de las piezas enteras y el costo de almacenamiento, asumiendo que sólo se permite guardar productos finales y que, por lo tanto, no se pueden almacenar piezas enteras de un período a otro.

En consecuencia con lo anterior, se define la siguiente notación para ser usada posteriormente en la formulación de los dos modelos que permiten abordar el problema de Cutting Stock y Pricing.

## Conjuntos

$T$  : Conjunto finito de períodos de igual tamaño en los que se divide el horizonte de planificación.

$F$  : Conjunto finito de productos finales (o indistintamente sub-productos) que son el resultado del trozado de una pieza entera. Incluye a la pieza entera no trozada como producto final para la venta.

$K$  : Conjunto finito de patrones de corte (formas de trozar la pieza entera).

$\Omega$  : Conjunto finito de precios factibles para los productos finales.

## Índices

$t, u$  : Índice asociado a los períodos de tiempo  $t, u \in T$ .

$f$  : Índice asociado a los productos finales  $f \in F$ .

$k$  : Índice asociado a los patrones de corte  $k \in K$ .

## Parámetros

- $\delta$  : Vida útil, en días, para cada producto final.
- $c_k$  : Costo en el que se incurre al utilizar el patrón de corte  $k$  para trozar una pieza entera.
- $a_{kf}$  : Cantidad de producto final  $f$  que se obtiene al utilizar el patrón de corte  $k$  para trozar una pieza entera.
- $q^t$  : Cantidad de piezas enteras que deben ser trozadas durante el período  $t$ .
- $h_f^t$  : Costo de almacenar una unidad del producto final  $f$  desde el período  $t$  al período  $t + 1$ . Notar que  $-h_f^{|T|}$  corresponde al valor residual de una unidad del producto final  $f$  al finalizar el horizonte de planificación.
- $\alpha_f$  : Máximo valor que puede alcanzar la demanda por el producto final  $f$ .
- $\beta_f$  : Tasa de cambio de la demanda ante una variación unitaria en el precio del producto final  $f$  (es la sensibilidad precio-demanda).
- $S_f^{0u}$  : Inventario inicial de producto final  $f$  con vencimiento en el período  $u$ .

En este trabajo consideramos que las demandas por los productos finales son una función de los precios de los productos finales. De este modo,  $D : \mathbb{R}_+^P \rightarrow \mathbb{Z}_+^F$  está dada por  $D(p) = (d_1(p), d_2(p), \dots, d_{|F|}(p))^T$  la cual está bien definida sobre el conjunto  $\Omega \subseteq \mathbb{R}_+^F$  de *precios factibles*. Por simplicidad asumimos que las demandas e inventarios son cantidades continuas, medidas por ejemplo en kilogramos, reservando la naturaleza discreta para las cantidades de materia prima. Las demandas individuales se consideran conocidas y de la forma  $d_f(p^t) = \alpha_f - \beta_f p_f^t$ , con  $\alpha_f, \beta_f > 0$ . En otras palabras, asumimos que las demandas individuales son lineales con respecto al precio correspondiente. El parámetro  $\alpha_f$  corresponde al valor máximo de la demanda y el parámetro  $\beta_f$  a la tasa con la que el consumo del producto cambia ante una variación unitaria del precio del mismo.

En el marco de las definiciones anteriores, se considera que en cada período de tiempo la empresa agroindustrial cuenta con una nueva cantidad de piezas enteras (por ejemplo, pollos faenados no trozados). Las decisiones a tomar en cada período son: cuántas de estas piezas enteras deben ser trozadas, qué patrón de corte debe ser utilizado para cada unidad, y qué precio cobrar por cada producto final. Todo lo anterior de manera de maximizar la diferencia entre los ingresos por ventas y los costos de producción y almacenamiento (los

únicos costos considerados en este problema), describiendo de esta manera la utilidad económica de la compañía agroindustrial.

A continuación se presentan dos modelos combinados de Cutting Stock, Lot Sizing y Pricing, en el contexto de productos perecibles. Para ambos modelos se asume conocida la cantidad de días de vida útil de cada producto final y se considera que los productos finales se encuentran relacionados entre sí a través del proceso de producción de los mismos (que corresponde a un problema de Cutting Stock).

---

## 4. Enfoque de Solución

---

El enfoque de solución está basado en la formulación del problema a través de modelos de programación no lineal entera mixta. La no linealidad del modelo está dada por la dependencia que existe entre la demanda de cada producto final y el precio del mismo. Como se menciona en la sección anterior, en este trabajo utilizaremos demandas lineales en los precios, generándose de esta manera problemas cuadráticos convexos enteros. La naturaleza entera está dada por las decisiones de cómo trozar las piezas enteras que llegan como insumo inicial. En la aplicación del enfoque construimos dos modelos diferentes, y realizamos diferentes experimentos sobre ellos para evaluar y comparar su desempeño.

En la literatura es común considerar el problema de producción como una variante del problema de Lot Sizing [16] ya que es un problema de producción e inventarios de varios productos en un horizonte finito de tiempo. En otras palabras, ya que existe un problema de “multi-product Lot Sizing” subyacente. Sin embargo, a diferencia del problema clásico de Lot Sizing, en este caso la producción no es simplemente el tamaño de un lote sino que es el resultado de las decisiones de corte de la materia prima. Además, la demanda no es fija o conocida de antemano sino que depende de la decisión de precio.

Los dos modelos construidos para enfrentar el problema descrito tienen las siguientes características. El primero utiliza variables de producción para cada período, almacenando en variables de inventario la cantidad de producto final que no fue vendida. A diferencia de este primer modelo, el segundo modelo considera variables de asignación de producción a demandas futuras. Ambos modelos utilizan la misma variable para la fijación de precios de los productos finales. El detalle y la formulación matemática de ambos modelos se presenta en las subsecciones 4.1 y 4.2, respectivamente.

### 4.1. Modelo con variables de producción total e inventarios

En el primer modelo se consideran variables de producción total e inventarios. En otras palabras, las variables de decisión se enfocan en la planificación de la producción en cada período. Las decisiones de fijación de precios están representadas por variables específicas las cuales son definidas para cada producto y en cada período.

Es así como las variables de decisión son:

- $x_k^t \in \mathbb{Z}_+$ : cantidad de cortes tipo  $k$  a realizar en el período  $t$ .
- $s_f^{tu} \in \mathbb{R}_+$ : cantidad del producto final  $f$  que vence en el período  $u$  a almacenar desde el período  $t$  al período  $t + 1$ .
- $d_f^{tu} \in \mathbb{R}_+$ : cantidad de la demanda por el producto final  $f$  en el período  $t$  que se satisface con productos que vencen en el período  $u$ .
- $p_f^t \in \mathbb{R}_+$ : precio de cada unidad de producto final  $f$  en el período  $t$ .
- $l_f^t \in \mathbb{R}_+$ : cantidad de productos final  $f$  que se declara como perdida por vencimiento (obsolescencia) en el período  $t$ .

Con estas definiciones se formula el siguiente modelo, que denominamos *Problema de Cutting Stock and Pricing (CSPP)*:

$$\text{máx} \sum_{t \in T} \sum_{u=t}^{t+\delta} \sum_{f \in F} p_f^t d_f^{tu} - \sum_{t \in T} \sum_{u \geq t} \sum_{f \in F} h_f^t s_f^{tu} - \sum_{t \in T} \sum_{k \in K} c_k x_k^t \tag{1}$$

s. a.

$$\sum_{k \in K} x_k^t = q^t \quad t \in T \tag{2}$$

$$s_f^{1u} = S_f^{0u} - d_f^{1u} \quad f \in F, u = 1, \dots, \delta \tag{3}$$

$$s_f^{tu} = s_f^{t-1,u} - d_f^{tu} \quad f \in F, t = 2, \dots, |T|, u = t, \dots, t + \delta - 1 \tag{4}$$

$$s_f^{t,t+\delta} = \sum_{k \in K} a_{kf} x_k^t - d_f^{t,t+\delta} \quad f \in F, t \in T \tag{5}$$

$$\sum_{u=t}^{t+\delta} d_f^{tu} = \alpha_f - \beta_f p_f^t \quad f \in F, t \in T \tag{6}$$

$$d_f^{11} = S_f^{01} - l_f^1 \quad f \in F \tag{7}$$

$$d_f^{tt} = s_f^{t-1,t} - l_f^t \quad f \in F, t = 2, \dots, |T| - 1 \quad (8)$$

$$p_f^t \leq 2p_{f+1}^t \quad f = 1, \dots, |F| - 1, t \in T \quad (9)$$

$$p \in \Omega \quad t \in T \quad (10)$$

La función objetivo busca maximizar el beneficio del productor. El primer término de la función objetivo representa los ingresos percibidos al satisfacer la demanda por ventas. A este se le restan el segundo y tercer término que representan los costos de inventario y de producción, respectivamente. La primera restricción de este modelo (2) impone que en cada período sean procesadas las unidades de materia prima (piezas enteras) recibidas. Las restricciones (3), (4) y (5) describen la relación entre los inventarios de cada período, los del período anterior y las demandas. Estas restricciones aseguran que los productos del inventario sólo se puedan usar hasta su período de vencimiento. La restricción (3) modela la dinámica del inventario del primer período tomando en cuenta la demanda y los vencimientos del inventario inicial. Por otra parte, la restricción (4) modela este mismo balance entre los inventarios y la demanda tomando en cuenta los distintos vencimientos que puede tener el inventario. La restricción (5) modela la dinámica de los inventarios de los productos finales tomando en cuenta la demanda en este balance. La restricción (6) impone que la suma de las demandas parciales sean igual a la demanda total de cada período. Las restricciones (7) y (8) relacionan las demandas parciales con los inventarios y las pérdidas por caducidad que se pueden tener en cada período. Para finalizar, la restricción (9) simplemente impone una relación de precio para evitar el canibalismo entre productos y la restricción (10) garantiza que los precios fijados cumplan con las políticas específicas de precio que la empresa ha determinado para cada uno de ellos.

## 4.2. Modelo con variables de producción en una fecha para consumo en otra fecha

Un modelo alternativo al anterior, considera producciones asignadas a demandas específicas en el futuro y sus implicancias en el precio [26]. De este modo, las variables de decisión sobre la producción definen una red de flujo en la que cada unidad producida es inmediatamente asignada a una demanda que se desea satisfacer en un período próximo.

Es así como las variables de decisión son:

- $x_k^t \in \mathbb{Z}_+$ : cantidad de cortes tipo  $k$  a realizar en el período  $t$ .
- $w_f^{st} \in \mathbb{Z}_+$ : cantidad de productos finales  $f$  producidos en el período  $s$  para satisfacer la demanda en el período  $t$ .

- $w_f^{0ts} \in \mathbb{R}_+$ : cantidad de productos finales  $f$  del inventario inicial que vence en el período  $s$  destinada a satisfacer la demanda en el período  $t$ .
- $p_f^t \in \mathbb{R}_+$ : precio de cada unidad de producto final  $f$  en el período  $t$ .
- $l_f^t \in \mathbb{R}_+$ : cantidad de productos finales  $f$  que se declara como perdida por vencimiento en el período  $t$ .

Con estas definiciones se formula el siguiente modelo, que denominaremos *Problema de Cutting Stock y Pricing en Flujo (CSPFP)*:

$$\begin{aligned} \text{máx} \sum_{t \in T} \sum_{f \in F} p_f^t (\alpha_f - \beta_f p_f^t) - \sum_{t \in T} \sum_{s \leq t} \sum_{u > t} \sum_{f \in F} h_f^t w_f^{su} \\ - \sum_{t \in T} \sum_{s=t} \sum_{f \in F} h_f^t w_f^{0ts} \end{aligned} \tag{11}$$

s. a.

$$\sum_{k \in K} x_k^t = q^t \quad t \in T \tag{12}$$

$$\sum_{k \in K} a_{kf} x_k^t = \sum_{u=t}^{t+\delta} w_f^{tu} \quad f \in F, t \in T \tag{13}$$

$$l_f^t + \alpha_f - \beta_f p_f^t = \sum_{u=t}^{\delta} w_f^{0tu} + \sum_{s=1}^t w_f^{st} \quad f \in F, t = 1, \dots, \delta \tag{14}$$

$$l_f^t + \alpha_f - \beta_f p_f^t = \sum_{s=t-\delta}^t w_f^{st} \quad f \in F, t = \delta + 1, \dots, |T| \tag{15}$$

$$\sum_{u \leq \delta} \sum_{t \leq u} w_f^{0tu} = S_f^0 \quad t \in T \tag{16}$$

$$l_f^t \leq \sum_{u \leq t} w_f^{0ut} \quad f \in F, t = 1, \dots, \delta \tag{17}$$

$$l_f^t \leq w_f^{t-\delta, t} \quad f \in F, t = \delta + 1, \dots, |T| \tag{18}$$

$$p_f^t \leq 2 p_{f+1}^t \quad f \in F, t \in T \tag{19}$$

$$p \in \Omega^T \quad t \in T \tag{20}$$

La función objetivo, igual que en el modelo anterior, maximiza la diferencia entre los ingresos por ventas y los costos de inventario y de producción,

respectivamente. La restricción (12) impone que en cada período sean procesadas las unidades de materia prima programadas. La restricción (13) asigna la producción de productos finales a los períodos en que se van a usar para satisfacer las demandas. Las restricciones (14) y (15) modelan la satisfacción de la demanda usando los productos finales que se han asignado para cada fecha. Esta restricción también garantiza la consistencia de los inventarios y de los productos perdidos por caducidad. La restricción (14) contempla el caso del principio del horizonte de planificación cuando existen productos disponibles en el inventario inicial mientras que la restricción (15) corresponde a la segunda parte del horizonte donde ya han caducado todas las existencias iniciales. La restricción (16) asegura que el inventario inicial  $S_f^0$  del producto final  $f$  es consumido antes de su caducidad. Las restricciones (17) y (18) establecen un límite superior al nivel de pérdidas por caducidad que se pueden tener en cada período. La restricción (19) establece una relación entre los precios de los productos para evitar el canibalismo entre ellos. Finalmente, la restricción (20) garantiza que los precios fijados cumplen con las políticas específicas de precio que la empresa ha determinado para cada uno de ellos.

## 5. Experimentos computacionales

Por temas de simplicidad, dentro de esta sección llamaremos Modelo 1 al modelo CSPP (presentado en la Sección 4.1) y Modelo 2 al modelo CSPFP (presentado en la Sección 4.2). A continuación describimos dos experimentos computacionales realizados sobre estos modelos. Los experimentos fueron realizados con el propósito de: (1) comparar los tiempos de ejecución de ambos modelos al resolver a optimalidad las mismas instancias; (2) analizar la evolución de los precios e inventarios óptimos a lo largo del horizonte de planificación; y (3) determinar el impacto de los inventarios iniciales en la evolución de los precios óptimos.

### 5.1. Instancias de prueba

Los casos de prueba contruidos para este trabajo se basan en el problema práctico que lo motiva, la planificación del trozado de aves para la venta de carne de pollo.

Se definió que una unidad de insumo podía ser dividida en partes correspondientes a los productos finales: “entero”, “medio”, “cuarto”, “octavo” y “dieciseisavo” de pollo. Con esto se construyeron dos familias de instancias, unas que incluyen hasta los octavos de pollo y otras que incluyen todos los productos finales. Se consideraron todas las combinaciones posibles para definir

los patrones de corte. En la Tabla 1 se muestran los valores de los parámetros  $a_{kf}$  correspondientes a la cantidad de productos finales que se obtiene al aplicar los patrones de corte antes definidos. En las tablas 2 y 3 se presentan los costos  $c_k$  para cada uno de estos patrones de corte. La Tabla 4 presenta los valores utilizados para los parámetros con que se construyen las funciones de demanda de los productos finales y el valor de los costos de inventario.

Se estableció para este caso una vida útil de 14 períodos ( $\delta = 14$ ) y que se deben procesar  $q_t = 7200$  unidades de piezas enteras en cada uno de los períodos del horizonte de planificación. Respecto a los inventarios iniciales, se consideran los inventarios presentados en la Tabla 5.

Tabla 1: Costos y unidades obtenidas con cada patrón de corte para patrones que no incluyen dieciseisavos de pollo.

Productos	Patrones									
	1	2	3	4	5	6	7	8	9	10
Entero	0	0	0	0	0	0	0	0	0	1
Medio	0	0	0	0	0	1	1	1	2	0
Cuarto	0	1	2	3	4	0	1	2	0	0
Octavo	8	6	4	2	0	4	2	0	0	0
Costo	355	352.5	350	347.5	345	323.5	321	319	292.5	240

Tabla 2: Costos y unidades obtenidas con cada patrón de corte para patrones que incluyen dieciseisavos de pollo (primera parte).

Productos	Patrones										
	11	12	13	14	15	16	17	18	19	20	21
Entero	0	0	0	0	0	0	0	0	0	0	0
Medio	0	0	0	0	0	0	1	0	0	1	0
Cuarto	0	0	1	0	0	2	0	0	1	0	0
Octavo	0	1	0	2	4	0	0	3	2	1	5
Dieciseisavo	16	14	12	12	8	8	8	10	8	6	6
Costo	480	465	447.5	450	420	415	385	435	417.5	370	405

Tabla 3: Costos y unidades obtenidas con cada patrón de corte para patrones que incluyen dieciseisavos de pollo (segunda parte).

Productos	Patrones										
	22	23	24	25	26	27	28	29	30	31	32
Entero	0	0	0	0	0	0	0	0	0	0	0
Medio	0	0	1	1	0	0	0	1	0	1	0
Cuarto	1	2	1	0	1	0	0	0	2	1	1
Octavo	3	1	0	2	4	6	7	3	3	1	5
Dieciseisavo	6	6	4	4	4	4	2	2	2	2	2
Costo	402.5	400	352.5	355	387.5	390	375	340	370	337.5	372.5

Tabla 4: Costos de inventario y parámetros de las funciones de demanda por productos finales.

Productos	Costo de inventario	Demanda	
		$\alpha_f$	$\beta_f$
Entero	200	20000	15
Medio	145	21000	20
Cuarto	87.5	23500	25
Octavo	45	25000	35
Dieciseisavo	30	27500	40

Tabla 5: Inventarios iniciales según producto y vida útil remanente. No hay productos en inventario con más de 7 períodos de vida útil remanente.

Productos	Períodos de vida útil remanente						
	1	2	3	4	5	6	7
Entero	1500	2000	2500	3000	3500	3500	3500
Medio	1500	2000	2500	3000	3500	3500	3500
Cuarto	1500	2000	2500	3000	3500	3500	3500
Octavo	1500	2000	2500	3000	3500	3500	3500
Dieciseisavo	1500	2000	2500	3000	3500	3500	3500

## 5.2. Resultados

Los experimentos fueron realizados en un computador con procesador Intel Core i5 de 1.6GHz con 4GB de RAM y sistema operativo Windows 10. Para la implementación computacional se utilizó el lenguaje Python 3.6. Los modelos fueron resueltos utilizando Gurobi 8.0 con una condición de parada de un *gap* relativo menor a  $10^{-6}$ .

### 5.2.1. Tiempos de solución

Para comparar el desempeño de ambos modelos se analizaron los tiempos de solución para instancias de diferente tamaño definidas por la duración del horizonte de planificación. Se consideraron horizontes de planificación de 30 hasta 180 períodos de duración.

Las figuras 1 y 2 muestran los tiempos de solución para ambos modelos y las curvas de tendencia de estos tiempos en función del tamaño de las instancias resueltas. En la Figura 1 se consideran las instancias sin incluir dieciseisavos de pollo como producto final, mientras que en la Figura 2 se consideran todos los productos finales.

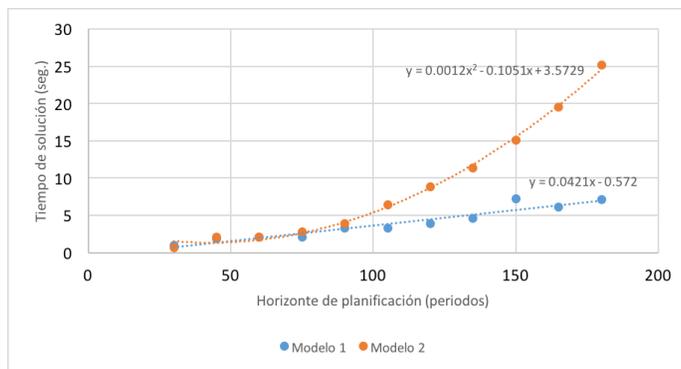


Figura 1: Comparación de tiempos de solución para instancias sin dieciseisavos de pollo como producto final.

De las figuras 1 y 2 se puede apreciar que el tiempo de solución del Modelo 1 crece aproximadamente de manera lineal al aumentar el tamaño de las instancias, mientras que el tiempo de solución del Modelo 2 lo hace de manera cuadrática. Esto sugiere que para instancias de mayor tamaño es más conveniente utilizar el Modelo 1. Como ambos modelos resuelven las instancias de prueba a optimalidad, los experimentos reportados en las próximas subsecciones fueron realizados utilizando solamente el Modelo 1 ya que requiere menos tiempo de ejecución.

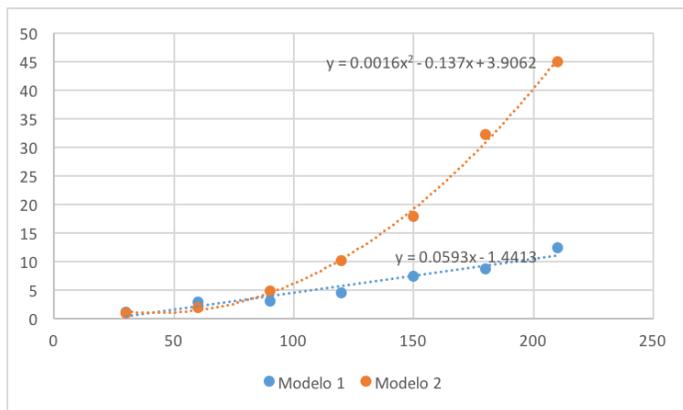


Figura 2: Comparación de tiempos de solución para instancias con dieciseisavos de pollo como producto final.

### 5.2.2. Consumo de inventario inicial y estabilización de precios

Para entender las características de las decisiones óptimas, se analizó la evolución del inventario a lo largo del horizonte de planificación. En prácticamente todas las instancias evaluadas el comportamiento fue similar.

A modo ilustrativo, la Figura 3 muestra la evolución de los inventarios para la instancia sin dieciseisavos de pollo como producto final, considerando un horizonte de 60 períodos. Se observa en la figura que luego de una etapa inicial, donde los inventarios iniciales de los productos finales se van consumiendo, la planificación óptima considera inventarios nulos para todos los productos.

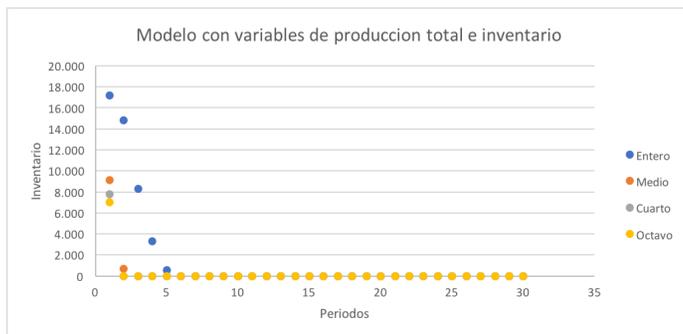


Figura 3: Evolución de los niveles de inventario óptimo a lo largo del horizonte de planificación.

Durante los primeros períodos se debe almacenar inventario debido a que existe inventario inicial que se intenta vender lo antes posible (por su perecibilidad próxima), el cual no se pueden consumir inmediatamente. Para conseguir que los inventarios se reduzcan rápidamente, los precios fijados en los prime-

ros períodos parten en un nivel “bajo”. Los precios luego suben hasta alcanzar los niveles óptimos fijos (más altos que los precios iniciales) que acompañan al fenómeno de los inventarios nulos en los períodos finales del horizonte. La Figura 4 muestra la evolución de los precios óptimos para la misma instancia.

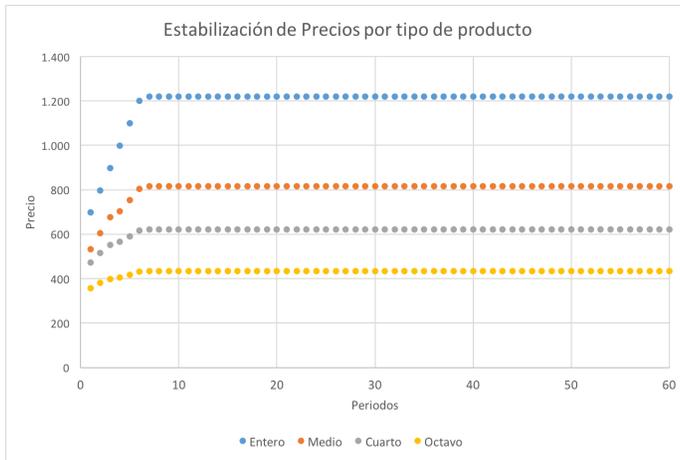


Figura 4: Estabilización de precios por tipo de producto final.

