TÉCNICAS DE CLUSTERIZACIÓN APLICADAS EN LA PROGRAMACIÓN DEL BLOQUE QUIRÚRGICO EN HOSPITALES DE ALTA COMPLEJIDAD

ANDRÉS FELIPE VEGA SIERRA

UNIVERSIDAD SANTO TOMÁS

DIVISIÓN DE INGENIERIAS

FACULTAD DE INGENIERIA INDUTRIAL

BOGOTA

2019

TÉCNICAS DE CLUSTERIZACIÓN APLICADAS EN LA PROGRAMACIÓN DEL BLOQUE QUIRÚRGICO EN HOSPITALES DE ALTA COMPLEJIDAD

ANDRES FELIPE VEGA SIERRA

Proyecto de investigación para optar al título de ingeniería Industrial

Asesor

Helien Parra Riveros

UNIVERSIDAD SANTO TOMÁS

INGENIERIA INDUSTRIAL

SEMINARIO DE GRADO

BOGOTÁ

2019

AGRADECIMIENTOS

Esta tesis se la dedico a mi padre y a mi madre por su apoyo incondicional, por estar siempre a mi lado, sujetando mi mano y por no permitirme jamás desviarme del camino y a cada persona que me apoyo durante mi formación académica, mi abuela mis amigos, y hermanos, pero sobre todo a Dios por guiarme durante todo este proceso y no dejarme caer.

RESUMEN

En los últimos años los índices de intervenciones quirúrgicas programadas han tenido un ascenso considerable, el cual se puede ver reflejado en los RIPS, este ascenso ha tenido impacto en diferentes frentes dentro de los cuales se encuentra aquel que da inicio al proceso interno que ocurre en todo este ciclo de atención, este hace referencia aquel en el cual llegan todas las entidades y esperan ser asignadas, esta asignación más conocida como programación del bloque quirúrgico, en su gran mayoría es ejecutada de forma manual o semiautomática. Uno de los principales problemas existentes para el agendamiento el bloque quirúrgico es la rigidez existente en diferentes algoritmos y la incertidumbre de los tiempos de intervención, existen diferentes formas de abarcar el problema para la asignación del quirófano, dentro de la literatura existen diversos métodos, unos mas enfocados al tema de la asignación como problema de óptimos y otros mas enfocados hacia la reducción de las incertidumbre existentes en el tiempo de cirugía para la asignación. En este caso se ha tomado el problema bajo un enfoque mas complejo planteando una forma de reducir la incertidumbre en la variable tiempo por medio de métodos de clusterizacion sugiriendo la adición de variables externas para generar un modelo integral, y por otro lado, una segunda parte que se encarga de un planteamiento para adaptar diferentes algoritmos de agendamiento, en este caso un particular que surge como un planteamiento novedoso.

Palabras clave: Clusterización, programación, intervenciones, quirurgico.

ABSTRACT

In recent years, the rates of scheduled surgical interventions have had a considerable rise, which can be seen reflected in the RIPS, this rise has had an impact on different fronts, including the one that starts the internal process that occurs in This entire cycle of care, this refers to the one in which all entities arrive and wait to be assigned, this assignment better known as programming of the surgical block, the vast majority is executed manually or semi-automatically. One of the main problems for scheduling the surgical block is the existing rigidity in different algorithms and the uncertainty of the intervention times, there are different ways of covering the problem for the allocation of the operating room, within the literature there are various methods, some more focused on the issue of allocation as a problem of optimal and others more focused on reducing the existing uncertainties in the surgery time for allocation. In this case, the problem has been taken under a more complex approach, proposing a way to reduce the uncertainty in the time variable by means of clustering methods, suggesting the addition of external variables to generate an integral model, and on the other hand, a second part which is in charge of an approach to adapt different scheduling algorithms, in this case a particular one that emerges as a novel approach.

Key words: Clustering, programming, interventions, surgical.

CONTENIDO

ÍNDICE DE TABLAS

ÍNDICE DE GRÁFICOS

ÍNDICE DE FIGURAS

ÍNDICE DE ECUACIONES

1. DEFINICIÓN DEL PROBLEMA

ESTADO ACTUAL DEL SISTEMA DE SALUD EN COLOMBIA

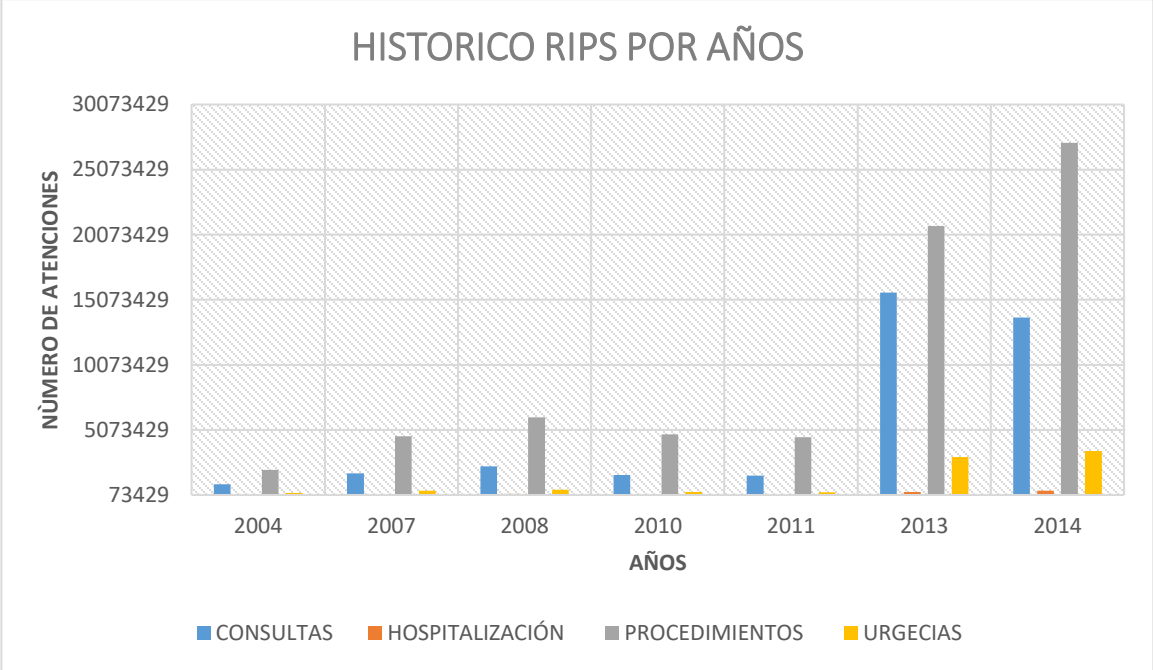
El estado de salud actual en Colombia ha tenido una serie de cambios relevantes los cuales pueden ser descritos en primer lugar por el incremento en cuanto a demanda y en segundo lugar por las reformas a la salud, en este caso específicamente la ley 100, ha sufrido una serie de transformaciones en relación a la prestación del servicio, enmarcadas en una serie de decretos y manuales los cuales detallan y definen los respectivos procedimientos del plan obligatorio de salud (POS). Algunos cambios importantes en cuanto al actual sistema de salud, han sido la descentralización en la administración de la salud respecto al modelo anterior el cual era administrado por una sola entidad (seguro social), la libre elección del cotizante a la entidad que le prestará el servicio, es decir diversidad entre las entidades administradoras; en cuanto a los cambios referentes al POS estos son registrados en los manuales de actividades los cuales se actualizan de forma periódica cada dos años (Ley Nº 1438, 2011). El contenido de dicho manual presenta algunas similitudes respecto a su antecesor; y otras diferencias en cuanto a su financiación, siendo esta efectuada por medio del pago de "cuotas moderadoras". El modelo actual del POS reproduce un modelo de aseguramiento donde todos los afiliados tienen acceso al plan integral de salud según el artículo 156 de la presente ley (Bernal Oscar, 2012) "[…] Todos los afiliados al sistema general de seguridad social en salud recibirán un plan integral de protección de la salud, con atención preventiva, médico-quirúrgica y medicamentos esenciales, que será denominado el plan obligatorio de salud" (Ley Nº 100, 1993). La serie de problemáticas alrededor del sistema de salud en el país en cuanto a la oportunidad con la cual se presta el servicio se relaciona con la demanda y la gestión de los recursos hospitalarios para suplir la misma, la cual va de la mano con la serie de cambios que ha tenido en cuanto a la regulación el mismo.

Como se puede observar en la *Gráfico1* el área que más crecimiento ha tenido en los últimos años según el reporte de RIPS ha sido el área de intervenciones quirúrgicas, cuyo desempeño se ve afectado por la gestión adecuada de los recursos, según (Estupiñán et, 2016) el departamento de cirugías en Colombia genera más del 40% de la las ganancias totales en los hospitales "este a su vez es uno de los más costosos debido a que representa el 9% del presupuesto anual del

hospital" (Estupiñán et al., 2016) por ello la adecuada programación de los recursos que serán destinados al bloque quirúrgico se hace un evento crítico.

Según lo descrito por (Calderón, 1996) sistema actual de salud en Colombia se encuentra afrontando nuevos retos, debido a la demanda que ha ido en incremento, como se puede observar en la *Gráfico1* la cual presenta para los últimos periodos una demanda en ascenso, que en cada avanza duplicando los años anteriores; lo cual quiere decir que el sistema de salud se enfrenta a nuevos desafíos en cuanto a prestación, por su alta demanda, teniendo que adaptarse a los nuevos requerimientos del servicio en términos de calidad y oportunidad. (Bernal Oscar, 2012)

Gráfico 1 Histórico del incremento de la demanda



Fuente: elaboración propia. Datos: Secretaria de salud.

Los datos presentados en la tabla anterior constituyen la demanda general atendida de la población vinculada los cuales hacen parte del plan obligatorio y cotizante; esta información es validada por la SDS en número de atenciones. Con base en la anterior ilustración y los datos suministrados por la Secretaria de Salud se puede inferir que, para cada uno de los periodos las intervenciones o procedimientos son una de las variables que más crecimiento ha tenido durante los últimos años, con

un total atendido de 1’249.153 pacientes entre 2004, 2007, 2008, 2010, 2011, 2013 y 2014.

DESCRIPCIÓN DEL PROBLEMA

Debido a que los centros médicos no responden a menudo satisfactoriamente con la demanda requerida (Durán, Rey, & Wolff, 2017) e incluyendo el incremento que ha tenido la misma durante los últimos periodos, tal como se puede observar en la *ilustración 1*, se ha hecho critico el proceso de programación para el bloque quirúrgico, el cual debe asegurar el debido cumplimiento de todas las restricciones de personal y equipo, las cuales garanticen la atención oportuna de los usuarios (Spratt & Kozan, 2016).

La programación de procedimientos del bloque quirúrgico es un conjunto de elementos, los cuales deben coordinarse de forma adecuada teniendo en cuenta cada uno de los factores que pueden llegar a afectar la prestación en calidad del mismo (Latorre-Núñez et al., 2016), es por esto que dentro del sistema existe una serie de variables y restricciones, que afectan la programación del bloque quirúrgico como lo dice (Marty, 2019) La realización de una intervención quirúrgica necesita la presencia simultánea de varios profesionales, para (Zhang, Dridi, & El Moudni, 2019) y (Santoso, Sudiarso, & Herliansyah, 2017) es importante tener en cuenta las limitaciones en el problema de programación del bloque quirúrgico, dentro de las cuales la que más se destaca es la incertidumbre de la duración; por estos y otros factores se hace un sistema complejo a la hora de organizar. La utilización inadecuada de los recursos puede generar retrasos los cuales se ven reflejados en un mayor tiempo de espera del paciente (Akbarzadeh, Moslehi, Reisi-Nafchi, & Maenhout, 2019); el objetivo al realizar la programación del bloque quirúrgico es lograr el equilibrio de la capacidad y la demanda, ya que es necesaria la prestación del servicio de forma oportuna, esta debe minimizar los costos relacionados a la gestión de dichos elementos (Yin, Zhou, & Lu, 2016), puesto que se ha generado un incremento al no mejorar el tiempo extra del personal (Al Hasan, Guéret, Lemoine, & Rivreau, 2019) o como lo es para (Shafaei & Mozdgir, 2019) uno de los factores más importantes, el cual debe también tenerse en cuenta, es el tiempo libre dentro de las salas de cirugía el cual está relacionado al porcentaje de utilización de los quirófanos.

Por lo tanto, debido a la creciente demanda y a la serie de restricciones existentes en la programación del bloque quirúrgico se hace necesario brindar una atención de

alta calidad de manera más eficiente la cual tenga en cuenta para su organización los recursos limitados como lo describe (Abedini, Li, & Ye, 2017) y (Zhou & Yin, 2016).

En la mayoría de las ocasiones la mala gestión de los recursos ha generado bloqueos por la falta de disponibilidad (Abedini, Li, & Ye, 2017) está a la hora de la programación de las intervenciones quirúrgicas debe tener en cuenta diferentes aspectos como ya se ha dicho anteriormente para el desarrollo de un horario quirúrgico maestro (M'Hallah & Visintin, 2019). Las variables que afectan y restringen el problema generan un complejo mecanismo en el cual algunas de las soluciones que parecen ser ineficientes, en algunas ocasiones comparando ciertos algoritmos para la solución que se adapten y disminuyan la variabilidad y ayuden en la prestación oportuna del servicio.

En hospitales de alta complejidad se hace crucial realizar todo este procedimiento de forma óptima ya que la gestión de ciertos recursos es aún más limitada, como ocurre en el caso del hospital San José, para el cual la asignación del horario maestro de quirófano se realiza de forma manual, utilizando métodos heurísticos para la secuenciación del MSS.

PREGUNTA PROBLEMA

La programación de las intervenciones quirúrgicas es una tarea que depende de todo el equipo administrativo el cual está sujeto a factores como, disponibilidad de salas y personal, tiempos de procesamiento, tiempos de intervenciones, etc. Es decir, diferentes factores que intervienen en la programación, muchos de los cuales presentan una alta incertidumbre (Santoso, Sudiarso, Masruroh, & Herliansyah, 2017), por lo tanto el uso deficiente de la información cuantitativa para la gestión de recursos físicos y humanos en servicios quirúrgicos ha generado una serie de problemas a la hora de disponer del quirófano, algunas de las causantes de la mala gestión de la información cuantitativa son asociadas a la metodología utilizada para la recolección de datos, junto con la forma en la cual son procesados; es decir, el modo en el cual dichos elementos son tratados para la solución del problema de programación quirúrgica, por lo tanto, dentro de la entidad se generan una serie de congestiones, tiempos muertos e incidencia en costos innecesarios, los cuales están ligados a dicha programación.

En casos de hospitales de alta complejidad como lo es el del hospital San José, en el cual este complejo proceso de gestión de los recursos en la asignación del bloque quirúrgico se hace de forma manual, hay una mayor probabilidad de cometer errores

al no contar algunas restricciones que hacen parte del sistema (Benchoff, Yano, & Newman, 2017). Existen diferentes métodos aplicados en la programación de quirófanos, y el método utilizado por el hospital San José es uno de los más recurrentes, aunque según lo encontrado en la literatura existen técnicas que generan un resultado más satisfactorio que la heurística, por lo tanto, hay una forma más adecuada de realizar el proceso de programación del bloque quirúrgico en el hospital San José, utilizando diferentes herramientas que se adapten a la demanda de sistema disminuyendo su variabilidad. Aunque uno de los métodos encontrados en la revisión sistemática es la programación lineal de enteros mixtos, en este caso se pretende usar una secuenciación por reglas de negocio utilizando la clusterización para la disminución de la incertidumbre; dando como resultado un modelo orientado a la naturaleza probabilística del sistema y su alta variabilidad, el cual se espera tendrá un mejor comportamiento que el método empleado en la actualidad.

