

Trabajo de grado en modalidad de aplicación

Diseño de una técnica de Simulación-Optimización para la solución del problema de asignación de citas del programa de gestión de riesgos en la ciudad de Bogotá

María Camila Valencia Romero^{a,c}, Arístides Millán Sabogal^{a,c}, Ivan Esteban Tejada Vega^{a,c},

Eliana María González Neira^{b,c}, Oscar David Barrera Ferro^{b,c}

^aEstudiante de Ingeniería Industrial

^bProfesor, Director del Proyecto de Grado, Departamento de Ingeniería Industrial

^cPontificia Universidad Javeriana, Bogotá, Colombia

Resumen de diseño en Ingeniería

This thesis aims to find the solution to the Patient Scheduling Problem (PSP) faced by Colombian's Health Secretary (Secretaría Distrital de Salud), taking into consideration constraints such as: patients' stochastic probability of assistance, maximum capacities of the medical centers, and different schedule arrangements. This is accomplished by the design and implementation of a Simheuristic that combines a Greedy algorithm and the principles of special scheduling (overbooking, prioritization, etc.) which guarantees a high-quality schedule and reduction of lost resources. Initially, the design process required the construction and analysis of the deterministic approach which served to measure the quality of the schedule in terms of prioritization. Finally, the stochastic design was used in order to measure the number of patients programmed that effectively showed up to the appointment. The results of the Simheuristic were compared with the results of the actual scheduling method to evaluate its quality and effectiveness. Results show that in all the instances performed, the method proposed presented a significantly better result than the generated by the actual method.

Key Words: Simheuristics, Patient Scheduling Problem, Greedy Algorithm, Healthcare, Scheduling

1. Justificación y Planteamiento del Problema

Actualmente los sistemas de salud enfrentan un sinnúmero de desafíos, factores como las nuevas tecnologías, fármacos e inexistencia de tratamientos efectivos contra las enfermedades crónicas, incrementan constantemente los costos de estos (Chakraborty, Muthuraman y Lawley, 2013). Frente a este contexto, junto con infraestructuras limitadas y demanda cada vez más alta, la mayoría de las instalaciones de salud han usado principios de investigación de operaciones con el fin de hacer más eficiente la prestación de sus servicios. Dentro de esas medidas se encuentran el optar por períodos de estadía de pacientes más cortos y disminuir las admisiones en hospitales, reemplazándolas por instalaciones para pacientes ambulatorios (clínicas externas) (Chakraborty et al., 2013). Esto a su vez está forzando a las instituciones de salud a reevaluar sus operaciones y capacidades, con el objetivo de mejorar el acceso a la atención médica (Chakraborty et al., 2013). El Sistema

General de Seguridad en Salud de Colombia (SGSS) no es ajeno a este contexto, al cual se le debe adicionar otros factores, como la corrupción, los altos índices de morbilidad, mortalidad y absentismo (Semana, 2015).

A pesar de esto, el SGSS ha aumentado su cobertura de manera excepcional pasando del 23.5% en 1993 al 96.6% en 2014 (Ministerio de Salud Colombia, 2016). Adicionalmente, la cobertura creció con mayor rapidez para las personas en el 20% más pobre de la población (desde el 4% hasta el 89% en el mismo periodo) (Ministerio de Salud Colombia, 2016). Sin embargo, el país enfrenta retos para mejorar la calidad del servicio, aumentar la eficiencia del sistema y eliminar barreras de acceso (Revista Semana, 2017). Bogotá, con una población estimada de 8 millones de habitantes, no es ajena a estas problemáticas (El Espectador, 2017). Aproximadamente el 15% de la población no tiene capacidad de pago para afiliarse a compañías privadas de seguro en salud (Ayala García, 2015). En un modelo de competencia como el colombiano, a pesar de múltiples esfuerzos de regulación, esta inequidad de ingreso aumenta las barreras de acceso a los servicios (Ayala García, 2015). En ese contexto, la Secretaría Distrital de Salud (SDS) diseñó un programa para eliminar las barreras de acceso al sistema de salud que afectan a la población de bajos ingresos de la ciudad. Como parte de una estrategia de promoción y prevención, el programa comprende un conjunto de acciones dirigidas a la población de mayor vulnerabilidad. El objetivo es disminuir la probabilidad de ocurrencia de un evento no deseado, evitable y negativo para la salud de la persona (Secretaría de Salud, 2016).

El proceso de atención puede resumirse en tres fases. En la primera, se identifican y georreferencian los pacientes usando bases de datos. En la segunda, un conjunto de gestores visita a los pacientes y los clasifican en función de su nivel de riesgo asignándoles una alerta alta, media o baja. Finalmente, en la tercera, se define una ruta de atención para los pacientes de acuerdo con su nivel de riesgo. A los pacientes con alerta alta, se les asigna cita médica para el mismo día en el centro de atención más cercano. Los pacientes clasificados en alerta media, entran en línea de espera para ser atendidos en un tiempo máximo de 3 días. Los pacientes de alerta baja, reciben capacitación sobre prevención de factores de riesgo y hábitos saludables, entre otros. Después de esta fase, la barrera se considera superada y el paciente debe iniciar tratamiento usando los servicios de la EPS. En este orden de ideas, la programación de las citas juega un papel fundamental en el cumplimiento del objetivo del programa. Considerando que el esfuerzo para medir los niveles de riesgo está orientado a eliminar las barreras de acceso, es importante garantizar oportunidad en la atención médica.

Es regla general que cuando la demanda supera las capacidades de las instalaciones, el nivel de servicio disminuye (Chakraborty et al., 2013). Este programa no se encuentra por fuera de dicho enunciado, lo cual se puede evidenciar en el nivel de cobertura alcanzado durante el 2017 (Ilustración 1) que ha llegado a tener niveles poco mayores a la mitad de la meta presupuestada (57.03% en diciembre), lo cual es prueba de que requiere reevaluar sus procesos con el objetivo de mejorar el servicio.

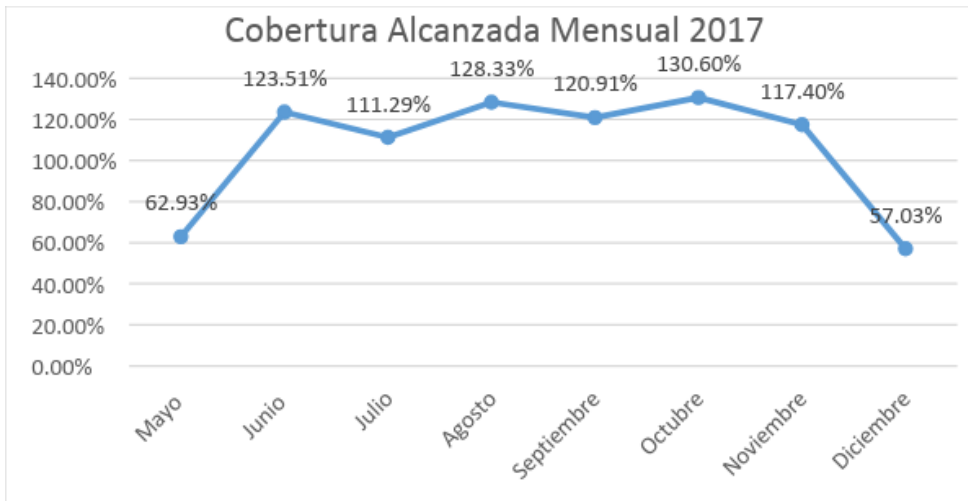


Ilustración 1 Cobertura Alcanzada 2017

En este contexto, debe decidirse cuántos pacientes programar en un intervalo de tiempo. Tomar esta decisión implica un esfuerzo para balancear la eficiencia y el acceso oportuno a los servicios (Gupta & Denton, 2008). Por ejemplo, asignar un único paciente a cada intervalo de tiempo puede mejorar los tiempos de espera en el centro de salud. Sin embargo, en un contexto de recursos limitados, puede incrementar el tiempo que un paciente debe esperar para que le asignen la cita. Así mismo, no tener en cuenta que algunos pacientes incumplirán la cita va en detrimento de la eficiencia del sistema. Por esta razón, los problemas de programación de recursos han sido ampliamente estudiados en el área de investigación de operaciones para el sector salud (Vargas, 2009). La hipótesis es que el uso de herramientas formales puede facilitar la gestión de los servicios y mejorar variables críticas como costo, cobertura y tiempos de respuesta (Ayala, 2015).