Dado el impacto del inventario inicial en los precios de los primeros períodos, se realizó un análisis para evaluar su influencia en la estabilización de los precios óptimos determinados por el modelo. El gráfico presentado en la Figura 5 permite observar que a medida que aumenta la cantidad de períodos de duración del inventario inicial, el período en el cual se estabilizan los precios también aumenta.

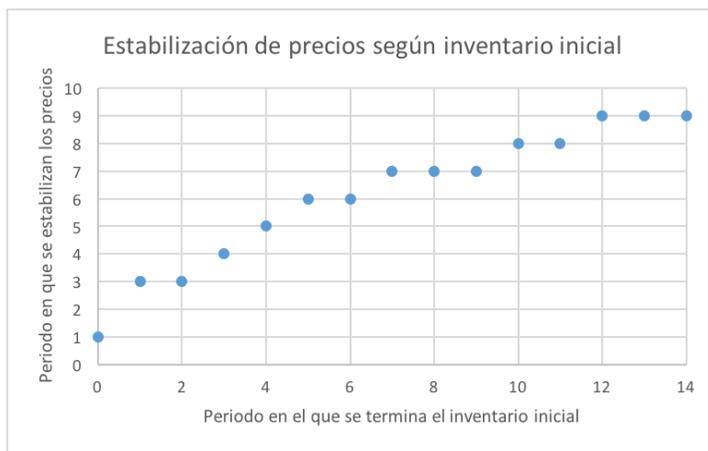


Figura 5: Estabilización de los precios según nivel de inventario inicial.

Es interesante destacar que en todas las instancias analizadas el inventario óptimo de los productos finales se hace cero a partir de un período específico en adelante, lo que también estabiliza el precio y la producción en algún periodo. Por ejemplo, cuando el inventario inicial se acaba en los primeros períodos (por ejemplo, entre el periodo 0 y el período 5) es posible observar que los precios se estabilizan en el período en que se termina el inventario o en el período posterior. Sin embargo, cuando el inventario inicial se acaba entre los períodos 6 y 14, los precios se estabilizan antes de que se acabe el inventario inicial. Esto se observó para distintos valores del intervalo de perecibilidad.

---

## 6. Conclusiones y comentarios finales

---

En este trabajo consideramos un problema de planificación de abastecimiento y trozado de pollos con fijación óptima y dinámica de precios, en el cual las principales decisiones son sobre cómo trozar las piezas enteras (insumo), cómo conservar los productor finales y qué precios colocarles. En las instancias analizadas se consideró un horizonte de planificación de 30 a 180 días.

Se propone un enfoque de solución basado en modelos de programación no lineal entera mixta, mediante los cuales se busca resolver en forma simultánea los problemas de *Cutting Stock* y *Lot Sizing* y *Pricing* bajo condiciones de perecibilidad de los productos finales. Lo anterior, con el fin de maximizar el beneficio de los productores. El primer modelo se basa en variables de producción total e inventario. En otras palabras, a través de este modelo se busca decidir en cada período cuánto producir y cuánto de la demanda satisfacer, transformando lo que sobre en inventario. Este modelo considera las decisiones de cada período como un problema “aislado”, en el que dadas las condiciones existentes se decide cómo satisfacer la demanda. El segundo modelo se basa en variables de producción en un período para consumo futuro. Con este tipo de variables se define una ventana productiva, determinada por la vida útil de los productos, la que permite considerar lo que pasa en el futuro.

Al contrario de lo que sucede en el caso de problemas de *Lot Sizing* lineales clásicos (i.e., con formulaciones lineales), los modelos con variables de producción en una fecha para consumo en otra fecha son menos eficientes en términos de tiempo de resolución que los modelos con variables de producción total e inventarios. Esto se puede deber a que las soluciones del problema estudiado no necesariamente corresponden a vértices del conjunto factible, dada la naturaleza cuadrática de la función objetivo. Nuestros resultados parecen indicar

que el menor número de variables y restricciones en el primer modelo (CSPP) hacen que éste se pueda resolver computacionalmente en menor tiempo. Esta observación está basada en un número reducido de instancias de prueba y por lo tanto debe ser confirmada con experimentos más extensivos. Los resultados obtenidos sugieren también que una parte importante de la optimización consiste en reducir y eliminar rápidamente el inventario inicial y de esta forma entrar en un período estacionario de inventarios nulos y precios fijos.

Las extensiones naturales al problema conjunto, con el fin de acercarlo más al problema real de planificación, son: (1) la incorporación de incertidumbre en la demanda a través de un modelo de programación dinámica estocástica; y (2) la incorporación de distintas dimensiones temporales en la toma de decisiones de precio. Por ejemplo, muchas veces se definen “precios de lista” para ciertos intervalos de tiempo (por ejemplo, una o dos semanas) los que sólo pueden ser modificados levemente período a período.

## Referencias

- [1] P. L. Abad. Optimal pricing and lot-sizing under conditions of perishability, finite production and partial backordering and lost sale. *European Journal of Operational Research*, 144(3):677–685, 2003.
- [2] M. Abdollahpour, A. Rusta, M. R. Hamidizadeh, y G. Khorshidi. Optimal pricing policies for perishable products. *Journal UMP Social Sciences and Technology Management*, 3(3):219–226, 2015.
- [3] V. M. Albornoz, M. Gonzalez-Araya, M. C. Gripe, S. V. Rodriguez, y E. J. Treviño. An optimization model for planning operations in a meat packing plant. In *International Conference on Operations Research and Enterprise Systems*, páginas 136–146. Springer, 2015.
- [4] P. Amorim, B.-F. Meyr, F. Toledo, C. Márcio, y B. Almada-Lobo. Lot sizing versus batching in the production and distribution planning of perishable goods. *International Journal of Production Economics*, 146:208–218, 2013.
- [5] P. Amorim, H. Meyr, C. Almeder, y B. Almada-Lobo. Managing perishability in production-distribution planning: a discussion and review. *Flexible Services and Manufacturing Journal*, 25(3):389–413, 2013.
- [6] M. F. Anjos, R. C. Cheng, y C. S. Currie. Optimal pricing policies for perishable products. *European Journal of Operational Research*, 166(1):246–254, 2005.

- [7] H. Ben Amor y J. Valério de Carvalho. *Cutting Stock Problems*, páginas 131–161. Springer US, Boston, MA, 2005.
- [8] M. A. Bushuev, A. Guiffrida, M. Jaber, y M. Khan. A review of inventory lot sizing review papers. *Management Research Review*, 38(3):283–298, 2015.
- [9] A. Cataldo y J.-C. Ferrer. Optimal pricing and composition of multiple bundles: A two-step approach. *European Journal of Operational Research*, 259(2):766–777, 2017.
- [10] L. M. Chan, Z. M. Shen, D. Simchi-Levi, y J. L. Swann. Coordination of pricing and inventory decisions: A survey and classification. In *Handbook of quantitative supply chain analysis*, páginas 335–392. Springer, 2004.
- [11] X. Chen y D. Simchi-Levi. Pricing and inventory management. páginas 784–824. Citeseer, 2011.
- [12] C. Cheng, B. Feiring, y T. Cheng. The cutting stock problem—a survey. *International Journal of Production Economics*, 36(3):291–305, 1994.
- [13] A. C. Cherri, M. N. Arenales, H. H. Yanasse, K. C. Poldi, y A. C. G. Viana. The one-dimensional cutting stock problem with usable leftovers—a survey. *European Journal of Operational Research*, 236(2):395–402, 2014.
- [14] E. P. Chew, C. Lee, R. Liu, K.-s. Hong, y A. Zhang. Optimal dynamic pricing and ordering decisions for perishable products. *International Journal of Production Economics*, 157:39–48, 2014.
- [15] L. Y. Chu, V. N. Hsu, y Z.-J. M. Shen. An economic lot-sizing problem with perishable inventory and economies of scale costs: Approximation solutions and worst case analysis. *Naval Research Logistics (NRL)*, 52:536–548, 2005.
- [16] K. Copil, M. Wörbelauer, H. Meyr, y H. Tempelmeier. Simultaneous lotsizing and scheduling problems: a classification and review of models. *OR spectrum*, 39(1):1–64, 2017.
- [17] M. Delorme, M. Iori, y S. Martello. Bin packing and cutting stock problems: Mathematical models and exact algorithms. *European Journal of Operational Research*, 255(1):1–20, 2016.
- [18] C.-Y. Dye y L.-Y. Ouyang. A particle swarm optimization for solving joint pricing and lot-sizing problem with fluctuating demand and trade credit financing. *Computers & Industrial Engineering*, 60(1):127–137, 2011.

- [19] L. Feng, Y.-L. Chan, y L. E. Cárdenas-Barrón. Pricing and lot-sizing policies for perishable goods when the demand depends on selling price, displayed stocks, and expiration date. *International Journal of Production Economics*, 185:11–20, 2017.
- [20] I. Fragkos, Z. Degraeve, y B. De Reyck. A horizon decomposition approach for the capacitated lot-sizing problem with setup times. *INFORMS JOURNAL on Computing*, 28(3):465–482, 2016.
- [21] N. Giannelos y M. Georgiadis. A model for scheduling cutting operations in paper-converting processes. *Industrial & Engineering Chemistry Research*, 40(24):5752–5757, 2001.
- [22] P. C. Gilmore y R. E. Gomory. A linear programming approach to the cutting-stock problem. *Operations research*, 9(6):849–859, 1961.
- [23] M. Gramani y P. França. The combined cutting stock and lot-sizing problem in industrial processes. *European Journal of Operational Research*, (1):509–521, 2006.
- [24] M. C. N. Gramani, P. M. França, y M. Arenales. A lagrangian relaxation approach to a coupled lot-sizing and cutting stock problem. *International Journal of Production Economics*, 119(2):219–227, 2009.
- [25] A. Herbon y E. Khmelnitsky. Optimal dynamic pricing and ordering of a perishable product under additive effects of price and time on demand. *European Journal of Operational Research*, 260(2):546–556, 2017.
- [26] B. Herskovic. Networks in production: Asset pricing implications. *Browser Download This Paper*, 2015.
- [27] M. Jiangtao, C. Guimei, F. Ting, y M. Hong. Optimal ordering policies for perishable multi-item under stock-dependent demand and two-level trade credit. *Applied Mathematical Modelling*, 38(9):2522–2532, 2014.
- [28] J. Kallrath, S. Rebennack, J. Kallrath, y R. Kusche. Solving real-world cutting stock-problems in the paper industry: Mathematical approaches, experience and challenges. *European Journal of Operational Research*, 238(1):374–389, 2014.
- [29] B. Karimi, S. F. Ghomi, y J. Wilson. The capacitated lot sizing problem: a review of models and algorithms. *Omega*, 31(5):365–378, 2003.
- [30] R. Levi, A. Lodi, y M. Sviridenko. Approximation algorithms for the capacitated multi-item lot-sizing problem via flow-cover inequalities. *Mathematics of Operations Research*, 33(2):461–474, 2008.

- [31] W. Luo. An integrated inventory system for perishable goods with backordering. *Computers & Industrial Engineering*, 34(3):685–693, 1998.
- [32] N. Ma, Y. Liu, Z. Zhou, y C. Chu. Combined cutting stock and lot-sizing problem with pattern setup. *Computers & Operations Research*, 95:44–55, 2018.
- [33] G. M. Melega, S. A. de Araujo, y R. Jans. Classification and literature review of integrated lot-sizing and cutting stock problems. *European Journal of Operational Research*, 2018.
- [34] G. M. Melega, S. A. de Araujo, y R. Jans. Classification and literature review of integred lot sizing and cutting stock problems. *European Journal of Operational Research*, 271(1):1–19, 2018.
- [35] M. Noghabai. A study on the manufacturing planning and control in a beef cutting plant of a swedish company. *Engineering Costs and Production Economics*, 17(1-4):43–53, 1989.
- [36] J. Pahl y S. Voß. Discrete lot-sizing and scheduling including deterioration and perishability constraints. In *Advanced manufacturing and sustainable logistics*, páginas 345–357. Springer, 2010.
- [37] J. Perez, H. Lopez-Ospina, A. Cataldo, y J.-C. Ferrer. Pricing and composition of bundles with constrained multinomial logit. *International Journal of Production Research*, 54(13):3994–4007, 2016.
- [38] S. S. Sana. Optimal selling price and lotsize with time varying deterioration and partial backlogging. *Applied Mathematics and Computation*, 217(1):185–194, 2010.
- [39] S. S. Sana. Optimal pricing strategy for livestock of fishery and poultry. *Economic Modelling*, 29(4):1024–1034, 2012.
- [40] D. Whitaker y S. Cammell. A partitioned cutting-stock problem applied in the meat industry. *Journal of the Operational Research Society*, 41(9):801–807, 1990.
- [41] T. Wu, K. Akartunali, R. Jans, y Z. Liang. Progressive selection method for the coupled lot-sizing and cutting-stock problem. *INFORMS Journal on Computing*, 29(3):523–543, 2017.

---

# ANÁLISIS DE REDES SOCIALES PARA MEJORAR LA IDENTIFICACIÓN DE PATRONES DE ROBO DE VEHÍCULOS

---

ALEJANDRO VÁSQUEZ \*

RODRIGO JOANNON \*

RICHARD WEBER \*\*

## Resumen

Este artículo presenta una metodología para extraer, preprocesar y analizar datos de dos fuentes que captan denuncias de robos de vehículos. Una de las fuentes es la base de datos de la Asociación de Aseguradoras de Chile; la otra son los tuits que se publican en Twitter. Se muestra los sesgos que ambas bases tienen respecto del fenómeno de robo de autos y cómo los tuits podrán ayudar en la prevención de este delito. Dentro de los principales resultados encontrados, se identifica que existen diferencias en la proporción de denuncias dependiendo del valor del vehículo en las fuentes utilizadas. Sin embargo, para los modelos más robados, ambas fuentes presentan tasas de denuncia similares y, a su vez, los vehículos denunciados en la Twitter presentan mayor tasa de hallazgo que aquellos que no lo son independiente de su valor monetario. Este trabajo es uno de los fundamentos para la creación de un observatorio digital de apoyo a la industria automotriz chilena.

**Palabras Clave:** Patrones de crimen, Análisis de redes sociales, Robo de vehículos.

---

\*Asociación de Aseguradores de Chile, Santiago, Chile

\*\*Departamento de Ingeniería Industrial, Facultad de Ciencias Físicas y Matemáticas, Universidad de Chile, Santiago, Chile

---

## 1. Introducción

---

En Chile el robo de vehículos presenta tasas bastante elevadas, redondeando los 30.000 al año, lo que significa que un vehículo es sustraído cada 17 minutos. Este artículo pretende ayudar a entender este fenómeno y desarrollar un modelo predictivo del robo de vehículos utilizando dos fuentes de información: denuncias de vehículos y Twitter.

Twitter es una red social donde se comparten opiniones o declaraciones en tiempo real, los cuales son llamados “Tweets”<sup>1</sup>. Sus características han captado la atención de distintos investigadores que utilizan esta información para explicar fenómenos sociales [7].

El objetivo principal de este trabajo es identificar las relaciones existentes entre el robo de vehículos y denuncias realizadas a través de Twitter con el fin de establecer las diferencias y similitudes entre ambas fuentes de datos.

En general, se pudo mostrar que un análisis de los tuits aporta información al reconocimiento de patrones del fenómeno de robo de vehículos. Además se recreó las etapas del proceso por las cuales pasa la mayoría de los robos de vehículo y se pudo mostrar que la tasa de hallazgo es mayor en el caso de aquellos robos que fueron tuiteados.

El Capítulo 2 de este artículo describe tanto el problema de robo de vehículos como el uso de redes sociales para la prevención del crimen. En el Capítulo 3 se muestra cómo el análisis de tuits podrá ayudar a una mejor comprensión del fenómeno de robo de vehículos y presenta los resultados obtenidos. Finalmente el Capítulo 4 concluye el trabajo y señala trabajos futuros.

---

## 2. Definición del problema

---

El delito es una problemática social que se busca prevenir de la mejor forma posible. En Chile distintas instituciones gubernamentales están dedicadas a esta labor, como Carabineros de Chile, Policía de Investigaciones entre otros.

Existen más de 200 categorías de delitos, sin embargo se ha definido una categoría especial la cual incluye a las categorías de delitos que son más frecuentes en la sociedad, esta se llama “Delitos de Mayor Connotación Social” (DMCS) la cual incluye las siguientes categorías de delitos: Homicidio, Hurto, Lesiones, Violación, Robo con fuerza (Robo de accesorios de vehículos; Robo

---

<sup>1</sup>De aquí en adelante “tuits”; según la Real Academia Española.

de vehículo motorizado; Robo en lugar habitado; Otros Robos con fuerza), Robo con Violencia (Robo con intimidación; Robo con Violencia; Robo por sorpresa, Otros Robos con Violencia) [2].

Los DMCS son analizados y estudiados por el Departamento de Análisis Criminal de Carabineros de Chile (DAC). El 95 % de las denuncias son recibidas por Carabineros de Chile [20], el 5 % restantes son realizadas en Policía de Investigaciones (PDI) en su brigada especializada de robos.

Al momento de cuantificar la cantidad de delitos de alguna categoría existe el problema de la “cifra negra”, definida como los delitos que son cometidos pero no denunciados. Para disminuir la incertidumbre de esta cifra negra Carabineros de Chile realiza encuestas periódicas en la población, además existe la Encuesta Nacional Urbana de Seguridad Ciudadana (ENUSC) [3], la cual tiene como objetivo obtener información sobre la percepción de inseguridad, la reacción frente al delito y la victimización de personas y hogares, a partir de una muestra representativa de zonas urbanas a nivel nacional y regional.

Según la ENUSC, el delito que menor tasa de cifra negra tiene es el “Robo o Hurto de vehículos” (3,6 %), por lo tanto es posible considerar que las cifras obtenidas de las denuncias para ese tipo de delito son representativas en cantidad y en fluctuaciones. La siguiente figura presenta el número de denuncias de robos de vehículos en un período de 5 años.

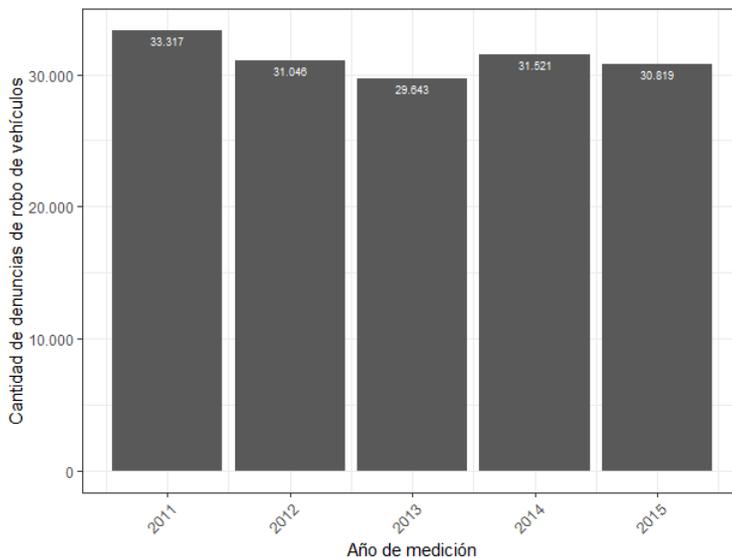


Figura 1: Denuncias de Robos de Vehículos. Elaboración propia con datos de “Estadísticas delitos de mayor connotación social [14].

El número de vehículos en circulación en Chile ha ido en constante creci-

miento, aumentando de 3.654.727 unidades en 2011 a 4.751.130 unidades en 2015 [5]. Cabe señalar que sólo el 30% del parque automotriz cuenta con seguro particular, los cuales cubren principalmente daños y robo del vehículo. El robo de vehículos es el que involucra más costos para la aseguradora y es por eso que es de gran interés para estas compañías el poder evitar estos delitos.

## 2.1. Investigación del crimen usando Twitter

Twitter [19] es una plataforma de mensajes compartidos que es usado extensamente por las personas. Es una fuente de información muy enriquecida ya que los usuarios de la plataforma discuten públicamente hechos, emociones, y varios otros tópicos. La Figura 2 muestra un ejemplo de un tuit.



Figura 2: Ejemplo tuit. Fuente: Twitter [19].

El tuit considera los siguientes campos:

- Usuario: Nombre del usuario creador del tuit.
- Fecha: Fecha y hora en la cual fue elaborado el tuit.
- Contenido: El contenido expresado en el tuit contemplado dentro de los 140 caracteres máximos. Desde el año 2018 se permite 280 caracteres en un tuit. Dado que los experimentos descritos en este trabajo fueron ejecutados antes del año 2018 se usa a lo largo de este artículo 140 como número máximo de caracteres de un tuit.
- Acciones: Los tuits tienen acciones ejecutables por otros usuarios que visualizan dicho tuit. Estas acciones son: Reply, ReTweet y Like. La

primera contiene el número de respuestas que tuvo dicho tuit, el segundo es la cantidad de veces que fue compartido el tuit por otros usuarios y el tercero corresponde a la cantidad de usuarios que expresaron gustarle el contenido del tuit.

- Geolocalización: Ubicación GPS, desde dónde es emitido el tuit, y por ende es posible de rastrear desde qué ubicación geográfica el usuario escribió el mensaje. Sin embargo, este dato es opcional, los usuarios pueden optar por no entregar su posición GPS, y emitir el mensaje sólo con los otros campos.

Varios trabajos destinados a la investigación del crimen usando Twitter han mostrado que esta red social puede ser utilizada para describir o predecir comportamientos criminales en las ciudades. A continuación se revisará tres trabajos donde se propone usar Twitter para la identificación de fenómenos del crimen. Luego se mostrarán las conclusiones de este estudio bibliográfico para el trabajo presentado en el capítulo 3.

### 2.1.1. Revisión bibliográfica

En [11] se estudia el impacto del uso de redes sociales para estimar la tasa de crimen, en particular para hot spots dinámicos. En la investigación se analiza el riesgo de victimización criminal desde una dimensión espacial, es decir dependiendo de su localización. Para realizar el estudio sólo incluyeron mensajes en Twitter que tienen coordenadas GPS incorporadas.

Utilizaron dos cálculos de poblaciones para medir la población en riesgo de crimen, la población residencial y la población móvil. La que se utiliza más frecuentemente es la población residencial pero es muy poco probable que esta medida sea la adecuada para estimar la población en riesgo de crimen en aquellos delitos que involucran población móvil. Lo primero que realizan es utilizar la densidad de mensajes para marcarlos en un mapa y poder identificar rápidamente las zonas con más y menos cantidad de delitos. Luego comparan los resultados de ambas fuentes de datos.

Los resultados más llamativos son los de Leeds City Center, zona que tiene un gran volumen de eventos criminales violentos. Este no exhibe una tasa estadística significativa cuando se usa la población ambiente como la medida de población en riesgo. Una conclusión es que deben ser precavidos con el uso de los datos de Twitter y en hacer generalizaciones. Se sabe que algunos grupos socioeconómicos están sobre representados en los datos de Twitter. Adicionalmente, a pesar de que las tasas de usuarios de medios sociales están incrementando, el porcentaje de mensajes que incluyen información geográfica son cercanos al 1%-2%.

En [15] se usa Twitter para identificar hotspots del crimen. Realizaron una recopilación inicial de un mes de tuits, para comenzar el proyecto, y poder trabajar con datos de prueba, luego fueron ampliando la recopilación de tuits durante varios meses (enero a abril 2014). Utilizando los metadatos de localización que vienen vinculados con los tuits, distribuyeron geo-espacialmente en donde apuntaban esas referencias criminales para detectar donde hay puntos críticos de actos criminales. Para ubicar geo-espacialmente las referencias utilizaron la API de Google Maps para trazar ubicaciones. Por otro lado trazaron en un mapa los delitos reales, información obtenida por la Policía, ambos conjuntos de datos de la misma área de Inglaterra.

Realizaron ambos trazados con el fin de explorar la existencia de correlación entre ambas fuentes de datos. Se consideraron diferentes categorías de delitos, entre los cuales están: Comportamiento anti-social, robo de bicicletas, robo en propiedad, posesión de armas, orden público, hurto, robo a personas, robo de vehículos, crímenes violentos. Para cada tipo de crimen, identificaron palabras y frases claves relacionadas, las cuales eran relevantes para cada tipo de crimen en particular. Luego realizaron diferentes cálculos estadísticos.

En términos generales Twitter sub-estimaba considerablemente la cantidad de delitos, como por ejemplo en la categoría de “Comportamiento anti-social”, la cual tenía una de las tasas más altas de denuncias. Si bien Twitter localizaba las mismas zonas con alta densidad, estas eran sub estimadas considerablemente, es decir en vez de abarcar una gran zona con alta densidad, detectaba varios puntos de esa zona. Otro ejemplo son los “Daños criminales”, donde sucede lo mismo, y en donde los autores mencionan que en este caso esperaban mayor tasa de denuncias por Twitter ya que estos delitos son muy populares y están presentes en varias áreas. En el caso del robo de vehículos, prácticamente no detectaba zonas de alta densidad de denuncias, pero los autores explican que les fue muy difícil definir las palabras claves para identificar este tipo de crimen en particular y que puede ser que por haber escogido mal estas palabras es que no hayan detectado bien estas denuncias. Otro tema que menciona es la importancia de discriminar o categorizar de manera correcta los tuits, ya que influye directamente en los resultados. De hecho en el caso del robo de vehículos en particular, podría decirse que los resultados no son concluyentes ya que no pudieron captar una gran cantidad de tuits en esta categoría, y como mencionan los autores la justificación es que tuvieron dificultades para identificar cuáles tuits correspondían a esta categoría.