¿La clusterización logra reducir el grado de incertidumbre en cuanto a tiempos de intervención para la realización del agendamiento del bloque quirúrgico y como puede aplicarse esto directamente sobre el agendamiento? Esta será la pregunta que se espera resolver por medio del presente proyecto.

2. JUSTIFICACIÓN

La ingeniería industrial ha tomado relevancia en el campo de la salud (Healthcare Engineering), en el cual se han aplicado diferentes técnicas con el objetivo de mejorar los tiempos de servicio, gestión interna de los datos o procesos, e incluso en el aseguramiento en la prestación de un servicio de calidad, esto ha generado que surjan nuevas áreas de investigación en ingeniería industrial orientadas en salud, tales como lo es la ciencia de datos aplicada en diferentes contextos. Por lo tanto en la presente tesis se presentara y aplicaran algunas de estas herramientas, las cuales son suministradas por la ingeniería industrial, de la mano con otros campos de la ingeniería, para el desarrollo de una solución al problema descrito anteriormente, asumiendo los nuevos retos de la ingeniería industrial en el área de la salud, en la cual la Universidad Santo Tomás debería iniciar su proceso de investigación asumiendo los retos que propone el futuro de la ingeniería industrial.

El presente proyecto tendrá un impacto en la forma en la cual se realiza la programación de cirugías en el hospital San José, a través de la integración de diferentes herramientas por medio de las cuales se pretende simplificar los procesos para la programación de las intervenciones quirúrgicas, desde el punto de vista del hospital este disminuirá las horas hombre aplicadas para la ejecución de todo el proceso de programación, también se espera que disminuya la congestión en el procesamiento de solicitudes, realizando una gestión adecuada de los recursos limitados.

La ejecución del proyecto contribuirá a la asignación y preparación del bloque quirúrgico puesto que el elemento central de la misma es la reducción de la incertidumbre en los tiempos de intervenciones quirúrgicas, aplicando técnicas desde la ingeniería industrial de manera transversal a otras áreas aplicadas del conocimiento.

3. OBJETIVOS

3.1. General

- Desarrollar un algoritmo de estimación de la variabilidad que afecta los tiempos quirúrgicos en las intervenciones, basado en técnicas de Clusterización para generar una programación eficiente del bloque quirúrgico.

3.2. Específicos

- Comprender la naturaleza de la información por medio de un análisis estadístico estructurado con el fin de proyectar el algoritmo
- Desarrollar un algoritmo que tenga que tenga la capacidad de agrupar procedimientos quirúrgicos con el fin de estimar la variabilidad haciendo uso de técnicas de Clusterización.
- Evaluar la utilidad da la Clusterización para el desarrollo un agendamiento del bloque quirúrgico por medio del análisis de los procesos y sus tiempos.

4. MARCO REFERENCIAL

4.1. MARCO CONCEPTUAL

Para comprender la complejidad del sistema es necesario tener claridad respecto a algunos conceptos en el contexto del presente proyecto definidos por diferentes autores los cuales servirán de apoyo para el desarrollo de la metodología.

Prestación oportuna del servicio: la oportunidad es la posibilidad que tiene el usuario de obtener los servicios que requieren sin que se presenten retrasos que pongan en riesgo su vida o salud (Ley N° 100, 1993).

Bloqueo en quirófanos: es definido como la acción de una asignación ineficiente de los recursos para la prestación oportuna del servicio el cual genera bloqueos debido a las limitaciones del sistema (Abedini, Li, & Ye, 2017).

Programa maestro de cirugía (MSS): o por sus siglas en ingles Master schedule surgeri según (Akbarzadeh, Moslehi, Reisi-Nafchi, & Maenhout, 2019) el programa maestro de cirugía MMS se usa en aquellos casos en los cuales se requiere construir un sistema de reserva de bloque el cual sirve como base para la programación de los casos quirúrgicos.

Paciente electivo: los pacientes electivos según lo descrito por (Shafaei & Mozdgir, 2019) son aquellos que son programados con anticipación, los cuales obedecen a un tiempo de espera el cual es determinado por la congestión dentro del sistema y la prioridad de atención; son aquellos pacientes que no tienen algún tipo de urgencia quirúrgica.

Tiempos quirúrgicos: para el caso de estudio se definirán los tiempos quirúrgicos según lo describe el autor (Choque López, 2011), estos inician con la preparación del paciente que abarca todo el proceso preoperatorio, en el cual se tienen en cuenta las características físicas y psicológicas del mismo; abarcando todos los procesos de preparación de quirófano y del personal hasta el momento en el cual el sujeto es trasladado para su recuperación, es decir no es incluido el tiempo de recuperación dentro del tiempo quirúrgico.

Programación lineal: este tipo de problemas tienen como objetivo la asignación de recursos entre los sujetos que compiten entre sí por estos, para ello el modelo de programación lineal planea los recursos de la forma más óptima, teniendo como característica principal el modelado de funciones lineales (Hillier, Lieberman, & Osuna, 1997).

Modelo de programación lineal de enteros mixtos: en este caso se plantea el modelo para la programación lineal como se desarrolla normalmente en el cual a la asignación se realizará por medio de la combinatoria de diferentes valores tomando como restricción únicamente los números enteros (Bravo Bastidas, 2013).

Programación por objetivos: en la cual se jerarquiza cada uno de ellos realizando una ponderación de los mismos, el enfoque básico de este tipo de programación es establecer un objetivo numérico para cada uno de los objetivos, formulando la función objetivo de tal forma que la desviación de las funciones para las metas específicas sea minimizada (Chavez Quisbert, 2011)

Estocástico: para (Hillier, Lieberman, & Osuna, 1997) los procesos estocásticos son "todos aquellos que evolucionan con el tiempo de una manera probabilística", por ejemplo, la llegada de pacientes al área de urgencias es un proceso estocástico y el fon es predecir la forma en la cual evolucionara dicho proceso.

Cadena de Markov: las cadenas de Markov según (Hillier, Lieberman, & Osuna, 1997) hacen parte de los procesos estocásticos, las cuales tienen la propiedad de predecir el comportamiento futuro teniendo dependiendo solo del estado actual del proceso, por ejemplo la predicción del clima basándose en el comportamiento de este el día antes, por lo tanto si mañana llueve o hace sol el proceso será independiente de las demás ocurrencias pasadas; este tipo de herramientas pueden ser usados en modelos heurísticos.

Heurística: en primer lugar se iniciará con la definición de heurística puesto que gran parte de los modelos que se han desarrollado para la solución del problema de programación del bloque quirúrgico parten de este método para el diseño una solución; heurístico es una técnica utilizada para la solución del problema haciendo uso del razonamiento estadístico para realizar una estimación o predicción (Novo, Arce, Fariña, & a de Vega, 2003), cuando se hace referencia a un modelo heurístico se habla de la simplificación de dicha cuestión por medio de atajos cognitivos, los cuales ayudan en el análisis de una gran cantidad de información (Novo, Arce, Fariña, a de Vega, 2003). En el caso de la programación de quirófanos, el método heurístico es utilizado por los responsables de la asignación del bloque quirúrgico, los cuales basados en el proceso histórico desarrollan la solución más simple, que en algunas ocasiones no es la óptima.

Algoritmo de búsqueda local: este tipo de algoritmos están diseñados para encontrar el óptimo dentro de un área específica, es decir, en caso que exista más de un punto óptimo sea máximo o mínimo dentro de una serie de datos, el algoritmo solo arrojara el punto óptimo en el área que realice la primera iteración al azar y esto ocurrirá pero de forma manual cada vez que se itere, es la forma en la cual funcionan algoritmos como descenso gradiente.

Algoritmos Metaheurísticos: para (Morillo, Moreno, Díaz, 2014) una meta heurística es un método heurístico en el cual se añade una inteligencia que escape a los óptimos locales, con el objeto de proporcionar una solución general por medio de diferentes iteraciones en la cual busca el punto más óptimo (Hillier, Lieberman; Osuna, 1997).

Búsqueda tabú: la búsqueda tabú es una de las Metaheurísticas más utilizadas puesto que utiliza algunas de las ideas del sentido común para permitir un proceso de búsqueda de escape del óptimo local (Hillier, Lieberman, Osuna, 1997) mediante el uso de la información recolectada, este algoritmo inicia con una solución inicial aleatoria, posteriormente define un vecindario local y realiza el nuero de iteraciones que sea necesario para llegar al optimo realizando el recorrido de toda la función (Morillo, Moreno, & Díaz, 2014).

Algoritmos genéticos: según (Morillo, Moreno, & Díaz, 2014) se basa en el concepto de evolución de Darwin en el cual las especies que cumplan con las características que requiere el espacio sobrevivirán; en este tipo de algoritmos en lugar de ser evaluadas cada una de las posibles soluciones de forma individual y su evolución en la estructura en la cual se comporta el sistema, como ocurre en el algoritmo de templado, en el cual las estructuras se reorganizan al hacer el temple, en este caso son evaluado todo el conjunto de posibles soluciones como familias en el cual solo sobreviven aquellas con las características óptimas para la solución del problema.

Clusterización: según (Gallardo, 2010) es un método estadístico multivariante que inicia con un conjunto de datos contenido sobre una muestra de entidades para su organización en o grupos relativamente homogéneos a los que se llamaran Clusters

4.2. ANTECEDENTES

Al realizar la revisión de la literatura se encontraron dos métodos aplicados con mayor frecuencia para la solución de este tipo de problemas. Métodos heurísticos y metaheurísticos, cada uno de los cuales incluye diferentes herramientas y soluciones para el mismo problema; estos se muestran en la siguiente tabla.

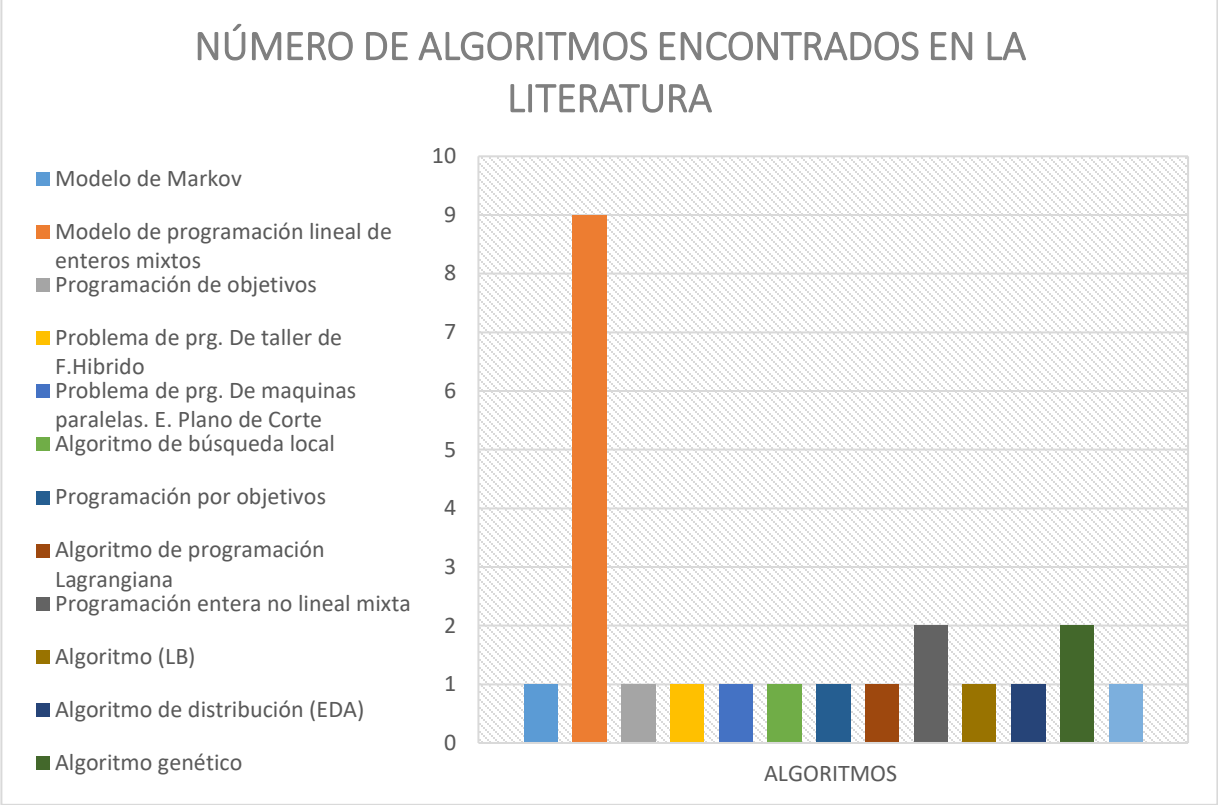
Gráfico 2 Métodos encontrados en la literatura

Fuente: elaboración propia. Datos: Scopus.

Dentro de la literatura encontrada se hallaron dos métodos comúnmente utilizados en la solución de problemas cuyo espacio de posibles soluciones se reduce a calcular un óptimo (véase *anexo1 Ecuaciones de búsqueda*); en el *gráfico2* se puede observar que dentro de los métodos encontrados en la revisión sistemática, los utilizados con mayor frecuencia para la solución de este tipo de problemas, están concentrados en métodos heurísticos, que en comparación con los procesos Metaheurísticos presentan una desarrollo que reduce los costos de procesamiento, aunque debido a la naturaleza del problema los métodos Metaheurísticos son una solución más adecuada.

Dentro de los dos métodos descritos anteriormente existen diversas formas de plantear el problema como modelo para abordar la mejor solución, en la literatura se encontró que los métodos más usados para abordar la programación de quirófanos es programación lineal de enteros mixtos, el cual presenta una solución más sencilla al problema, aunque no disminuye la variabilidad del sistema según los resultados comparados con algunos autores.

Gráfico 3 Algoritmos usados con más frecuencia en la literatura



Fuente: elaboración propia. Datos: Scopus.