Dadas las características de la presente situación surge un reto en particular, el cual consiste en determinar una aproximación adecuada de la probabilidad estocástica de absentismo por paciente. Recientemente, Ahmadi-Javid, Jalali, y Klassen (2017) publicaron una revisión de la literatura de los estudios de programación de pacientes desde la perspectiva de la gestión de operaciones. De acuerdo con los autores, las investigaciones del área pueden clasificarse en función del horizonte de planeación en: estratégicas, tácticas y operativas. En los últimos años, las publicaciones están fundamentalmente concentradas en las decisiones operativas y se evidencian, entre otras, las oportunidades de estudio: el desarrollo de técnicas de solución para problemas complejos, el diseño de modelos que incorporen las condiciones particulares de sistemas de salud y el modelado de aspectos como la probabilidad de ausentismo.

Algunos casos de estudio dan cuenta del potencial de impacto de las investigaciones. Gavriloff et al., 2017 presentaron un enfoque basado en la aplicación de los principios de Lean Six Sigma (específicamente el Heijunka o cargas de nivel) para establecer el número óptimo de pacientes a tratar por franja horaria. Dentro de los resultados obtenidos se encuentra un aumento en el nivel de pacientes de 3.39 a 3.72 y una disminución del tiempo promedio de espera desde 96 a 91 minutos. Otra solución fue propuesta por Chakraborty et al. (2013) la cual tomó como base un modelo en el cual no había franjas horarias predefinidas para asignar citas, es decir, cada cita y su duración se iba creando en cuanto un paciente lo necesitara y compararon sus resultados con diferentes franjas preestablecidas (1 hora, 30 minutos y 15 minutos). Dentro de los resultados se encuentran un aumento del 20%, 16% y 14% respectivamente del beneficio económico esperado, así como un aumento de pacientes tratados al día: 8, 4 y 2, respectivamente.

Otro de los aspectos relevantes para la solución del Patient Scheduling Problem (PSP) es el absentismo, definido como el evento en el que un paciente no se presenta a la cita que le fue asignada, ya que ocasiona

incremento de costos para el centro de salud, disminución de la capacidad de atención de pacientes e indicadores de rendimiento de recursos. Chen et al. (2015) modelaron el efecto que tenían diferentes políticas de sobrecupo como solución al problema de absentismo sobre los indicadores de un hospital en Macau mediante el uso de software de simulación. Los resultados obtenidos en este modelo fueron la reducción del tiempo extra del personal en 58,32%, el tiempo ocioso en 23,65% y el aumento del número de pacientes en 15,9%. Por otra parte, Nuti, Lawley, Turkcan, Tian, Zhang, Chang, Willis y Sands (2012) propusieron una política de sobrecupo con el fin de compensar el absentismo que se presentaba en los centros de salud, esto con el fin de maximizar las utilidades ya que disminuir el absentismo suponía una minimización de costos generados por el mismo. Dentro de los resultados que fueron obtenidos se identificó el comportamiento de las utilidades del centro de salud en función al número de pacientes programados, así como una serie de recomendaciones en cuanto a la metodología de programación en función de la probabilidad de no asistencia de cada paciente.

En este contexto, este trabajo está orientado a responder la siguiente pregunta:

¿Cómo diseñar una técnica de simulación-optimización para resolver el problema de programación de citas de consulta externa para el programa de gestión de riesgo de la SDS?

Es por esto que, para dar solución a dicha pregunta, se propone una heurística o metaheurística como método de solución para cubrir los requerimientos actuales del programa, que tenga versatilidad para que sea posible realizar modificaciones que cubran nuevas necesidades que se presenten a futuro o cambios en algunas ya existentes.

2. Antecedentes

El problema de la asignación de citas ha sido tratado por diferentes autores con un amplio rango de soluciones y supuestos. Dada la complejidad del problema, se han formulado distintos modelos como el de Chakraborty et al. (2013) en el cual se tiene en cuenta una combinación de costos (Ingresos-Espera-Ocioso) como función objetivo unimodal a maximizar, y en donde se agrupan los pacientes en tres categorías dependiendo de la probabilidad de asistir a una cita. Adicional a esto Erdogan, Gose y Denton (2015), propusieron un modelo estocástico de dos etapas en donde se tienen en cuenta la probabilidad de asistir de cada paciente (otra vez, agrupados en tres categorías) así como la entrada de pacientes en estado crítico que necesitan asignación inmediata.

Un estudio hecho por Nguyen, Sivakumar y Graves (2017) desarrolla reglas heurísticas con el fin de maximizar el nivel de utilización de los recursos con los que cuenta el centro médico y además maximizar el número de pacientes atendidos. Estas reglas heurísticas implican dos pasos: el primero, consiste en determinar si se puede aceptar la solicitud de programar la cita por parte del paciente y, el segundo es que, en caso de aceptar la solicitud, se debe buscar el espacio óptimo de acuerdo con una serie de prioridades en la que se encuentran el tiempo de espera y mantener un nivel de servicio cercano a la mediana.

Lin, Ling y Yeung (2017) estudiaron problemáticas como el gasto de recursos debido a los tiempos extras incurridos debido a ineficiencias en el sistema y el tiempo de espera de los pacientes debido a la alta congestión de las salas de espera. Sin embargo, es necesario tener en cuenta que no todos los pacientes muestran las mismas necesidades, esto haciendo que varíen factores como: la puntualidad del paciente, tiempos de llegada, duración del procedimiento y cantidad de visitantes acompañados. Como método de solución los autores presentaron un algoritmo heurístico basado en la simulación que consta de dos partes. En la primera se realiza un plan de asignación de recursos y en la segunda, se realiza separación por bloques que representaban un tipo de paciente en particular y la disciplina de servicio que se debe utilizar para atenderlo.

De manera similar Alaeddini, Yang, Reddy y Yu (2011) obtienen soluciones a este problema mediante el uso de un sistema probabilístico híbrido entre la regresión logística y la inferencia Bayesiana. Dichas soluciones fueron utilizadas para evaluar la optimalidad de una política de sobrecupo. Chen, Kuo, Balasubramanian y Wen (2015) desarrollan un sistema de simulación empírico auto calibrado que se basa en el concepto de

multiservidores y sobrecupo. Las soluciones obtenidas son contrastadas con diferentes políticas de sobrecupo, en donde se obtienen mejoras sustanciales en las medidas de desempeño de las instalaciones de salud.