El artículo [4] muestra el uso de Twitter y “Kernel Density Estimation” (KDE) para predecir crimen. Recopilaron tuits entre enero y marzo del 2013, filtraron para aquellos que son emitidos dentro de la ciudad de Chicago y que tienen la opción de GPS activado. Consideraron 25 tipos de crímenes, entre los

cuales se considera robo, daños criminales, violación del uso de armas, asaltos, robo de propiedad privada, robo de vehículos, homicidios, entre otros.

El modelo utilizado para fue KDE, una técnica que ajusta un espacio bi-dimensional de función de densidad de probabilidad a un registro histórico de crímenes. Compararon dos tipos de modelos, el primero es el modelo que utiliza solamente características de KDE, el segundo es el modelo que combina KDE con los datos de Twitter.

Usaron una técnica de minería de textos para separar el tuit en palabras o tokens. Luego analizando esas palabras identificaron el tópico del tuit, asignando un tipo de crimen de las categorías opcionales. Para medir el rendimiento de los modelos, se grafica la capacidad de predicción del modelo, es decir se calcula AUC (area under curve) [1].

De los 25 tipos de crímenes considerados, 19 mostraron mejoras en la medición de AUC cuando se añadieron los tópicos de Twitter al modelo KDE. Dentro de los cuales está el robo de vehículos motorizados, el cual mostró una mejora de 0,69 a 0,71 en la medida AUC.

El modelo que cataloga los tuits en tópicos, es no supervisado, es decir que el humano no interviene en el aprendizaje del modelo, y según los autores esto dificulta el entender por qué algunos tipos de crímenes tienen mejores resultados que otros.

### **2.1.2. Conclusión de la investigación bibliográfica**

Diferentes estudios e investigaciones han sido realizados en torno al tema central del crimen en las ciudades. Todas coinciden en la importancia que tiene categorizar bien los tuits en los tipos de delitos. Otra de las conclusiones más frecuentes en estas investigaciones es que es importante identificar los sesgos que presenta la plataforma social Twitter al ser utilizada como fuente de información de un modelo predictivo del crimen.

No pueden tampoco identificar cuánto de los errores cometidos en las predicciones pueden ser atribuidas a la red social por problemas de sobre exposición de ciertos perfiles de usuarios, ya que no hacen el análisis de los sesgos que presenta Twitter, pero mencionan que identificando estos sesgos y solucionando problemas mencionados en la investigación, Twitter sería una valiosa fuente de información para modelos predictivos.

Finalmente, al complementar una fuente de datos principal de denuncias reales de delitos con la fuente de datos compuesta por tuits, se concluye que los modelos predictivos mejoran su capacidad predictiva significativamente al incorporar a los tuits como una fuente de datos complementaria. Si bien al ser utilizada por si sola como fuente de datos principal no muestra la misma

capacidad predictiva que las denuncias reales efectuadas en los organismos responsables correspondientes, sí es una fuente de datos valiosa al ser considerada como complemento de otra fuente de datos principal, mejorando la capacidad predictiva de los modelos. Es así que dadas estas conclusiones, la motivación de este trabajo es avanzar en descubrir aquellas relaciones existentes entre una fuente de datos principal de denuncias realizadas formalmente y una fuente de datos secundaria basada en denuncias realizadas en Twitter con la finalidad de colaborar para el desarrollo futuro de modelos predictivos.

---

### 3. Usando Twitter para una mejor identificación del patrón de robo de vehículos

---

A continuación se mostrará la extracción de tuits. Luego se presentará los pasos del proceso Knowledge Discovery in Databases para preprocesar y analizar estos tuits.

#### 3.1. Extracción de tuits

Para la extracción de tuits existen diferentes herramientas a utilizar, por un lado están las APIs ofrecidas por Twitter.

- AdsAPI (avisos): Destinada a la extracción de datos relacionados con los avisos publicitarios mostrados en Twitter [18].
- REST API (históricos): Permite extraer datos de Twitter históricos [16].
- Streaming API: Permite obtener tuits en tiempo real a través de fijar palabras filtros, todos los tuits escritos que contengan las palabras filtros serán recibidos por la API [17].

Por otro lado están los mecanismos de scraping: proceso que permite la extracción de una gran cantidad de datos de páginas web, el cual puede incluir únicamente el texto, todo el HTML de la página o incluso las imágenes dispuestas en la misma. Luego de un análisis de todas las opciones (ver [20]), se decidió usar la técnica de scraping para extraer los tuits relevantes en el periodo 2012 – 2016 donde se escogieron los siguientes términos de búsqueda:

*robo patente OR robado patente OR robaron patente*

La razón de haber escogido esos términos es porque en una primera etapa exploratoria se visualizaron diferentes tuits relacionados con robo de vehículos, y en todos se mencionaba la patente.

El objetivo de la extracción de tuits es obtener una base de datos de Twitter que represente denuncias de vehículos robados realizadas por la red social con el propósito de ser comparada con la base de datos de denuncias de AACH. Al utilizar los términos escogidos se podrá luego extraer de los tuits la patente del vehículo denunciado, la cual será un identificador único del vehículo denunciado por robo y a la vez será el dato clave para hacer la conexión con la base de datos de AACH. La patente no sólo es un dato importante por las ventajas que ofrece de poder obtener mayores datos con ella, sino que también es necesaria por las características de los tuits, los cuales presentan dos características que limitan la extracción de información correspondiente a las características de los vehículos.

La primera es la acotación del tuit a 140 caracteres, lo que provoca que las denuncias hagan referencias principalmente a características básicas del vehículo sin mayores detalles de cómo ocurre el acto delictual. Lo segundo es que al ser texto libre, muchas veces los términos usados para describir al vehículo son escritos de manera incorrecta o acotada. Sin embargo con la extracción de la patente denunciada es posible recuperar las características del vehículo, como el año, el modelo, el color o el tipo de vehículo.

### 3.2. Selección de datos

La base de datos de AACH consolida la información referente a las denuncias de robos de vehículos realizadas por las personas que sufrieron la sustracción de su vehículo entre los años 2012 y 2016 y que además cuentan con seguro particular (aproximadamente 30% del parque automotriz).

La base original considera 45.018 registros con 74 atributos distintos, varios de los cuales remiten a la misma información sólo que usando nomenclaturas distintas por ejemplo comuna, y comuna id. De todos los atributos de la base, a priori se consideran los siguientes para la investigación:

- Aseguradora: Nombre de la aseguradora a la que está suscrita el vehículo sustraído, 13 categorías.
- Patente: Patente única del vehículo sustraído.
- Color: Color del vehículo, 16 categorías.
- Marca: Marca del vehículo, 107 categorías.
- Año: Año del vehículo.

- Estado: Busca o Encontró, indica si el vehículo fue encontrado o si aún está en búsqueda.
- Fecha último estado: Fecha del último estado, si el vehículo fue encontrado, este valor registra la fecha en que se encontró.
- Modelo vehículo: Registro del modelo, según el padrón del vehículo.
- Fecha denuncia: Fecha de cuando fue realizada la denuncia.
- Tipo de vehículo: Tipo de vehículo, 20 categorías. Separadas entre vehículos pesados, vehículos livianos u otros.
- Fecha Siniestro: Fecha y hora aproximada del robo del vehículo.
- Fecha AACH: Fecha de registro del robo en AACH.

Para realizar la extracción de tuits se aplicó la metodología descrita en el capítulo 3.1 considerando también el periodo entre el año 2012 y el año 2016 obteniendo 18.112 tuits.

La información de Twitter fue extraída de la red social como texto simple, por ende para obtener una base de datos con esa información es necesario distribuirla en variables como se muestra en el siguiente ejemplo de un tuit extraído.

“Silvia PaillánCampos @kvyen 17 dic. 2016 Favor RT! robo de Mazfa Artis, 1999, color rojo burdeo, patente SU 3765. Si lo ven avisen al XXXXXXXXXX o a carabineros al 133 RT! Responder Retwittear 163 Me gusta 9”.

La minería de datos se aplica sobre el tuit, identificando caracteres especiales ocultos que segmentan la información almacenándola en variables. En el caso del ejemplo mostrado, las variables tendrían los siguientes datos:

Usuario: “kvyen”. Fecha: “17 dic. 2016”. Hora: “7:37”. (Este dato se obtiene de un dato oculto en formato time stamp). Texto tuit: “Favor RT! robo de Mazda Artis, 1999, color rojo burdeo, patente SU 3765. Si lo ven avisen al XXXXXXXXXX o a carabineros al 133 RT!”. Reply: “”. ReTweet: “163”. Like: “9”.

Para realizar la extracción de las patentes contenidas en el texto de la variable “Texto tuit” se aplicó técnicas de minería de texto que se detalla en [20].

### 3.2.1. Limpieza y pre procesamiento

De los 45.018 registros de la AACH, 433 presentan registros irrecuperables, ya que no cuentan con la patente del vehículo ni modelo, por ende fueron eliminados, implicando que la base se redujera a 44.585 registros.

De los 44.585 datos, 1.744 registros presentaban datos faltantes como modelo del vehículo, tipo, año del vehículo o color. Para recuperar los datos faltantes se utiliza información obtenida por dos medios distintos: por la Ley N° 20.285 de Transparencia y por información de las plantas de revisión técnica.

Con Ley de Transparencia se entregó un listado de patentes y se obtuvieron datos de tipo de vehículo, año, marca y modelo. Con la información de Plantas de Revisión Técnica se obtuvieron los datos de color del vehículo. Entre los datos extraídos de Twitter que contienen la palabra “patente” se identificó 6.113 registros no válidos para este análisis ya que no tenían relación con robos de vehículos; por ejemplo:

“Con la nueva patente de Apple en caso de robo el iPhone podrá recompilar información sobre el autor del mismo. . .”.

Se aprecia que el tuit contenía las palabras filtros (robo + patente), sin embargo al aplicar el código de extracción de patentes hubiese extraído el término “de Apple” como posible patente de vehículo, la cual no es una patente válida, por lo tanto no corresponde a una denuncia de un robo de vehículo. Todos los tuits que fueron identificados con patentes no válidas fueron eliminados quedando 11.999 tuits válidos.

Paso importante en el pre procesamiento es la imputación de datos en el cual los valores faltantes (“missing values”) o que han fallado alguna regla de edición del conjunto de datos son reemplazados por valores aceptables conocidos. La principal razón por la cual se realiza la imputación es para obtener un conjunto de datos completos y consistentes al cual se le pueda aplicar las técnicas de estadística clásica. Las razones para utilizar estos procedimientos en el análisis de datos son:

- Reducir el sesgo de las estimaciones.
- Facilitar procesos posteriores de análisis de datos.
- Facilitar la consistencia de los resultados entre distintos tipos de análisis.
- Mantener la estructura de asociación entre las variables.
- Mantener intervalos de confianza más robustos [12].

Al momento de imputar datos, los valores perdidos son llenados y la base de datos ya completada es analizada por métodos estandarizados. Los métodos comúnmente usados incluyen Hot Deck, Imputación por promedio e imputación por regresión [10].

Otro de los pasos fundamentales en el proceso de transformación, es la homologación de variables, es decir transformar la codificación de los datos de la variable de tal modo que los datos representen una categoría en particular que facilite su posterior análisis en el proceso de minería de datos. En ambas bases de datos se homologaron las variables con la misma codificación para que puedan ser relacionadas en el posterior análisis. A continuación se muestra algunos ejemplos de la homologación.

#### **Color: Reducción a 1 palabra:**

La variable color incorporaba distintas clasificaciones de colores, en más de la mitad de los registros estaba compuesto por 2 palabras por ejemplo “Gris Metalizado”. Se decidió restringir el color a sólo 1 palabra, ya que ensuciaría el posterior análisis de esta variable. Gris Metalizado - Gris.

#### **Modelos y Marcas: Codificación**

Algunos modelos y marcas de vehículos presentaban 2 maneras distintas de referirse a una misma categoría por lo que se cambió su codificación por la categoría que presentaba mayor cantidad de registros. Por ejemplo: KIA - KIA MOTORS.

#### **Tipo de vehículo: Codificación**

Para la variable del tipo de vehículo fue necesario realizar distintos tipos de modificación en las categorías, principalmente agruparlas. Para definir las categorías principales se recurrió al archivo de tasación de vehículos del Servicio de Impuestos Internos (SII) [13].

#### **Categoría AUTOMOVIL:**

“AUTOMÓVIL” o “AMBULANCIA” o “COCHE MORTUORIO” o “TAXI BÁSICO” o “LIMUSINA” o “JEEP” ⇒ “AUTOMOVIL”.

#### **Categoría CAMIONES:**

“TRACTO CAMION” o “TRACTOCAMION” o “CAMION” o “CAMIÓN” o “CHASIS CABINADO” o “CHASISCABINADO” o “CHASIS” ⇒ “CAMIONES”.

La siguiente parte del proceso corresponde a la creación de nuevas variables que permiten realizar un mejor análisis de las relaciones existentes entre ambas bases de datos o mejorar las conclusiones posteriores.

#### **Creación variable modelo:**

Los modelos de los vehículos pueden presentar diferentes escrituras haciendo referencia al mismo vehículo, ya que para el proceso de inscripción no se sigue una nomenclatura. Además un modelo de vehículo puede presentar diferentes versiones ya sea con equipamiento “básico”, o “top de línea”. Otra complicación es que el modelo de vehículo puede presentar distintas versiones de potencia de motor, es así donde en los registros se encuentra por ejemplo:

CAMIONETA- NISSAN - 2010 - NAVARA 4X4 D/C DIES CUERO  
CAMIONETA - NISSAN - 2010 - NAVARA 4X4 C/S DIESEL  
CAMIONETA - NISSAN - 2010 - NAVARA 4X4  
CAMIONETA - NISSAN - 2010 - NAVARA 4X4 D/C DIES AT CU

En donde los 4 vehículos son del mismo tipo “CAMIONETA”, la misma marca “NISSAN”, del mismo año, “2010”, pero el modelo está escrito en 4 formas distintas ya sea por inscripción distinta o por ser distintas versiones. El principal problema de mantener estos 4 tipos de modelos es que dificulta el cálculo estadístico posterior, como tasas de frecuencias, ranking de modelos robados, o correlación con otras variables. Se toma la decisión de crear una variable restringida del modelo a sólo 2 palabras, de tal manera que los 4 ejemplos de modelos mencionados anteriormente quedarían registrados así:

CAMIONETA -NISSAN- 2010- NAVARA 4X4

Por lo tanto los 4 casos harían referencia a un mismo vehículo, si bien se pierden algunas características de la versión del vehículo, se asume el costo por el beneficio que se puede obtener en el análisis estadístico posterior.

#### **Creación variable tasación:**

Se decide crear una nueva variable con información de la tasación fiscal del vehículo, ya que esto permitiría ahondar en un mejor análisis respecto a segmentación de vehículos según tramos de valor de vehículo, y así encontrar los sesgos entre ambas bases de datos en cuanto a la valorización de los vehículos denunciados.

Para obtener la valorización fiscal de los vehículos se obtuvo la información histórica de valorización del Servicio de Impuestos Internos, el cual valoriza

anualmente a cada modelo de vehículo. La información se obtiene de manera online descargando los archivos del SII [13], los cuales contienen las siguientes variables: Tipo Vehículo – Marca – Modelo – Año – Valorización según año.

### **Creación variable grupo tasación:**

Para evidenciar sesgos en cuanto a la tasación de los vehículos denunciados en ambas bases de datos era necesario obtener categorías comparativas o segmentos, es por esto que se decidió crear 4 segmentos o grupos de tasación. Considerando la base de datos de AACH como la base de datos de referencia, se calcularon los cuartiles de tasación, obteniendo:

1° cuartil (25 % de los registros): Valorización  $\leq$  \$4.160.000

2° cuartil (25 % de los registros): \$4.160.000 < Valorización  $\leq$  \$5.480.000

3° cuartil (25 % de los registros): \$5.480.000 < Valorización  $\leq$  \$7.910.000

4° cuartil (25 % de los registros): \$7.910.000 < Valorización

Con las reglas para obtener los cuartiles en AACH se creó una nueva variable que puede tener los valores 1, 2, 3 o 4. Según el cuartil al que corresponda. Esta variable fue creada tanto en la base de datos de AACH como en la base de datos de Twitter, pero utilizando en ambos casos el criterio recién mencionado, es decir el que define los cuartiles en la base de datos de AACH, considerando a esta base de datos como la base de datos de referencia.

### **3.2.2. Reducción de los datos**

Según las características de las variables, los valores presentes en ellas, las dificultades posteriores que se podrían enfrentar si se mantiene la misma cantidad de datos, etc. se decidió eliminar determinados registros.

La primera reducción de datos realizada en las bases de datos fue eliminar los registros que no corresponden al tipo de vehículos “livianos”, es decir todos aquellos registros en donde la variable “Tipo Vehículo” no contenga una de las siguientes opciones:

Automóvil - Camioneta - Furgón – Minibus - Moto - StationWagon – Todo Terreno.

Finalmente la decisión fue tomada considerando que los vehículos no livianos representan un 9 % de los datos tanto en la fuente de datos AACH como en la fuente de datos Twitter. Esta reducción de datos significó reducir de 44.585 registros a 41.692 en la base de datos de AACH y de 11.999 tuits a 11.025.

La base de datos de registros de denuncias de AACH presenta en este momento del proceso KDD 1.947 registros duplicados, los cuales fueron eliminados, manteniendo el que presentaba actualización más reciente.

En el caso de la base de datos de registros de denuncias realizadas por Twitter, se comprobó que también existen tuits que denuncian el mismo vehículo, esto se pudo comprobar a través de la variable patente, la cual fue extraída del tuit. Si bien el tuit puede no ser idéntico, están denunciando al mismo vehículo por lo que en este caso se mantuvo aquel tuit que fue emitido antes, es decir aquel que presenta mayor antigüedad. Según este procedimiento se identificaron 3.758 tuits que denuncian un vehículo ya denunciado anteriormente en otro tuit. La reducción de registros duplicados significó reducir de 41.692 registros a 39.745 (AACH) y de 11.025 tuits a 7.267.

En las tablas 1 y 2 se muestra un extracto de la base de datos de AACH y de Twitter (las patentes fueron modificadas).

### 3.2.3. Análisis estadístico

A continuación en el proceso KDD se aplicarán diferentes cálculos estadísticos que permitirán descubrir las relaciones entre ambas bases de datos, patrones y sesgos. Lo primero es realizar una inspección visual de las tasas de frecuencias históricas entre ambas bases de datos. Para confeccionar el gráfico de las tasas de frecuencias de denuncias, se acumularon mensualmente, con el fin de poder evidenciar con mayor claridad las relaciones existentes en cuanto al comportamiento de ambas fuentes de datos.

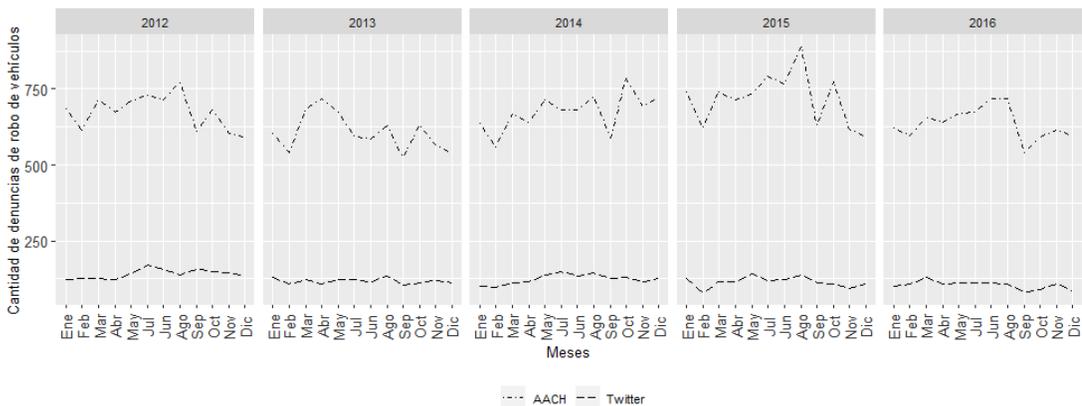


Figura 3: Gráfico Cantidad de denuncias 2012-2016.

Tabla 1: Base de datos AACH

Patente	Tipo Vehículo	Marca	Modelo	Año	Color	Tasación	Grupo	Fecha Sin	Fecha AACH	Fecha Hallaz	Si tweet
IJ1234	AUTOMOVIL	TOYOTA	YARIS GLI	2010	BLANCO	\$5.090.000	2	27-01-2014	05-02-2014	29-01-2014	1
JK1234	STATION WAGON	KIA MOTORS	GRAND CARNIVAL	2012	BLANCO	\$8.420.000	4	24-01-2014	03-02-2014	10-02-2014	0
KL1234	AUTOMOVIL	CHEVROLET	AVEO LT HB 5P1.4	2007	GRIS	\$2.980.000	1	10-02-2014	19-02-2014	NA	0

Tabla 2: Base de datos Twitter

Patente	Tipo Vehículo	Marca	Modelo	Año	Color	Tasación	Grupo tasación	Reply	Retweet	Like	User	Fecha
AB1234	AUTOMOVIL	HONDA	CIVIC LSI 1.5	1995	VERDE	\$1.250.000	1	1	41	6	Quinta.coquimbo	29-12-2016 9:39:36
BC1234	CAMIONETA	GREAT WALL	DEER 22	2008	BLANCO	\$2.120.000	1	0	0	0	SilvPadilla	28-12-2016 19:13:12
CD1234	AUTOMOVIL	NISSAN	V16 SENTRA	2008	PLATEADO	\$2.650.000	1	0	1	1	laarristaddet	27-12-2016 9:56:24

En  $Y$  están las cantidades de denuncias según la fuente de datos. En las frecuencias históricas se aprecia una cierta correlación en las tasas de frecuencias de ambas bases de datos, en particular en el periodo correspondiente a los años 2015-2016, en donde el crecimiento o decrecimiento de ambas bases de datos aparentan estar más correlacionados.

Un buen método para entender de manera numérica el grado de relación que tienen dos fuentes de datos distintas es aplicar un test de correlación. El coeficiente de correlación para el periodo 2012-2016 muestra 0,42 de correlación y 0,73 para el periodo 2015-2016.

Un segundo análisis está enfocado en identificar sesgos entre las bases de datos respecto de las características de los vehículos denunciados por robos. Para poder identificar los sesgos se realizan gráficos con las tasas de frecuencias de robos segmentado por grupo de tasación y modelo de vehículo. Las fuentes de datos consideradas son “AACH”, Twitter, con todos sus registros, y por último “Twitter Aseg.” que son los vehículos denunciados por Twitter que cuentan con seguros particular.

A continuación se presenta el gráfico de porcentaje de robos por **grupo de tasación** de vehículos, en donde se muestran las frecuencias que tienen cada una de las fuentes de datos.

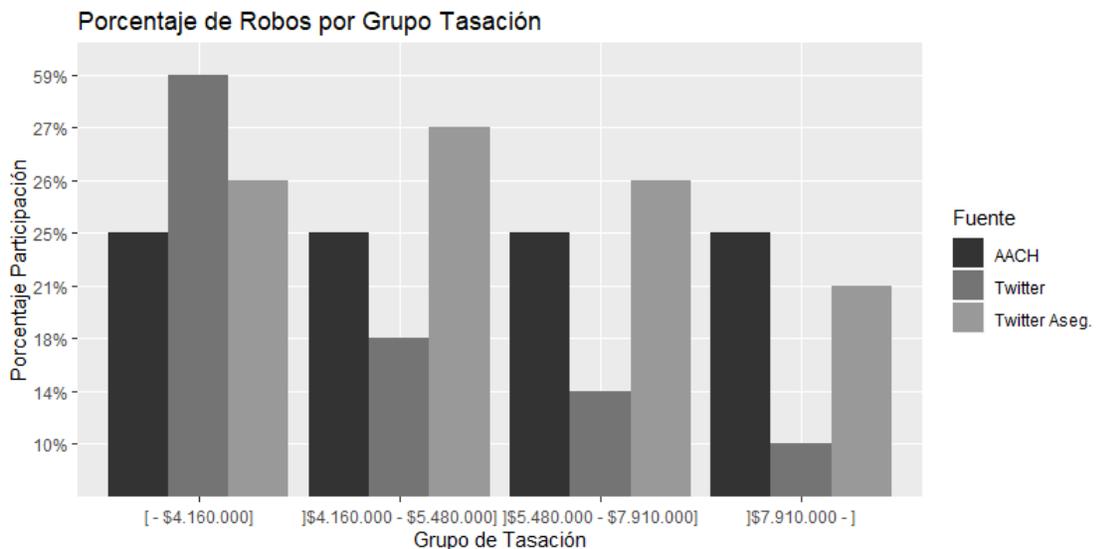


Figura 4: Gráfico Porcentaje robos por grupo tasación.