Para (Wang, Meskens, & Duvivier, 2015) la programación lineal por enteros mixtos tiene un mejor comportamiento que la programación por restricciones, a partir de esto se puede inferir que debido a la incidencia de estos en la literatura los métodos heurísticos tienen un mejor desempeño respecto a los Metaheurísticos en algunos casos. Para este caso el autor puso a prueba los dos tipos de algoritmos los cuales se sitúan entre heurísticos y metaheurísticos, encontrando un mejor desarrollo del problema al aplicar el primero, también se puede encontrar esta misma hipótesis en (Castro & Marques, 2015) donde el autor hace una prueba cuyo objetivo es el mismo entre algoritmos de programación lineal de enteros mixtos y algoritmos genéticos, los resultados encontrados fueron los mismos para cada caso de programación del bloque quirúrgico, en el cual el desempeño de los algoritmos heurísticos en la solución de este tipo de problemas supera de cierta manera los Metaheurísticos

A priori se podría llegar a la conclusión de que para la solución de este tipo de problemas lo mas acertado es optar por los algoritmos heurísticos, los cuales presentan un modelo más simple de desarrollar, al ser formulados a través de un

algoritmo que se ajuste de forma adecuada a la demanda del sistema reduciendo la variabilidad, a pesar de esto en los pocos casos encontrados en la literatura en los cuales se utilizó un método Metaheurístico, se encontró que para el planteamiento del modelo, el mismo era descompuesto en varias partes las, cuales componen un conjunto para la solución óptima; aunque se hace más compleja de estructurar, presenta resultados que se reducen a la variabilidad del sistema por lo tanto la incertidumbre del tiempo de intervención, como lo es el caso de (Soudi & Heydari, 2019) en el cual el problema es descompuesto con el objeto de implementar búsqueda tabú para el desarrollo del problema. O como es el caso de (Cappanera, Visintin, & Banditori, 2018) en donde el autor disgrega el problema para ejecutar un análisis de correlación entre cada una de las variables con el objeto priorizar aquellas que tienen mayor incidencia para la realización del MSS.

En otros casos se halló la hibridación entre un modelo de programación lineal por enteros mixtos con algoritmos genéticos para la asignación de quirófano, es decir dependiendo de la forma en la cual fuera abarcado el problema desde la Metaheurística pueden concebirse diferentes soluciones, que, aunque presenten un mayor nivel de complejidad tendrán una mejor adaptabilidad a las condiciones que requiere el sistema. Dentro de la literatura hasta el momento se ha encontrado un solo caso en el cual se utilice la Clusterización en el desarrollo de este tipo de problemas, en este caso se asumió que la mayor incertidumbre existente es el tiempo de intervención quirúrgica, para lo cual se realizó un análisis jerárquico de las mismas con el fin de determinar aquellas variables que tendrían más peso a la hora de estimar el tiempo de intervención para la programación del bloque quirúrgico.

Los algoritmos genéticos resultan ser una solución muy efectiva a los problemas en los cuales las variables y la asignación de recursos se hacen extensivos, para estos se encuentran varios óptimos locales, ya que este tipo de algoritmos evalúan todas las posibles soluciones al mismo tiempo evaluando cada uno de los individuos o variables y su adaptabilidad (Kuri & Galaviz, 2002). Los algoritmos de aprendizaje automático son una solución poco explorada para el problema de la asignación quirófano, aunque según los resultados y las conclusiones obtenidas por (Santoso, Sudiarso, Masruroh, & Herliansyah, 2017) el esperado es satisfactorio, al realizar un clúster para cada uno de los grupos priorizando aquellos que tienen mayor efecto sobre los tiempos.

En conclusión el método que se utilizara en el presente proyecto, es un método que según la literatura, ha sido poco explorado, y los métodos tradicionales algunos hibridados con los no tradicionales dan como resultado una solución satisfactoria al sistema, pero esto no resuelve en totalidad los problemas de la alta variabilidad de la demanda, por lo que los modelos tradicionales deben ser expuestos a ajustes, limitando sus opciones para definir diferentes soluciones sin incurrir en el problema

de los mínimos locales; por otro lado aquellos que si presentan de forma un mejor nivel de adaptabilidad son los metaheurísticos, que hasta el momento han sido los menos explorados en cuanto a algoritmos genéticos y de aprendizaje no supervisado.

4.3. MÉTODOS APLICADOS

El método será planteado como un problema metaheurístico de aprendizaje no supervisado el cual hará uso de la Clusterización para el desarrollo del problema; paso seguido se espera reducir la **incertidumbre** del proceso, se pretende agrupar los objetos en clases o Clusters, de forma que estos tengan una alta similaridad entre ellos y una muy baja entre objetos de otros Clusters, para esto se tendrá cuenta la estructura de datos la cual corresponde a una forma matricial con el objeto de encontrar la distancia mínima entre los componentes de dicha matriz; por ejemplo las variables tiempo de cirugía y especialidad, pueden ser representadas respectivamente como vectores los cuales componen la matriz A cuyas dimensiones serán *x* por *y* (Martínez, 2011) estas están ordenadas en *y* columnas por *x* filas donde cada una de las variables representa respectivamente las especialidades y los tiempos de intervenciones quirúrgicas.

A partir de dicha matriz de datos lo siguiente será calcular la matriz de distancias sobre la cual se realizará la relación entre cada una de las variables por las cuales esta compuesta la base de datos para ello es necesario realizar un análisis entre las distancias, lo primero será calcular la desviación respecto la media entre cada uno de los datos que componen la matriz

Ecuación 1 Cálculo de la desviación respecto a la media

$$S_f = \frac{1}{n}\left(\left|x_{1f} - m_f\right| + \left|x_{2f} - m_f\right| + \cdots + \left|x_{nf} - m_f\right|\right)$$

Fuente (Millan, 2010)

Donde $\left|x_{nf} - m_f\right|$ representa el número de *n* medidas de *f* y $\left|m_f\right|$ es el valor medio de la medida. Donde m_f se define como:

Ecuación 2 Calculo del valor medio

$$m_f = \frac{1}{n}\left(x_{1f} + x_{2f} + \cdots + x_{nf}\right)$$

Fuente (Millan, 2010)

Ecuación 3 Calculo de la media estándar

$$Z_{if} = \frac{x_{if} - m_f}{S_f}$$

Fuente (Millan, 2010)

Las ecuaciones que serán presentadas a continuación son utilizadas para para medir las distancias entre cada uno de los objetos donde se describen dos tipos elementos que componen la matriz descrita en el inicio del ejemplo, utilizando la distancia Euclídea.

Ecuación 4 Calculo de la distancia Euclidiana

$$d(i,j) = \sqrt[q]{\left(\left|x_{i1} - x_{j1}\right|^q + \left|x_{i2} - x_{j2}\right|^q + \cdots + \left|x_{ip} - x_{jp}\right|^q\right)}$$

Fuente (Millan, 2010)

Donde para este caso $(q \in \mathbb{Z} \geq 0)$

Las ecuaciones que se presentan a continuación son utilizadas en aquellos casos en los cuales se espera obtener dos clases de dicha variable, estas clases se pueden representar como 1 y 0.

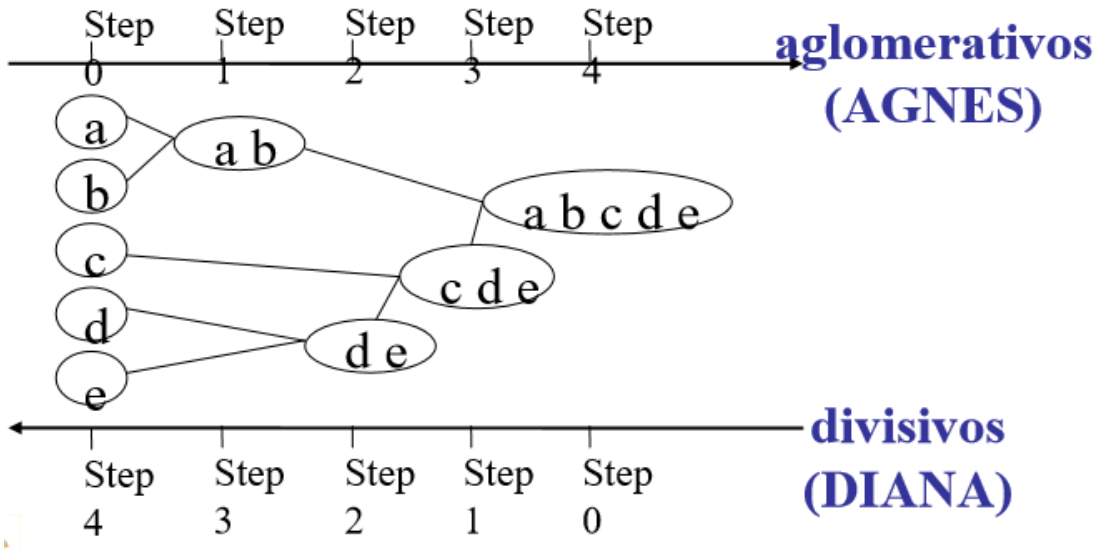
Ecuación 5 Coeficiente de Jaccard para enteros binarios

$$d(i,j) = \frac{b+c}{a+b+c+d}$$

Fuente (Millan, 2010)

Una vez determinada la distancia entre cada una de las variables se utilizará el método de k-means en donde cada clúster se representa por la medida central la cual en la primera iteración es seleccionada de forma aleatoria, es decir, por medio de la búsqueda de la distancia mínima entre cada uno de los datos se realizará el agrupamiento teniendo en cuenta los aquellos que presentan una mayor distancia entre cada uno de los grupos y una menor entre ellos; este valor será comparado con el método de clúster jerárquico en el cual no se requiere determinar el numero de Clusters o grupos que serán generados pero si la condición de terminación que presentará el número de clústers ; para este caso los datos son agrupados de forma jerárquica, de tal manera que al iniciar de abajo hacia arriba los mismos son descompuestos en variables que conforman el problema, de iniciar de arriba hacia abajo se hace una composición de cada una de las variables las cuales son la composición de las que tendrán un mayor efecto sobre los tiempos intervención.

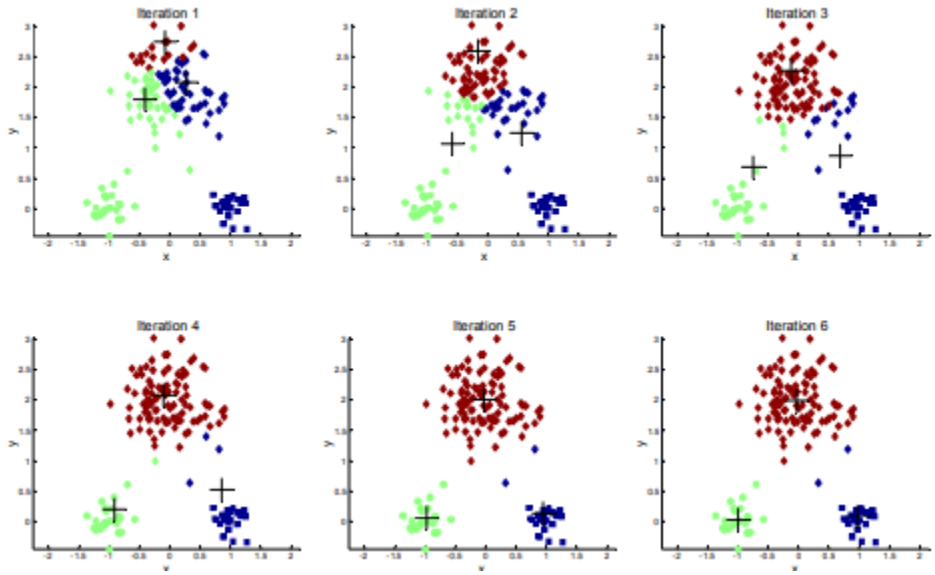
Figura 1 Ejemplo de clúster Jerárquico



Fuente (Millan, 2010)

En la tabla anterior se puede observar la forma en la cual opera el clustering por análisis jerárquico, el cual agrupa dependiendo de la incidencia que tienen los datos sobre las variables más generales.

Figura 2 Ejemplo de cluster por metodo K-Means



Fuente: (Guzmán, 2016)

Como es posible observar en la ilustración por medio de una serie de iteraciones el algoritmo de k-means llega finalmente al punto optimo por medio de una serie de iteraciones en las cuales reducen la distancia existente dentro de cada uno de los datos y aumenta la distancia entre los grupos que tienen menor similaridad, finalmente al obtener este grupo de datos es preciso que un experto lleve a cabo el etiquetado de cada uno de los grupos, las etiquetas para estos se definen a continuación.

Tabla 1 Características de las etiquetas

ETIQUETA	**CARACTERÍSTICA DE LA ETIQUETA**
Tiempo de intervención alto	Grupo estándar dentro de un tiempo superior clasificado entre la media de dicho grupo (*ecuación2*)

Espacio para tiempos intermedios de error	Espacio de error determinado para los datos atípicos los cuales están entre el valor de la media alta y media.
Tiempo de intervención medio	Grupo estándar dentro de un tiempo superior clasificado entre la media de dicho grupo (*ecuación2*)
Espacio para tiempos intermedios de error	Espacio de error determinado para los datos atípicos los cuales están entre el valor de la media media y baja.
Tiempo de intervención bajo	Grupo estándar dentro de un tiempo superior clasificado entre la media de dicho grupo (*ecuación2*)

Fuente: elaboración propia

El espacio para los datos atípicos se estructura dentro del cuadro como una medida de contingencia puesto que se espera que la misma clusterización se encargue de incrementar la distancia entre cada uno de los grupos y reducir la misma entre aquellos que presenten una mayor similaridad para la agrupación generando los clústers.

Para la evaluación de desempeño de cada uno de los métodos de clusterización (jerárquico y K.- Means) se utilizaran dos medidas de desempeño; aunque dentro de la literatura se encontraron métodos de validación externa y métodos de validación interna se han seleccionado las siguientes metodologías para la validación de la solución: el índice de Davis Boulding véase la ecuación 6, donde para k es el un número de clústers "σi es la distancia promedio entre cada punto en el clúster i y el centroide del clúster, σj es la distancia promedio entre cada punto del clúster j y el centroide del clúster, y $d(ci\ ,\ cj)$ es la distancia entre los centroides de los 2 clústeres" (Guzmán, 2016)

Ecuación 6 Coeficiente de Davis Boulding

$$DB = \frac{1}{k}\sum_{i=1,i\neq j}^{k} max\left(\frac{\sigma_i + \sigma_j}{d(c_i, c_i)}\right)$$

Fuente: (Guzmán, 2016)

.

La siguiente medida de desempeño es el coeficiente de Silhouette el cual mide la distancia promedio de x a todos los demás puntos en el mismo clúster y la distancia promedio de este punto aleatorio al clúster mas cercano

Ecuación 7 Desviación de la distancia respecto a x coeficiente de Silhouette

$$S(x) = \frac{b(x) - a(x)}{max\{b(x), a(x)\}}$$

Fuente: (Guzmán, 2016)

Ecuación 8 Coeficiente de Silhouette

$$SC = \frac{1}{N}\sum_{i=1}^{N} S(x)$$

Fuente: (Guzmán, 2016)

Tal que para cada uno de los casos se cumple que $(S|1 \geq S \geq -1)$ donde los resultados esperados responden a:

Tabla 2 Clasificación de resultados Silhouette

S= 1	S= 0	S= -1
Bueno	Indiferente	Malo

Fuente: (Guzmán, 2016)

La evaluación de los Clusters es necesaria para llevar a cabo el proceso de etiquetado y posteriormente la secuenciación a partir de reglas de negocio las cuales se definirán con base en los requerimientos de la organización, y del juicio de expertos.