Asimismo, un estudio realizado por Chakraborty et al. (2013) se centró en el impacto que tenían las franjas horarias predeterminadas sobre el desempeño del sistema de salud. Para esto propusieron un modelo en donde no se tenían franjas horarias fijas, sino que éstas podían cambiar de acuerdo con la estimación de la duración de la cita de un paciente (teniendo en cuenta distintas características del mismo, como el motivo para requerir la cita). Adicional a esto, compararon los resultados obtenidos con distintas duraciones fijas de franjas (15, 30 y 60 minutos) para evaluar la pertinencia de implementación de esta medida.

Otros estudios, como el de Srinivas y Ravindran (2018) proponen la implementación de un modelo de tres etapas en la Clínica de Medicina Familiar en Pensilvania para minimizar los costos producto de los malos sistemas de programación. Dichas etapas consisten en: 1) recolección de información de los pacientes, 2) aprendizaje basado en la predicción del absentismo en pacientes y 3) programación inteligente. Los autores proponen un algoritmo capaz de aprender de la información suministrada y clasificar a los pacientes de acuerdo con una serie de variables (clima, distancia al centro de salud, número de días hasta la cita, etc.) para posteriormente programarlos con base en unas reglas previamente establecidas (sobrecupo, secuenciación, desempeño, etc.).

Como se puede evidenciar, la mayoría de los artículos en donde se formulan modelos que dan solución al PSP se enfocan en minimizar los costos en los que se incurre al atender un paciente. Adicional a esto, recurren a varias suposiciones que pueden alejar drásticamente dicho modelo de la realidad, como la puntualidad de la totalidad de pacientes, o la inexistencia de un tiempo de preparación (definido como el tiempo que le toma al profesional de la salud alistar los insumos necesarios para llevar a cabo la cita). Además, cuando se hace referencia a la posibilidad de absentismo, la mayoría de los modelos tienden a agrupar a los pacientes en categorías de acuerdo a unas probabilidades predeterminadas lo cual, al trasladarlo a la realidad puede afectar la calidad y objetividad de los resultados, esto debido a que cada persona posee una serie de características que configuran distintos valores de absentismo.

La técnica propuesta difiere de enfoques presentados por otros autores, en que se pretende tener en cuenta de manera simultánea los dos objetivos: calidad de horario (medido en la cantidad de pacientes con prioridad alta que efectivamente son programados) y cantidad de pacientes programados que en realidad asisten a la cita. Así, se pretende dar una solución que permita al tomador de decisiones analizar el *trade-off* entre ambos criterios. Sin embargo, también se realizan algunas de las suposiciones que se hacen en otros modelos estudiados con el fin de reducir la complejidad del algoritmo y el tiempo computacional que es usado para obtener una solución factible. No obstante, es importante recalcar que la técnica busca individualizar la probabilidad de ausentismo por paciente para eliminar el sesgo en el que incurren los demás modelos de homogeneizar a todos los pacientes en este aspecto. De esta manera, se espera que los resultados obtenidos sean mucho más cercanos a la situación actual colombiana permitiendo tomar medidas pertinentes para su mejoramiento.

3. Objetivos

General:

Diseñar una técnica de simulación-optimización para resolver el problema de programación de citas de consulta externa en el contexto del programa de gestión de riego de Bogotá

Específicos:

1. Calcular la probabilidad de no asistencia de los beneficiarios del SISBEN de manera individual teniendo en cuenta los diferentes factores que afectan dicha asistencia.
2. Diseñar e implementar un algoritmo para la asignación de citas de consulta externa considerando todos los parámetros de entrada como determinísticos.

3. Diseñar e implementar modelo de simulación-optimización con la metaheurística para la asignación de citas de consulta externa considerando las probabilidades de no asistencia de los beneficiarios del SISBEN.
4. Comparar los resultados del modelo propuesto con las condiciones actuales para medir el impacto de la implementación del mismo.

3.1. PROBABILIDAD DE NO ASISTENCIA

Con el fin de calcular la probabilidad de no asistencia se llevaron a cabo 3 pasos. Como primer paso se realizó un análisis estadístico de los datos proporcionados por la SDS para determinar cuál era la mejor forma de organizarlos. Este análisis permitió generar grupos de información para ser utilizados en la siguiente fase. Como segundo paso se realizó una regresión logística para determinar cuáles variables tenían efecto significativo en la asistencia de los pacientes a la cita y establecer si la ecuación de regresión permitiría predecir el comportamiento de asistencia que tendrían los pacientes. Como último paso se calculó la probabilidad condicional de asistencia dado un clúster de pertenencia del paciente.

- **Análisis estadístico (ANEXO 1)**

En primera instancia se realizó un análisis estadístico de las diferentes subredes (Norte, Sur, Centro y Suroccidente), con el fin de extraer información pertinente de las bases de datos.

El primer paso de dicho análisis fue implementar un diagrama de Pareto en cada subred, identificando cuáles de las especialidades eran las más representativas en cada una de éstas. Se obtuvo como resultado que las especialidades de Programa CCU, Higiene oral, Programa del Joven y Crecimiento y desarrollo deberían ser el foco central del análisis ya que son el 20% de las especialidades que ocupan el 80% de las frecuencias de citas.

Como segundo paso se realizaron pruebas de hipótesis de diferencia de proporciones de las especialidades seleccionadas entre las subredes para determinar si estas podían ser tratadas como conjunto. Es decir, en caso de que no hubiera diferencia de proporción para cada especialidad específica entre las distintas subredes, poder tratar todos los datos de las subredes como un único conjunto de datos. Los resultados de las pruebas mostraron que las proporciones de las especialidades seleccionadas en la subred centro, eran significativamente diferentes a las de las otras tres subredes (Norte, Sur y Suroccidente). Sin embargo, las proporciones de las especialidades seleccionadas de las subredes Norte, Sur y Suroccidente no eran significativamente diferentes entre sí. Por tanto, se podían agrupar los datos de estas tres subredes, por especialidad, para contar con una mayor cantidad de información y por ende mayor obtener precisión en los análisis. Dichas pruebas se realizaron siguiendo las ecuaciones 1, 2, 3 y 4.

$$Z_0 = \frac{P_1 - P_2}{\sqrt{P(1-P)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \quad \text{Ecuación (1)}$$

$$P = \frac{n_1 P_1 + P_2 n_2}{n_1 + n_2} \quad \text{Ecuación (2)}$$

$$P_1 = \frac{A}{n_1} \quad \text{Ecuación (3)}$$

$$P_2 = \frac{B}{n_2} \quad \text{Ecuación (4)}$$

En el Anexo 1 se presentan los resultados de todas las pruebas para cada especialidad.

- **Regresión logística y análisis de clusters (ANEXO 2 y 3)**

En segunda instancia se realizó una regresión logística mediante los programas R-Studio, SPSS y Minitab, con el fin de determinar las variables que influyen en la asistencia o no de los pacientes a una cita, y también evaluar si la ecuación de regresión permitiría predecir con un alto porcentaje de exactitud la asistencia o absentismo de un paciente a la consulta. Para dicho análisis se tuvieron en cuenta las variables de Sexo, Edad, Capitados, Especialidad, Prioridad y Subred, tomando como variable respuesta el Estado de la cita, la cual se tomó como variable binaria, siendo 0-Incumplido y 1-Cumplido. Según esto, se calcularon los coeficientes de cada una de las variables y se determinaron los factores significativos para el modelo analizado, los resultados obtenidos se presentan en la Tabla 1 y la

Tabla 2.