En el primer grupo de tasación, los de menor valor, Twitter presenta una altísima participación. Una de las justificaciones posibles, es que puede ser asociado a un segmento de menor poder adquisitivo.

La Figura 5 muestra el porcentaje de robos por **modelos de vehículos**,

en donde se presentan las frecuencias que tienen cada una de las fuentes de datos.

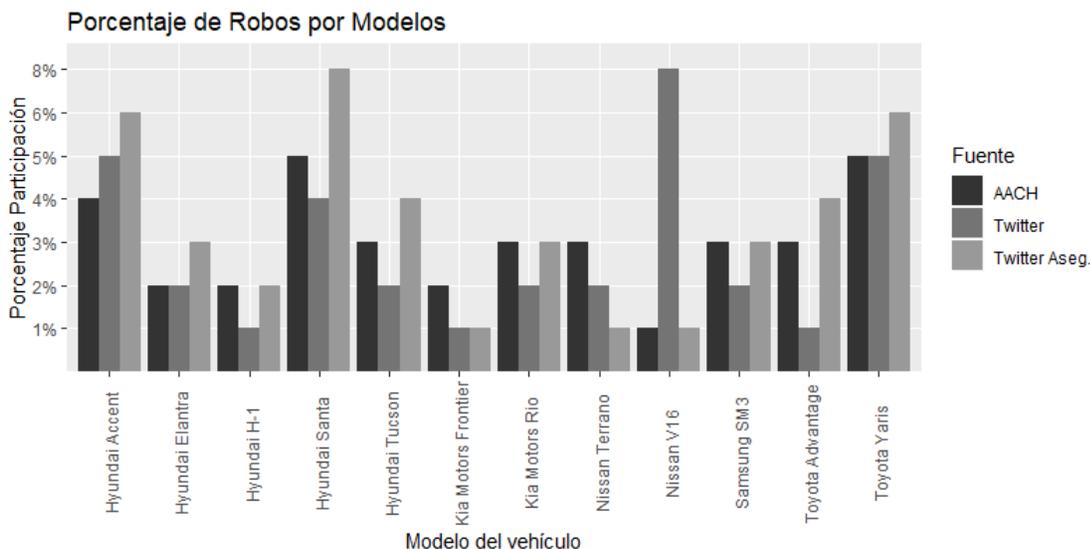


Figura 5: Gráfico Porcentaje de robos por modelos.

Una de las preguntas de investigación que estaba presente al momento de iniciar la investigación es de saber si el hecho de denunciar el robo de un vehículo por Twitter tiene un efecto positivo en la probabilidad de encontrar el vehículo.

Se verifica si el tamaño de muestra es suficiente para que sea representativo de acuerdo a la fórmula de Cochran [6, 8] y la distribución de los datos distribuye normal con el test prueba Kolmogorov-Smirnov [9].

Para esto se realiza un Test de Igualdad de Proporciones sobre las proporciones de vehículos robados que fueron encontrados, comparando entre aquellos que sólo fueron denunciados por las instituciones formales y aquellos que, además de hacerlo por esta vía, lo realizaron también por Twitter. Se define:

$p_1$ : La proporción de vehículos robados que No fueron denunciados por Twitter y fueron encontrados.

$p_2$ : La proporción de vehículos robados que Sí fueron denunciados por Twitter y fueron encontrados.

*Hipótesis nula*  $H_0 : p_1 = p_2$

*Hipótesis alternativa*  $H_1 : p_1 \leq p_2$

Como hipótesis alternativa se expresa que la proporción de hallazgos de aquellos vehículos robados que fueron denunciados por Twitter es mayor que aquellos que no lo fueron. Para analizar las tasas de hallazgo se consideraron en la base de datos de Twitter aquellos vehículos que están asegurados, es decir que las patentes coinciden en la base de datos de AACH, los cuales son 1.763 vehículos (24 % de los datos totales). Las tasas de hallazgos de ambas bases de datos son: 59 % (AACH) y 70 % (Twitter).

Al realizar un t-test de proporciones a un 99 % de confianza se obtiene como resultado un p-valor  $2,2 \times 10^{-16}$ . Algo importante de descartar es el hecho de que el sesgo en los grupos de vehículos denunciados en Twitter expliquen las diferencias en las proporciones, es decir que las tasas de hallazgos no mejoren en todos los grupos, sino que más bien se mantengan constantes pero que en Twitter se denuncien con mayor proporción aquellos vehículos que tienen mayor tasa de recuperación haciendo que la proporción general de hallazgo suba.

Por lo anterior es que se analiza las tasas de hallazgos entre grupos de tasación, para verificar si la proporción de hallazgos de los vehículos robados se ve afectada en todos los grupos. Para realizar esta comparación se consideran los registros de Twitter que también están presentes en AACH.

A continuación se presenta un resumen de los resultados obtenidos respecto al análisis de la frecuencia de robos y hallazgos de los vehículos robados según grupo de tasación y si fue denunciado en Twitter o no.

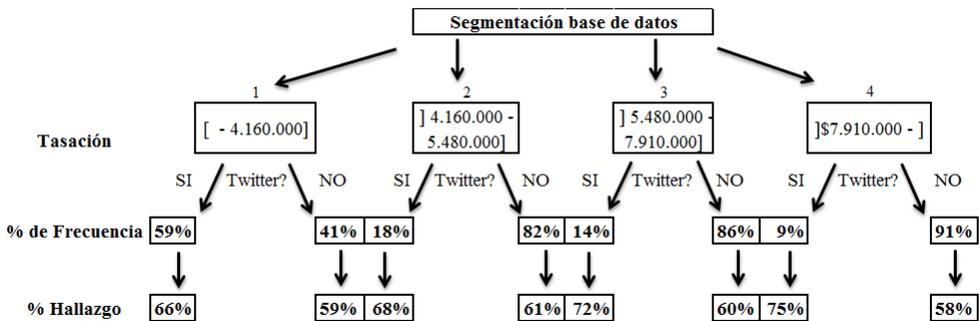


Figura 6: Resumen árbol de análisis según grupo de tasación.

La Figura 6 muestra cómo se generó una segmentación de las denuncias en 4 grupos de acuerdo a la tasación del vehículo, para luego seguir una segmentación según si el vehículo fue denunciado en Twitter o no, y según esta decisión medir la frecuencia con la que aparece este tipo de segmento en la fuente de datos y su correspondiente porcentaje de hallazgo. Recordar que el grupo de tasación se escogió dividiendo en 4 cuantiles la fuente de datos de

AACH de tal manera de que cada grupo representara un cuarto de la base de datos. Otra inquietud presente en esta investigación es saber si hay variación en las tasas de robos según el tipo de vehículo dependiendo el día de semana, lo cual hablaría de las intenciones detrás del robo. Para poder visualizar la variación se calculó la tasa de participación de robos de cada tipo de vehículo, y luego se calculó como esa tasa variaba dependiendo el día de la semana. Primero se analiza la fuente de datos de AACH y posteriormente la fuente de datos de Twitter.

En los resultados de AACH dos tipos de vehículos presentan variaciones importantes, el tipo de vehículo “Automóvil” y el tipo “Camioneta”. La explicación podría residir en la intención del uso del vehículo robado, según Carabineros de Chile, algunos robos son realizados para ser utilizados para fiestas, en este caso privilegian los del tipo “Automóvil”. Por otra parte algunos robos de vehículos se realizan con la intención de utilizar el vehículo para perpetrar un robo de otro tipo, como asaltos, o robos de bienes comerciales, para este caso es posible que los vehículos del tipo “Camioneta” sean más adecuados para cumplir la intención, además es común que las camionetas sean utilizadas con fines laborales y por ende con mayor frecuencia los días de la semana aumentando sus tasas de robos en estos días al estar más expuestas. Obtenido estos resultados provenientes de los datos de AACH, se mostrarán los resultados de Twitter para poder observar si estos muestran los mismos.

Los resultados de Twitter son similares, sólo que incorpora también el día viernes, pero el análisis sigue siendo el mismo, el fin de semana un tipo de vehículo “Automóvil” aumenta considerablemente y por el contrario el tipo de vehículo “Camioneta” decrece en los mismos días.

#### 3.2.4. Interpretación de patrones

En este paso final se interpretan los patrones y relaciones encontrados en el proceso KDD. Se han aplicado diferentes técnicas que han permitido entender las relaciones entre ambas fuentes de datos. Previo a interpretar los resultados expuestos en el paso anterior se hará una reconstrucción del hecho, es decir, se ordenarán las fases por las que pasa un vehículo robado, concentrándose en vehículos asegurados.

- **Robo del Vehículo:** Corresponde al momento en que el vehículo fue robado. La hora de este evento es declarada por el dueño del vehículo al momento de denunciar en Carabineros de Chile, por lo que en algunos casos corresponde a una hora aproximada, ya que no siempre el dueño del vehículo puede observar cuándo el delito es realizado.

Tabla 3: Variación tasa de robo de vehículo según tipo, en fuente de datos Twitter.

Día	Automóvil	Station Wagon	Camioneta	Todo Terreno	Furgón	Minibus
lunes	-2,3 %	0,9 %	0,0 %	0,8 %	0,7 %	-0,2 %
martes	-0,1 %	-1,2 %	0,9 %	0,9 %	-0,3 %	-0,2 %
miércoles	-1,8 %	0,9 %	1,0 %	-0,2 %	0,2 %	-0,1 %
jueves	-0,7 %	-0,5 %	1,2 %	-0,1 %	-0,2 %	0,3 %
viernes	0,6 %	-0,3 %	-1,3 %	-0,3 %	0,7 %	0,5 %
sábado	4,3 %	-0,4 %	-1,7 %	-1,3 %	-0,6 %	-0,2 %
domingo	0,0 %	0,6 %	-0,2 %	0,2 %	-0,5 %	0,0 %
<b>Promedio Anual</b>	61,1 %	18,9 %	11,1 %	6,1 %	2,0 %	0,8 %

Tabla 4: Variación tasa de robo de vehículo según tipo, en fuente de datos AACH.

Día	Automóvil	Station Wagon	Camioneta	Todo Terreno	Furgón	Minibus
lunes	-1,7 %	-0,7 %	1,9 %	-0,3 %	0,9 %	0,0 %
martes	-2,0 %	-0,5 %	2,0 %	0,4 %	0,1 %	0,0 %
miércoles	-0,9 %	-0,7 %	1,5 %	0,0 %	0,0 %	0,1 %
jueves	-0,7 %	0,2 %	1,2 %	-0,3 %	-0,3 %	0,0 %
viernes	-0,8 %	0,6 %	0,5 %	-0,5 %	0,2 %	0,1 %
sábado	3,9 %	1,3 %	-4,6 %	-0,1 %	-0,4 %	0,0 %
domingo	2,7 %	-0,3 %	-3,0 %	1,0 %	-0,5 %	0,0 %
<b>Promedio Anual</b>	46,3 %	24,0 %	19,5 %	7,4 %	2,6 %	0,1 %

- **Envío del tuit:** Es la hora exacta en la que el tuit denunciando el robo del vehículo fue enviado.
- **Denuncia en Carabineros de Chile:** Es la hora en la que se realiza la denuncia formalmente en Carabineros de Chile por el robo del vehículo.
- **Validación AACH:** Es la fecha y hora en que AACH almacena los datos de las denuncias del robo del vehículo, la cual fue realizada en la aseguradora en la que el cliente está suscrito. Las aseguradoras no envían

los datos todos los días, sino que consolidan varias denuncias y luego se las transfieren a AACH, quien las valida y almacena.

- **Hallazgo del Vehículo:** Corresponde a la fecha y hora en la que el vehículo denunciado por robo fue hallado. Como no todos los vehículos son encontrados, para este cálculo se consideraron solamente aquellos que efectivamente fueron hallados.

Para medir los tiempos entre las fases se calculó la mediana de cada una. Asumiendo los tiempos medianos se puede reconstruir la secuencia de los hechos relacionados con un robo como lo muestra la Figura 7.

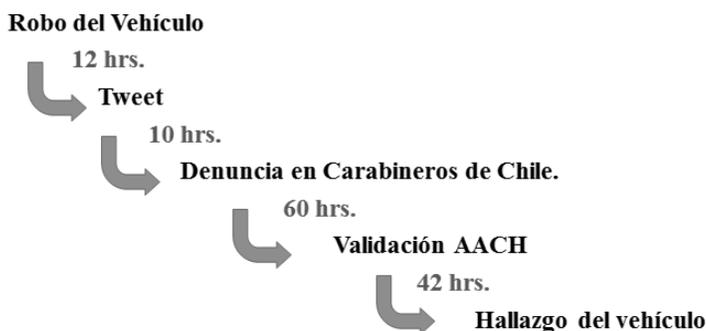


Figura 7: Reconstrucción orden de los hechos en un robo de vehículo.

Esto expresa el valor que presenta Twitter, ya que permite obtener información de manera anticipada, de hecho la mediana indica que el tuit se genera con 10 hrs. de anticipación frente a la denuncia formal. Además hay un largo periodo de tiempo transcurrido entre que el robo es realizado y AACH obtiene esa información, información que debe ser organizada y tratada para extraer los campos de interés. Además si bien en la generalidad de los casos, son aproximadamente 3 a 4 días los que transcurren desde que el vehículo es robado hasta que AACH obtiene la información, en muchos casos la obtención de esos datos se aproxima a las 2 semanas [20], por lo que en esos casos Twitter presenta un valor aun mayor al poder recibir información relacionada de manera anticipada. Respecto al tiempo transcurrido desde el robo hasta que es encontrado el vehículo, según la información de AACH, en la mayoría de los casos es cercana a los 5 días.

---

## 4. Conclusiones y Trabajo Futuro

---

Este trabajo presenta de obtención de tuits, extracción de información de ellos con técnicas de minería de texto, y aplicación de la metodología KDD sobre la fuente de AACH (datos de robos de vehículos asegurados) y sobre los datos de Twitter (tuits extraídos con un script).

Lo primero que se realizó en el proceso de minería de datos fue analizar la correlación entre las frecuencias de las denuncias de Twitter y AACH, el resultado mostró que tienen un comportamiento correlacionado entre el año 2012-2016, y que se acentúa más en el periodo 2015-2016 alcanzando un coeficiente de correlación de 0,73.

En cuanto a la valorización de los vehículos denunciados se expresó un gran sesgo, ya que en Twitter cerca del 60 % de los vehículos denunciados corresponden al grupo de menor valor, es decir aquellos que tienen una valorización menor a \$4.160.000. Esto puede estar relacionado con el hecho de que quienes emplean mayor tiempo de navegación en internet son los más jóvenes, a quienes se les puede atribuir una menor disposición a pago por un vehículo. Lo importante de este hallazgo es considerar que al momento de analizar el robo de vehículos de alto valor, estos presentarían frecuencias considerablemente menores en Twitter, incluso pudiendo no ser factible de utilizar Twitter para predecir comportamientos delictuales para estos vehículos.

En cuanto a los modelos de vehículos denunciados, ambas fuentes de datos consideran tasas similares de robo.

Lo relevante del hallazgo es que para los modelos más robados, Twitter presenta tasas muy similares a las denuncias realizadas en AACH, y por lo tanto las estimaciones basadas en la red social deberían ser bastantes confiables para estos casos. Sin embargo, para los modelos de los vehículos utilizados para transporte de pasajeros no presentan correlación. Es fundamental ir monitoreando los modelos utilizados para estos fines ya que al momento de hacer las mediciones o utilizar el modelo del vehículo como variable predictiva se generarán conclusiones erróneas si no se considera.

Se descubrió que los vehículos denunciados en la red social presentan mayor tasa de hallazgo que aquellos que no lo son, en donde esta tasa de hallazgo no dependía de un sesgo asociado a la valorización del vehículo. La diferencia significativa de 11 puntos porcentuales puede indicar que el uso de las redes sociales ayuda a masificar las denuncias e incluso a recuperar sus pertenencias en caso de robo.

Este descubrimiento es altamente valioso para generar políticas públicas respecto al robo de vehículos, en donde se utilizan recursos de Carabineros de Chile para la recuperación de vehículos, ya que estimulando el uso de Twitter y otras redes sociales para denunciar los robos y posteriormente hallarlos se podría disminuir el costo en los recursos empleados actualmente para estos fines. Incluso las aseguradoras podrían estimular o promocionar el uso de esta red social para aumentar la probabilidad de encontrar un vehículo robado.

Se descubrió que el fin de semana el robo de vehículos de la categoría “Automóvil” aumenta y por otro lado disminuye aquellos de la categoría “Camioneta”. Este patrón se evidenció en ambas fuentes de datos, lo cual es importante porque indicaría que Twitter no presenta sesgo en ese comportamiento. Finalmente se realizó una reconstrucción de las fases por las cuales pasa un vehículo robado en donde el resultado fue que primero el vehículo es robado, luego es denunciado por Twitter, para luego ser denunciado en Carabineros, siguiendo con la recepción de los datos por la institución AACH y finalmente el vehículo es hallado. Este ordenamiento mostró el valor que tiene Twitter como una fuente de información anticipada, ya que es el primer evento que sucede luego del robo del vehículo, de hecho es realizado antes que la denuncia formal en Carabineros de Chile. El hecho de que el tuit se origine como el primer evento luego de la sustracción del vehículo puede ser aprovechado por distintas entidades como AACH, municipios o las aseguradoras, para monitorear en todo momento los vehículos denunciados recientemente y aumentar las probabilidades de hallazgo al focalizarse en los que fueron denunciados en algún lugar en particular.

Tal como mostraban investigaciones del robo de vehículo que incluyen modelos predictivos, el no saber los sesgos en Twitter, dificultaba el entendimiento de algunos patrones de comportamientos, y dificultaba la optimización o mejoras en el rendimiento del modelo. Los descubrimientos aquí expuestos entregan información que permitirá estar conscientes de estos sesgos, pudiendo reparar errores en medición o en proporciones, por ejemplo en probabilidades de robos de algunos modelos.

En el trabajo futuro se construirá sobre la metodología desarrollada y los resultados presentados en este trabajo para construir un observatorio digital de robos de vehículos. El observatorio será alimentado con varias fuentes de información, entre ellas los siniestros guardados en las bases de datos de AACH y los datos publicados en redes sociales como por ejemplo Twitter.

**Agradecimientos:** Este trabajo fue financiado por el proyecto FONDEF ID16I10222 – CONICYT y el Instituto Sistemas Complejos de Ingeniería (ISCI) CONICYT FB0816. Se agradece el apoyo de la Asociación de Aseguradores de Chile (AACH) y de Carabineros de Chile.

## Referencias

- [1] B. Baesens. *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons, 2014.
- [2] Carabineros de Chile. Algunas definiciones : Delitos de mayor connotación social. <http://dac.carabineros.cl/datos.php>. En línea; Último acceso Diciembre 2016.
- [3] ENUSC. Presentación de resultado xii encuesta nacional urbana de seguridad ciudadana, 2015. [Último acceso: Diciembre 2016].
- [4] M. S. Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.
- [5] INE. Ine anuarios parque de vehículos en circulación 2015. "[http://www.ine.cl/canales/chile\\_estadistico/estadisticas\\_economicas/transporte\\_y\\_comunicaciones/parquevehiculos.php#](http://www.ine.cl/canales/chile_estadistico/estadisticas_economicas/transporte_y_comunicaciones/parquevehiculos.php#)", apr 2017. [En línea; Último acceso Agosto 2018].
- [6] G. Israel. Determining sample size, arlington: Program evaluation and organizational development, ifas, university of florida. peod-6. *National Science Foundation, Research and Development in Insutry*, 1992.
- [7] J. Kalyanam, M. Quezada, B. Poblete, y G. Lanckriet. Prediction and characterization of high-activity events in social media triggered by real-world news. *PloS one*, 11(12):e0166694, 2016.
- [8] E. Lehmann y J. Romano. *Testing statistical hypotheses*. Springer Science & Business Media, 2006.
- [9] H. W. Lilliefors. On the kolmogorov-smirnov test for normality with mean and variance unknown. *Journal of the American statistical Association*, 62(318):399–402, 1967.
- [10] R. J. Little y D. B. Rubin. *Statistical analysis with missing data*. John Wiley & Sons, 2002.

- [11] N. Malleson y M. A. Andresen. The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 42(2):112–121, 2015.
- [12] D. Padula y L. Debera. Técnicas de imputación, una aplicación para medir el ingreso. In *Décimo Congreso Latinoamericano de Sociedades de Estadística. Córdoba*, 2012.
- [13] Servicio de Impuestos Internos. Sii tasación fiscal de vehículos. "[http://www.sii.cl/pagina/actualizada/noticias/tasacion\\_vehiculos.htm](http://www.sii.cl/pagina/actualizada/noticias/tasacion_vehiculos.htm)". [En línea; Último acceso: Abril 2017].
- [14] Subsecretaría de Prevención del Delito. Delitos de mayor connotación social - series de datos 2001 - 2016, apr 2017.
- [15] R. Todd-Bennet. Identifying crime hotspots using twitter: School of Computer Science and Informatics Cardiff University, 2015. [En línea; Último acceso Agosto 2018].
- [16] Twitter. REST API. "<https://developer.twitter.com/en/docs>". [En línea; Último acceso: Agosto 2018].
- [17] Twitter. Streaming API's. "<https://developer.twitter.com/en/docs/accounts-and-users/subscribe-account-activity/guides/account-activity-data-objects1>". [En línea; Último acceso Agosto 2018].
- [18] Twitter. Twitter ads API. "<https://developer.twitter.com/en/docs/ads/general/overview>". [En línea; Último acceso: Agosto 2018].
- [19] Twitter. Blog twitter. "<https://twitter.com/twittersupport/status/555076845293432834>", jan 2015. [En línea; Último acceso: Agosto 2018].
- [20] A. Vásquez. Análisis de relaciones existentes entre datos de robos de vehículos e información extraída de twitter aplicando kdd. Memoria para optar al título de ingeniero industrial, Universidad de Chile, 2017.

---

# EFECTOS DE LA CONTAMINACIÓN DEL AIRE EN ATENCIONES DE URGENCIA DE LA REGIÓN METROPOLITANA

---

VALERIA SCAPINI \*  
CARLA CARRASCO \*\*  
CINTHYA VERGARA \*\*

## Resumen

La contaminación del aire genera diversos efectos sobre la salud humana. Los últimos años se evidencia a nivel mundial alrededor de 4,2 millones de muertes prematuras anuales como resultado de la exposición a la contaminación del aire. Esta investigación busca medir el impacto de las concentraciones de ozono, material particulado  $MP_{10}$  y  $MP_{2,5}$  en las atenciones de urgencia del sector público de salud en la Región Metropolitana de Chile. Se utilizaron 2 fuentes de datos dentro del período 2010 al 2017; para capturar parámetros contaminantes se utilizó el Sistema de Información Nacional de Calidad del Aire (SINCA) y, para los datos de atenciones de urgencia por causas respiratorias, los datos del Departamento de Estadísticas e Información en Salud (DEIS) dependiente del Ministerio de Salud. La estimación de modelos se llevó a cabo con regresiones lineales log-log para calcular la elasticidad. Los resultados obtenidos son mixtos, existe una relación entre las concentraciones de contaminantes y la cantidad total de atenciones de urgencias en la Región Metropolitana que varía dependiendo del tipo de enfermedad respiratoria y la estación del año. Cuantificar los efectos de los distintos parámetros contaminantes es de gran utilidad para priorizar medidas ambientales y planificar, incluso adelantarse y reducir costos relacionados a las atenciones de salud.

**Palabras Clave:** Políticas de salud, Medio ambiente, Contaminación, Salud pública, Parámetros contaminantes, Contaminación.

---

\* Escuela de Ingeniería Comercial, Universidad de Valparaíso, Santiago, Chile

\*\* CRS Hospital Provincia Cordillera, Santiago, Chile

---

## 1. Motivación y contexto

---

Los efectos de la contaminación atmosférica en la salud es un área del conocimiento que ha sido ampliamente discutido. Distintos estudios [3, 8, 10, 16, 18, 22, 26] muestran que los contaminantes ambientales pueden contribuir a problemas en la salud y calidad de vida como la disminución de la función pulmonar, el nivel de tolerancia al ejercicio, un aumento en la reactividad bronquial, el riesgo de bronquitis obstructiva crónica, exacerbación del asma bronquial, cáncer pulmonar, entre otros.

La exposición a la contaminación del aire se ha asociado con una serie de efectos adversos para la salud humana [8, 16]. No sólo hay evidencia que afecta al sistema respiratorio sino también al cardíaco [2, 21], trae consecuencias neurológicas [5, 6, 23] e incluso efectos sobre la reproducción [7, 11, 15].

De acuerdo con datos de la Organización Mundial de la Salud (OMS) la contaminación del aire exterior es la principal causa medioambiental de riesgo a la salud generando alrededor del 23 % de las muertes a nivel mundial. La contaminación del aire está asociada a cerca de 4,2 millones de muertes prematuras, donde el 58 % de las muertes prematuras está asociada a cardiopatía isquémica y accidentes cerebro vasculares, el 18 % a enfermedad pulmonar obstructiva crónica e infecciones respiratorias inferiores y el 6 % a cáncer de pulmón [18, 17].