4.4. MARCO LEGAL

4.4.1. Ley 100

El sistema de seguridad social integral en Colombia fue instituido por la ley 100 de 1993, el cual comprende un conjunto de normas, entidades y procedimientos, orientados a garantizar la calidad de vida, este es el sistema de protección social integral, junto con otra serie de normas acordes al aseguramiento de la dignidad y calidad humana. El sistema de seguridad social en Colombia está conformado por los sistemas de pensiones, salud y de riesgos laborales, junto con los servicios sociales complementarios (Ley N° 100, 1993)

La constitución colombiana reconoce la seguridad social como un servicio que está a cargo del Estado, es decir, su labor es asegurar el acceso de la población en general al sistema integrado, garantizando la promoción, protección y recuperación en salud. De esta manera la reforma de salud en el país se orientó en tres direcciones, la desmonopolización de la seguridad social en salud para los trabajadores del sector privado, dando la opción a los asalariados de elegir el prestador de salud; la presencia del sector privado en el área de pensiones dando la cual brinda al cotizante la capacidad de elegir una aparte del Instituto de Seguros Sociales (ISS), y la creación de un fondo de solidaridad, en el cual “los recursos provenientes del erario público en el sistema de seguridad aplicaran siempre a los grupos de población más vulnerables” (Ley N° 100, 1993); los principios generales de esta ley, los cuales son expresados en el capítulo 1 establecen que “el servicio público esencial de seguridad social se prestará con sujeción a los principios de eficiencia, universalidad, solidaridad, integralidad, unidad y participación” por lo tanto uno de los principales cambios para este fue el nuevo modelo de aseguramiento que parte de los principios de equidad, obligatoriedad, protección integral, libre elección, autonomía de instituciones, descentralización administrativa, participación social, concertación y calidad, asegurando que la población en general participe de este sistema en una de varias formas. Tomado de: (Calderón, 1996) adaptado de: (Ley N° 100, 1993)

4.4.2. Plan decenal

El plan decenal es producto del plan nacional de desarrollo y busca la disminución de la desigualdad en la prestación del servicio de salud, sin hacer distinción ante cualquier situación diferencial; para lograr este fin el plan decenal se plantea 8 dimensiones las cuales están dividas entre componentes sectoriales y transectoriales, estos hacen parte de un mismo conjunto, los cuales se dividen en,

dimensión salud ambiental, vida saludable y condiciones no transmisibles, convivencia social y salud mental, seguridad alimentaria y nutricional, sexualidad y derechos sexuales y reproductivos, vida saludable y enfermedades transmisibles, salud pública en emergencias y desastres, dimensión de salud y ámbito laboral. El plan decenal a su vez para el cumplimento de lo planteado está conformado por los siguientes objetivos: (Plan decenal, 2012)

- Avanzar hacia el goce efectivo del derecho a la salud.
- Mejorar las condiciones de vida que modifican la situación de salud y disminuyen la carga de enfermedad existente.
- Mantener cero tolerancias frente a la mortalidad y la movilidad junto con la discapacidad evitable.

En este se define que la salud pública es un pacto y mandato ciudadano el cual hace parte de un trabajo en conjunto con diferentes actores públicos y privados con el fin de garantizar el bienestar integral y la calidad de vida en Colombia los cuales están a fon con los objetivos de desarrollo del nuevo milenio de la ONU el cual está integrado por otros sectores junto con el sector salud. (Plan decenal, 2012)

4.4.3 Ley estatutaria

La ley estatutaria 1751 de 2015 tiene por objeto garantizar el derecho fundamental en la salud, en la cual se establecen los métodos de regulación y de garantías para la protección del derecho fundamental de la salud, referente a los derechos de la población en general y las obligaciones del estado en cuanto a las disposiciones para el acceso al sistema de salud, se instaura la salud como un derecho autónomo e irrenunciable en lo individual y en lo colectivo, por lo tanto, es deber del estado brindar el acceso al sistema de salud de manera oportuna asegurando la igualdad de trato y de “oportunidades en el acceso a las actividades de promoción, prevención, diagnóstico, tratamiento rehabilitación y paliación para todas las personas” (Ley N° 100, 1993), también es deber del estado garantizar la protección de sujetos de trato especial como se establece en el **artículo 11**. Este aplica a todas las entidades que puedan influir de forma directa o indirecta en la garantía del derecho fundamental de la salud, por lo tanto es deber de los sujetos acatar a las normas del sistema, contribuir solidariamente al financiamiento de los gastos demandados por la atención en salud y seguridad social. (Ley N° 100, 1993)

En lo referente a las garantías del Estado hacia la comunidad debe cumplir con una serie de obligaciones para garantizar los derechos de la población en general, los

cuales responden a los principios anteriormente nombrados por la ley 100 (Artículo 10-1751-2015). (Ley Estatutaria N° 1751, 2015)

4.4.4. Manual de actividades y procedimientos

El manual de procedimientos el cual es establecido a partir de la resolución 5261 de 1994, instaura que “se hace necesario expedir los manuales de Actividades, Intervenciones y Procedimientos con el fin de que sea utilizado en el Sistema de Seguridad Social en Salud, para garantizar, el acceso a los contenidos específicos del Plan Obligatorio de Salud, la calidad de los servicios y el uso racional de los mismos”, el cual es expedido con el objetivo de unificar los criterios en la prestación de servicios de salud dentro de la seguridad social en salud. (Ministerio de salud, 1994).

Dentro del manual de actividades y procedimientos se encuentra establecido la definición y contenidos de las actividades, intervenciones y procedimientos contemplados en el plan obligatorio de salud (capitulo 2) en él se encuentran contenidas la definiciones para determinar la calidad del servicio (artículo 22) y las definiciones para la aplicación del manual (artículo 23) dentro de esta se define como actividad o procedimiento quirúrgico a “[…] operación instrumental total o parcial de lesiones causadas por enfermedades o accidentes, con fines diagnósticos, de tratamiento o rehabilitación de secuelas”; del reconocimiento para la repetición de una intervención quirúrgica acerca de la cual se establece que para casos de complicación se realizará la intervención siendo está supeditada a orden previa por la autoridad medica competente de la EPS; respecto a la definición de cuidados intensivos y de las características de aquellos pacientes que serán admitidos en dicha unidad se establecen los criterios y términos para llevar a cabo dicha actividad (Res N° 5261, 1994).

En cuanto a tiempos y definiciones de internación y recuperación ante un evento quirúrgico; en primera instancia se define como internación al “l ingreso a una institución para recibir tratamiento médico y/o quirúrgico con una duración superior a veinticuatro (24) horas.” En consideración cuando el ingreso es inferior a este tiempo es considerado como una atención ambulatoria, para dicho acceso a intervención “[...] Salvo en los casos de urgencia, para la utilización de este servicio deberá existir la respectiva remisión del profesional médico”. Respecto a las salas de cirugías son establecidos los derechos de las mismas (artículo 46), en el cual se definen los derechos respectivos a las salas de cirugías en cuanto a “[…]la dotación básica del quirófano, los implementos, instrumental, ropas reutilizables o desechables, materiales, medicamentos y soluciones, oxígeno, agentes y gases anestésicos, los servicios de enfermería, esterilización, instrumentación, circulantes para el acto quirúrgico y anestésico, se incluye recuperación hasta por seis horas.”

5. MARCO METODOLÓGICO

5.1. DEFINICIÓN DE LA INVESTIGACIÓN

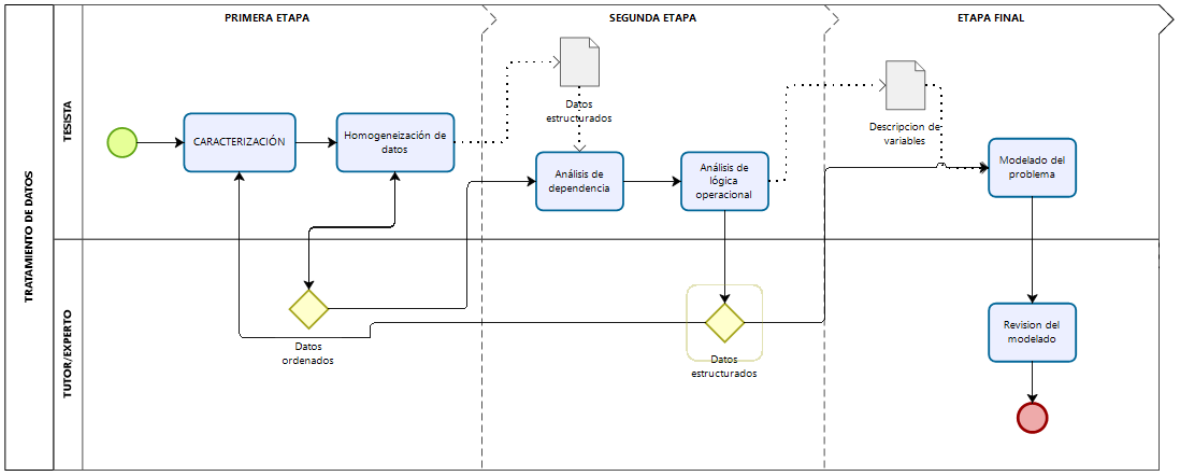
De acuerdo con lo encontrado durante la revisión sistemática se concluyó que existen una serie de variables que pueden afectar el comportamiento del sistema, la metodología sobre la cual estas son modeladas para la asignación de quirófano depende de la complejidad que se desea dar a la asignación de cada uno de los recursos y del manejo de las restricciones que ayudaran en el proceso de modelado, por lo tanto se define como una investigación correlacional puesto que se desea predecir el aproximado para una adecuada asignación del bloque quirúrgico, teniendo en cuenta la relación existente entre las variables, es decir, el efecto de cada una de ellas sobre el tiempo quirúrgico y como se correlacionan; por ejemplo, determinar el nivel de correlación existente entre la especialidad y el tipo de procedimiento, y el efecto de esta sobre el tiempo de intervención. En este caso se hará el modelamiento del problema partiendo de soluciones poco convencionales, poco probadas en la literatura, por lo tanto no hay una forma adecuada de evaluar en igualdad de condiciones, frente a otros métodos aplicados como los algoritmos metaheurísticos, por ejemplo, *Makespam* para la secuenciación y heurísticos como el *algoritmo de relajación lagrangiano* los cuales han aplicados en el contexto de asignación del bloque quirúrgico contra la clusterización para reducir la incertidumbre de tiempo, lo que hace a este tipo de sistemas aún más complejos y la acumulación de error es un factor crítico en cada una de las iteraciones para los clústers.

En la literatura se han encontrado algunos elementos que pueden ser de ayuda para la programación del bloque quirúrgico, dentro de ellos se encuentran algunas variables y restricciones que deberían ser tenidos en cuenta, las mismas corresponden a un tratamiento especial dependiendo del modelado, según lo describe (Spratt & Kozan, 2016) uno de los factores más importantes en el desarrollo del MSS son los tiempos quirúrgicos. Puesto que se analizarán una serie de datos que son el resultado de la colección dentro de una población definida y la prevalencia de dichos resultados sobre la misma, se ha definido como tipo transversal – analítico. Dando como resultado una investigación de tipo correlacional transversal analítica debido a cada una de las características previamente descritas.

5.2. RECOLECCIÓN DE LA INFORMACIÓN

Para este proceso se ha definido la entrada de datos con una serie de características o requerimientos mínimos para el caso de estudio, estos atravesaran por un proceso de evaluación y rectificación, con el objeto de asegurar una estructura correcta para que cada uno de ellos cuente con las variables estandarizadas requeridas, esto se encuentra sujeto al acuerdo de confidencialidad con la sociedad de cirugía del hospital San José.

Figura 3 Metodología para el tratamiento de datos



Fuente: elaboración propia

En el *cuadro1* se define el proceso para el tratamiento de los datos, estos son suministrados por el hospital, si los datos cumplen con una serie de características mínimas las cuales son requeridas para el desarrollo del modelo, véase la *tabla1*; una vez se ha realizada la recepción de los datos, estos pasaran por un proceso en el cual se garantiza que todos cumplen con las mismas características (homogeneización), es decir que no existe la repetición de términos que puedan pertenecer a una misma categoría y que los formatos correspondientes entre cada uno de los datos son homogéneos; posteriormente se realiza un análisis de dependencia entre cada una de las variables, ya que todas las bases de datos debido a la naturaleza del sistema tienen un comportamiento diferente, con los resultados del comportamiento de cada una de las variables se procederá con el análisis de la lógica operacional, dentro del cual se realiza el análisis de aquellas

variables que interactúan entre sí para su selección, suministrando finalmente los datos al algoritmo de clusterización.

El proceso previo a la recolección y tratamiento de los datos se espera lo realizará el hospital, ya que este es el encargado de alimentar sus bases de datos para la programación heurística que se lleva a cabo de forma interna, durante este proceso no se hará ninguna intervención con el objeto de no tener incidencia sobre los datos. El proceso de recolección de datos es realizado por parte del hospital en Excel por lo tanto el tratamiento de los mismos será realizado en este programa o en su defecto en SPSS.

Tabla 3 Descripción de variables para recolección de datos

VARIABLE	CARACTERÍSTICA
Hora de llegada del paciente.	La hora de llegada del paciente corresponde al momento en el cual es sujeto llega por el área de admisiones para su preparación en hh/mm/ss.
Hora de anestesia	Hora de inicio de la anestesia en hh/mm/ss.
Hora de ingreso al quirófano	Hora de entrada a sala de qx en hh/mm/ss.
Hora de limpieza de quirófano	Tiempo de limpieza en hh/mm/ss.
Genero del paciente.	Masculino/Femenino.
Medico	Caracterizado con un numero o serial.
Mortalidad/Mortandad	Falleció durante la intervención – SI/NO.
Sala asignada	Referencia de la sala a la cual se asigno el paciente para la intervención.
Programada	Si la intervención corresponde a una urgencia, o se trata de una cirugía programada.
Tipo de intervención.	Descripción cualitativa de la intervención realizada.
Especialidad	¦especialidad a la cual pertenece la intervención.
Oportuna.	Este campo permite saber si la programación fue oportuna – SI/NO.

Fuente: elaboración propia

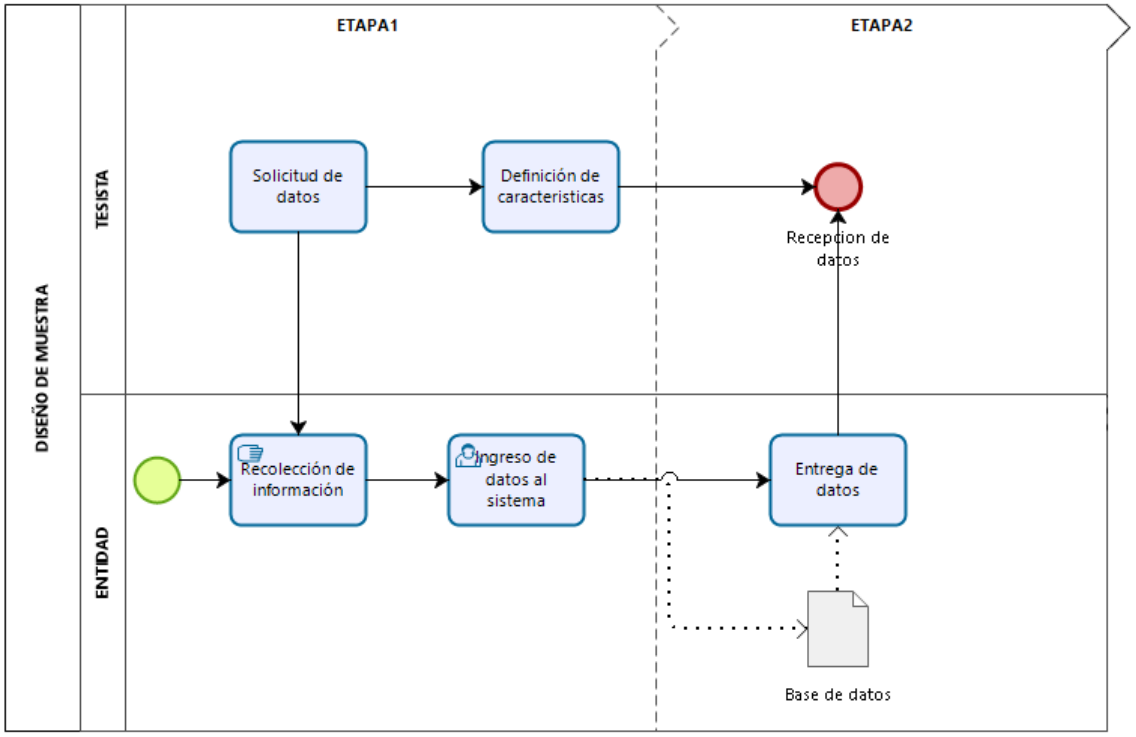
En la *tabla1* se presenta una descripción de las características de cada una de las variables dentro de la cual se especifica el formato en el cual la misma será admitida para el proceso de Clusterización, y las variables que serán permitidas para el estudio; durante el proceso de tratamiento de datos se hará una eliminación de aquellas que no correspondan o que no cumplan ciertas características.