Tabla 1 Coeficientes variables

Factor	Coefficiente
Intercepto	2,847
Sexo	
F	-0,096
Edad	
Gestantes	-0,193
Adulto (30-59)	-0,109
Joven (10-29)	-0,411
Infancia (0-9)	-0,307
Capitado	
NO	-0,539
Especialidad	
Control Prenatal	0,814
Crecimiento y desarrollo	0,344
Higiene Oral	-0,160
Cáncer de Mama	0,350
Prioritaria	0,301
Programa CCU	-0,094
Programa del Adulto	0,364
Programa del Joven	0,112
Tamizaje	0,460
Prioridad	
Alta	0,291
Subred	
Sur	-2,12
Norte	-1,950

Tabla 2 P-value variables

Factor	P-Value
Sexo	0,265
Edad	0,021
Capitado	0,000
Especialidad	0,006
Prioridad	0,442
Subred	0,000

Cabe aclarar que el modelo de regresión logística fue obtenido usando como base una selección aleatoria del 70% de los datos. Posterior a eso, se predijo la asistencia de los pacientes del 30% restante con el modelo obtenido y se contrastó con el dato real, obteniendo un porcentaje de eficacia del 99.71% (2431 obtenido con la simulación utilizando la regresión logística vs 2438 pacientes que asistieron realmente).

Paralelo a esto, se llevó a cabo un análisis de clúster como método adicional para contrastar con la regresión logística. En este análisis, cada Clúster -o grupo de pacientes- era una combinación de las características previamente nombradas. (Ej. Un clúster eran los hombres adultos que pertenecen a la especialidad higiene oral de la subred norte y que capitán en el programa de gestión de riesgo, mientras que otro clúster eran las mujeres jóvenes que pertenecen a la especialidad de crecimiento y desarrollo de la subred sur y que no capitán en el programa de gestión de riesgo). De esta manera, cada clúster contaba con una frecuencia de ocurrencia (que es la cantidad de pacientes del total que pertenecen a ese clúster) y una intersección de probabilidades (que es la cantidad de pacientes dentro del clúster que asistieron a la cita de entre el total de este último).

Con el fin de calcular la probabilidad de asistencia de cada uno de los pacientes de acuerdo con este análisis se realizó el cálculo de la probabilidad condicional, la cual funciona bajo el principio de que ocurra un evento A, dado que el evento B sucede con anterioridad.

Para este caso A representa la variable respuesta que indica si el paciente del clúster asiste o no a la cita y B es el evento en el que el paciente pertenece a dicho clúster. Esta probabilidad se puede ver reflejada en la ecuación 5:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad \text{Ecuación (5)}$$

De esta manera, si solo existieran los dos clúster mencionados anteriormente, y el primero tuviese 30 pacientes (de los cuales 8 asistieron), mientras que el segundo estuviera conformado por 80 (de los cuales 50 asistieron), la probabilidad de ocurrencia de cada clúster - $P(B)$ - sería de 27.7% (30 de 110) y 72.3% (80 de 110) respectivamente. Mientras que la probabilidad de que un paciente asista y que sea del clúster - $P(A \cap B)$ - es de 7.2% (8 de 110) y 45.5% (50 de 110) respectivamente. Así las cosas, las probabilidades de asistencia condicionales para cada clúster se obtendrían dividiendo las dos probabilidades anteriormente encontradas y se obtendría un total de 25.99% y 62.93% .

Cabe aclarar que para este análisis se tomó el mismo principio de tomar el 70% de los datos como base y predecir la asistencia de los pacientes del 30% restante. Este análisis arrojó un nivel de eficiencia del 98.96% (2412 de pacientes usando el análisis de clúster vs 2438 de pacientes que asistieron realmente).

Según los datos anteriores se identificó que la metodología que posee un mejor rendimiento y, que por ende, es mejor usar para estimar la probabilidad de asistencia de cada paciente es la regresión logística.

En el Anexo 3 se presentan los resultados de las probabilidades condicionales para cada clúster y los cálculos de efectividad para la regresión logística y el análisis de clúster

3.2. ASIGNACIÓN DE CITAS: CASO DETERMINÍSTICO (ANEXO 4 - HOJAS 1 Y 2)

En esta parte se llevó a cabo el diseño y la implementación de un algoritmo para la asignación de citas de consulta externa, en el cual se consideraron todos los parámetros de entrada como determinísticos. Esto quiere decir que la probabilidad de asistencia de los beneficiarios es 1. La función objetivo analizada es la maximización de la cantidad de pacientes atendidos en los horarios disponibles al mismo tiempo que se maximiza la cantidad de pacientes de alta prioridad atendidos. Para la elaboración del algoritmo se tuvieron en cuenta tres diferentes políticas de asignación. Además, se realizó para cada una de éstas un análisis que tenía en cuenta a qué subred y a qué especialidad pertenecía cada paciente y que prioridad presentaba.

- **Política 1**

Una vez se tienen todos los pacientes solicitantes de cita de un mes (jornada laboral), estos son clasificados dentro de dos grupos: aquellos que tienen una probabilidad de asistencia mayor o igual al 60%, quienes se catalogan como asignación de citas “con prioridad”. Los demás pacientes se clasifican dentro del grupo de asignación “no prioritaria y aleatoria”.

La dinámica de asignación entonces se basa, en primera instancia, en agendar a los del primer grupo (“asignación con prioridad”) para el horario disponible más cercano a la fecha, de manera que se priorice a este grupo. Una vez finalizado esto, se procede a asignar a los pertenecientes del segundo grupo de manera aleatoria en los espacios libres que quedan después de realizar la primera asignación.

Adicional a esto, la política cuenta con sistemas de recordatorios para las personas que están asignadas en citas “no prioritarias y aleatorias”. Dicho sistema se basa en hacer una llamada antes del día de la cita para que estas confirmen o no su asistencia. En caso tal de que no puedan asistir se le cancela la cita y el horario en el que estaba programada su cita queda disponible para agendar a otro paciente. Cabe aclarar que la persona a la que se le cancela la cita vuelve a entrar al grupo de citas “no prioritarias y aleatorias”. En la Ilustración 2 se muestra el flujograma correspondiente.

- **Política 2**

Esta política consiste en ordenar de mayor a menor, de acuerdo con su probabilidad de asistencia, a todos los pacientes que solicitan cita en un mes determinado (jornada laboral). Posterior a eso, se procede a generar un grupo de baja prioridad con los pacientes que posean una probabilidad de asistencia menor de 0,2. Como siguiente paso, se asignan los horarios disponibles más cercanos a la fecha según la lista obtenida previamente sin tener en cuenta a los pacientes probabilidad de asistencia menor a 0,2. Como último paso se verifica si existen horarios disponibles, en caso de ser así se asignan las citas de estos pacientes, y para los que no alcancen a ser programados se realiza un *overbooking*, en donde se agendaron los primeros pacientes de la lista. Esta política se realiza para cada subred y cada especialidad al final de cada mes. En la Ilustración 3 se presenta el diagrama de flujo correspondiente.