Dentro de las principales partículas contaminantes está el material particulado ( $MP$ ) que corresponde a una mezcla de partículas, sólidas o líquidas, suspendidas en el aire, que difieren en tamaño, composición y origen y se clasifican de acuerdo a sus propiedades aerodinámicas. El  $MP_{10}$  comprende las partículas de diámetro aerodinámico menor a 10 micrómetros, siendo partículas que pueden penetrar el sistema respiratorio hasta los pulmones, produciendo irritaciones y aumentando la incidencia de enfermedades relacionadas a éste. El  $MP_{2,5}$  comprende las partículas menores a 2,5 micrómetros y estas son de especial atención debido a que pueden ingresar hasta los alvéolos y luego a la sangre. Estudios epidemiológicos dan cuenta de la asociación que hay entre la contaminación ambiental, cuando existen altas concentraciones de  $MP_{10}$  y  $MP_{2,5}$  y la incidencia en el aumento de patologías respiratorias y cardiovasculares en la población [2, 21], siendo los niños y adultos mayores los más vulnerables y con mayores índices de mortalidad por causa de estas enfermedades.

Otro contaminante de especial atención es el ozono ( $O_3$ ), un gas que se

forma y reacciona frente a la luz solar y que, al estar a nivel del suelo puede afectar a la salud de los pulmones, el tracto respiratorio, los ojos y aumentar la susceptibilidad a los alérgenos respiratorios. El ozono superficial se forma a partir de otros contaminantes y puede reaccionar con otras sustancias por la acción de la luz. Las concentraciones son normalmente bajas en las zonas urbanas con mucho tráfico y suelen ser superiores en las afueras y en zonas rurales anexas, especialmente en los días soleados de verano.

En la Región Metropolitana de Chile existe una relación espacial entre las variables climáticas y las concentraciones de  $MP$  debido a la geografía de la zona. Al encontrarse rodeada de montañas, se produce el fenómeno de *inversión térmica*, concentrándose las partículas contaminantes dentro de la cuenca y, frente a bajas temperaturas, se concentran y se forma una capa de aire que no permite la ventilación ni la dispersión de los contaminantes. En verano, el fenómeno se ubica aproximadamente a 1000 metros de altura y, en invierno, alcanza los 200 o 300 metros sobre el nivel del suelo, afectando en mayor medida la salud de las personas.

En esta región, en el año 2018, hay 13 estaciones de monitoreo operativas que miden las concentraciones de partículas en el aire de material particulado  $MP_{2,5}$ , material particulado  $MP_{10}$ , dióxido de azufre  $SO_2$ , dióxido de nitrógeno  $NO_2$ , óxidos de nitrógeno  $NO_X$ , monóxido de nitrógeno  $NO$ , monóxido de carbono  $CO$ , ozono  $O_3$ , metano  $CH_4$  e hidrocarburos no metánicos  $HCNM$ .

Por su parte, el sistema de salud chileno posee un sistema previsional mixto, donde el sector público se gestiona a través del Fondo Nacional de Salud (FONASA) quien se encarga de la cobertura financiera de sus afiliados. El seguro público cubre al 72,5% de población [14], por lo tanto, la mayoría de las atenciones se resuelven en recintos de salud pública.

El presente trabajo de investigación busca cuantificar cómo afectan los contaminantes del aire  $MP_{10}$ ,  $MP_{2,5}$  y  $O_3$  en la salud de la población tomando como referencia los datos de atenciones de urgencia del sistema de salud pública correspondientes a urgencias respiratorias dentro de la Región Metropolitana. Los datos de atenciones de salud se estructuraron de acuerdo a semanas estadísticas, estación del año (verano, otoño, invierno, primavera) y zonas con el promedio semanal de atenciones dentro del período 2010-2017. Se generaron series de tiempo para comprender el comportamiento tanto de las atenciones como de los contaminantes bajo estudio. Finalmente, se utilizó como método de análisis el modelo de regresiones lineal con la estrategia log-log para estimar las elasticidades entre las atenciones de urgencia respiratoria y las partículas contaminantes  $MP_{10}$ ,  $MP_{2,5}$  y  $O_3$  para cada tipo de problema respiratorio registrado en las series estadísticas.

La sección 2 muestra trabajos relacionados con el análisis de los efectos de la calidad del aire y las enfermedades respiratorias, la sección 3 analiza de manera cualitativas los datos obtenidos, la sección 4 detalla los modelos estimados y sus resultados, la sección 5 presenta las principales conclusiones.

---

## 2. Trabajos relacionados

---

Para abordar el problema de la contaminación del aire, se han desarrollado investigaciones que analizan desde distintas perspectivas los efectos de parámetros contaminantes en la salud de las personas. Por un lado nos encontramos con análisis que buscan explicar las causas de la contaminación y sus efectos en la salud de las personas y, por otro, modelos que cuantifican los efectos de los distintos agentes contaminantes del aire así como también modelos de predicción de los niveles de dichos contaminantes.

Según estudios recientes [10] sobre el efecto de las partículas contaminantes en la salud, el ozono y el  $MP$  son los que más impacto tienen sobre problemas cardiopulmonares. En [1] estudian los efectos de la exposición al ozono y otros contaminantes atmosféricos en la Ciudad de México entre los años 1990-1992, donde se utilizó una regresión Poisson para modelar el número de muertes diarias en función de parámetros de contaminación del aire, y se observó una asociación con la concentración máxima diaria de ozono y con las concentraciones medias de dióxido de azufre y  $MP$ . Por el contrario, otra línea de estudios no revelan efectos en la salud [19]. Se ha discutido respecto al nivel de exposición y a la fuente contaminante, encontrando quienes concuerdan con los altos niveles de material en suspensión o gases y su incidencia en la salud [25], mientras que para otros los efectos se evidencian incluso con bajas concentraciones [12]. En [13] se investiga los niveles elevados de contaminación del aire en China encontrando que éstos resultan ser significativos y adversos para la salud de la población en zonas urbanas.

Para el caso chileno, en [20] se utilizó un modelo de regresión múltiple con controles de temperatura mínima diaria, temperaturas extremas, día de la semana, mes y año. En él se identifica que un aumento de 10 *microgramos*/ $m^3$  en el  $MP_{10}$  diario se relaciona con un aumento del 1% en la mortalidad en Santiago de Chile. Otro estudio muestra que los casos de pacientes con insuficiencia cardíaca presentan altos riesgos de hospitalización ante incrementos en las concentraciones de  $MP_{2,5}$  [4]. En [9] se estudia el efecto de  $MP_{2,5}$  y otros contaminantes sobre la cantidad de atenciones de urgencias respiratorias en un hospital infantil entre 1995 y 1996, concluyendo que la contaminación,

especialmente asociada a partículas finas, afecta negativamente la salud de los niños en la ciudad de Santiago.

Finalmente es posible ver que existen distintas aproximaciones así como técnicas para abordar los problemas de contaminación del aire y su relación con los problemas de salud.

---

### 3. Descripción de los datos

---

#### 3.1. Datos recolectados

Para realizar el estudio se utilizaron 2 fuentes de datos independientes dentro del período 2010 al 2017. Para capturar parámetros contaminantes se utilizó el Sistema de Información Nacional de Calidad del Aire (SINCA) y, para los datos de atenciones de urgencia por causas respiratorias, los datos del Departamento de Estadísticas e Información en Salud (DEIS).

Los datos de atenciones de salud se estructuraron de acuerdo a 53 semanas estadísticas<sup>1</sup>, estación del año (verano, otoño, invierno, primavera) y 6 zonas geográficas (central, norte, occidente, oriente, sur y suroriente) de la Región Metropolitana.

Tabla 1: Asignación de zona para las estaciones de monitoreo SINCA

Nº	Estación	Zona Asignada
1	Cerrillos	Norte
2	Cerrillos I	Norte
3	Cerro Navia	Occidente
4	El Bosque	Sur
5	Independencia	Norte
6	La Florida	Suroriente
7	Las Condes	Oriente
8	Pudahuel	Occidente
9	Puente Alto	Suroriente
10	Quilicura	Norte
11	Quilicura I	Norte
12	Parque O'Higgins	Central
13	Talagante	Occidente

---

<sup>1</sup>Una semana estadística corresponde a cada una de las semanas del año, desde la semana 1 a la 53, con sus correspondientes datos estadísticos de atención.

Las causas respiratorias se obtuvieron del registro estadístico del DEIS<sup>2</sup> desde donde es posible obtener el promedio semanal de atenciones de urgencia para las 6 zonas de la Región Metropolitana.

Tabla 2: Causas respiratorias disponibles en el registro estadístico del DEIS

Causa	Abreviación	Códigos CIE10
Infección respiratoria aguda	IRA	J00-J06
Influenza	-	J09-J11
Neumonía	-	J12-J18
Bronquitis	-	J20-J21
Crisis bronquial obstructiva	CBO	J40-J46
Otra causa respiratoria	Otras	J22-J30, J39-J47-J60, J98

De acuerdo a esta estructura se recolectaron los datos del SINCA desde 13 estaciones de monitoreo de la Región Metropolitana, asociando a las atenciones de urgencia respiratorias, los niveles de las partículas contaminantes  $MP_{10}$ ,  $MP_{2,5}$  y ozono. Las estaciones fueron clasificadas de acuerdo a las zonas determinada en el DEIS y para la consolidación de datos se calculó el promedio por semana estadística. También se generaron 9 variables indicando el promedio de emisión de la primera, segunda y tercera semana anterior de cada semana como variables de rezago.

Finalmente se transformó el valor de las variables a su valor logarítmico para estimar la elasticidad de la cantidad de atenciones con respecto a los contaminantes seleccionados.

### 3.2. Análisis de series de tiempo

Durante el período analizado hubo un total de 16.141.716 atenciones de urgencias respiratorias en establecimientos públicos de urgencia de la Región Metropolitana, con un promedio semanal de 38.709 atenciones (ver Figura 1).

De acuerdo con los datos consolidados se identifica que, para la serie de tiempo de número de atenciones por semana estadística, existe un patrón estacional de ciclo anual con un importante componente aleatorio (ver Figura 2).

<sup>2</sup>CIE10 (Clasificación Internacional de Enfermedades, 10<sup>a</sup> versión) de la Organización Mundial de la Salud (OMS)

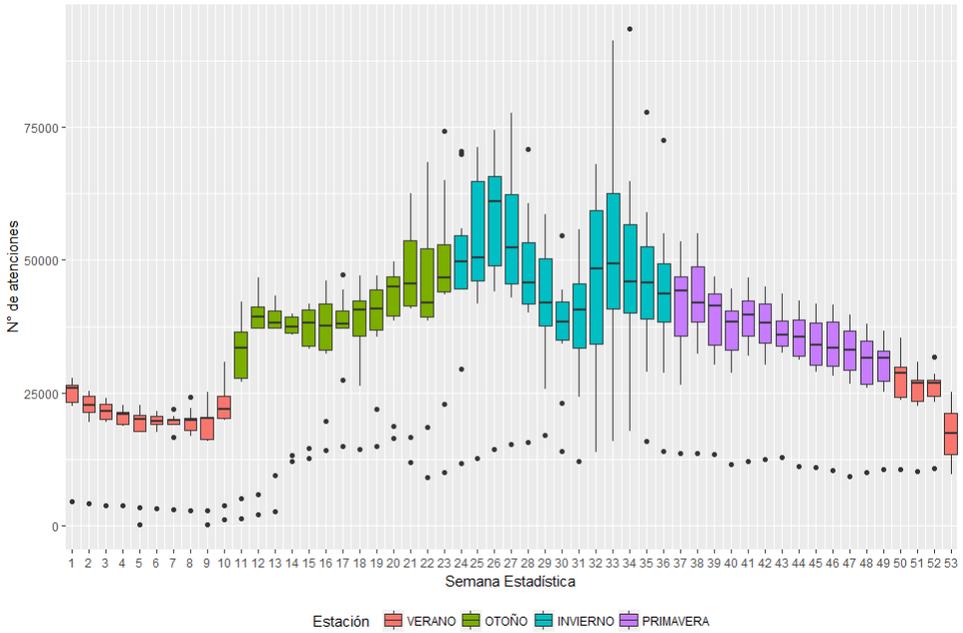


Figura 1: Diagrama de caja comportamiento semanal atenciones de urgencia respiratorias período 2010-2017

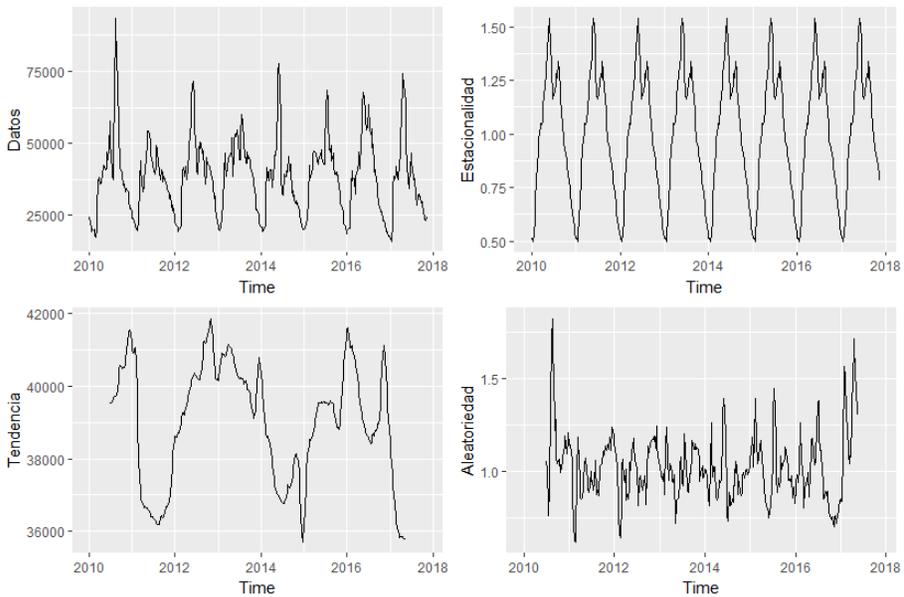


Figura 2: Descomposición de los multiplicativos de series de tiempo de atenciones de urgencia respiratorias

Existe mayor presión asistencial debido a enfermedades respiratorias en el período invernal caracterizado por menor temperatura atmosférica y mayor cantidad de precipitaciones que los otros períodos del año. Separando la serie por cada una de las zonas de la Región Metropolitana se identifican diferencias en la magnitud de atenciones manteniendo el mismo comportamiento semanal.

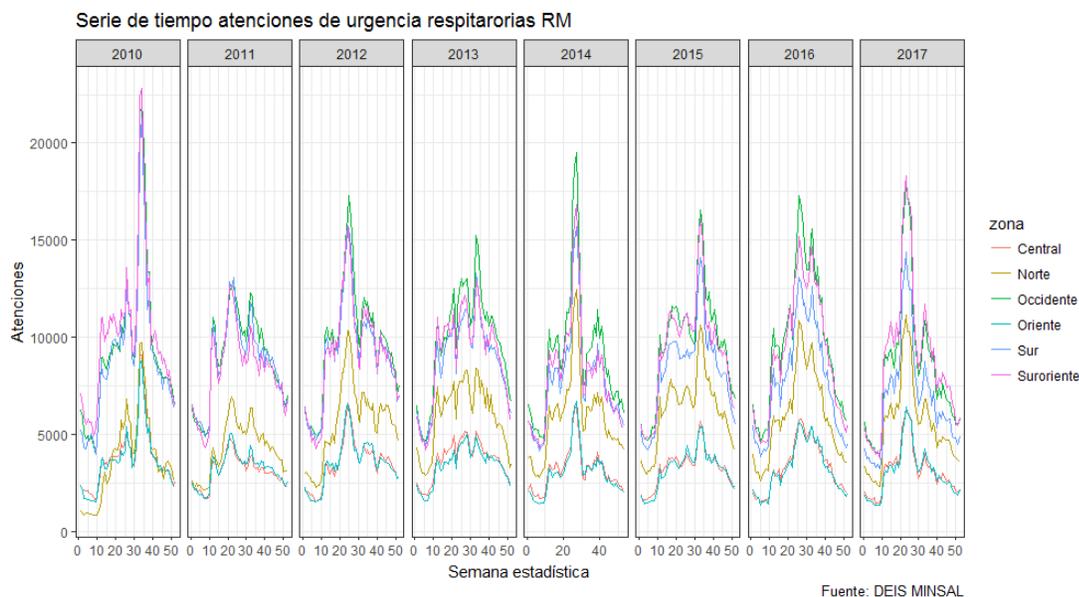


Figura 3: Serie de tiempo atenciones de urgencia respiratorias en la Región Metropolitana entre los años 2010-2017 por zona

Por el lado de los contaminantes del aire  $MP_{10}$ ,  $MP_{2,5}$  y ozono, se identifica un patrón estacional de ciclo anual donde el  $MP_{10}$  y  $MP_{2,5}$  aumentan significativamente durante el período de invierno mientras que el ozono disminuye.

El comportamiento simultáneo entre contaminantes, atenciones de urgencia respiratoria y estaciones del año muestra que cuando hay mayor número de atenciones hay mayor nivel de contaminantes  $MP_{2,5}$  y  $MP_{10}$  y menor nivel de ozono, este último aumenta cuando hay mayor temperatura ambiental y, por el contrario, el  $MP_{10}$  y el  $MP_{2,5}$  aumentan frente a bajas temperaturas en la cuenca de Santiago debido al fenómeno de inversión térmica. A su vez, las enfermedades respiratorias aumentan cuando hay menor temperatura y mayor humedad, escenario que se presenta en el período de invierno. Todas las variables tienen un marcado componente estacional.

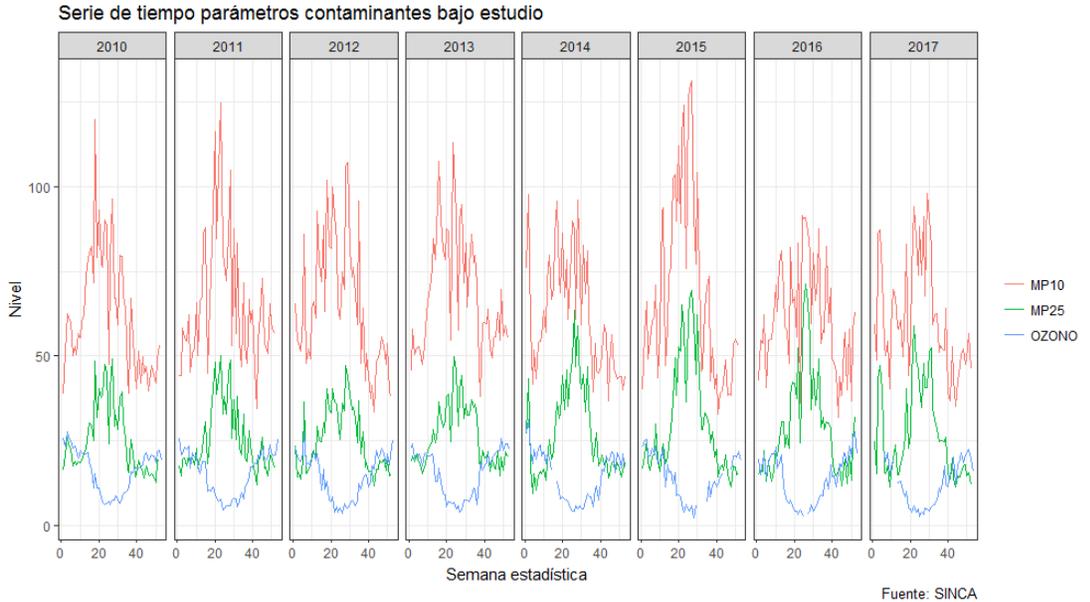


Figura 4: Niveles de los materiales contaminantes  $MP_{2,5}$  y  $MP_{10}$  y ozono en la Región Metropolitana entre los años 2010-2017

---

## 4. Estimación de modelos y resultados

---

De acuerdo al análisis de las series de tiempo de las variables bajo estudio se identifica que existe efectos estacionales significativos para cada una de las variables, así como diferencias de magnitud por zona. Frente a este escenario se utilizó la *semana* como efecto fijo para todos los modelos. La notación utilizada será la siguiente:

### Índices

$t$ : Año, con  $t \in \{2008, \dots, 2017\}$ .

$s$ : Semana estadística, con  $s \in \{1, \dots, 53\}$ .

$e$ : Estación del año, con  $e \in \{\text{verano}, \text{otoño}, \text{invierno}, \text{primavera}\}$

$s_i$ : Semana anterior  $i$ , con  $i \in \{0, 1, 2, 3\}$ .

$r$ : Enfermedad respiratoria específica, con  $r \in \{\text{IRA}, \text{Influenza}, \text{Neumonía}, \text{Bronquitis}, \text{CBO}, \text{Otras}\}$ .

## Variables

$AU$ : Logaritmo del número de atenciones de urgencia respiratorias de la Región Metropolitana.

$MP_{2,5}$ : Logaritmo del nivel de material particulado  $MP_{2,5}$ .

$MP_{10}$ : Logaritmo del nivel de material particulado  $MP_{10}$ .

$O_3$ : Logaritmo del nivel de ozono.

$EF^s$ : Efecto fijo semana estadística.

$EF^e$ : Efecto fijo estación del año.

El primer modelo estimado tiene por objetivo analizar de manera general los efectos de los contaminantes sobre el total de atenciones de urgencia respiratorias de acuerdo a la siguiente estructura:

$$AU^{tse} = EF^s + EF^e + \beta_1 \times MP_{2,5}^{ts} + \beta_2 \times MP_{10}^{ts} + \beta_3 \times O_3^{ts} + \epsilon^{ts} \quad (1)$$

Sobre este modelo se agregaron los niveles de contaminación de las semanas anteriores (período de rezago), generando la siguiente estructura:

$$AU^{tse} = EF^s + EF^e + \beta_0 + \beta_1 \times MP_{2,5}^{ts_i} + \beta_2 \times MP_{10}^{ts_i} + \beta_3 \times O_3^{ts_i} + \epsilon^{ts_i} \quad (2)$$

Tomando el modelo general se puede establecer que un aumento en un 1 % del  $MP_{2,5}$  se relaciona con un aumento de 0,04 % en el número de atenciones de urgencia, mientras que un aumento en un 1 % del  $MP_{10}$  se relaciona con un aumento en un 0,05 % en el número de atenciones. Los resultados muestran que los contaminantes bajo estudio no tienen efectos significativos lineales sobre las atenciones de urgencia totales.

Para el caso del ozono ( $O_3$ ), como se muestra en la sección 3.2, su comportamiento es contracíclico respecto a las atenciones de urgencia respiratorias, lo que explica que muestre ser significativo con signo negativo. Sin embargo la significancia que muestra se puede explicar de acuerdo con su comportamiento estacional y no necesariamente indica una relación causal con las atenciones de urgencia respiratorias.

Al agregar los rezagos, con los niveles de contaminantes en la primera, segunda y tercera semana anterior al registro, tampoco se encuentran resultados significativos de los contaminantes sobre las atenciones de urgencia. El contaminante  $MP_{10}$  en el modelo dado por (2) muestra ser significativo y, de acuerdo con la revisión de literatura, podría significar una relación causal positiva donde un aumento de un 1 % de  $MP_{10}$  se relaciona con un aumento

Tabla 3: Resultado modelos generales

	<i>Variable Dependiente AU<sup>ts<sub>e</sub></sup></i>	
	Modelo general (1)	Modelo general con rezago (2)
$MP_{2,5}$	0.040	-0.079
$MP_{10}$	0.054	0.136*
$O_3$	-0.152***	0.005
$MP_{2,5}^1$	-	0.025
$MP_{2,5}^2$	-	0.025
$MP_{2,5}^3$	-	0.033
$MP_{10}^1$	-	-0.024
$MP_{10}^2$	-	-0.040
$MP_{10}^3$	-	-0.066
$O_3^1$	-	-0.043
$O_3^2$	-	-0.025
$O_3^3$	-	-0.035
EF - Estación	Sí	Sí
EF - Semana	Sí	Sí
Semanas	417	414
R <sup>2</sup>	0.794	0.848
R <sup>2</sup> ajustado	0.763	0.821
Residual Std. Error	0.153 (df = 361)	0.147 (df = 349)

*Códigos de significancia:* \* p < 0.1; \*\* p < 0.05; \*\*\* p < 0.01

del 0,14% en las atenciones respiratorias. En los modelos generales obtenidos, los factores estacionales son los que muestran mayor significancia en relación con la cantidad de atenciones respiratorias de urgencia. Adicionalmente se realizó un análisis general utilizando la ecuación (1) para cada estación del año  $e$  donde ninguno de los contaminantes bajo estudio, de manera aislada, mostró tener efectos significativos sobre las atenciones de urgencia.