5.3. DISEÑO DE MUESTRA

A pesar que los datos son de orden cualitativo existen variables que corresponden a un orden cuantitativo, las características definidas para la recolección de los datos dependen directamente de la entidad, estas son discriminadas al momento de la recolección de los datos, en el momento en el cual se realiza la extracción de la información del paciente para realizar la asignación de quirófano dependiendo de las características del mismo; por tanto se definirán únicamente las características que tendrá la muestra al momento de ser recibida para el proceso de tratamiento el cual se ha descrito anteriormente, es por esto que el tipo de muestra es de orden cualitativo, el cual corresponde a un muestreo determinístico o no probabilístico puesto que las bases de datos deben cumplir con una serie de criterios, los cuales se han definido en la *tabla1*; estas están sujetas a juicios y criterios subjetivos del evaluador o en este caso de quien hace la recolección de los datos del paciente, por ello para este caso se ha definido una modalidad intencional o por juicio, ya que para determinadas unidades de la población se concentra un alto porcentaje del valor de dicha variable.

El número de etapas que se ha definido para la muestra puede observarse en el *cuadro2* para el cual se han definido dos etapas, por lo tanto, esta selección tiene la característica de ser un muestreo polietápico. (INEGI, 2011)

Figura 4 Etapas para el diseño de la muestra



Fuente: elaboración propia

5.4. PROCEDIMIENTOS

5.4.1 Tratamiento de datos

El procedimiento para el análisis de los datos se ha descrito en la *figura1* en la cual se especifica el proceso para el tratamiento de los datos, estos deben cumplir con una serie de características de salida, al ser integrados en el algoritmo tales características responden al orden de sintaxis, orden y clasificación de cada una de las variables, los datos deben tener una estructura homogénea que evite que el algoritmo a diseñar genere la acumulación de errores que no corresponden al funcionamiento del mismo, estos serán dispuestos en forma de un archivo plano para el proceso de aprendizaje de la máquina.

Durante el proceso de caracterización se busca analizar a priori el comportamiento de los datos, para el diseño de la siguiente etapa, es decir, conocer las características que componen la base de datos, desde su formato hasta la estructura gramatical utilizada para alimentarla; posteriormente el proceso de

homogeneización es el momento en el cual se define la estructura fija que tendrán todos los datos de entrada para la integración en el algoritmo, al realizar este tratamiento es importante definir un estándar el cual servirá como guía para los demás datos de entrenamiento.

El análisis de dependencia ayuda en la definición de las variables que se tendrán en cuenta y la forma en la cual estas tendrán incidencia sobre el modelo, es decir, puede existir el caso en el cual hay etapas en las cuales algunas variables tengan mayor incidencia sobre otras debido a la naturaleza estocástica del problema, por ello nuevamente para no incurrir en errores que puedan afectar el aprendizaje del algoritmo se debe definir un estándar para concluir si el uso de la muestra es pertinente o no. Finalmente, al desarrollar este proceso se procederá a integrar los datos dentro del problema planteado, el proceso anteriormente descrito se repetirá con cada una de las bases de datos suministradas.

5.4.2 Calculo de clústers

El siguiente paso será calcular la media estándar con el objeto de hallar la distancia entre los datos para la elaboración del clúster, utilizando la *ecuación1* y *cuación2* por medio de las cuales se hará el cálculo de la media y de la desviación de la media, es decir el centroide y la distancia entre cada uno de los demás puntos para el agrupamiento.

Con el objeto de determinar cuál será el clúster que tenga mejor comportamiento, se seleccionaron dos tipos de clusterización Jerárquica y K-means. El primer paso para llevar a cabo dicha clusterización será calcular la distancia Euclidiana haciendo uso de la *ecuación4,* En el caso de los datos cualitativos tales como sexo, complicación o mortandad/mortalidad los cuales pueden ser representados como variables binarias en primer lugar se debe determinar si son binarias simétricas o asimétricas, es decir si ambos estados tienen la misma ponderación o si por el contrario ambos estados no son igualmente importantes, en cuyo caso el valor para la asignación de proximidad estará dado por el coeficiente de Jaccard. Una vez que se ha llevado a cabo el etiquetado de cada uno de los grupos generados, el cual se muestra en la *tabla2,* posteriormente se realizara la evaluación de desempeño aplicando los coeficientes de Davis Boulding y de Silhouette, los cuales se presentan dentro de la *ecuación6* y *ecuación7* respectivamente

Para la aplicación del coeficiente de Silhouette el resultado de desempeño del clúster se evalúa respecto a la *tabla3*.

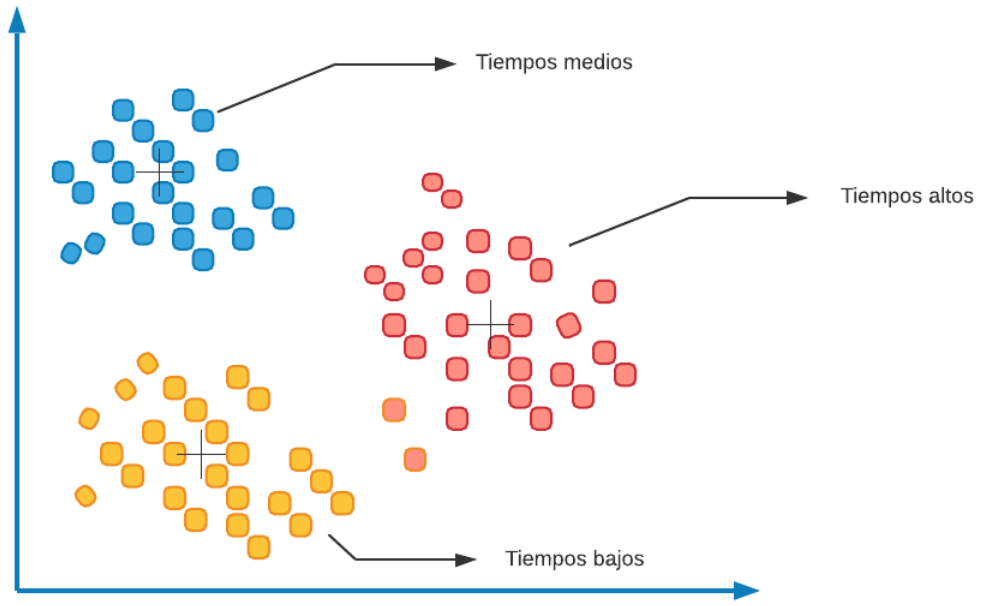
Se espera que la clusterización resuelva uno de los problemas más comunes encontrados en la revisión sistemática en cuanto a la incertidumbre del tiempo de intervención quirúrgica. Una vez obtenido el conjunto de datos a partir de la

clusterización se procede a la segunda etapa que consiste en el desarrollo de la secuenciación para la programación de intervenciones quirúrgicas por medio de reglas de negocio las cuales dependerán de los requerimientos del hospital y del juicio de expertos como ya se ha dicho, estos serán determinados durante el proceso de estudio y dependerán de las variaciones que tenga el sistema respecto al tiempo de desarrollo, este diseñado en C++. Se espera que esta haga la asignación de los recursos teniendo en cuenta las restricciones del mismo

5.4.3.Etiquetado

Para el proceso de etiquetado según cada uno de los grupos formados por el clúster se tendrá en cuenta el juicio de un experto con el fin de etiquetar cada uno de los grupos respecto a lo presentado en la *tabla2*.

Gráfico 4 Ejemplo de etiquetado

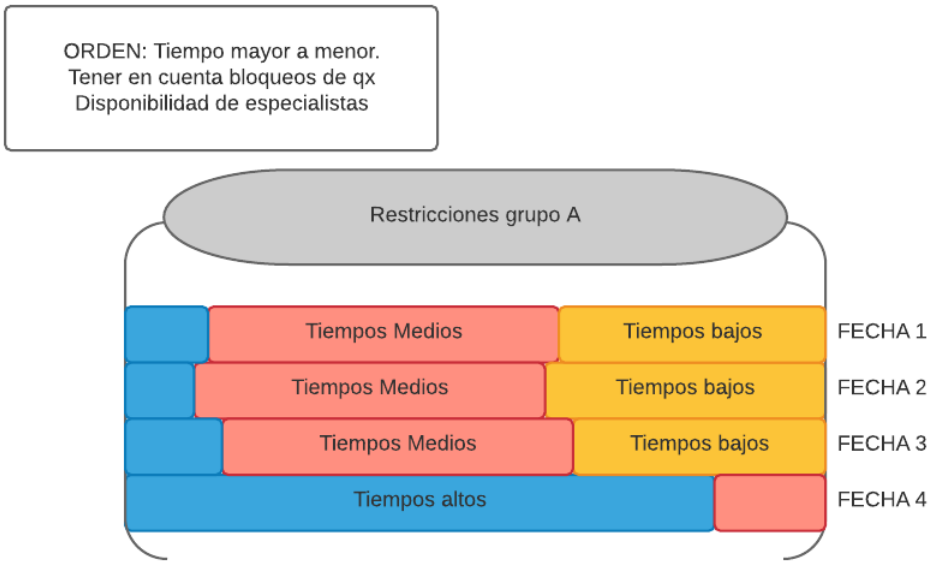


Fuente: Elaboración propia.

Tal y como se presenta en el grafico dependiendo de las características que tengan cada uno de los datos se realizará el proceso de etiquetado; partiendo de estas etiquetas se secuenciará a partir de reglas de negocio las cuales se explican de forma más clara dentro del *grafico5.* Para este ejemplo se puede observar como se han determinado una serie de variables que son tenidas en cuenta, en este caso se

programan de los tiempos de intervención más largos a los tiempos de intervención más cortos, este proceso de realizar una estimación de tiempo no está establecido bajo ningún estándar en el cual se instaure que para ciertas especialidades y ciertos procedimientos hay un tiempo estándar. El proceso de etiquetado por lo tanto es un paso crucial dentro del desarrollo del algoritmo puesto que determinara antes de la cirugía un estimado de tiempo para la asignación del bloque quirúrgico llamado comúnmente como agendamiento

Gráfico 5 Ejemplo de secuenciación



Fuente: Elaboración propia

En este ejemplo se ha determinado un orden y una serie de restricciones que son definidas por los requerimientos del hospital y de los recursos disponibles al momento de la asignación del bloque quirúrgico.

6. PRESUPUESTO

RUBROS	DESCRIPCIÓN	CANTIDAD	COSTO UND	MONTO TLT
Transporte	Se tendrán en cuenta solo aquellos pasajes que son de ida y regreso a la universidad, dependiendo de las ocasiones que sean requeridas	10	$ 2.400	$ 24.000
Equipos	El costo de procesamiento es demasiado alto, para ello se tendrá que comprar un equipo que cuente con la capacidad suficiente para procesar algunos datos	1	$ 3.450.000	$ 3.450.000
Fotocopias	Fotocopias y elementos de interés y para la presentación del proyecto que sean necesarias	250	$ 50	$ 12.500
Papelería	Cuaderno para el desarrollo matemático y apuntes para el proyecto	2	$ 3.800	$ 7.600
Servicios tecnicos	Google Colab es un ambiente de programación en el cual se ejecuta el código, en la nube almacena el código que se va desarrollando, cuenta con algunas librerías entre otros, debido a la alimentación del programa se hace necesario comprar más espacio y algunas extensiones	1024	$ 192	$ 196.608
Total		1287	3456442	$ 3.690.708

7. ANÁLISIS Y RESULTADOS

7.1 PROCESAMIENTO DE LA INFORMACIÓN

7.1.1 Anonimidad y preparación de la información

El proceso de anonimizacion de la data corresponde al proceso de tomar cada una de los componentes que puedan poner en evidencia datos particulares que atenten contra la privacidad de las personas que se encuentren registradas en dicha base, tales como: médico y paciente. En este caso es necesario hacer una diferenciación entre cada una de las componentes de estas variables sin que las mismas pongan en evidencia de forma discriminatoria un caso en particular.

En segunda estancia para la utilización de los datos es indiferente componentes de las variables tales como, nombre completo del paciente, en este caso será de utilidad una discriminación que haga referencia al sexo del paciente, ya que este si puede ser un dato de gran utilidad. Para el caso de médico tratante, por lo anteriormente ya descrito en cuanto al tratamiento de la información y el respeto a la privacidad existente en los registros, fue necesario realizar una codificación que no hace alusión al nombre de un médico en particular si no simplemente discrimina que el paciente “A” fue tratado por el médico especialista x, y o z, que es una diferenciación para los estadísticos y resultados, con el cual será posible observar que existen N número de médicos especialistas diferentes.

7.1.2 Preparación de datos del paciente

Como ya se ha expresado anteriormente por cuestiones de tratamiento de la información y de la sensibilidad de la misma, fue necesario transformar cada uno de los nombres de los pacientes y discriminarlos solo por sexo; puesto que cuando se

desea alimentar el algoritmo, tener una variable que está compuesta por tantas características particulares no representa valor; puesto que dentro de la base suministrada por el hospital no estaba discriminado este dato del paciente, para ello el proceso seguido consta de:

1. **Alimentación por librerías:** Se tomo una librería la cual contenía más de 20.000 nombres ordenados de forma alfabética, esta librería estaba compuesta por la variable nombre y sexo, esto con el objetivo de realizar una comparativa entre el nombre del paciente y el sexo al cual el mismo podía corresponder.
2. **Comparación por composición:** Puesto que cada nombre tiene una composición particular de tres tipos
    a. **Nombre M & F:** Esta composición hacer referencia a los nombres que tienen según la librería, un primer nombre de tipo masculino, y en su segundo nombre uno de tipo femenino, en este caso el nombre será tomado como masculino, por ejemplo: José María = M.
    b. **Nombre F & M:** Esta segunda composición hace referencia a los nombres compuestos por dos tipos, el primero un nombre de tipo femenino y el segundo de tipo masculino, en cuyo caso el nombre será clasificado como tipo femenino, por ejemplo: María José = F
    c. **Nombre simple:** Este caso es el más simple de clasificar puesto que solo toma el nombre del paciente y comparándolo con la librería que se encuentra clasificada según género, determina si se trata de una persona de sexo masculino o femenino.
    d. **Nombre igual:** El caso de nombre igual es aquel en que los nombres que componen la identificación del paciente corresponden al mismo tipo, es decir, M & M o F &F; en cuyo caso la clasificación se hará basado en el sexo que defina.
3. **Generación de reglas:** Finalmente las reglas se encuentran resumidas dentro del anexo **8.3 informe técnico tratamiento de información** en forma

de código en lenguaje Python, aunque el procesamiento finamente se encuentre expresado en un archivo xlsx; por medio de condicionales se realizó la parametrización para cada uno de los casos descritos anteriormente.