- **Política 3**

Los pacientes se agrupan en n grupos. La cantidad n de grupos dependerá de la cantidad de pacientes por mes y la probabilidad de asistencia de estos. El principio que debe seguir cada uno de los grupos es que el valor esperado de pacientes que van a asistir en una franja de atención con un médico específico sea mayor o igual a 1 pero estrictamente menor a 1,2, ya que al manejar este valor se asegura una baja probabilidad de que dos pacientes asistan en un mismo horario (este valor esperado se obtiene mediante la suma de las probabilidades de cada uno de los pacientes que conforman un grupo). Una vez finalizado este paso, cada uno de los grupos es asignado a un horario de manera aleatoria. Esta política se realiza para cada subred y cada especialidad al final de cada mes. El diagrama de flujo se presenta en la Ilustración 4.

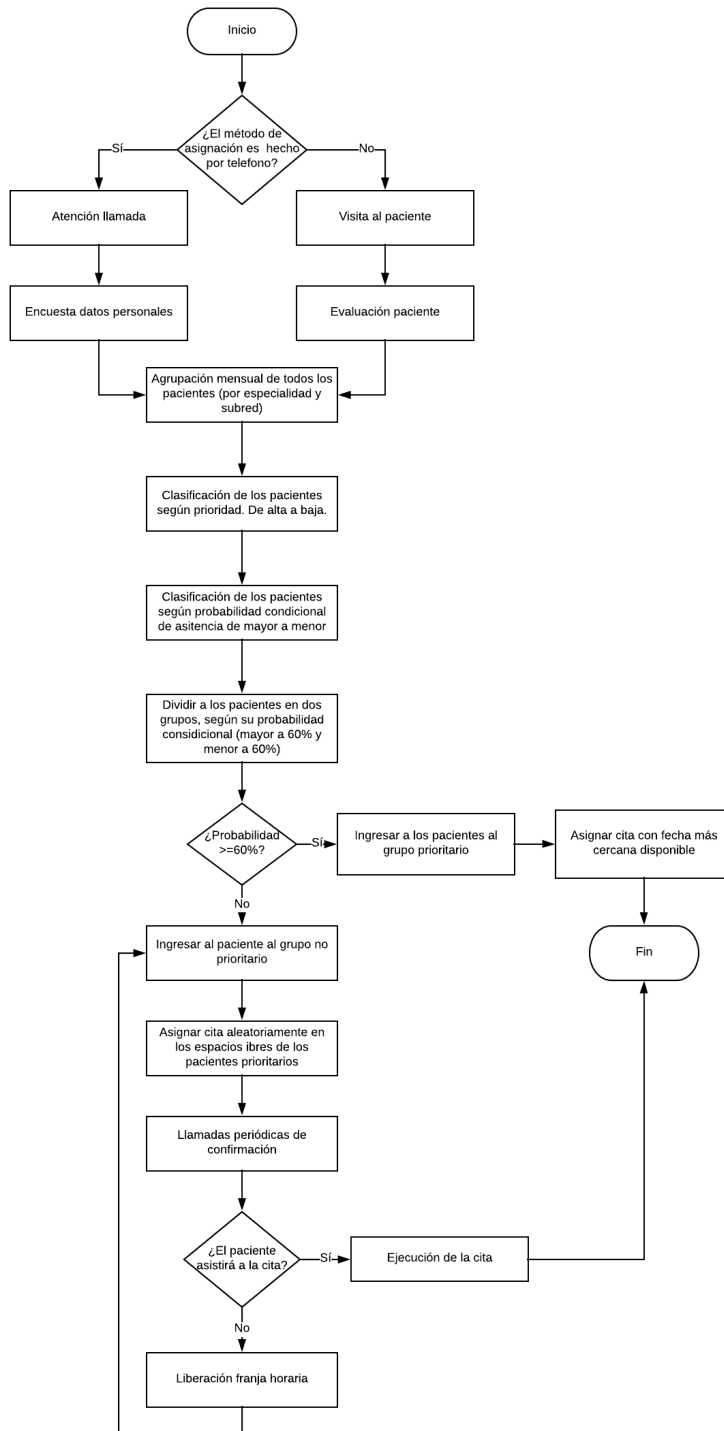


Ilustración 2 Flujograma Política 1

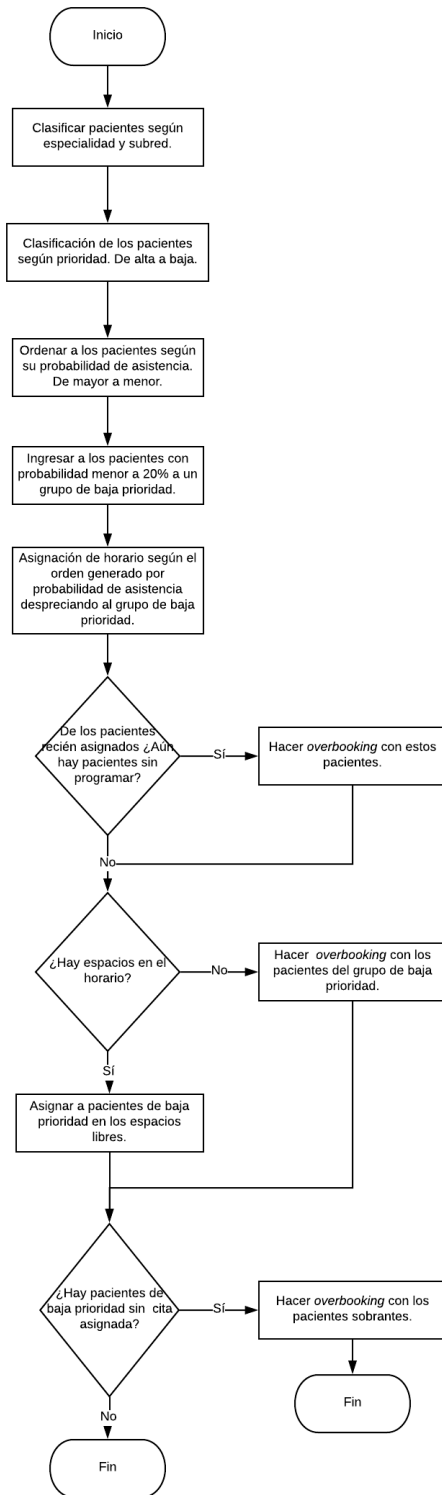


Ilustración 3 Flujograma Política 2

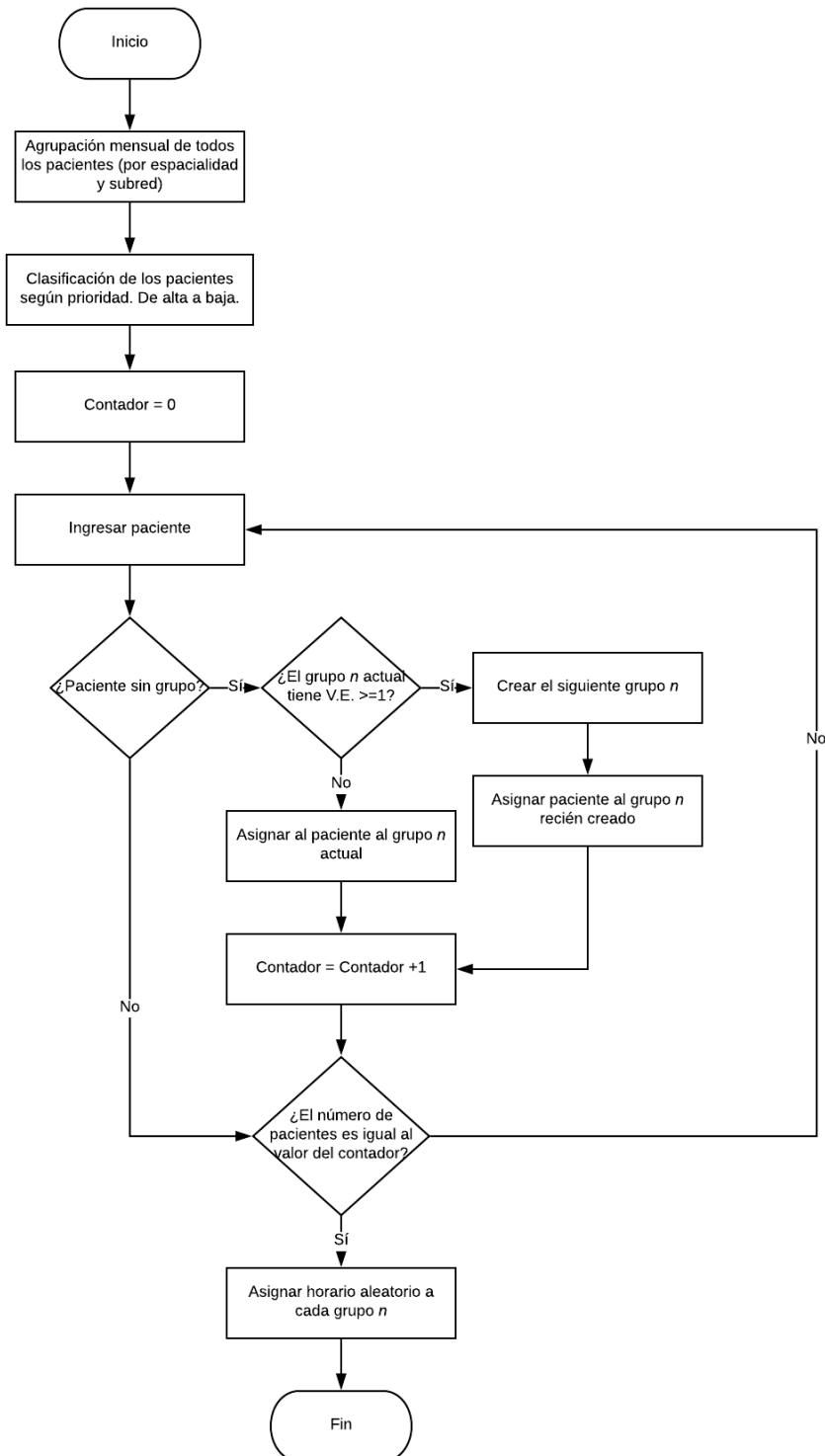


Ilustración 4 Flujograma Política 3

- **Resultados**

Se compararon las políticas en escenarios determinísticos. Para ello se tomó como base la programación obtenida de cada política. Partiendo de esto se supuso que todos los pacientes asistían y se midió la calidad de programación en términos de la cantidad de pacientes con prioridad alta que se encontraban programados. Se encontró que, tomando como enfoque esta función objetivo las políticas 1 y 2 tuvieron mejor desempeño que la 3, alcanzando un total de 22 pacientes cubiertos (igual al total de pacientes prioritarios en la instancia). Estas dos políticas mostraron un igual desempeño al de la política actual de la SDS, como se evidencia en la **¡Error! No se encuentra el origen de la referencia..**

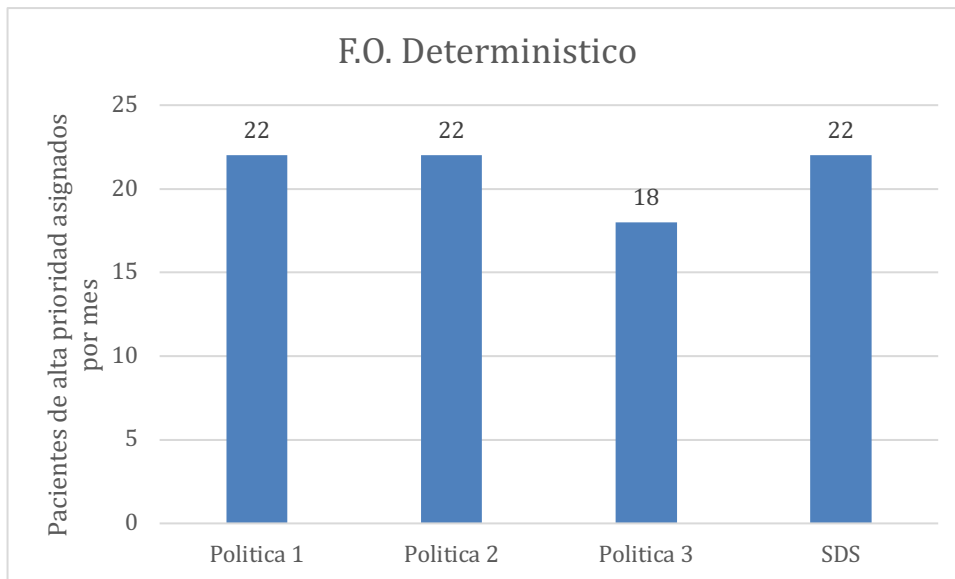


Ilustración 5 Cubrimiento de pacientes de alta prioridad de cada política vs cubrimiento de la SDS.

Por consiguiente, se recomienda adoptar cualquiera de las dos políticas si el objetivo es mejorar la calidad del horario.