Estos resultados llevaron a la generación de modelos por tipo de enfermedad respiratoria para identificar de manera aislada si alguna de las enfermedades respiratorias se ve afectada por los contaminantes bajo estudio en cada unas de las estaciones de acuerdo a la siguiente fórmula:

$$AU_{re}^{ts} = EF^s + \beta_1 \times MP_{2,5}^{ts} + \beta_2 \times MP_{10}^{ts} + \beta_3 \times O_3^{ts} + \epsilon^{ts} \quad (3)$$

En este último modelo el signo de los valores estimados podría indicar que los contaminantes están relacionados con ciertas condiciones ambientales específicas que producen su aumento o disminución. Para las IRA Alta se identifica que en las estaciones de verano y otoño existe relación significativa respecto a los contaminantes  $MP_{10}$  y  $O_3$  y en invierno con los contaminantes  $MP_{2,5}$  y  $MP_{10}$ . Para la influenza sólo una relación significativa en otoño con los contaminantes  $MP_{10}$  y  $O_3$ . Para la neumonía no se encuentra relación significativa con ninguno de los contaminantes y en ninguna de las épocas. La crisis obstructiva bronquial muestra relación significativa en el periodo de otoño con los tres contaminantes,  $MP_{2,5}$ ,  $MP_{10}$  y  $O_3$  y en invierno con  $MP_{2,5}$  y  $MP_{10}$  y en primavera con el  $O_3$ . Finalmente, otras causas respiratorias muestran relación significativa en el periodo de verano sólo con  $MP_{10}$ , en otoño con los tres contaminantes y en primavera con  $O_3$ . Los resultados de los modelos estimados se muestran en la Tabla 4.

Tabla 4: Modelos por tipo de urgencia

	Estación del año			
	Verano	Otoño	Invierno	Primavera
<i>3.1 IRA Alta</i>				
$MP_{2,5}$	-0.074	-0.057	0.293*	0.153
$MP_{10}$	0.287**	0.267**	-0.331*	-0.028
$O_3$	-0.277**	-0.214***	-0.114	-0.078
$R^2$	0.627	0.380	0.310	0.378
$R^2$ ajustado	0.560	0.274	0.193	0.272
<i>3.2 Influenza</i>				
$MP_{2,5}$	-0.061	-0.469	-0.626	-0.296
$MP_{10}$	0.585	0.521*	-0.170	0.275
$O_3$	-0.565	-0.314*	-0.299	0.457
$R^2$	0.294	0.467	0.097	0.480
$R^2$ ajustado	0.165	0.376	-0.057	0.391
<i>3.3 Neumonía</i>				
$MP_{2,5}$	0.038	-0.006	0.091	0.040
$MP_{10}$	-0.001	0.161	-0.228	0.033
$O_3$	-0.061	-0.065	-0.116	0.087
$R^2$	0.669	0.761	0.299	0.746
$R^2$ ajustado	0.609	0.721	0.180	0.702
<i>3.4 Crisis obstructiva bronquial</i>				
$MP_{2,5}$	0.039	-0.369**	-0.563***	-0.251
$MP_{10}$	-0.004	0.467***	0.429**	0.215
$O_3$	-0.096	0.255***	0.133	0.420**
$R^2$	0.455	0.642	0.389	0.368
$R^2$ ajustado	0.355	0.581	0.285	0.260
<i>3.5 Otra causa respiratoria</i>				
$MP_{2,5}$	0.158	-0.201**	-0.017	-0.057
$MP_{10}$	-0.386**	0.261**	-0.172	0.102
$O_3$	0.237	0.203***	0.123	0.303**
$R^2$	0.339	0.496	0.198	0.316
$R^2$ ajustado	0.219	0.410	0.061	0.199
<i>Sig. códigos:</i>	* $p < 0.1$ ; ** $p < 0.05$ ; *** $p < 0.01$			

---

## 5. Discusión

---

Los resultados muestran una relación directa entre el material particulado y el número de atenciones de urgencia, sin embargo, los coeficientes estimados no son estadísticamente significativos. En el caso del ozono, se puede ver una correlación negativa estadísticamente significativa, que puede ser explicado por su comportamiento contracíclico y no necesariamente indica causalidad.

Al desagregar el análisis por tipo de enfermedad respiratoria y por estación del año (ver ecuación 3), los resultados también tienen un comportamiento mixtos respecto al número de enfermedades respiratorias por cuadros de urgencia.

Los resultados, si bien muestran un buen ajuste ( $R^2$ ) con las variables seleccionadas, al no ser del todo concluyentes, nos llevan a discutir la inclusión de variables complementarias que permitan elaborar un nuevo modelo. Además, el haber realizado los modelos incluyendo variables contracíclicas, como el ozono, podrían ser la causa de los resultados mixtos obtenidos. La hipótesis inicial tenía como premisa que los ciclos podrían aislarse agregando las semanas estadísticas y las estaciones del año y, por el hecho de estar presente todo el año, era importante mantener dentro del mismo modelo todos los contaminantes independiente de la estacionalidad. Sin embargo, no se obtuvieron resultados que permitiesen comprender cabalmente el fenómeno de la contaminación del aire con un modelo lineal general.

Para identificar el efecto causal de cada contaminante y/o cuantificar su contribución al número de atenciones de urgencia, pareciera ser necesario incluir otras variables que permitan aislar el efecto individual o, de manera alternativa, realizar análisis de acuerdo a cada ciclo estacional o a cada patología de manera aislada. La alta correlación entre estaciones y semanas estadísticas de cada unas de las variables bajo estudio dificulta la posibilidad en gran medida la posibilidad de capturar el efecto causal, lo que genera que finalmente nos encontremos frente a resultados mixtos respecto al número de enfermedades respiratorias por cuadros de urgencia.

Adicionalmente, sería de gran interés comparar estos resultados con otro tipo de modelos, agregar nuevas variables y, convenientemente, incluir un panel con una mayor cantidad de datos climáticos y mayor detalle temporal, que permita identificar con exactitud el efecto causal de cada contaminante en las enfermedades estudiadas.

---

## 6. Conclusiones y trabajo futuro

---

Este trabajo tuvo como finalidad medir el impacto de las concentraciones de ozono, material particulado  $MP_{10}$  y  $MP_{2,5}$  en las atenciones de urgencia del sector público de la Región Metropolitana teniendo como hipótesis un efecto causal entre los contaminantes del aire y el número de atenciones de urgencia respiratoria en la Región Metropolitana. El análisis realizado muestra que los resultados no son lineales con respecto a los cambios semanales de la cantidad de partículas contaminantes del aire y que cada tipo de partícula contaminante afecta de manera distinta a los distintos tipos de enfermedades respiratorias. Además hay una alta correlación entre los distintos tipos de partículas contaminantes y las estaciones del año en la Región Metropolitana.

Los resultados mostraron tener un componente principalmente estacional tanto para la variable dependiente como para las variables independientes. Además, la Región Metropolitana tiene asociadas características geográficas que acentúan los efectos estacionales vinculados a la concentración de las partículas contaminantes. Por ejemplo, la concentración del material particulado en invierno producto de las bajas temperaturas, mala ventilación y fenómeno de inversión térmica [24]. Con ello el número de atenciones de urgencia respiratoria podría estar principalmente relacionado con variables climáticas, entre las cuales encontramos el nivel de precipitaciones, la temperatura, viento y humedad ambiental. Para generar modelos que efectivamente sean capaces de capturar el impacto de los contaminantes del aire en el estado de salud de la población, es necesario contar con datos que incluyan estas variables climáticas y así ampliar el estudio, por tanto, y tomando en cuenta los resultados obtenidos, queda esto como desafío futuro para los investigadores.

Dado lo anterior, es importante señalar que la obtención e integración de datos para realizar análisis cuantitativos más precisos es un desafío para los investigadores puesto que actualmente no existe información pública en Chile que permita construir un conjunto de datos con mayor detalle.

Otras variables de interés a considerar, que puedan incidir en la salud de las personas y, por consecuencia, en el número de atenciones de urgencia, son el número de vehículos en circulación, el número de habitantes, nivel de hacinamiento por zona, la geografía del lugar en estudio y la pertenencia a grupos de riesgo. Como las atenciones de urgencia tienen un comportamiento principalmente relacionado con enfermedades agudas, el análisis podría mejorar considerablemente si la información estuviese disponible con el detalle diario

de cada una de las variables.

Por otro lado, en la Región Metropolitana la contaminación atmosférica varía según sea la hora del día, por lo cual sería interesante conocer la dinámica de la contaminación diaria y replicar el estudio para distintas horas, con el fin de determinar zonas más vulnerables. Por último, se deberían considerar los grupos de riesgo, enfermos crónicos y personas de la tercera edad, así como las campañas de vacunación preventivas, ya que todo ello tiene relación con la incidencia de los distintos tipos de enfermedad respiratoria. Del mismo modo, se podría estudiar el efecto de las variables consideradas y propuestas sobre las atenciones de urgencia relacionadas a otro tipo de enfermedades no necesariamente respiratorias, pues como la revisión bibliográfica evidenció, los contaminantes también influyen en enfermedades de otra naturaleza como las cardiológicas.

Finalmente, es importante recalcar la importancia de las políticas públicas en prevención y control del nivel de contaminantes del aire para mejorar el funcionamiento del sistema de salud. Las normas, los planes de prevención y descontaminación generan beneficios significativos para la salud pública tanto económicos como sociales. En la actualidad, Chile cuenta con instrumentos de comando y control que ayudan a mejorar la calidad del aire, entregando solución a los principales contaminantes del aire como la combustión a leña, la actividad industrial,  $MP_{2,5}$ ,  $SO_2$  y  $NO_x$ . Es necesario fortalecer el vínculo de las políticas de salud con las políticas de medio ambiente tanto para mejoras en la gestión de establecimientos de salud como para asegurar un cuidado integral y mejor calidad de vida de los ciudadanos.

**Agradecimientos:** Los autores agradecen al CRS Hospital Provincia Cordillera, por su apoyo en el desarrollo de la investigación. Los resultados preliminares de este trabajo fueron presentados en las XXXIII Jornadas Chilenas de Salud Pública.

## Referencias

- [1] V. H. Borja-Aburto, D. P. Loomis, S. I. Bangdiwala, C. M. Shy, y R. A. Rascon-Pacheco. Ozone, suspended particulates, and daily mortality in Mexico City. *American journal of epidemiology*, 145(3):258–268, 1997.
- [2] R. D. Brook, S. Rajagopalan, C. A. Pope, J. R. Brook, A. Bhatnagar, A. V. Diez-Roux, F. Holguin, Y. Hong, R. V. Luepker, M. A. Mittleman, et al. Particulate matter air pollution and cardiovascular disease: an

- update to the scientific statement from the american heart association. *Circulation*, 121(21):2331–2378, 2010.
- [3] B. Brunekreef y S. T. Holgate. Air pollution and health. *The lancet*, 360(9341):1233–1242, 2002.
- [4] P. Castro, J. Vera, L. Cifuentes, G. Wellenius, H. Verdejo, L. Sepúlveda, J. L. Vukasovic, y S. Llevaneras. Polución por material particulado fino (pm 2, 5) incrementa las hospitalizaciones por insuficiencia cardiaca. *Revista chilena de cardiología*, 29(3):306–314, 2010.
- [5] A. Clifford, L. Lang, R. Chen, K. J. Anstey, y A. Seaton. Exposure to air pollution and cognitive functioning across the life course—a systematic literature review. *Environmental research*, 147:383–398, 2016.
- [6] E. Fuertes, M. Standl, J. Forns, D. Berdel, J. Garcia-Aymerich, I. Markevych, G. Schulte-Koerne, D. Sugiri, T. Schikowski, C. M. Tiesler, et al. Traffic-related air pollution and hyperactivity/inattention, dyslexia and dyscalculia in adolescents of the german ginipus and lisaplu birth cohorts. *Environment international*, 97:85–92, 2016.
- [7] S. Ha, R. Sundaram, G. M. B. Louis, C. Nobles, I. Seeni, S. Sherman, y P. Mendola. Ambient air pollution and the risk of pregnancy loss: a prospective cohort study. *Fertility and sterility*, 109(1):148–153, 2018.
- [8] J. Huang, X. Pan, X. Guo, y G. Li. Impacts of air pollution wave on years of life lost: A crucial way to communicate the health risks of air pollution to the public. *Environment international*, 113:42–49, 2018. <https://www.sciencedirect.com/science/article/pii/S0160412017>.
- [9] M. Ilabaca, I. Olaeta, E. Campos, J. Villaire, M. M. Tellez-Rojo, y I. Romieu. Association between levels of fine particulate and emergency visits for pneumonia and other respiratory illnesses among children in Santiago, Chile. *Journal of the Air & Waste Management Association*, 49(9):154–163, 1999.
- [10] O. K. Kurt, J. Zhang, y K. E. Pinkerton. Pulmonary health effects of air pollution. *Current opinion in pulmonary medicine*, 22(2):138, 2016.
- [11] R. Lafuente, N. García-Blàquez, B. Jacquemin, y M. A. Checa. Outdoor air pollution and sperm quality. *Fertility and sterility*, 106(4):880–896, 2016.
- [12] F. Lipfert y T. Hammerstrom. Temporal patterns in air pollution and hospital admissions. *Environmental research*, 59(2):374–399, 1992.

- [13] K. J. Matus. *Health impacts from urban air pollution in China: the burden to the economy and the benefits of policy*. PhD thesis, Massachusetts Institute of Technology, 2005.
- [14] MINSAL. Sustentabilidad financiera del sistema público de salud: Estableciendo las bases de un nuevo sistema de financiamiento y de gestión. departamento de asignación de recursos división de presupuesto subsecretaría de redes asistenciales. <http://www.minsal.cl/wp-content/uploads/2018/03/Sustentabilidad-Financiera-del-Sistem>, feb 2018. En línea; Último acceso Mayo 2010.
- [15] M. J. Nieuwenhuijsen, X. Basagaña, P. Dadvand, D. Martínez, M. Ci-rach, R. Beelen, y B. Jacquemin. Air pollution and human fertility rates. *Environment international*, 70:9–14, 2014.
- [16] W. H. Organization. Contaminación del aire de interiores y salud nota descriptiva n°292 febrero de 2016. *WHO Media centre*, 2016.
- [17] W. H. Organization. Ambient (outdoor) air quality and health. [http://www.who.int/en/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](http://www.who.int/en/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health), may 2018. En línea; Último acceso Abril 2018.
- [18] W. H. Organization. *World Health Statistics 2018: Monitoring health for the SDGs*. World Health Organization, 2018.
- [19] W. H. Organization et al. Guías de calidad del aire de la OMS relativas al material particulado, el ozono, el dióxido de nitrógeno y el dióxido de azufre: actualización mundial 2005. 2006.
- [20] B. Ostro, J. M. Sanchez, C. Aranda, y G. S. Eskeland. Air pollution and mortality: results from a study of santiago, chile. *Journal of exposure analysis and environmental epidemiology*, 6(1):97–114, 1996.
- [21] C. A. Pope, M. C. Turner, R. Burnett, M. Jerrett, S. M. Gapstur, W. R. Diver, D. Krewski, y R. D. Brook. Relationships between fine particulate air pollution, cardiometabolic disorders and cardiovascular mortality. *Circulation research*, páginas 108–115, 2014.
- [22] C. Pope III y D. Dockery. Health effects of fine particulate air pollution: lines that connect. *Journal of the air & waste management association*, 56(6):709–742, 2006.

- [23] M. Power, S. Adar, J. Yanosky, y J. Weuve. Exposure to air pollution as a potential contributor to cognitive function, cognitive decline, brain imaging, and dementia: a systematic review of epidemiologic research. *Neurotoxicology*, 56:235–253, 2016.
- [24] H. Romero, F. Irrázaval, D. Opazo, M. Salgado, y P. Smith. Climas urbanos y contaminación atmosférica en santiago de chile. *EURE (Santiago)*, 36(109):35–62, 2010.
- [25] C. Shy. Epidemiologic evidence and the united states air quality standards. *American journal of epidemiology*, 110(6):661–671, 1979.
- [26] K. R. Smith, C. F. Corvalán, T. Kjellstrom, et al. How much global ill health is attributable to environmental factors? *Epidemiology-Baltimore*, 10(5):573–584, 1999.



---

# ASIGNACIÓN DE ÁRBITROS PARA LAS LIGAS PROFESIONALES DE BÁSQUET DE ARGENTINA MEDIANTE INVESTIGACIÓN DE OPERACIONES

---

GUILLERMO DURÁN \*

MARIO GUAJARDO \*\*

FACUNDO GUTIÉRREZ \*\*\*

## Resumen

A partir de la temporada 2014-2015, las ligas profesionales de básquet de Argentina transformaron su programación de un formato en que los partidos se jugaban solamente durante los fines de semana a uno similar NBA, en que se disputan partidos a lo largo de toda la semana. Esta transformación ha generado nuevos desafíos para los organizadores de la liga. Uno de estos desafíos es la asignación de árbitros a partidos. En este trabajo, abordamos este problema de asignación, mediante una herramienta basada en programación lineal entera. El objetivo es minimizar el costo total de los viajes realizados por los árbitros y al mismo tiempo satisfacer una serie de condiciones. La resolución del problema se realiza en una serie de etapas, resolviendo subproblemas más pequeños, cuyas soluciones forman parte de la entrada en siguientes etapas. En las comparaciones realizadas con la fase regional de la temporada 2015-2016, donde la asignación de árbitros fue realizada de forma manual, el uso de la técnica de resolución presentada reporta una reducción superior al 25 % en los costos de los viajes efectuados. Posteriormente, los resultados de este trabajo han sido utilizados para la asignación de árbitros de la Primera División durante las temporadas 2016-2017 y 2017-2018. Además,

---

\* Departamento de Ingeniería Industrial, Universidad de Chile. Departamento de Matemática e Instituto de Cálculo, FCEyN, Universidad de Buenos Aires and CONICET, Argentina

\*\* Department of Business and Management Science, NHH Norwegian School of Economics, Bergen, Norway

\*\*\* Departamento de Matemática e Instituto de Cálculo, FCEyN, Universidad de Buenos Aires

en la última temporada nuestro enfoque también ha sido adoptado por la Segunda División.

**Palabras Clave:** Asignación de Árbitros, Básquet, Programación Lineal Entera, Traveling Umpire Problem, Sports Scheduling

---

## 1. Introducción

---

El básquet es uno de los deportes más populares de la Argentina. La creación de la Liga Nacional en 1985 significó un gran impulso para este deporte en todo el país. Hoy la Liga Nacional de Básquet (LNB), la primera división del básquet de la Argentina, es altamente profesional y está muy expandida a lo largo y a lo ancho de todo el país. Prueba de ello es que los 20 equipos de la temporada actual (2017-2018), provienen de 10 provincias distintas. Esto conlleva a que las distancias a recorrer por los equipos sean potencialmente muy extensas y, por lo tanto, es importante considerarlas al realizar la programación de partidos. En la Figura 1 se muestra la ubicación geográfica de los 20 equipos de la LNB.

Desde la temporada 2014-2015, la programación de los partidos es realizada mediante el uso de técnicas de Investigación Operativa [5]. Los equipos eligen un conjunto de *giras deseables* de modo de disputar partidos de manera consecutiva al jugar de visitante. Por ejemplo, un equipo de Formosa (ciudad situada a 1200 km al norte de Buenos Aires) podría elegir como una gira deseable viajar a Buenos Aires y jugar de visitante contra tres equipos allí. En lo que sigue del trabajo, se asume que la programación de los partidos ya fue realizada y es un dato de entrada para el modelo. Notemos que al hacer esto, se está priorizando optimizar los viajes de los equipos por el del cuerpo arbitral, pues no se tienen en cuenta los viajes de los árbitros a la hora de realizar la programación de los partidos. Esto se justifica dado que la cantidad de personas que deben movilizarse cuando viajan los integrantes del plantel de un equipo, es sustancialmente mayor que la cantidad total de árbitros.

La programación de torneos ha generado una corriente de literatura que ha proliferado intensamente en los últimos años, conocida como *sports scheduling* (ver por ejemplo, revisiones literarias en [9] y [10]). Gran parte de la atención de la literatura se ha enfocado en aplicaciones a la programación de partidos de un torneo. Por ejemplo, en Sudamérica se ha reportado implementaciones



Figura 1: Mapa de Argentina con los equipos de La Liga Nacional, temporada 2017-2018.

para la programación de partidos en ligas de fútbol [2, 6, 8, 7, 11, 12], básquet [5] y vóley [3]. En contraste, hay pocos trabajos enfocados en la programación de árbitros. Aparte de excepcionales aplicaciones que incluyen torneos de cricket en Inglaterra [14], béisbol en Estados Unidos [13], y fútbol en Chile [1], otras publicaciones se han limitado a abordar aspectos metodológicos y experimentos computacionales. En [13], particularmente, se formuló un problema conocido como *Traveling Umpire Problem* (TUP). Este problema busca capturar los aspectos más esenciales que aparecen al realizar una asignación arbitral eficiente. Formalmente, el problema está formulado de la siguiente manera: Dado un conjunto de  $n$  árbitros y  $2n$  equipos, los equipos juegan un *double-round robin* compacto, lo cual da un total de  $4n - 2$  fechas. Se busca en cada fecha asignar un árbitro a cada partido (notar que todos los árbitros son necesarios en cada fecha). Además hay cinco restricciones que definen el problema.

1. Cada juego debe ser asignado a un árbitro.

A su vez cada árbitro debe:

2. Dirigir un partido por fecha
3. Visitar todas las localías al menos una vez a lo largo del campeonato
4. Visitar a lo sumo una vez una localía en toda ventana de  $n - d_1$  fechas consecutivas
5. Dirigir a lo sumo una vez a un equipo en  $\lfloor \frac{n}{2} \rfloor - d_2$  fechas consecutivas.

Un resultado derivado en [4] muestra que el *Traveling Umpire Problem* resulta NP-completo para  $0 \leq d_1 \leq \lfloor \frac{n}{2} \rfloor$  y  $d_2 = \lfloor \frac{n}{2} \rfloor - 1$ .

Si bien el TUP captura aspectos generales del problema de asignación de árbitros a partidos, las aplicaciones en la práctica incorporan otras condiciones ad hoc a las particularidades de cada liga. Nuestro trabajo se enfoca en el modelamiento y resolución del problema de asignación de árbitros a la Liga Nacional de Básquet de Argentina. Los resultados de este trabajo han sido adoptados en las dos temporadas más recientes de esta liga, además de la última temporada de la Segunda División, conocida como Liga Argentina. El resto del artículo se organiza de la siguiente forma. En la Sección 2, presentamos una descripción del problema a resolver, incluyendo detalles sobre el formato de los torneos del básquet profesional argentino. En la Sección 3 formulamos un modelo de programación lineal entera para el problema y un enfoque de resolución diseñado para la generación de soluciones para su adopción en la práctica. En la Sección 4 analizamos los resultados obtenidos en comparación a la solución precedente. La Sección 5 concluye el artículo y delinea alternativas para trabajo futuro.

---

## 2. Descripción del problema

---

A partir de la temporada 2015-2016, un total de 20 equipos conforman la Liga Nacional, divididos en dos conferencias (norte y sur) de 10 equipos cada una. En esa temporada la serie regular se disputó en dos fases (regional y nacional), ambas de todos contra todos ida y vuelta. En la fase regional sólo

los equipos de una misma conferencia disputaron partidos entre sí, mientras que en la fase nacional juegan todos contra todos. En total, el torneo entonces cuenta en su serie regular con 560 partidos (180 en la fase regional y 380 en la fase nacional). Al finalizar la serie regular, los dos primeros de cada conferencia clasificaron directamente a los *play-off* de semifinales, mientras que del tercero al sexto se enfrentaron en los *play-off* de cuartos de final al mejor de cinco partidos. Además, los últimos de cada conferencia disputaron la serie de permanencia entre sí. La temporada 2016-2017 se disputó de manera similar, y fue la primera en la que la herramienta desarrollada en este artículo fue utilizada para la asignación de árbitros.

Cada partido de la serie regular fue arbitrado por 2 árbitros, y los partidos de *play-off* fueron dirigidos por 3 árbitros. En ese entonces el plantel arbitral para los partidos de la Liga estaba conformado por 20 árbitros, 10 de categoría A (la principal) y 10 de categoría A1. Para cada árbitro, se conoce la ciudad de residencia de cada árbitro, dato necesario a la hora de calcular los costos de los viajes que realizan.

El formato en la temporada 2017-2018 para la Primera División varió levemente disminuyendo el número de partidos global en alrededor de 50. Además, a partir de esta temporada, la herramienta fue utilizada para la asignación de los partidos de la Segunda División, sumando consecuentemente al plantel arbitral a 26 árbitros nuevos divididos en categorías A2 y A3. Por disposición de la Asociación que organiza los torneos (AdC), en la primera parte de la competencia los jueces de categoría A y A1 no estaban habilitados para dirigir a los equipos de la Liga Argentina. Como veremos, esto reduce el espacio de búsqueda factible, lo que ayuda a mejorar los tiempos de corrida del modelo. Para la segunda parte de la competencia, los árbitros de categoría A y A1 fueron habilitados a dirigir también partidos de la Liga Argentina.

A partir de 2017-2018, el arbitraje pasó a ser de tres árbitros para la Liga Nacional, aunque siguió siendo de 2 árbitros para la Liga Argentina. Salvo una pequeña cantidad de excepciones a pedido del comisionado técnico, la terna arbitral en los partidos de la Liga debía ser conformada por un árbitro de categoría A, uno de categoría A1, y otro de categoría A2 o A3 que actúa como tercer árbitro.