7.1.3 Anonimidad y preparación de la información

El proceso de anonimizacion de la data corresponde al proceso de tomar cada una de los componentes que puedan poner en evidencia datos particulares que atenten contra la privacidad de las personas que se encuentren registradas en dicha base, tales como: médico y paciente. En este caso es necesario hacer una diferenciación entre cada una de las componentes de estas variables sin que las mismas pongan en evidencia de forma discriminatoria un caso en particular.

En segunda estancia para la utilización de los datos es indiferente componentes de las variables tales como, nombre completo del paciente, en este caso será de utilidad una discriminación que haga referencia al sexo del paciente, ya que este si puede ser un dato de gran utilidad. Para el caso de médico tratante, por lo anteriormente ya descrito en cuanto al tratamiento de la información y el respeto a la privacidad existente en los registros, fue necesario realizar una codificación que no hace alusión al nombre de un médico en particular si no simplemente discrimina que el paciente “A” fue tratado por el médico especialista x, y o z, que es una diferenciación para los estadísticos y resultados, con el cual será posible observar que existen N número de médicos especialistas diferentes.

7.1.4 Preparación de datos del paciente

Como ya se ha expresado anteriormente por cuestiones de tratamiento de la información y de la sensibilidad de la misma, fue necesario transformar cada uno de

los nombres de los pacientes y discriminarlos solo por sexo; puesto que cuando se desea alimentar el algoritmo, tener una variable que está compuesta por tantas características particulares no representa valor; puesto que dentro de la base suministrada por el hospital no estaba discriminado este dato del paciente, para ello el proceso seguido consta de:

7.4.1 Alimentación por librerías: Se tomo una librería la cual contenía más de 20.000 nombres ordenados de forma alfabética, esta librería estaba compuesta por la variable nombre y sexo, esto con el objetivo de realizar una comparativa entre el nombre del paciente y el sexo al cual el mismo podía corresponder. Dicha alimentación será posible observarla en el anexo **8.4 Procesamiento de la información.excel**

7.4.2 Comparación por composición: Puesto que cada nombre tiene una composición particular de tres tipos

a. **Nombre M & F:** Esta composición hacer referencia a los nombres que tienen según la librería, un primer nombre de tipo masculino, y en su segundo nombre uno de tipo femenino, en este caso el nombre será tomado como masculino, por ejemplo: José María = M.
b. **Nombre F & M:** Esta segunda composición hace referencia a los nombres compuestos por dos tipos, el primero un nombre de tipo femenino y el segundo de tipo masculino, en cuyo caso el nombre será clasificado como tipo femenino, por ejemplo: María José = F
c. **Nombre simple:** Este caso es el más simple de clasificar puesto que solo toma el nombre del paciente y comparándolo con la librería que se encuentra clasificada según género, determina si se trata de una persona de sexo masculino o femenino.
d. **Nombre igual:** El caso de nombre igual es aquel en que los nombres que componen la identificación del paciente corresponden al mismo tipo, es decir, M & M o F &F; en cuyo caso la clasificación se hará basado en el sexo que defina.

7.4.3 Generación de reglas: Finalmente las reglas se encuentran resumidas a continuación en forma de código en lenguaje Python, aunque el procesamiento finamente se encuentre expresado en un archivo xlsx; por medio de condicionales se realizó la parametrización para cada uno de los casos descritos anteriormente.

Codificación por médico especialista: Cada uno de los médicos especialistas tienen una codificación la cual está compuesta por tres variables, cada una de ellas en el orden que se presentan separadas por un punto“.”:

1. **Sexo:** Esta es la primera letra que se encuentra dentro de la codificación y hace referencia al género que identifica al médico especialista (M para masculino y F para femenino)
2. **Especialidad:** Las especialidades han sido resumidas de la siguiente manera, y es el segundo digito que se encuentra en la codificación:
    a. Cdva: Cirugía Cardio Vascular
    b. ClDl: Clínica del dolor.
    c. CxGn: Cirugía general.
    d. CxPl: Cirugía plástica.
    e. Dmt: Dermatología.
    f. Gast: Gastroenterología.
    g. Ginc: Ginecología.
    h. Neum: Neumología.
    i. Neur: Neurología.
    j. Odnt: Odontología.
    k. Oftm: Oftalmología.
    l. Ortp: Ortopedia.
    m. Otgia: Otorrinolaringología.
    n. Rdg: Radiología.
    o. Rsn: Resonancia.
    p. Urlg: Urología.
    q. Hemt: Hematología.

3. **Nombre:** El nombre completo del médico especialista se ha resumido por medio de un consecutivo que va desde el 1 al 180, esto no quiere decir que existan exactamente 180 doctores dentro de la base de datos, puesto que los mismo tuvieron que ser modificados debido a la variabilidad de cada uno de los casos.

Finalmente, la combinación de cada una de estas variables da como resultado una codificación compuesta de la siguiente forma: Sexo/Especialidad/Cd nombre, en cada uno de los casos se encontrará la misma de la siguiente manera, por ejemplo:

Tabla 4 codificacion de médicos especialistas

Cirujano 1	sexo	especialidad	serie
MARQUEZ AZUERO	M	Cdva	33
ALVARO GOMEZ	M	Otgia	63

Dando como resultado para los casos mostrados en la tabla la siguiente codificación respectivamente:

M.Cdva.33

M.Otgia.63

Existencia de más de una especialidad para Cx: Dentro de los casos se encontró que para un mismo procedimiento se encontraba más de un médico especialista en la sala, para ello la codificación se vio combinada, pero no mezclada, separándola por medio de una barra “/” quedando las codificaciones de la siguiente forma:

M.Cdva.33/M.Otgia.63

7.1.5 Tratamiento de tiempos

Para el tratamiento de los tiempos se definió en primera instancia la escala en la cual se iban a tratar los mismo, que para este caso se ha tomado la decisión de que sea en minutos, seguidamente se discriminaron las particularidades respecto a los tiempos y a partir de ahí se definió el proceso correspondiente a cada uno de estos casos para realizar la transformación de horas en formato hh:mm:ss a minutos como entero.

7.5.1 Tiempo de inicio > a tiempo de finalización: Este caso depende directamente del formato en el cual se ha estado trabajando la data, es decir, los tiempos al ser trabajos en formatos de 24 hora los casos para los cuales la hora de inicio superaba la hora de finalización como entero. Por ejemplo:

Hora de inicio Cx: 23:30

Hora fin Cx: 1:30

Para este caso como es posible notar, la hora de inicio es las 11 de la noche y la intervención concluye a la 1:30 de la madrugada lo que da como resultado dos horas de cirugía.

7.5.2. Tiempo de inicio < a tiempo de finalización: Como en el caso anterior depende directamente del formato, aunque este caso dentro es de los más comunes dentro de la base. Por ejemplo:

Hora de inicio Cx: 13:30

Hora fin Cx: 15:30

La hora de inicio para este caso es la 1 de la tarde y su hora de finalización son las 3:30 de la tarde lo que da como resultado 2 horas de cirugía.

7.5.3 Tiempo de inicio = tiempo de finalización: Existen casos dentro de la base en los cuales por cuestiones de digitación la hora de inicio es igual a la hora de finalización en estos casos al hacerse imposible obtener un tiempo de intervención lógico, por lo tanto, estos casos fueron eliminados.

7.5.4 Formato no correspondiente imposible de transformar: Durante el proceso de digitación y de compilación de todos los años dentro de una misma base, algunos de los datos perdieron su formato de origen, y al tratar de regresar los mismos a un formato en el cual se pueda realizar la transformación no fue posible, estos datos también se eliminaron.

Los dos últimos casos no representan un porcentaje considerable de la base, aunque es importante nombrar la particularidad de los mismos.

Teniendo en cuenta los casos anteriormente nombrados se han definido las siguientes expresiones:

$$T_f > T_i \rightarrow T_f - T_i = \Delta T$$

$$T_f < T_i \rightarrow T_i - T_f = \Delta T$$

$$T_f = T_i \rightarrow no\ operable$$

Donde ΔT es el delta del tiempo de cirugía, que representa el tiempo real de intervención, el cual inicia desde que el paciente ingresa a la sala de Cx y finaliza con el momento en el cual se inicia la limpieza de la sala.

7.6 Organización de las variables

Para concluir se tomaron cada una de las variables organizadas y se ordenaron dentro de una tabla que contiene 5 diferentes variables:

1. **Fecha de registro:** Corresponde a un dato de control el cual indica la fecha en la cual se tomó dicho registro.
2. **Genero del paciente:** Teniendo en cuenta el tratamiento que se realizó anteriormente para tratar el género del paciente por medio de librerías y haciendo uso de condicionales.
3. **Código de especialista:** Este hace referencia a la clasificación que se le ha dado al médico especialista por medio de una codificación la cual se ha expuesto anteriormente.
4. **Tiempo de intervención:** Tiempo real que dura el procedimiento o intervención quirúrgica en minutos.
5. **Especialidad intervenida:** la especialidad intervenida hace referencia a la especialidad que prevalece sobre la intervención, esta se ha decidido dejar

para aquellos casos en los cuales hay más de un médico especialista registrado para dicha intervención.

Las anteriores han sido hasta el momento las variables que se ha considerado que son de alta importancia para alimentar el algoritmo que se desarrollará, se ha efectuado todo este proceso puesto que es necesario tener un espectro un poco más cerrado y claro de las variables que tienen más peso y el efecto de las mismas sobre el desarrollo de la programación,

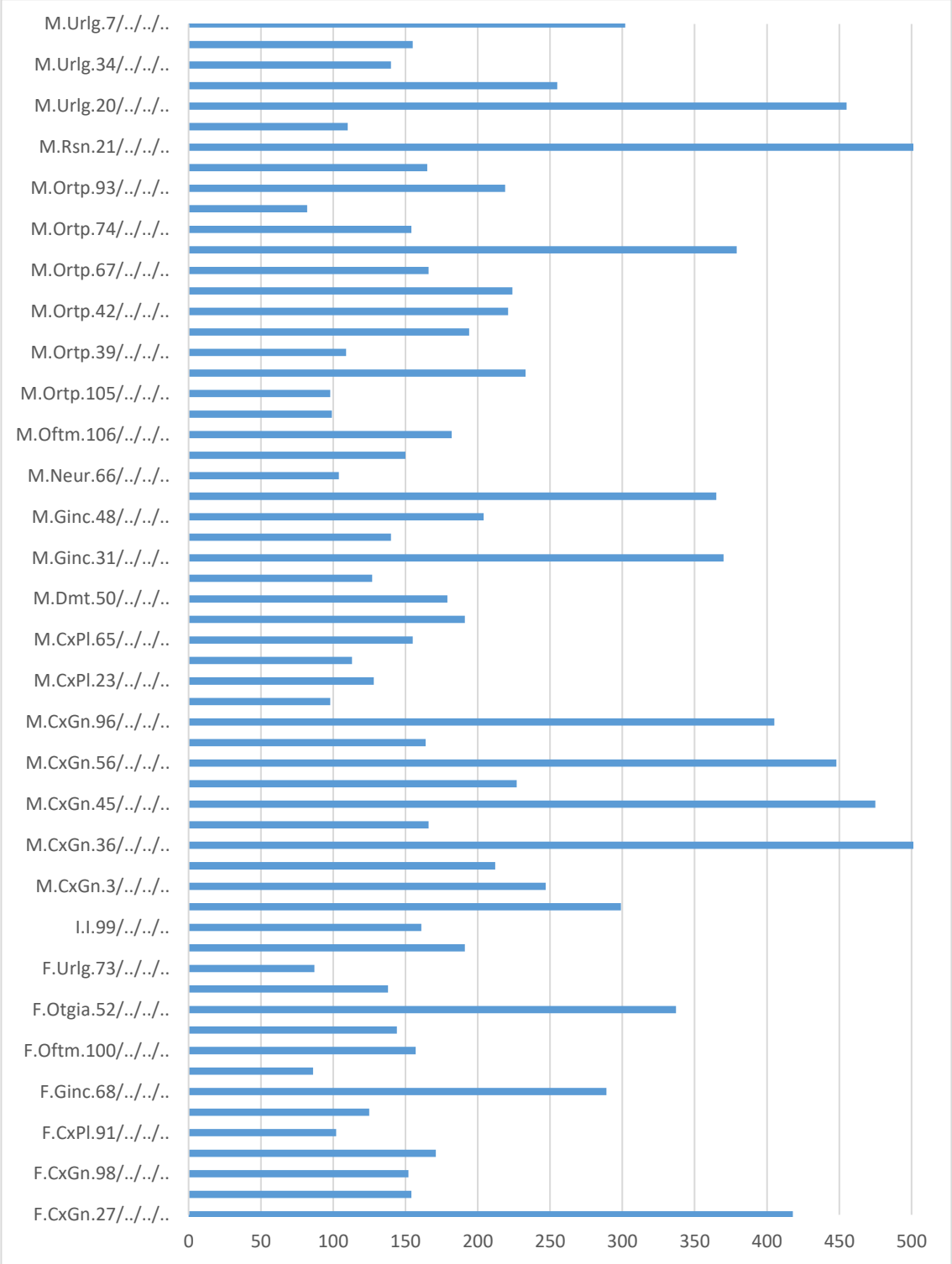
7.2 RESULTADOS DEL PROCESAMIENTO

A continuación, se presentan los resultados obtenidos a partir de una tabla resumen que se realizó con cada una de las variables ordenadas la cual se encuentra contenida en el anexo 8.. Los resultados arrojados presentan una alta frecuencia de intervenciones quirúrgicas para la especialidad de medicina general con una alta frecuencia de entre 70 a 105 minutos de intervención, es decir, que el tiempo de intervención que más se repite para intervenciones de cirugía general el intervalo ya mencionado, las siguientes intervenciones con mas frecuencia son ortopedia y ginecología con un intervalo para ambos casos de tiempo de intervención, entre 135 y 196 minutos para ambos casos. El comportamiento de las variables especialidad, se ha encontrado que solo algunas como la especialidad de oftalmología, ortopedia, otorrinolaringología cuentan con dicha distribución, lo que hace ocasiona que definir la probabilidad de ocurrencia de cada uno de los tiempos que la componen sea aún más fácil.