3.4 COMPARACION CON CASOS REALES (ANEXO 4 – HOJAS 4, 5 Y 6)

En esta parte se tuvieron en cuenta las 3 políticas explicadas en el capítulo 2 teniendo en cuenta las probabilidades de asistencia de los pacientes. Esto significa que se hibridizan los algoritmos *greedy* (políticas) con Simulación de Montecarlo, para determinar el valor esperado del número de pacientes que asistirán a la cita. Para calcularlo se toma la probabilidad de asistencia de cada paciente y se compara con un número aleatorio generado anteriormente. Si el segundo es mayor al primero, se dice que el paciente no asiste a la cita, mientras que, si es menor, el paciente asiste a su cita. En cada corrida se cuenta el número de pacientes que asisten a su cita, y el número que se obtiene al final se toma como la función objetivo de esa corrida.

En este apartado se realizaron 100 corridas para cada simulación de la asistencia de los pacientes programados de acuerdo con el horario arrojado por una política, tomando como función objetivo la cantidad efectiva de pacientes que asistían a la cita. Este número es elegido debido a que es mayor a los resultados obtenidos al calcular el número mínimo de corridas con la ecuación 6 para cada política, para este caso se utilizó un error del 0,5% alrededor de la media y una confianza del 99%, obteniendo como resultados los valores de 54, 87 y 97 respectivamente.

$$N = \left[\frac{\sigma * Z}{Error * X} \right]^2 \text{ Ecuación (6)}$$

- N = Número de corridas
 σ = Desviación estándar de la variable respuesta a analizar de las 50 corridas
X = Promedio de la variable de respuesta a analizar de las 50 corridas
Z = Estadístico normal estándar (para un nivel de confianza del 95% =1.96)

Al promedio de funciones objetivo de cada una de esas simulaciones se tomaba como un dato. Para comparar cuál política era mejor en términos de asistencia esperada se tomaron 10 datos (simulaciones completas) por política para posteriormente realizar un ANOVA (análisis de varianzas).