El objetivo es el de asignar a cada partido una dupla o terna arbitral (según corresponda), de forma de satisfacer todas las restricciones impuestas por el comisionado técnico de la AdC, minimizando los costos de viajes y hoteles incurridos por los árbitros. Estas restricciones están relacionadas con las mencionadas en el *Traveling Umpire Problem*. Por ejemplo, deben pasar tres

partidos intermedios de un equipo para que un árbitro pueda estar habilitado a volver a dirigirlo, motivado por razones deportivas.

El problema en la práctica está sujeto a cambios durante la temporada. Por ejemplo, un árbitro puede lesionarse, o ser seleccionado para dirigir en una competencia internacional, inhabilitándolo durante ese período para dirigir. También pueden haber cambios de categorías dentro del plantel arbitral, así como también nuevos árbitros deben poder ser fácilmente incorporados. Por ejemplo, esto ocurrió cuando por primera vez en la historia de la Liga Nacional, se incluyó en tres partidos distintos a tres mujeres dentro de la terna arbitral. También pueden haber reprogramaciones en la programación de los partidos, cambiando el día en que originalmente se habían programado.

Considerando la gran cantidad de partidos y múltiples condiciones, confeccionar una asignación arbitral para la temporada entera simultáneamente es un problema demasiado grande. Además, si la solución fuese computada para toda la temporada en su inicio, no será necesariamente implementable, pues durante el transcurso del campeonato suelen ocurrir eventualidades como las previamente mencionadas, que hacen que se tenga que cambiar la solución propuesta. Por eso, para resolver el problema procederemos primero a formular un modelo y luego optaremos por resolverlo secuencialmente para ventanas de tiempo limitadas (usualmente de dos semanas). De esta forma, el modelo puede ajustarse durante cada período según las eventualidades que hayan ocurrido. Una vez que se tiene la corrida para una ventana de tiempo anterior, estos datos pasan a ser utilizados como dato de entrada para la corrida en la siguiente ventana de tiempo.

---

### 3. Resolución del problema

---

En esta sección vamos a presentar el enfoque que se utiliza para llegar a la formulación del modelo y su consecuente resolución. Primeramente es necesario notar que las fechas y las localías de los partidos ya están definidas de antemano, y es una parte clave de la entrada del modelo. Además de las distancias que deben viajar los árbitros, es necesario conocer entre qué pares de sedes un árbitro puede dirigir en días consecutivos (con un día intermedio libre se considera posible para un árbitro dirigir entre cualquier par de sedes, posiblemente incurriendo en un gasto relacionado con la estadía en un hotel). Recordemos que en el básquet generalmente hay partidos todos los días de la

semana durante la fase regular (y no sólo los fines de semana como en otros deportes).

La idea central en la que ronda el modelo es en hacer un seguimiento de la ubicación de cada árbitro a lo largo del tiempo. Si en una fecha en particular hay un partido que debe ser dirigido por dos árbitros que enfrenta a  $A$  (local) contra  $B$  (visitante), el modelo deberá conseguir que dos árbitros del plantel estén en la ubicación del equipo  $A$  en esa fecha (penalizando en función del costo de movilizar a cada uno de los árbitros desde su ubicación anterior a la localía de  $A$ ). Para denotar que un árbitro no está dirigiendo y se encuentra en su domicilio, agregamos un equipo ficticio. Para modelar esta situación lo que hacemos es añadir un partido en la localía de este equipo ficticio en todos los días y no acotamos superiormente la cantidad de árbitros que pueden dirigir en este partido. Por lo tanto lo que decide el modelo es en qué partidos dirige cada árbitro (tanto real como ficticio) a lo largo del tiempo. Además, como los árbitros deben comenzar y finalizar la temporada en sus domicilios, se agregan dos días extras al principio y al final de la temporada.

Los modelos fueron implementados en el lenguaje de modelado OPL y resueltos usando Cplex 12.6.3, en una computadora con 4 procesadores corriendo a 3.4GHz y con una memoria RAM de 16GB.

### 3.1. Modelo para la asignación de árbitros

Planteamos a continuación un modelo de programación lineal entera para decidir qué árbitro debe dirigir en cada partido. El modelo realiza la asignación considerando los partidos sólo hasta un cierto tiempo que llamamos  $t_{\text{ventana}}$ . La función objetivo es el costo total de los viajes que realizan todos los árbitros hasta  $t_{\text{ventana}}$ , incluyendo los costos asociados a estadías en hotel.

Los datos de entrada del modelo utilizados son los siguientes:

- Conjunto  $\mathcal{E}$  de los equipos. Los equipos reales tienen todos índices positivos, y utilizamos el cero como índice del equipo ficticio que representa el domicilio de los árbitros. De esta forma, nos queda que  $\mathcal{E} = \{0\} \cup \mathcal{E}_1 \cup \mathcal{E}_2$ , donde  $\mathcal{E}_1$  denota a los equipos de la Liga (primera división) y  $\mathcal{E}_2$  denota a los equipos de la Liga Argentina (segunda división).
- Conjunto  $\mathcal{P}$  de partidos. Cada elemento de  $\mathcal{P}$  es un partido y está dado por una tupla de la forma  $(t, k, l) = (\text{Día}, \text{Local}, \text{Visitante})$ , que denotan el día (contando desde el comienzo del campeonato) en el que se juega

un partido, el índice del equipo que es local y el índice del equipo que hace de visitante respectivamente. Es decir,  $t \in \mathbb{Z}_{\geq 0}$  y  $k, l \in \mathcal{E}$ .

- Conjunto  $\mathcal{A} = \mathcal{A}_I \overset{d}{\cup} \mathcal{A}_{II} \overset{d}{\cup} \mathcal{A}_{III} \overset{d}{\cup} \mathcal{A}_{IV}$ , de los árbitros separados por su correspondiente categoría. Además, se tiene un subconjunto de árbitros  $\widehat{\mathcal{A}} \subseteq \mathcal{A}_{III} \cup \mathcal{A}_{IV}$ , formado por aquellos que pueden actuar de tercer árbitro en los partidos de la Primera División.
- Conjunto  $\mathcal{Q}_{\text{local}} = \{(i, k) : i \in \mathcal{A} \text{ no puede dirigir a } k \in \mathcal{E} \text{ en condición de local}\}$ . En la mayoría de los casos esto se debe a que un árbitro y un equipo residen en el mismo lugar. A su vez, se tiene el conjunto  $\mathcal{Q}_{\text{visitante}}$  definido de manera análoga.
- Conjunto  $\mathcal{V}$  de los viajes posibles a realizar. Cada elemento de  $\mathcal{V}$  representa un viaje posible y está dado por una tupla de la forma  $(s, t, k, m, n)$ , lo cual denota un viaje que se comienza en  $t \in \mathbb{Z}_{\geq 0}$  desde la ubicación del equipo  $k \in \mathcal{E}$ , y finaliza en la localía del equipo  $m \in \mathcal{E}$  en el día  $t + s$  que disputa su partido contra el rival  $n \in \mathcal{E}$ .

Para generar este conjunto se toma cada partido  $(t, k, l) \in \mathcal{P}$  y para cada  $s \in \{1, 2\}$ , se toman todos los partidos de la forma  $(t + s, m, n) \in \mathcal{P}$ , y se agrega el viaje  $(s, t, k, m, n)$  a  $\mathcal{V}$ . De la misma forma, se genera  $\widehat{\mathcal{V}}$ , agregando  $(s, t, k, m)$  a  $\widehat{\mathcal{V}}$  (esta definición tendrá sentido más adelante al describir las variables utilizadas). Notar que será necesario sólo considerar los viajes que son realmente factibles en la realidad, dado que generando  $\mathcal{V}$  de esta forma, para  $s = 1$  se están generando muchos viajes que no tienen por qué ser posibles de realizar en días consecutivos.

- Conjunto  $\mathcal{C}_E = \{(k, m) : \text{es posible dirigir en } k \in \mathcal{E} \setminus \{0\} \text{ y en la sede de } m \in \mathcal{E} \setminus \{0\} \text{ en días consecutivos}\}$ . Análogamente se define  $\mathcal{C}_A = \{(i, m) : \text{es posible dirigir en el domicilio del árbitro } i \in \mathcal{A} \text{ y en la localía del equipo } m \in \mathcal{E} \setminus \{0\} \text{ en días consecutivos}\}$
- Parámetros  $d_{km}$  que denotan el costo en el que se incurre al realizar un viaje desde la ubicación del equipo  $k \in \mathcal{E} \setminus \{0\}$  y la del equipo  $m \in \mathcal{E} \setminus \{0\}$ . Análogamente se definen  $r_{ik}$  que denotan el costo que realiza un árbitro  $i \in \mathcal{A}$  desde su domicilio a la ubicación del equipo  $k \in \mathcal{E}$ .
- Parámetros  $z_i^A$  que denotan la zona de la residencia del árbitro  $i \in \mathcal{A}$ . Análogamente se definen  $z_k^E$ , que denotan la zona del equipo  $k \in \mathcal{E} \setminus \{0\}$ . Notar que equipos distintos pueden ser de la misma zona (por ejemplo, los equipos que hacen de local en la Ciudad de Buenos Aires).

- Parámetro  $h$  que denota el costo diario que se incurre por el uso de un hotel.

Para la formulación del modelo, introducimos dos tipos de variables binarias. Primeramente, tenemos la variable  $x_{ip}$  para cada árbitro  $i \in \mathcal{A}$  y para cada partido  $p \in \mathcal{P}$ . De esta forma, nos queda que  $x_{ip} = 1$  si y sólo si el árbitro  $i \in \mathcal{A}$  es asignado a dirigir el partido  $p \in \mathcal{P}$ . Además, para modelar correctamente los viajes que se realizan, introducimos la variable  $z_{iv}$  para cada  $i \in \mathcal{A}$  y  $v \in \widehat{\mathcal{V}}$ , donde  $z_{iv} = 1$  si y sólo si el árbitro  $i \in \mathcal{A}$  realiza un viaje de largo  $s \in \{1, 2\}$ , el día  $t \in \mathbb{Z}_{\geq 0}$  desde la localía del equipo  $k \in \mathcal{E}$  hacia la localía de  $m \in \mathcal{E}$  (recordar que  $v = (s, t, k, m)$ ).

Con estos elementos, podemos plantear el modelo de programación lineal entera que resuelva el problema presentado de la siguiente forma:

1. La función objetivo busca minimizar el costo total proveniente de los viajes realizados por el plantel arbitral. Se asume en lo que sigue que al viajar entre dos destinos el costo no depende del sentido del viaje. Además, si un árbitro viaja entre la sede de dos equipos que no son de la zona en la que vive el árbitro en cuestión, entonces se incurre en un gasto de estadía de hotel. En lo que sigue, es conveniente tener en cuenta que el índice cero en  $\mathcal{E}$  se utiliza para denotar que un árbitro está en su casa, por lo tanto  $k > 0$  hace referencia a un equipo real. Tengamos presente que  $v \in \widehat{\mathcal{V}}$  es de la forma  $v = (s, t, k, m)$ .

$$\begin{aligned}
 \min : & \quad \underbrace{\sum_{i \in \mathcal{A}} \sum_{\substack{v \in \widehat{\mathcal{V}} \\ k > 0 \\ m > 0}} z_{iv} \cdot d_{km}}_{\text{viajes entre ciudades de equipos}} + \underbrace{\sum_{i \in \mathcal{A}} \sum_{\substack{v \in \widehat{\mathcal{V}} \\ k > 0 \\ m = 0}} z_{iv} \cdot r_{ik}}_{\text{viajes que llegan a un domicilio}} + \quad (1) \\
 & \quad + \underbrace{\sum_{i \in \mathcal{A}} \sum_{\substack{v \in \widehat{\mathcal{V}} \\ k = 0 \\ m > 0}} z_{iv} \cdot r_{im}}_{\text{viajes que salen de un domicilio}} + \underbrace{\sum_{i \in \mathcal{A}} \sum_{\substack{v \in \widehat{\mathcal{V}} \\ k > 0 \\ m > 0 \\ z_i^A \neq z_k^E}} z_{iv} \cdot s \cdot h}_{\text{costo de hotel}}
 \end{aligned}$$

2. Durante un período podemos tener cotas inferiores de los partidos que puede dirigir un cierto árbitro. Recordemos también que  $p \in \mathcal{P}$  es de la forma  $p = (t, k, l)$ .

$$\sum_{\substack{p \in \mathcal{P}: \\ t \leq t_{\text{ventana}} \\ k > 0}} x_{ip} \geq \alpha_i \quad \forall i \in \mathcal{A} \quad (2)$$

3. Análogamente podemos tener cotas superiores para los partidos que dirige cada árbitro. En particular si un árbitro no puede dirigir durante un período (porque está lesionado por ejemplo), la cota superior será cero.

$$\sum_{\substack{p \in \mathcal{P}: \\ t \leq t_{\text{ventana}} \\ k > 0}} x_{ip} \leq \beta_i \quad \forall i \in \mathcal{A} \quad (3)$$

4. Para que un árbitro pueda volver a dirigir a un equipo deben pasar  $\gamma$  partidos intermedios del equipo donde ese árbitro no lo dirija (usualmente  $\gamma = 3$ ). Además, denotamos por  $\mathcal{P}_{k,q}$  a los partidos  $q, q + 1, \dots, q + \gamma$  inclusive del equipo  $k \in \mathcal{E}$  en orden cronológico. Llamamos  $T_k$  a la cantidad de partidos que juega el equipo  $k \in \mathcal{E}$  a lo largo del torneo.

$$\sum_{p \in \mathcal{P}_{k,q}} x_{ip} \leq 1 \quad \forall i \in \mathcal{A}, \forall k \in \mathcal{E} \setminus \{0\}, \forall q \in \{1, \dots, T_k - \gamma\} \quad (4)$$

5. De manera similar, para que un árbitro pueda volver a dirigir a un equipo en condición de local, debe pasar un partido intermedio del equipo jugando de local donde ese árbitro no lo dirija. Análogamente, llamando  $\mathcal{P}_{k,q}^{\text{local}}$  a los partidos de local del equipo  $k \in \mathcal{E}$  en orden cronológico, tenemos la siguiente restricción.

$$\sum_{p \in \mathcal{P}_{k,q}^{\text{local}}} x_{ip} \leq 1 \quad \forall i \in \mathcal{A}, \forall k \in \mathcal{E} \setminus \{0\}, \forall q \in \{1, \dots, T_k - \gamma_{\text{local}}\} \quad (5)$$

6. Por cuestiones de descanso, un árbitro  $i$  no puede estar más de  $\delta_i$  días consecutivos fuera de su domicilio, usualmente  $\delta_i = 14$  para todo  $i$ , pero puede variar dependiendo del árbitro. Sea también  $p_t^0 = (t, 0, 0) \in \mathcal{P}$ , que es el partido ficticio agregado en cada día calendario para modelar que un árbitro está en su domicilio.

$$\sum_{q \leq t \leq q + \delta_i - 1} x_{ip_t^0} \geq 1 \quad \forall i \in \mathcal{A}, \forall q \in \{1, \dots, t_{\text{ventana}} - \delta_i + 1\} \quad (6)$$

7. También relacionado con el descanso necesario, se impone un máximo de tres partidos dirigidos por un árbitro en una ventana de cinco días.

$$\sum_{\substack{p \in \mathcal{P}: \\ q \leq t \leq q + 4 \\ k > 0}} x_{ip} \leq 3 \quad \forall i \in \mathcal{A}, \forall q \in \{1, \dots, t_{\text{ventana}} - 4\} \quad (7)$$

8. Algunos árbitros tienen prohibido dirigir a ciertos equipos. Usualmente esto se debe a que un árbitro de una cierta zona no puede dirigir a los equipos que se encuentran en la misma zona (por ejemplo, un árbitro de la ciudad de Córdoba, no puede dirigir a Atenas o a Instituto).

$$\sum_{(t,k,l)=p \in \mathcal{P}} x_{ip} = 0 \quad \forall (i, k) \in \mathcal{Q}_{\text{local}} \quad (8)$$

Análogamente, en condición de visitante se tiene la siguiente restricción.

$$\sum_{(t,k,l)=p \in \mathcal{P}} x_{ip} = 0 \quad \forall (i, l) \in \mathcal{Q}_{\text{visitante}} \quad (9)$$

9. En los partidos de la Primera División, actualmente se arbitra con tres árbitros. La composición de la terna arbitral suele estar conformada por un árbitro de  $\mathcal{A}_I$ , uno de  $\mathcal{A}_{II}$  y otro de  $\widehat{\mathcal{A}}$ .

$$\sum_{i \in \mathcal{A}'} x_{ip} = 1 \quad \forall (t, k, l) = p \in \mathcal{P} : k \in \mathcal{E}_1, t \leq t_{\text{ventana}} \quad (10)$$

Tenemos tres restricciones de este tipo, según si  $\mathcal{A}' = \mathcal{A}_I, \mathcal{A}_{II}$  o  $\widehat{\mathcal{A}}$

10. En los partidos de la Segunda División, actualmente se arbitra con dos árbitros. La composición de la dupla arbitral suele estar conformada por un árbitro de  $\mathcal{A}_I \cup \mathcal{A}_{II} \cup \mathcal{A}_{III}$  y otro de categoría  $\mathcal{A}_{III} \cup \mathcal{A}_{IV}$ .

$$\sum_{i \in \mathcal{A}'} x_{ip} = 1 \quad \forall (t, k, l) = p \in \mathcal{P} : k \in \mathcal{E}_2, t \leq t_{\text{ventana}} \quad (11)$$

Tenemos dos restricciones de este tipo, según si  $\mathcal{A}' = \mathcal{A}_I \cup \mathcal{A}_{II} \cup \mathcal{A}_{III}$  o bien  $\mathcal{A}_{III} \cup \mathcal{A}_{IV}$

11. Ningún árbitro puede dirigir dos partidos en un mismo día.

$$\sum_{\substack{(t,k,l)=p \in \mathcal{P}: \\ t=q}} x_{ip} \leq 1 \quad \forall i \in \mathcal{A}, q \in \{1, \dots, t_{\text{ventana}}\} \quad (12)$$

12. De manera similar, ningún árbitro puede emprender más de un viaje en un mismo día.

$$\sum_{\substack{(s,t,k,m)=v \in \widehat{\mathcal{V}} \\ t=q}} z_{iv} \leq 1 \quad \forall i \in \mathcal{A}, q \in \{1, \dots, t_{\text{ventana}}\} \quad (13)$$

13. En caso de que un árbitro realice un viaje que involucre dos días, es decir, con un día intermedio, entonces no puede emprender un viaje en ese día intermedio.

$$\sum_{\substack{(s,t,k,m)=v \in \widehat{\mathcal{V}} \\ t=q \\ s=2}} z_{iv} + \sum_{\substack{(s,t,k,m)=v \in \widehat{\mathcal{V}} \\ t=q+1}} z_{iv} \leq 1 \quad \forall i \in \mathcal{A}, q \in \{1, \dots, t_{\text{ventana}}\} \quad (14)$$

14. Si un árbitro dirige los partidos  $p_1 = (t, k, l)$  y  $p_2 = (t+1, m, n)$  (es decir,  $x_{ip_1} = x_{ip_2} = 1$ ), queremos que la variable  $z_{i\hat{v}}$  con  $\hat{v} = (1, t, k, m)$  valga 1 para indicar que ese viaje fue realizado. Análogamente, buscamos que si  $z_{i\hat{v}} = 1$ , entonces valga que  $x_{ip_1} = x_{ip_2} = 1$ .

$$\begin{cases} x_{ip_1} + x_{ip_2} & \leq & 1 + z_{i\hat{v}} \\ 2 \cdot z_{i\hat{v}} & \leq & x_{ip_1} + x_{ip_2} \end{cases} \quad (15)$$

$$\forall i \in \mathcal{A}, \forall q \leq (t_{\text{ventana}} - 1), p_1 = (q, k, l) \in \mathcal{P}, v = (1, q, k, m, n) \in \mathcal{V}$$

Donde  $\hat{v} = (1, q, k, m)$  y  $p_2 = (q+1, m, n)$

15. Si un árbitro dirige los partidos  $p_1 = (t, k, l)$  y  $p_2 = (t+2, m, n)$  (es decir,  $x_{ip_1} = x_{ip_2} = 1$ ), queremos que la variable  $z_{i\hat{v}}$  con  $v = (2, t, k, m)$  valga 1 para indicar que ese viaje fue realizado, a no ser que se deba a que el árbitro dirigió en tres días consecutivos (para lo cual debe hacer un viaje de  $s = 1$  en el día intermedio).

$$\begin{cases} x_{ip_1} + x_{ip_2} & \leq & 1 + z_{i\hat{v}} + \sum_{\substack{(s,t,k,a)=\bar{v} \in \widehat{\mathcal{V}} \\ t=q \\ s=1}} z_{i\bar{v}} \\ 2 \cdot z_{i\hat{v}} & \leq & x_{ip_1} + x_{ip_2} \end{cases} \quad (16)$$

$$\forall i \in \mathcal{A}, \forall q \leq (t_{\text{ventana}} - 2), p_1 = (q, k, l) \in \mathcal{P}, v = (2, q, k, m, n) \in \mathcal{V}$$

Donde  $\hat{v} = (2, q, k, m)$  y  $p_2 = (q+2, m, n)$

16. Para poder seguir la ubicación de los árbitros, necesitamos que arbitren al menos un partido (ya sea real o ficticio) cada dos días.

$$\sum_{\substack{(t,k,l)=p \in \mathcal{P} \\ t=q}} x_{ip} + \sum_{\substack{(t,k,l)=p \in \mathcal{P} \\ t=q+1}} x_{ip} \geq 1 \quad \forall i \in \mathcal{A}, q \in \{1, \dots, t_{\text{ventana}} - 1\} \quad (17)$$

17. Debemos tener en cuenta en que si se realiza un viaje sin un día intermedio, entonces este viaje debe ser factible.

$$z_{i\hat{v}} = 0 \quad \forall i \in \mathcal{A}, \forall \hat{v} = (1, t, k, m) \in \hat{\mathcal{V}} : (k, m) \in \mathcal{C}_E \quad (18)$$

$$z_{i\hat{v}} = 0 \quad \forall i \in \mathcal{A}, \forall \hat{v} = (1, t, k, m) \in \hat{\mathcal{V}} : k = 0, (i, m) \in \mathcal{C}_A \quad (19)$$

$$z_{i\hat{v}} = 0 \quad \forall i \in \mathcal{A}, \forall \hat{v} = (1, t, k, m) \in \hat{\mathcal{V}} : m = 0, (i, k) \in \mathcal{C}_A \quad (20)$$

El esquema que representa el seguimiento de un árbitro en el tiempo puede verse en la Figura 2.

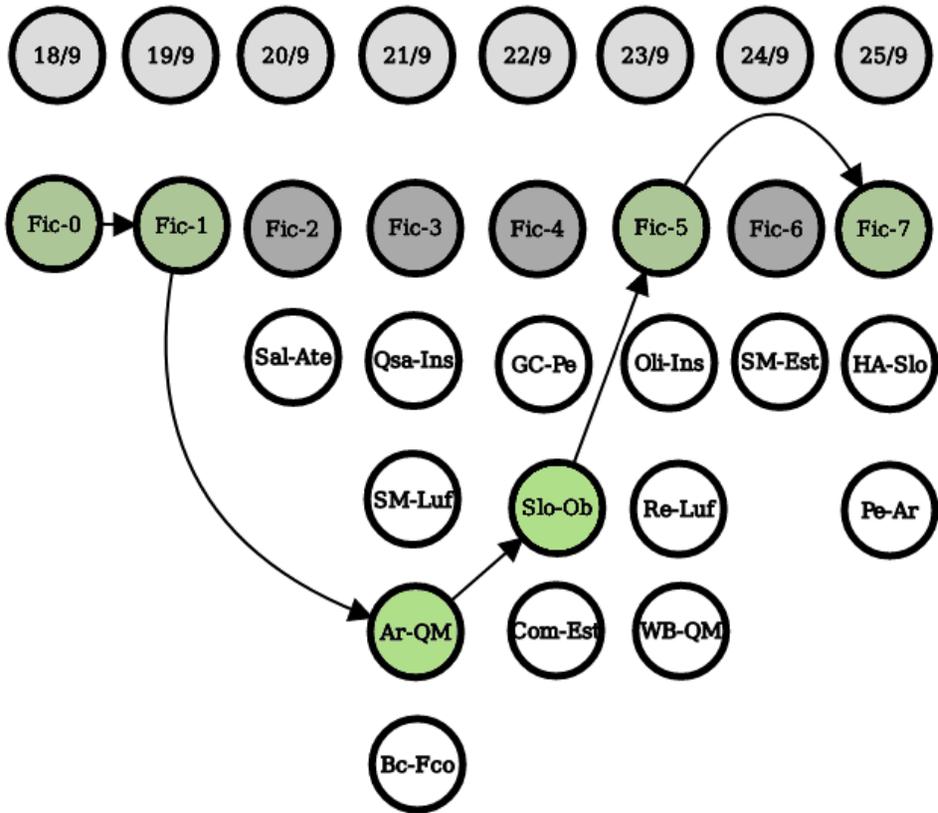


Figura 2: Representación visual del seguimiento de un árbitro en el tiempo

### 3.2. Resolución secuencial del modo por períodos

Para la obtención de una solución, consideramos horizontes de tiempo reducido y resolvemos los subproblemas correspondientes. La solución en la práctica es implementada en forma de horizonte rodante. A continuación explicamos este enfoque.