Existen otra serie de particularidades que se pueden inferir de la visualización de los datos, por ejemplo, es posible a partir de la simple observación clasificar aquellas intervenciones que tienen una serie de similaridades entre los intervalos para los tiempos de intervención, como cirugía plástica, ginecología y ortopedia, que aunque no tienen nada que ver una con la otra en cuanto a sus procedimientos, tienen una gran similaridad en la forma en la cual se presenta el porcentaje de sus frecuencias, de dichos o intervalos de tiempos.

Posteriormente se pasará a visualizar uno por uno los casos que menor frecuencia tienen, su comportamiento y el comportamiento por sexo por código de doctor, entre otros datos que serán de ayuda para entender la forma en la cual se comportan las intervenciones.

Gráfico 4 Frecuencia de intervenciones por medico

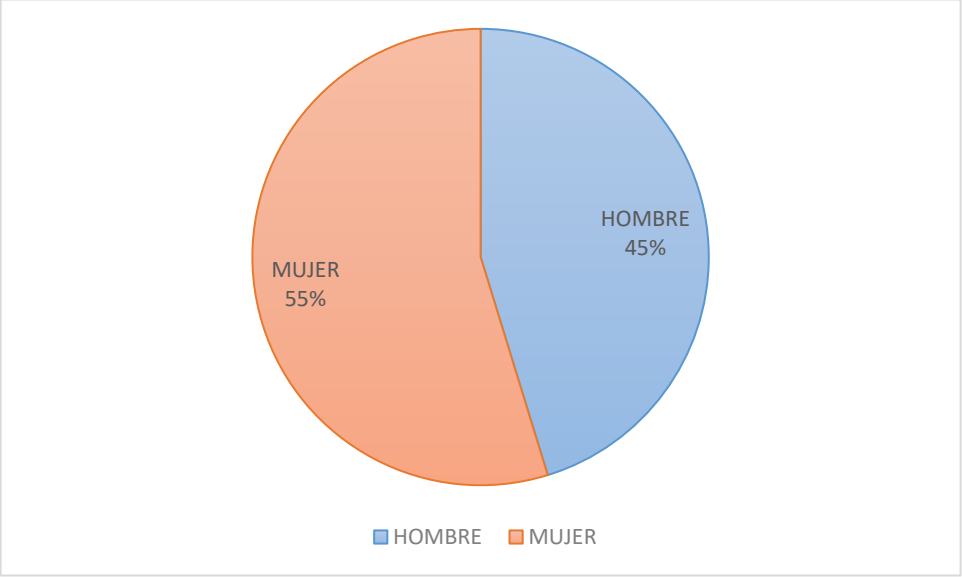


Fuente: Elaboración propia

La tabla de frecuencia de especialista representa la frecuencia, es decir, el número de veces que cada uno de los especialistas ha estado en una intervención quirúrgica; teniendo en cuenta que la forma en la cual se presentan cada uno de los especialistas corresponde a la codificación mostrada anteriormente para la anonimización de los datos del médico especialista.

Como es posible observar, existen una serie de médicos y de que sobresalen, como ya se ha dicho antes, por la cantidad de intervenciones realizadas, dentro de estos encontramos al medico identificado con el código M.CxGn.36 el cual es quien mas sobresale seguido por M.CxGn.45, M.CxGn.56 y M.CxGn.96, como es posible observar estos cuatro primeros especialistas pertenecen a una misma especialidad, cirugía general, por lo cual, el dato de los especialistas, respalda la visualización del grafico anterior, a partir de aquí se podría partir de la premisa que estos doctores se encuentran en una media de tiempo de intervención quirúrgica de entre 70 a 105 minutos, ya que esta es respaldada por el DashBoard presentando anteriormente , en donde se visualizan los intervalos de tiempos para cada una de las especialidades. Esta visualización de las frecuencias, también será útil para determinar la relación existente entre cada uno de los datos, los grupos con los cuales cada variable se puede relacionar y complementar la caracterización para el análisis de los Clusters.

Gráfico 5 Porcentaje de operaciones por sexo



Fuente: Elaboración propia

El numero de pacientes atendidos en su mayoría corresponden a hombres (sin caracterización de edad) quienes en su mayoría se someten a una intervención quirúrgica correspondiente a la especialidad de medicina general. Existen casos en los cuales no se especifica el campo de nombre, lo cual imposibilita. Que se realice la identificación del sexo.

7.3 METODOLOGÍA PARA LA GENERACIÓN DE CLÚSTERS

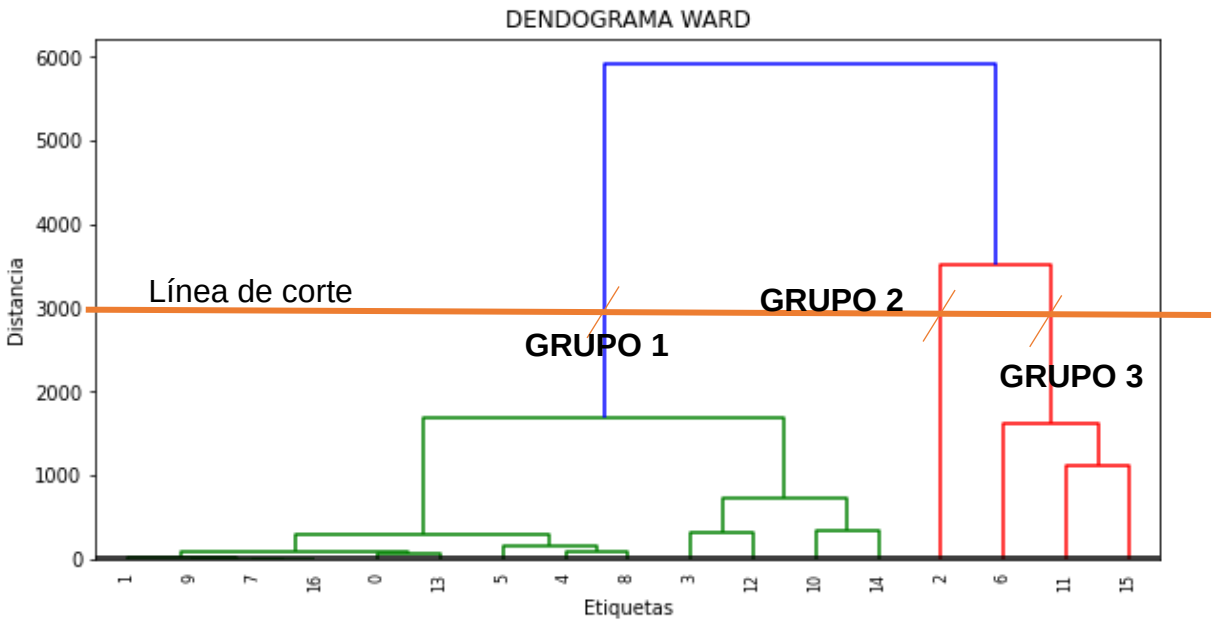
Para entender el procedimiento llevado a cabo para realizar los Clusters, y su importancia dentro del desarrollo del presente proyecto, lo primero será visualizar el problema en general. Una de las problemáticas existentes para generar la programación del bloque quirúrgico es la incertidumbre en el tiempo de utilización de cada sala, es decir, el tiempo de duración de la intervención; este tiempo es fundamental para determinar el momento en el cual la sala estará libre para el agendamiento de la próxima intervención.

La solución que se le ha dado a este problema es una ejecución de un algoritmo de dos fases, la primera consta de la reducción de dicha incertidumbre para, posteriormente continuar con el agendamiento de cada uno de los casos, planteando este como un problema de óptimos. La reducción de la incertidumbre consta de un proceso de clasificación de procedimientos por medio del cual se determina cuál de estos constituyen los tiempos más altos, o los más cortos, existiendo claramente, una serie de categorías dentro de la clasificación más alta y más baja.

Los Clústers presentados cuentan con las características de tiempos y de procedimiento, siendo etiquetados por especialidad, es decir, cada grupo tiene una serie de características que cumplen cierta similitud entre cada etiqueta del grupo, siendo estas similitudes el nivel de correlación existente entre las componentes de cada una de las variables. Por ejemplo: Suponga que existe un Grupo o clúster 3 compuesto por las especialidades 6, 11, 15 que corresponden Ginecología, Ortopedia y Urología respectivamente, tal y como se puede observar en el dendográma, este grupo al realizar el análisis o prueba de medias, da como resultado que las variables que mayor puntaje tienen, son aquellos grupos constituidos por 2 – 91 minutos de intervención seguido por, 92 – 181 y 182 – 271 minutos de intervención, por lo tanto, estas especialidades, (sin tener en cuenta los procedimientos) pueden ser clasificadas como las especialidades con tiempos cortos de intervención. Este proceso se repite con cada uno de los casos o para cada uno de los grupos, una vez determinadas las características de cada grupo por medio de la prueba de comparación de medias, es posible determinar cuáles son las intervenciones con los tiempos más altos y tiempos más bajos, en este caso

solo existirán tres clasificaciones, tiempos altos medios y bajos; ahora es posible continuar con la fase de programación aplicando los resultados del clúster aplicando teoría de grafos, la cual será explicada a profundidad en la última sección del presente.

Gráfico 6 Dendográma método de WARD



Fuente: elaboración propia (Python code – Google colaboratory)

7.3.1 Reducción de casos

La reducción de casos es el proceso por el cual se toman cada una de las variables que componen la base de datos y se realiza un conteo de frecuencias de cada una de ellas; cada uno de estos conteos se reducen por un número de casos; para esta ocasión el número de casos bajo el cual se desarrolló dicha reducción fue por medio del sexo, y rangos de tiempo, siendo estas etiquetadas por especialidad.

La reducción de los casos es la composición de cada una de las variables que hacen parte de la Clusterización, la comparación entre cada una de estas variables permite medir el nivel de correlación existente entre cada una de las etiquetas y por medio

de dicha correlación existente entre las diferentes variables o componentes de las variables para cada caso, es posible realizar la Clusterización.

A continuación, se presenta un ejemplo de cómo dicha reducción ha compuesto las variables de intervalos de tiempos para la construcción de este caso.

Tabla 5 Ejemplo de reducción de casos

ESPECIALIDAD	SxF	SxM	SxA	2 - 91	92 - 181
CdVa	35	55	0	6	10
CiDl	6	5	0	5	6
CxGn	2864	2242	3	2180	2259
CxPl	571	478	1	190	457
Dmt	130	87	0	45	114
Gast	75	44	0	101	12
Ginc	1713	168	0	607	947

Fuente: Elaboración propia

7.3.2 Reducción de rangos de tiempo

Los rangos de tiempo se determinaron estableciendo la longitud de cada una de las clases, esta longitud es esencial para determinar el rango y el número de intervalos que serán utilizados; como es posible observar en el ejemplo de la tabla anterior donde se encuentran dos rangos de tiempo, de 2 – 91 minutos y de 92 – 81 minutos de intervención quirúrgica, en total para este caso la reducción de casos es de 15 rangos, cuyas componentes están integradas por la frecuencia que determina la ocurrencia de dichos tiempos en ciertas etiquetas o especialidades.

Para determinar dichos rangos o número de casos fue necesario, en primer lugar determinar el rango de los datos, según (Valencia, 2012) dicho rango es posible calcularlo a partir de:

$$R = (Max - Min)$$

Fuente: (Valencia, 2012)

Donde Max es el numero máximo y Min será el número mínimo.

Una vez determinado el rango será necesario obtener el número de clases, el cual representa el número de casos o de variables que surgirán de las variables, por ejemplo, se ha dicho que el número de casos para los intervalos de tiempo fue de 15, este será calculado a partir de la regla de Sturges propuesta en 1926, la cual constata que:

Ecuación 9 calculo del numero de casos regla Sturges

$$k = 1 + 3.322 * Log_{10}(n)$$

(Valencia, 2012)

Donde n es el tamaño de la muestra, es decir el numero de datos que conforman las variables.

Ahora bien, dicho esto es posible construir cada uno de los rangos iniciando por el número mínimo y sumando la longitud de la clase de forma consecutiva al número de clase máximo del rango anterior. Por ejemplo, suponga que el número mínimo, es decir, donde iniciaran cada uno de los casos es el numero 2 (es decir inicia en el tiempo mínimo de intervención quirúrgica que es dos minutos) en este caso el resultado de la longitud de las clases es de 89 lo que quiere decir que este caso va desde 2 minutos hasta 91 minutos que es 2+89. El cálculo de la longitud de la clase se calcula de la siguiente manera:

Ecuación 10: calculo del numero de clases

$$c = \frac{R}{k}$$

(Valencia, 2012)

Donde R es el rango k es el numero de clases y c es la longitud de las clases.

7.3.3 Selección de método de clusterización

Existe una variedad de métodos para realizar la Clusterización, dentro de los más utilizados o nombrados comúnmente, se encuentra el método jerárquico y el método

de K – Means, cada uno de estos métodos presentan ciertas facilidades al momento de realizar los cálculos, lo que en cierta medida se traduce en un menor costo de procesamiento.

El método seleccionado es el método Jerárquico (aunque es posible contrastar entre un método jerárquico y un método de K – medias, por medio de un índice de Silhouette y un indicador de David Boulding para el contraste de desempeño entre cualquiera de ambos métodos), para la ejecución de la técnica jerárquica existen dos metodologías las cuales contienen una serie de procedimientos, estos afectan la forma en la cual se realizan los cálculos. Los métodos jerárquicos aglomerativos parten de un factor inicial y a partir de este generan una serie de grupos o clústeres dependiendo del nivel de correlación existente, para este método existen diferentes procedimientos dentro de los cuales se encuentran:

a. **Estrategia de la distancia mínima o similitud máxima:** también llamada método de amalgamiento simple, busca el nivel de similitud o distancia mínima existente entre el conjunto de datos, donde la distancia para cada uno de los Clusters estará determinada por:

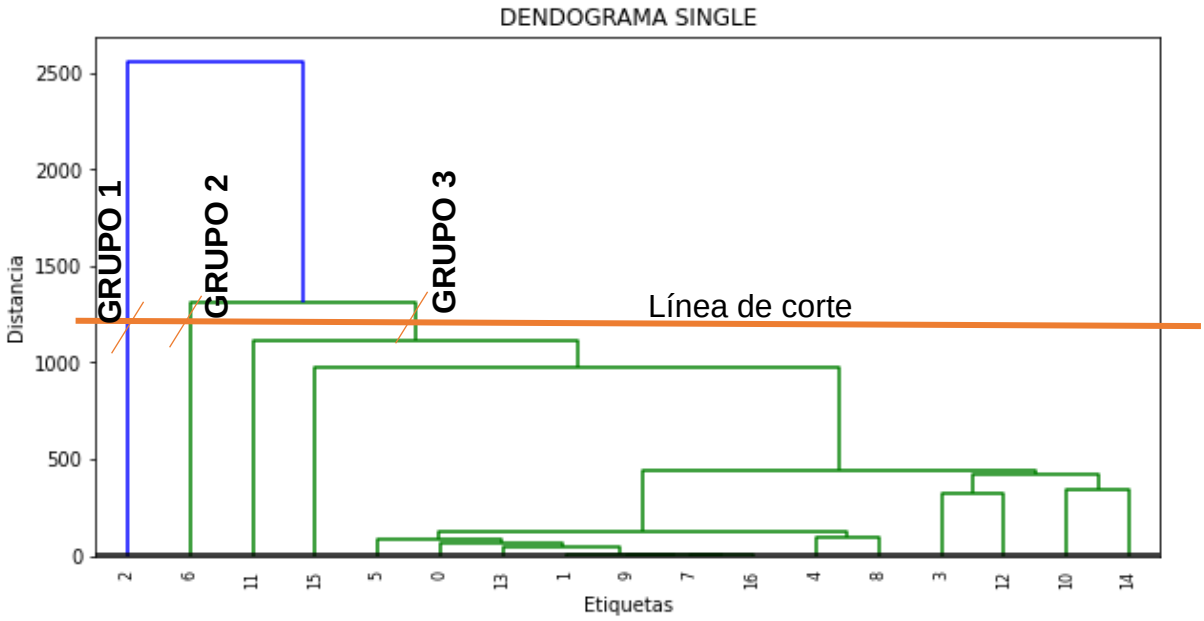
Ecuación 11: calculo de la distancia por amalgamiento simple

$$d(C_i, C_j) = MIN_{x_m \in C_j}^{x_l \in C_i}\{d(x_l x_m)\} \;\; l = 1 \dots n_i \;\; ; \;\; m = 1 \dots n_j$$

Fuente: (Valencia, 2012)

Donde la distancia entre los Clústers para $C_i \; (con \; n_i \; elementos) \;\; ; \;\; C_j \; (con \; n_j \; elementos)$
Para el cual los resultados obtenidos del clúster son los siguientes:

Gráfico 7 Dendográma método de amalgamiento simple



Fuente: elaboración propia (Python code – Google colaboratory)

Como es posible observar para este caso el número de clúster obtenidos tomando como punto de partida la distancia mas larga representada en el dendográma es igual a tres grupos, aunque como es posible observar, en el grupo número tres se encuentran contenidas gran mayoría de las etiquetas.