Para establecer la efectividad de las políticas ideadas en el caso real se tomaron los pacientes programados desde el 5 de septiembre hasta el 25 de septiembre de 2017 (20 días), se escogieron estos días debido a que aseguraban una mayor densidad de pacientes en la base de datos. En primera instancia se tomaron aquellos pacientes que la SDS pudo programar dentro de ese rango, así como los que quedaron en estado “Pendiente”, es decir, que se presentaron en septiembre pero que no pudieron ser programados. Este grupo de pacientes se programó con cada una de las tres políticas descritas en los apartados anteriores para obtener el horario correspondiente de acuerdo con las reglas de cada política. Posterior a eso, se tomó cada programación arrojada por cada política y se simuló por 100 corridas para obtener el número esperado de pacientes que asistían a las mismas.

Cada corrida consiste en evaluar si los pacientes programados en cada espacio que se tiene destinado dentro del horario (Ej. Lunes a las 10 A.M) asistirán o no. Esto se lleva a cabo mediante números aleatorios y la probabilidad de asistencia de cada paciente calculada usando la regresión logística. Para cada espacio se generan tantos números aleatorios como pacientes programados en dicha cita (si se tiene un paciente programado por horario entonces se genera un solo número aleatorio, si se tienen dos entonces se generan dos números, y así sucesivamente). Posterior a eso, se toma la probabilidad de asistencia de cada paciente y se compara con el número aleatorio generado. Si el segundo es mayor al primero, se dice que el paciente no asiste a la cita, mientras que, si es menor, el paciente asiste a su cita. En cada corrida se cuenta el número de pacientes que asisten a su cita, y el número que se obtiene al final se toma como la función objetivo de esa corrida.

Este proceso se realizó para cada una de las políticas, así como para la programación realizada por la SDS. Los resultados iniciales permiten suponer que las políticas 2 y 3 son superiores en desempeño a la programación realizada por la SDS. En la Ilustración 6 se pueden ver los resultados de las simulaciones:

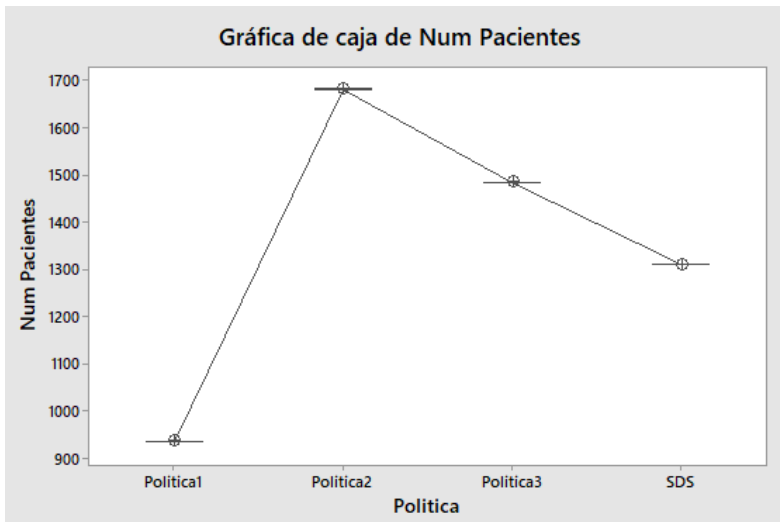


Ilustración 6 Diagrama de caja y bigotes con muestreo real.

Una vez finalizado esto, se realizó un ANOVA para establecer diferencias significativas entre los desempeños de las políticas. Ya comprobados los supuestos de independencia, homogeneidad y normalidad de los residuos, se obtuvo que por lo menos una política es significativamente diferente de las demás con un nivel de significancia del 5%. Esto indica que hay alguna política cuyo desempeño es mejor (o peor) que al de las demás. En la **¡Error! No se encuentra el origen de la referencia.** y Tabla 4 se muestran los resultados del ANOVA.

Tabla 3 Análisis de categorías objetivo 4.

Análisis de varianza de un factor				
Resumen				
Grupos	Cuenta	Suma	Promedio	Varianza
Política 1	10	9349.67	934.967	0.85
Política 2	10	16809.49	1680.949	2.69
Política 3	10	14833.25	1483.325	1.78
SDS	10	13098.44	1309.844	0.07

Tabla 4 Análisis de varianza Objetivo 4.

ANÁLISIS DE VARIANZA						
Origen de las variaciones	Suma de cuadrados	Grados de libertad	Promedio de los cuadrados	F	Probabilidad	Valor crítico para F
Políticas	3011470.573	3	1003823.524	744148.854	2.65602E-86	2.87
Error	48.56239	36	1.35			
Total	3011519.136	39				

Una vez finalizado este análisis, se hicieron las pruebas post-hoc para determinar cuáles medias eran diferentes y establecer la mejor política a seguir. A continuación, la Tabla 5, se muestran la comparación entre cada política realizada en la prueba Tukey.

Tabla 5 Resultado Prueba Tukey

Políticas	N	Media	Agrupación			
Política 2	10	1680,95	A			
Política 3	10	1483,32		B		
SDS	10	1309,84			C	
Política 1	10	934,967				D

Teniendo en cuenta que todas las políticas quedaron en una agrupación diferente en la prueba de Tukey, se puede determinar que la política que mejor desempeño tiene en términos del número esperado de pacientes es la política 2.

4. Conclusiones y recomendaciones

Para la resolución de problema de programación de pacientes de la secretaria de salud se realizaron análisis estadísticos con el fin de determinar una aproximación acertada para la resolución de dicho problema. Posterior a eso se formularon tres políticas de programación y se evaluó su desempeño frente a la situación actual.

Una vez realizados los análisis estadísticos de las características o aspectos que impactan significativamente la probabilidad de asistencia de los pacientes se encontró que las variables con mayor incidencia fueron: especialidad a la que asisten, capitados, ciclo de vida y subred a la que pertenecen.

Para el caso determinístico, las políticas propuestas 1 y 2 mostraron un cubrimiento que se adapta a los requerimientos necesarios para satisfacer la demanda presentada por los pacientes de alta prioridad del SISBEN, asegurando mejoras en la calidad de programación. Esto se debe a que las políticas propuestas realizan segmentación de los pacientes previo a la programación de los mismos. Por lo tanto, esta división de acuerdo con las características de los pacientes se puede usar para próximas políticas que se deseen crear.

Para el caso de la comparación contra el caso real, la política 2 obtuvo un mejor desempeño que la programación actual de la SDS tanto en el cubrimiento de pacientes con prioridad alta como en la cantidad de pacientes que efectivamente asisten a las citas. Por lo tanto, para el caso de los pacientes de la SDS en el programa de gestión de riesgo de la ciudad de Bogotá, lo más pertinente es aplicar segmentación de pacientes y overbooking para asegurar optimalidad en la programación de citas. Es importante tener en cuenta que para aplicar correctamente el principio de overbooking es necesario contar con un análisis a profundidad de las características de los pacientes. Esto debido a que si se aplican estos principios de manera errada se puede llegar a un cubrimiento más alto, pero a expensas de la calidad del servicio (tiempos de espera prolongados, horas extra de trabajo para el personal y sobrecupo en los centros médicos).