Al comienzo de cada período recibimos las indicaciones a tener en cuenta por parte del comisionado técnico de la AdC, así como también todo cambio que haya ocurrido respecto de lo que les presentamos en la corrida del período anterior. A continuación lo que se realiza es fijar la asignación recibida que finalmente se optó por utilizar en el período anterior (posiblemente relajando algunas restricciones en caso de que los cambios realizados por el comisionado técnico no cumplan las condiciones dadas).

Una vez recibidas las indicaciones para el período actual, hay que modelarlas debidamente. Por ejemplo, si se sabe que en el período en cuestión los árbitros de  $\mathcal{A}_I$  no pueden dirigir más de un partido en la Liga Argentina (segunda división) durante este período, entonces deberemos agregar restricciones del tipo:

$$\sum_{\substack{(t,k,l)=p \in \mathcal{P}: \\ t \leq t_{\text{ventana}} \\ k \in \mathcal{E}_2}} x_{ip} \leq 1 \quad \forall i \in \mathcal{A}_I \quad (21)$$

De manera similar debemos tratar toda restricción ad hoc que pueda ocurrir en las indicaciones recibidas. Junto con las indicaciones, tenemos el día en el que termina el período en cuestión, llamémoslo  $f$ . Luego, realizamos una asignación utilizando el modelo mencionado, tomando como  $t_{\text{ventana}} = f + \eta$ , es decir hasta el día  $f + \eta$ . En la realidad solemos tomar  $\eta = 5$ . La elección de este parámetro se explica a continuación.

#### Condición del “fin del mundo”

Como ya dijimos, no realizamos la asignación de la temporada completa porque ocurren a lo largo de la misma eventualidades (como reprogramaciones en el fixture o lesiones de árbitros), que harían que la asignación de árbitros realizada deba cambiar completamente. En el caso de las reprogramaciones que van apareciendo aún en nuestra asignación parcial, lo que se

hace es volver al último período que no contiene la reprogramación en cuestión y seguir el proceso de asignación del mismo modo desde allí, pero con los subsecuentes cambios en los conjuntos  $\mathcal{P}$  y  $\mathcal{V}$ .

Por lo tanto, al no asignar la temporada completa, se optó por resolver el problema sólo para un período más pequeño. De aquí resulta natural que si no se tiene algún cuidado, puedan existir problemas si simplemente se ubican las soluciones de dos períodos consecutivos una atrás de la otra, pues los problemas no son independientes (en particular, puede que sea necesario realizar viajes entre los últimos días de un período y los primeros del siguiente). Este fenómeno que ocurre al “pegar” soluciones óptimas es lo que llamamos condición del “fin del mundo”. El nombre refiere a que al resolver un período ignorando lo que ocurre en el siguiente, el modelo podría optar por decisiones que si bien para el período en cuestión pueden ser mejores, implican decisiones muy malas en los períodos siguientes (o incluso pueden resultar infactibles en los períodos que siguen), debido a que el modelo no está viendo más allá del fin del período en cuestión (como si el “mundo” terminara allí).

Teniendo esto en mente, el método de resolución por el que se opta es el de resolver el problema con el modelo hasta la fecha  $f + \eta$ , y luego presentar como solución final lo que arroja el modelo hasta la fecha  $f$ . Una vez tomada esta decisión, resta elegir adecuadamente el parámetro  $\eta$ . Para ello se optó por resolver los mismos 5 períodos para  $\eta \in \{1, 2, 3, 4, 5\}$ , donde el fixture de los cinco períodos corresponden a la Fase Regional de la temporada 2015-2016 (luego se repitió para la temporada 2016-2017).

Para  $\eta \in \{0, 1\}$ , en ninguno de los dos casos resulta factible resolver períodos intermedios culpa de las decisiones tomadas en períodos anteriores. Para  $\eta = 2$  resultó factible asignar en todos los períodos, aunque el valor de la función objetivo es considerablemente mayor que para  $\eta \in \{3, 4, 5\}$ . En este último caso, para  $\eta \in \{3, 4, 5\}$ , no hay diferencias considerables en cuanto al valor de la función objetivo al finalizar el quinto período, y cualquiera de las tres opciones creemos podría considerarse una elección sensata del parámetro  $\eta$ .

Las Figuras 3 y 4 muestran el esquema de solución para  $\eta = 0$  y  $\eta > 0$ , respectivamente.

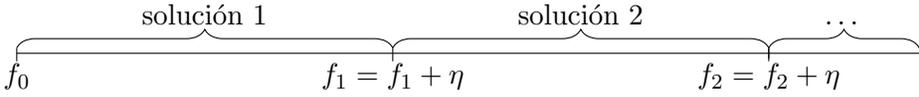


Figura 3: Solapamiento de soluciones con  $\eta = 0$

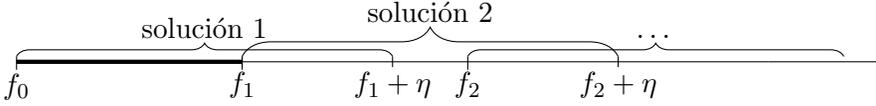


Figura 4: Solapamiento de soluciones con  $\eta > 0$

### Resolución de instancias grandes

Usualmente, los períodos a considerar son cercanos a 15 días (a veces unos días más, a veces unos días menos). Por lo tanto, utilizamos el modelo para resolver una instancia que ronda los 20 días (ya que tomamos  $\eta = 5$ ). Además de la cantidad de días a considerar, la cantidad de partidos en un período afecta considerablemente los tiempos de resolución.

En caso de que no se encuentre solución óptima al cabo de una hora, se parte el período en cuestión en subperíodos hasta que cada uno de estos se resuelva de forma óptima. Los subperíodos no son disjuntos sino que se solapan, más aún, se toma  $\eta = 3$  y se utiliza el mismo método que mencionamos antes para incrementalmente hallar una solución factible que corresponda a todo el período en cuestión.

Una vez que se tiene una solución factible de todo el período, se fijan las asignaciones realizadas que involucran a los árbitros de categorías  $\mathcal{A} \setminus \mathcal{A}_I$ . Finalmente, se busca la solución a este nuevo problema, que ahora involucra a todo el período de tiempo (pero no a todos los árbitros). Puede verse esto como una heurística de búsqueda local, donde el vecindario de una asignación, está compuesto por todas las asignaciones factibles que difieren de la asignación parcial en  $\mathcal{A}_I$  (en realidad, se relajan algunas asignaciones más donde un árbitro de  $\mathcal{A}_I$  podría officiar, por ejemplo en algunos partidos extra de la Liga Argentina).

Finalizada la corrida del modelo fijando los partidos de árbitros de  $\mathcal{A} \setminus \mathcal{A}_I$ , se toma esta nueva asignación y se fijan las asignaciones que involucran a árbitros de las categorías  $\mathcal{A} \setminus \mathcal{A}_{II}$ , y se procede análogamente. Una vez terminado este proceso con  $\mathcal{A}_{II}$ ,  $\mathcal{A}_{III} \cup \hat{\mathcal{A}}$ , y  $\mathcal{A}_{IV} \cup \hat{\mathcal{A}}$ , se da por terminada una iteración. Estas iteraciones se vuelven a repetir hasta que la variación en la solución final de

2 iteraciones consecutivas sea ínfima, y se da por terminada la solución del período con el resultado de la última iteración. En la práctica, este enfoque es razonable pues existe poca relación entre los problemas de las categorías  $\mathcal{A}_I, \mathcal{A}_{II}, \mathcal{A}_{III} \cup \widehat{\mathcal{A}}$ , y  $\mathcal{A}_{IV} \cup \widehat{\mathcal{A}}$ .

La Tabla 1 muestra la medición de tiempos para la resolución de una instancia conjunta de Primera y Segunda División, mientras que la Tabla 2 muestra lo mismo para una instancia sólo de Primera.

Tabla 1: Medición de tiempos para instancias comenzando el 1/2/2018 y finalizando el 14/2/2018. Se toman partidos de la Liga Nacional y la Liga Argentina. En cada columna se tiene el tiempo hasta alcanzar el valor óptimo para  $\mathcal{A}'$ , fijando los valores en  $\mathcal{A} \setminus \mathcal{A}'$ . Solución factible se encuentra en todos los casos en cuestión de segundos.

Días	Partidos	$\mathcal{A}_I$	$\mathcal{A}_{II}$	$\mathcal{A}_{III} \cup \widehat{\mathcal{A}}$	$\mathcal{A}_{IV} \cup \widehat{\mathcal{A}}$	$\mathcal{A}$
5	29	14 seg.	8 seg.	16 seg.	11 seg.	18 min.
6	35	18 seg.	53 seg.	30 seg.	29 seg.	> 1 h.
7	41	47 seg.	105 seg.	50 seg.	237 seg.	⋮
8	46	48 seg.	8 min.	104 seg.	22 min.	⋮
9	53	109 seg.	36 min.	13 min.	46 min.	⋮
10	56	11 min.	> 1 h.	30 min.	> 1 h.	⋮
11	65	19 min.	⋮	> 1 h.	⋮	⋮
12	70	> 1 h.	⋮	⋮	⋮	⋮
13	79	⋮	⋮	⋮	⋮	⋮
14	84	> 1 h.	> 1 h.	> 1 h.	> 1 h.	> 1 h.

Tabla 2: Medición de tiempos para instancias comenzando el 11/4/2018 y finalizando el 24/4/2018. Se toman sólo los partidos de la Liga Nacional. En cada columna se tiene el tiempo hasta alcanzar el valor óptimo para  $\mathcal{A}'$ , fijando los valores en  $\mathcal{A} \setminus \mathcal{A}'$ . Solución factible se encuentra en todos los casos en cuestión de segundos. Al no haber partidos de La Liga Argentina, sólo hace falta considerar  $\hat{\mathcal{A}}$ .

Días	Partidos	$\mathcal{A}_I$	$\mathcal{A}_{II}$	$\hat{\mathcal{A}}$	$\mathcal{A}$
5	15	10 seg.	11 seg.	14 seg.	25 seg.
6	19	12 seg.	15 seg.	21 seg.	54 seg.
7	20	13 seg.	16 seg.	23 seg.	85 seg.
8	24	16 seg.	26 seg.	35 seg.	159 seg.
9	26	17 seg.	28 seg.	42 seg.	11 min.
10	32	29 seg.	53 seg.	89 seg.	> 1 h.
11	35	48 seg.	88 seg.	103 seg.	⋮
12	39	11 min.	5 min.	20 min.	⋮
13	43	18 min.	26 min.	54 min.	⋮
14	46	> 1 h.	> 1 h.	> 1 h.	> 1 h.

---

## 4. Comparación con la asignación manual

---

El modelo que presentamos fue utilizado para ayudar a la confección de la asignación arbitral (todas las decisiones finales fueron tomadas por el comisionado técnico de la liga) durante las temporadas 2016-2017 y 2017-2018. En esta última temporada, además de utilizarse para los partidos de la Liga, también fue utilizada para la Liga Argentina.

Gracias a la colaboración del comisionado técnico, pudimos obtener la asignación realizada de forma manual en la Fase Regional 2015-2016. Esta temporada en particular nos es de interés porque en ella la programación de los partidos tiene en cuenta maximizar la elección de las giras de visitante preferidas por los equipos, que es la condición que se sigue usando hasta la actualidad para la realización del fixture. Por lo tanto, la estructura del problema es similar a la que tratamos actualmente, y resulta una buena instancia de estudio.

Utilizando el modelo para este período, se logra cumplir con las restricciones vistas anteriormente y se incurre en un costo estimado de U\$S 41.747, y una totalidad de 294.672 kilómetros viajados por el plantel de árbitros. Mientras que la designación arbitral real de la fase regional que nos fue facilitada por el comisionado técnico tiene un total de gastos estimado de U\$S 57.576 y un total de 419.400 kilómetros viajados por el plantel de árbitros, lo cual da una mejora a favor del modelo superior al 25 %, tanto en distancia viajada como en costos. Cabe destacar además que la asignación manual no cumplió con todas las restricciones que el problema tiene.

---

## 5. Conclusiones

---

Este trabajo ha reportado una nueva aplicación de sports scheduling, particularmente enfocado en la asignación de árbitros a partidos, un problema mucho menos estudiado en la literatura que el de la programación de partidos. Mediante un modelo de programación lineal entera y un enfoque de descomposición, nuestros resultados reportan una significativa reducción (alrededor de 25 %) en comparación a la solución obtenida manualmente por los programadores del torneo. Debido a los buenos resultados, nuestro enfoque ha sido utilizado para la asignación de árbitros de la Liga Nacional durante las temporadas 2016-2017 y 2017-2018, y por La Liga Argentina (segunda división) en la temporada 2017-2018.

Dado que para la programación del fixture se tiene en cuenta maximizar la cantidad de giras de visitante preferidas por los equipos [5], podemos notar que un árbitro jamás podrá realizar esa misma gira (ya que debe pasar una cierta cantidad de partidos de un equipo para que un árbitro pueda volver a dirigirlo). Por ello, algo que se podría intentar a futuro es tener en cuenta el problema de asignar árbitros a la hora de realizar la programación de partidos. Por ejemplo, supongamos que al programar los partidos se decide que un equipo haga una gira de 3 partidos, jugando de visitante en Mar del Plata un lunes y un miércoles, en Buenos Aires el viernes y finalizado ese partido vuelva a su ubicación. En este caso, un árbitro que dirige en Mar del Plata el lunes no puede dirigir en el partido del miércoles de ese equipo. Sin embargo, algo que podría tener sentido es que otros equipos hagan giras donde el martes y el jueves visitan a equipos en los cuales un mismo árbitro pueda dirigir en días consecutivos desde Mar del Plata (por ejemplo en Bahía Blanca o en Buenos Aires). De alguna manera, estamos evidenciando que habría giras que

dependiendo cómo son distribuidas en la semana, pueden ser complementarias y ayudar a la asignación de árbitros, sin modificar sustancialmente la primera etapa de la realización de la programación de los partidos (sino solamente teniendo criterios de esta índole a la hora de distribuir los partidos en el calendario).

Posiblemente, la integración del problema de asignación de árbitros con el de la programación de partidos implique mayor dificultad a problemas que ya por separado son difíciles, por lo cual la generación de métodos de resolución es también una línea de investigación futura relevante.

**Agradecimientos:** A la directiva y los clubes de la AdC por su compromiso con la implementación de este proyecto, especialmente a Fabián Borro, su presidente, a Sergio Guerrero, su Secretario Técnico, y a Eduardo Bellón y Luis Cornejo, comisionados técnicos responsables de la asignación arbitral. El primer autor está parcialmente financiado por el Instituto Sistemas Complejos de Ingeniería, ISCI, Chile (CONICYT PIA FB0816) y el subsidio ANPCyT PICT 2015-2218 (Argentina).

## Referencias

- [1] F. Alarcón, G. Durán, y M. Guajardo. Referee assignment in the chilean football league using integer programming and patterns. *International Transactions in Operational Research*, 21(3):415–438, 2014.
- [2] F. Alarcón, G. Durán, M. Guajardo, J. Miranda, H. Muñoz, L. Ramírez, M. Ramírez, D. Sauré, M. Siebert, S. Souyris, et al. Operations research transforms the scheduling of chilean soccer leagues and south american world cup qualifiers. *Interfaces*, 47(1):52–69, 2017.
- [3] F. Bonomo, A. Cardemil, G. Durán, J. Marenco, y D. Sabán. An application of the traveling tournament problem: The argentine volleyball league. *Interfaces*, 42(3):245–259, 2012.
- [4] L. de Oliveira, C. C. de Souza, y T. Yunes. On the complexity of the traveling umpire problem. *Theoretical Computer Science*, 562:101–111, 2015.

- [5] G. Durán, S. Durán, J. Marenco, F. Mascialino, y P. A. Rey. Programación matemática para los fixtures de los torneos profesionales del básquet de la argentina en un formato nba. *Revista de Ingeniería de Sistemas Volumen XXX*, 2016.
- [6] G. Durán, M. Guajardo, J. Miranda, D. Sauré, S. Souyris, A. Weintraub, y R. Wolf. Scheduling the chilean soccer league by integer programming. *Interfaces*, 37(6):539–552, 2007.
- [7] G. Durán, M. Guajardo, y D. Sauré. Scheduling the south american qualifiers to the 2018 fifa world cup by integer programming. *European Journal of Operational Research*, 262(3):1109–1115, 2017.
- [8] G. Durán, M. Guajardo, y R. Wolf-Yadlin. Operations research techniques for scheduling chile’s second division soccer league. *Interfaces*, 42(3):273–285, 2012.
- [9] G. Kendall, S. Knust, C. C. Ribeiro, y S. Urrutia. Scheduling in sports: An annotated bibliography. *Computers & Operations Research*, 37(1):1–19, 2010.
- [10] R. V. Rasmussen y M. A. Trick. Round robin scheduling—a survey. *European Journal of Operational Research*, 188(3):617–636, 2008.
- [11] D. Recalde, R. Torres, y P. Vaca. Scheduling the professional ecuadorian football league by integer programming. *Computers & Operations Research*, 40(10):2478–2484, 2013.
- [12] C. C. Ribeiro y S. Urrutia. Scheduling the brazilian soccer tournament: Solution approach and practice. *Interfaces*, 42(3):260–272, 2012.
- [13] M. A. Trick, H. Yildiz, y T. Yunes. Scheduling major league baseball umpires and the traveling umpire problem. *Interfaces*, 42(3):232–244, 2012.
- [14] M. Wright. Scheduling english cricket umpires. *Journal of the Operational Research Society*, 42(6):447–452, 1991.





# Instituto Sistemas Complejos de Ingeniería

## INSTITUTO SISTEMAS COMPLEJOS DE INGENIERÍA

**ISCI**  
Instituto Sistemas  
Complejos de Ingeniería

El Instituto Sistemas Complejos de Ingeniería (ISCI) trabaja en un continuo que va desde el desarrollo de metodologías de frontera hasta aplicaciones innovativas en Sistemas Complejos de Ingeniería, en los cuales interactúan infraestructura, comportamiento humano y gestión.

### GRUPOS DE INVESTIGACIÓN

- Recursos Naturales: Minería, Forestal y Acuicultura
- Transporte
- Localización y Logística
- Consumer Analytics
- Smartcities
- Data Science
- Organización Industrial
- Energía

PATROCINA:



INSTITUCIONES ALBERGANTES:



FAVULINA DE ESTUDIOS  
POSGRUPOS Y TERCEROS AÑOS  
UNIVERSIDAD DE CHILE



INSTITUCIÓN ASOCIADA:



contacto@isci.cl  
www.isci.cl

---

## Programas de Postgrado y Postítulos DII

---

### DOCTORADO

**Doctorado**  
en Sistemas de Ingeniería



Sólida formación en herramientas metodológicas necesarias para identificar, analizar, modelar y resolver problemas complejos en sistemas de ingeniería

Contacto:  
562-29784017 | [doctorado@sistemasdeingenieria.cl](mailto:doctorado@sistemasdeingenieria.cl)  
Informaciones y postulación en línea en:  
[www.dsi.uchile.cl](http://www.dsi.uchile.cl)

**fcfm** FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

### MAGÍSTERES

 INGENIERIA INDUSTRIAL  
UNIVERSIDAD DE CHILE



**MGO** | Magíster  
**Gestión de Operaciones**

Formar profesionales de excelencia en investigación de operaciones, quienes podrán enfrentar problemas complejos en gestión de operaciones, integrando herramientas matemáticas, económicas y tecnológicas.

**fcfm** FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS UNIVERSIDAD DE CHILE

Informaciones en: [www.mgo.uchile.cl](http://www.mgo.uchile.cl)  
Contacto: 562-29784073

# Programas de Postgrado y Postítulos DII



			
<p><b>Capacidad Analítica y Sólida Base en Economía</b></p> <p>Nuestros Graduados tienen sólida base en los tópicos fundamentales de la economía moderna y conocimientos profundos en Microeconomía, Macroeconomía, Métodos Empíricos, y Economía Financiera.</p>	<p><b>Cuerpo Docente de Excelencia</b></p> <p>Profesores que publican en las mejores revistas académicas y que cuentan con una amplia red de colaboradores académicos y profesionales.</p>	<p><b>Acreditado 8 años</b></p> <p>En el año 2016 el Magister en Economía Aplicada fue acreditado nuevamente y obtuvo 8 años. (2024)</p>	<p><b>Posicionamiento de sus Egresados</b></p> <p>Egresados trabajan en destacadas posiciones en empresas públicas y privadas. Alumnos ingresan a doctorados en economía en universidades como Stanford, Yale, Northwestern, NYU, y University of Pennsylvania.</p>



**8** AÑOS

Universidad Acreditada  
Docencia de Pregrado  
Gestión Institucional  
Docencia de Postgrado  
Investigación  
Vinculación con el medio

Sitio web: <http://www.magcea-uchile.cl/>

Av. Beauchef 851

Fono: (+562) 978 4072 | +(562) 9784073

Santiago - CHILE



**MGPP**  
UNIVERSIDAD DE CHILE  
INGENIERÍA INDUSTRIAL

MAGÍSTER EN GESTIÓN Y POLÍTICAS PÚBLICAS

MAGÍSTER EN  
GESTIÓN Y POLÍTICAS PÚBLICAS

 <p>Cuerpo Académico de Excelencia</p>	 <p>Formación multidisciplinaria y sello de gestión pública</p>	 <p>Internacionalización y vinculación con el quehacer público</p>	 <p>Reconocimiento y posicionamiento de los graduados</p>
---	--	---	---



**7** AÑOS

Universidad de Chile  
Acreditada por el máximo periodo en todas las áreas:  
Docencia de Pregrado  
Gestión Institucional  
Docencia de Postgrado  
Investigación  
Vinculación con el medio  
Diciembre 2011 a Diciembre 2018

**7** AÑOS

Magíster en Gestión y Políticas Públicas  
Acreditado por 7 años desde Octubre de 2011 a Octubre de 2018

## Más información

- Teléfonos (562) 29784067 / (562) 29784043
- [mgpp@dii.uchile.cl](mailto:mgpp@dii.uchile.cl)
- República 701 - Santiago, Chile.

---

## Programas de Postgrado y Postítulos DII

---



**Desarrolla  
una mirada  
estratégica de  
tu organización**

y lidera proyectos de  
innovación tecnológica  
en procesos de negocios

**15** AÑOS DE  
EXPERIENCIA



**MBE**

**MASTER IN BUSINESS ENGINEERING**  
INGENIERÍA INDUSTRIAL | UNIVERSIDAD DE CHILE

Duración: 3 semestres académicos

[www.mbe.cl](http://www.mbe.cl)

Contacto: [mbe@dii.uchile.cl](mailto:mbe@dii.uchile.cl) | +56 22 978 4835



INSTITUTO DE INVESTIGACIONES  
EN CIENCIAS FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

---

## Programas de Postgrado y Postítulos DII

---



**MBA** UNIVERSIDAD DE CHILE  
INGENIERÍA INDUSTRIAL

### PRIMEROS EN GESTIÓN

TALLERES DE HABILIDADES DIRECTIVAS  
QUE TE PREPARAN PARA CARGOS DE ALTO LIDERAZGO

ENFOQUE DEL MANAGEMENT SCIENCE PARA  
TOMAR MEJORES DECISIONES DE GESTIÓN

**MBA EXECUTIVE**

24 o 20 meses

Viernes (tarde) y sábado cada 2 semanas

**MBA PART TIME**

20 o 16 meses

3 a 4 veces a la semana (tardes)

8

AÑOS

REGISTRADO EN LA COMISIÓN NACIONAL DE POSTGRADOS Y POSTÍTULOS  
MAYO 2012 Y EN ABRIL 2013



MBA UChile  
Ingeniería Industrial | Universidad de Chile

Admisiones: +56 2 229784002

---

## Programas de Postgrado y Postítulos DII

---



**Generación  
de Redes**



**Carácter  
Internacional**



**Cuerpo Docente  
de Excelencia**

 +562 2978 4020

 mbamin@dii.uchile.cl

**[www.mbamin.cl](http://www.mbamin.cl)**



**fcfm**

FACULTAD DE CIENCIAS  
FÍSICAS Y MATEMÁTICAS  
UNIVERSIDAD DE CHILE

---

## Programas de Postgrado y Postítulos DII

---

### EDUCACIÓN EJECUTIVA



**EDUCACIÓN EJECUTIVA**  
UNIVERSIDAD DE CHILE  
INGENIERÍA INDUSTRIAL

# EXCELENCIA PARA IMPULSAR TU CARRERA

## DIPLOMADOS

En diferentes áreas y formatos

## CURSOS DE ESPECIALIZACIÓN

Más específicos y actualizados

## PROGRAMAS CORPORATIVOS

Creados a la medida de tu empresa

[diplomas@dii.uchile.cl](mailto:diplomas@dii.uchile.cl)  
+56 2 2978 4002  
[www.eeuchile.cl](http://www.eeuchile.cl)