Además, las políticas muestran diferencia en los resultados, especialmente la política 1 en comparación con la 2 y la 3, debido a los criterios que se tuvieron en cuenta para programar pacientes en cada una. Por ejemplo, en el caso de la política 2 se asegura la maximización de la atención de los pacientes programados al tener como criterio el overbooking de los pacientes que poseen una probabilidad de asistencia menor al 20% y a los que sobraron que tenían probabilidad mayor de 20%. Por otro lado, en la política 3 se conforman grupos de pacientes de probabilidad de asistencia alta y baja, haciendo que sea posible una mayor asistencia de los pacientes programados.

Ahora bien, este análisis de características de los pacientes se puede extender al considerar factores adicionales que puedan influir en la probabilidad de asistencia de tal manera que se pueda calcular una probabilidad más precisa. Asimismo, se recomienda realizar este tipo de análisis por lo menos dos veces al año, con el fin de evaluar las tendencias de las características que determinan esta probabilidad para poder ajustar el sistema de programación adecuadamente.

Usar una aproximación determinística permite establecer comparaciones entre la situación actual y nuevas políticas que se propongan para medir la calidad de la programación en cuanto a pacientes prioritarios programados. Esto se puede extender a la creación de parámetros de calidad (Ej. Porcentaje mínimo de cubrimiento que tiene que tener una política).

Finalmente, si se busca ahondar en las características de los pacientes, se pueden plantear diversas metodologías para estimación de probabilidades basándose en algoritmos como el genético, esto debido a que entre más precisa sea esta estimación, mejor será la calidad de la programación obtenida. Por otro lado, es pertinente aclarar que se pueden plantear técnicas de programación más robustas utilizando combinaciones de los principios mencionados a lo largo del documento, esta robustez permitiría mejorar los indicadores del programa y, por consiguiente, permitiría la optimización de recursos y tiempos.

5. Glosario

Algoritmo greedy: es un algoritmo que aplica la heurística de resolución de problemas haciendo una elección óptima localmente en cada etapa con el objetivo de encontrar un óptimo global. (Lu G, Xiong Y, Ding C, Wang Y, 2016)

Estocástico Un proceso estocástico es una colección de variables aleatorias $\{X_t, t \in T\}$ definidas sobre un espacio de probabilidad (Ω, A, P) . (Montes, 2007)

Determinístico: Según la definición generalmente aceptada, un modelo determinístico es aquel en el que no hay aleatoriedad involucrada en el desarrollo de estados futuros. (Toma, 2012)

Metaheurística: Son una clase eficiente de algoritmos de optimización que extienden los marcos metaheurísticos de una manera natural y flexible al integrar la simulación. (Hatami, S., Calvet, L., Fernández-Viagas, V., Framiñán, J. M., & Juan, A. A., 2018)

Bibliografía

- Ahmadi-Javid, A., Jalali, Z., & Klassen, K. J. (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research*, 258(1), 3-34.
- Alaeddini, A., Yang, K., Reddy, C., & Yu, S. (2011). A probabilistic model for predicting the probability of no-show in hospital appointments. *Health care management science*, 14(2), 146-157.
- Ayala García, J. (2015). La salud en Colombia: más cobertura pero menos acceso. *Banco de la República de Colombia Documentos de Trabajo*.
- Chakraborty, S., Muthuraman, K., & Lawley, M. (2013). Sequential clinical scheduling with patient no-show: The impact of pre-defined slot structures. *Socio-Economic Planning Sciences*, 47(3), 205-219.
- Chen, Y., Kuo, Y.-H., Balasubramanian, H., & Wen, C. (2015). *Using simulation to examine appointment overbooking schemes for a medical imaging center*. Paper presented at the Proceedings of the 2015 Winter Simulation Conference.
- Erdogan, S. A., Gose, A., & Denton, B. T. (2015). Online appointment sequencing and scheduling. *IIE Transactions*, 47(11), 1267-1286.

- Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE transactions*, 40(9), 800-819.
- Lin, C. K. Y., Ling, T. W. C., & Yeung, W. K. (2017). Resource Allocation and Outpatient Appointment Scheduling Using Simulation Optimization. *Journal of healthcare engineering*, 2017.
- Lynn A Nuti , Mark Lawley , Ayten Turkcan , Zhiyi Tian , Lingsong Zhang , Karen Chang , . . . Sands, L. P. (2012). No-shows to primary care appointments: subsequent acute care utilization among diabetic patients. *BMC health services research*, 12(1), 304.
- Nguyen, T.-B. T., Sivakumar, A. I., & Graves, S. C. (2017). Scheduling rules to achieve lead-time targets in outpatient appointment systems. *Health care management science*, 20(4), 578-589.
- Srinivas, S., & Ravindran, A. R. (2018). Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. *Expert Systems with Applications*, 102, 245-261.
- El Espectador, J. (2017). Reorganización de la salud en Bogotá: un año de aciertos y polémicas | ELESPECTADOR.COM. [online] ELESPECTADOR.COM. Available at: <https://www.elespectador.com/noticias/bogota/reorganizacion-de-la-salud-en-bogota-un-ano-de-aciertos-y-polemicas-articulo-686760> [Accessed 24 Mar. 2018].
- Lorenzo, I. V. (2009). Barreras en el acceso a la atención en salud en modelos de competencia gestionada: un estudio de caso en Colombia (Doctoral dissertation, Universitat Autònoma de Barcelona).
- Ministerio de Salud Colombia, M. (2016). En 2015 aumentó en 990.385 personas el número de afiliados al sistema de salud. [online] Minsalud.gov.co. Available at: <https://www.minsalud.gov.co/Paginas/En-2015-aumento-en-990.385-personas-el-numero-de-afiliados-al-sistema-de-salud.aspx> [Accessed 24 Mar. 2018].
- Revista Semana, S. (2017). Los cinco retos de la salud en Colombia. [online] Los cinco retos de la salud en Colombia. Available at: <http://www.semana.com/vida-moderna/articulo/los-cinco-retos-de-la-salud-en-colombia/537571> [Accessed 24 Mar. 2018].
- Revista Semana, U. (2015). Alarmantes cifras de morbilidad y mortalidad infantil. [online] Semana. Available at: <http://www.semana.com/opinion/articulo/uriel-ortiz-soto-alarmantes-cifras-de-morbilidad-mortalidad-infantil/426758-3> [Accessed 24 Mar. 2018].
- Secretaria de Salud (2016). [online] Saludcapital.gov.co. Available at: http://www.saludcapital.gov.co/Documents/Plan_Territorial_Salud_2016_2020.pdf [Accessed 24 Mar. 2018].
- Montes, F. (2007). Procesos estocásticos para ingenieros: Teoría y aplicaciones.
- Lu G, Xiong Y, Ding C, Wang Y (2016) An Optimal Schedule for Urban Road Network Repair Based on the Greedy Algorithm.
- Hatami, S., Calvet, L., Fernández-Viagas, V., Framiñán, J. M., & Juan, A. A. (2018). A simheuristic algorithm to set up starting times in the stochastic parallel flowshop problem. *Simulation Modelling Practice and Theory*, 86, 55-71.
- Toma, A. M. (2012). Deterministic Method for Assessing the Degree of Degradation for Masonry Condominium Structures in Romanian Urban Areas. *Buletinul Institutului Politehnic din Iasi. Sectia Constructii, Arhitectura*, 58(3), 113.



Facultad de Ingeniería
INGENIERÍA INDUSTRIAL

Trabajo de Grado –Primer Semestre 2019