Tabla 6 Clasificación de etiquetas para método de amalgamiento simple

CLUSTER	NUMERO DE VAR	ETIQUETA
Grupo1	2	CxGn
Grupo2	6	Ginc
Grupo3	8	Neur
Grupo3	5	Gast
Grupo3	0	CdVa
Grupo3	13	Rdg
Grupo3	1	CiDl

Grupo3	9	Odnt
Grupo3	7	Neum
Grupo3	16	Hemt
Grupo3	3	CxPl
Grupo3	12	Otgia
Grupo3	10	Oftm
Grupo3	14	Rsn
Grupo3	4	Dmt
Grupo3	11	Ortp
Grupo3	15	Urlg

Fuente: elaboración propia

b. **Estrategia de la distancia máxima o similitud mínima:** esta, al contrario de la anterior, la cual busca entrar el nivel de correlación más pequeño, este método, busca encontrar el nivel de correlación más grande entre cada una de las variables, de esta forma con cada iteración ira agrupando la matriz entre cada una de las variables; expresión contraria que ocurre en este caso.

Ecuación 12: calculo de la distancia para método de similitud minima

$$d(C_i, C_j) = MAX_{x_m \in C_j}^{x_l \in C_i}\{d(x_l x_m)\}\ \ l = 1 \dots n_i\ ;\ \ m = 1 \dots n_j$$

Donde la distancia entre los Clústers para $C_i\ (con\ n_i\ elementos)\ ;\ \ C_j\ (con\ n_j\ elementos)$

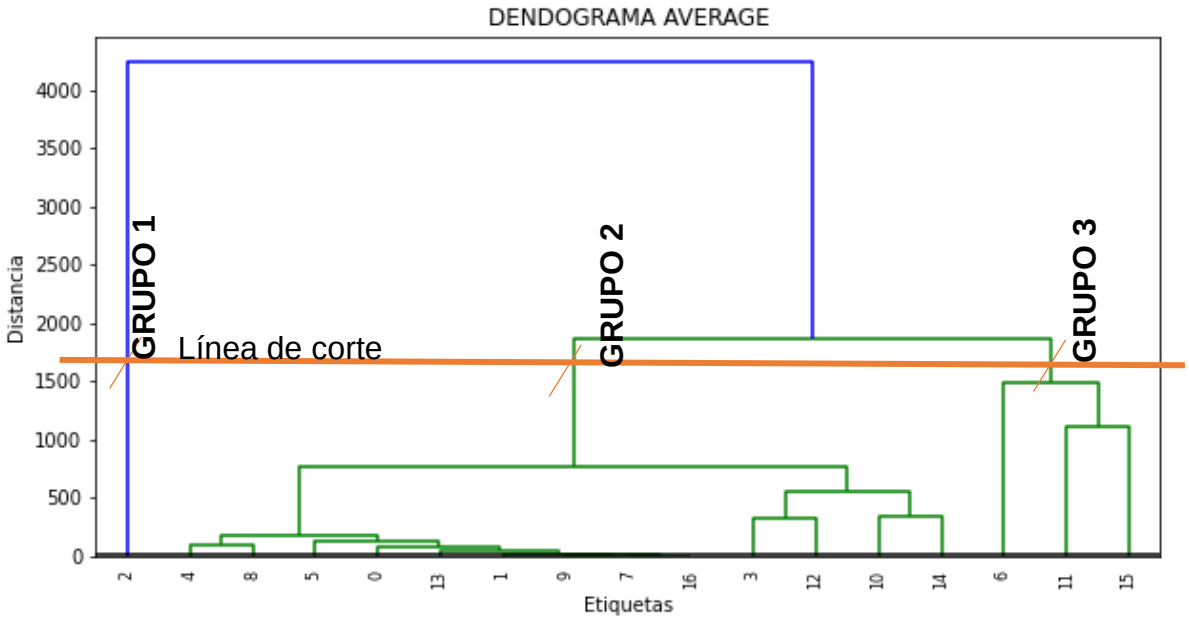
c. **Estrategia de la distancia, o similitud promedio no ponderada:** en esta estrategia la distancia se obtiene por medio de la media aritmética dando como resultado las distancias de los Clusters (C_i, C_j) relacionando los mismos con la similitud existente para cada una de las matrices, es decir, la distancia entre $d(C_i, C_j)$ será:

Ecuación 13: cálculo del clúster por promedio no ponderado

$$d(C_i C_j) = \frac{d(C_{i1}, C_j) + d(C_{i2}, C_j)}{2}$$

Para el cual los resultados obtenidos del clúster son los siguientes:

Gráfico 8 Dendográma para método similitud promedio



Fuente: elaboración propia (Python code – Google colaboratory)

Como es posible observar para este caso el número de clúster obtenidos tomando como punto de partida la distancia más larga representada en el dendográma es igual a tres grupos, y para este caso es posible observar más claramente la forma en la cual se encuentran compuestos estos tres grupos los cuales contienen.

Tabla 7 Clasificación de etiquetas para método de similitud promedio

CLÚSTER	NUMERO DE VAR	ETIQUETA
Grupo1	2	CxGn
Grupo2	4	Dmt
Grupo2	8	Neur
Grupo2	5	Gast
Grupo2	0	CdVa
Grupo2	13	Rdg
Grupo2	1	CiDl
Grupo2	9	Odnt
Grupo2	7	Neum
Grupo2	16	Hemt
Grupo2	3	CxPl
Grupo2	12	Otgia
Grupo2	10	Oftm
Grupo2	14	Rsn
Grupo3	6	Ginc
Grupo3	11	Ortp
Grupo3	15	Urlg

Fuente: Elaboración propia

d. Métodos basados en el centroide: en estos Clusters se tiene en cuenta la medida dada por la semejanza entre sus centroides, es decir el centroide entre (C_i, C_j); entre estos métodos se distinguen dos:

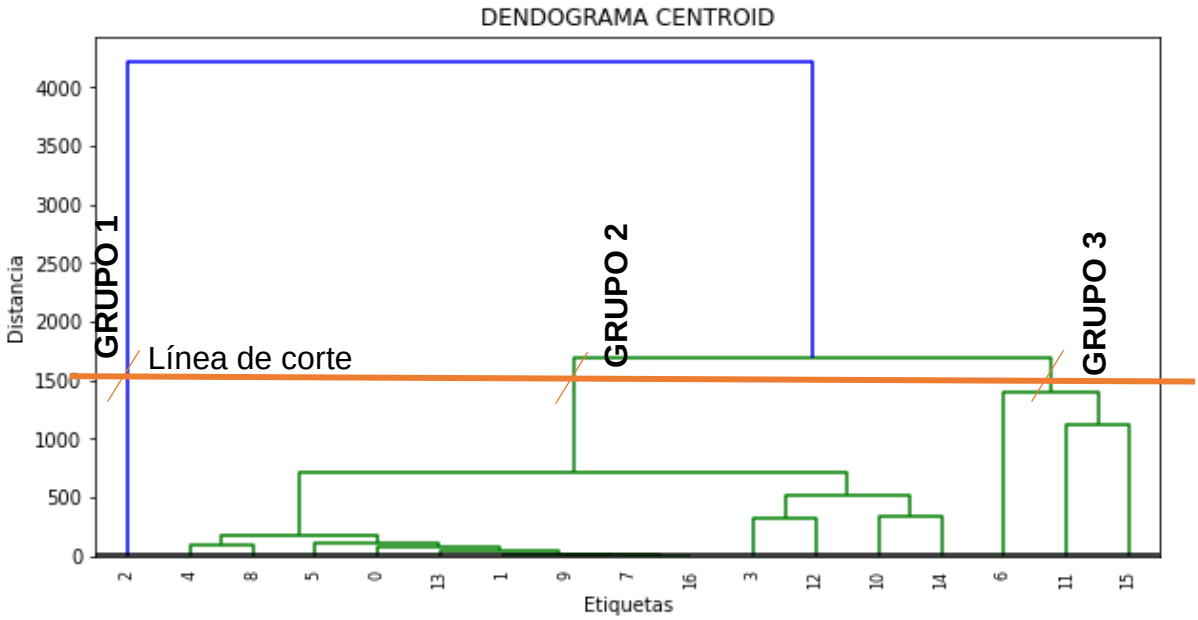
a. Método de centroide ponderado.

b. Método de la mediana.

Cada uno de estos métodos para la generación de los Clusters y el desarrollo de los cálculos cuenta con ciertas ventajas, por ejemplo, tendrá en cuenta el ajuste entre cada iteración para ir disminuyendo el error existente en el centroide; pero este ajuste podría no tener en cuenta los valores existentes que sean demasiado pequeños.

En este caso el método evaluado es el método de la mediana para el cual se obtuvieron los siguientes resultados:

Gráfico 9 Dendográma para método del centroide



Fuente: elaboración propia (Python code – Google colaboratory)

Para este caso nuevamente se encuentran tres grupos claramente diferenciados, los cuales se encuentran caracterizados de la siguiente manera:

Tabla 8 Clasificación de etiquetas para método del centroide

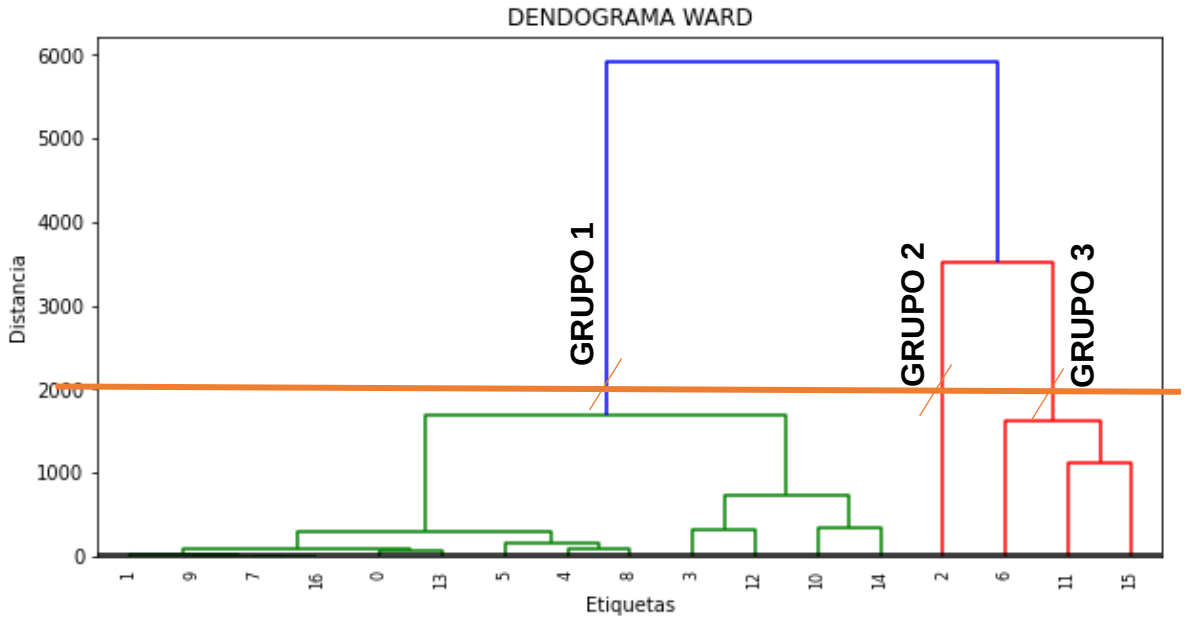
CLUSTER	NUMERO DE VAR	ETIQUETA
Grupo1	2	CxGn
Grupo2	4	Dmt
Grupo2	8	Neur
Grupo2	5	Gast
Grupo2	0	CdVa
Grupo2	13	Rdg
Grupo2	1	CiDl
Grupo2	9	Odnt
Grupo2	7	Neum
Grupo2	16	Hemt
Grupo2	3	CxPl
Grupo2	12	Otgia

Grupo2	10	Oftm
Grupo2	14	Rsn
Grupo3	6	Ginc
Grupo3	11	Ortp
Grupo3	15	Urlg

Fuente: Elaboración propia.

e. **Método de WARD:** este es quizá de los más conocidos y utilizados. Este método consiste en unir los dos Clusters de los cuales se tenga el menor incremento de la suma de sus cuadrados.
Este proceso inicia con m Clusters, cada uno de los cuales este compuesto por un solo elemento, cada uno de estos elementos será tenido en cuenta para determinar cuál de ellos coincide con el centroide y será agrupado, con cada iteración se realizará un proceso similar hasta que sea imposible encontrar dos Clusters cuya unión proporcione el menor incremento de la suma de errores.
Los resultaos arrojados por el clúster al ser evaluado este por método de Ward se presentan a continuación

Gráfico 10 Dendográma para método de WARD



Fuente: elaboración propia (Python code – Google colaboratory)

Nuevamente se encuentran tres grupos claramente diferenciados entre si, en este caso para los todos los dendográmas se ha encontrado que la etiqueta numero 2 la cual corresponde a Cirugía General queda nuevamente fuera de cualquier grupo. Cada uno de estos grupos se encuentra caracterizado de la siguiente forma:

Tabla 9 Clasificación de etiquetas para método de WARD

CLUSTER	NUMERO DE VAR	ETIQUETA
Grupo1	1	CiDl
Grupo1	9	Odnt
Grupo1	7	Neum
Grupo1	16	Hemt
Grupo1	0	CdVa
Grupo1	13	Rdg
Grupo1	5	Gast
Grupo1	4	Dmt
Grupo1	8	Neur
Grupo1	3	CxPl
Grupo1	12	Otgia
Grupo1	10	Oftm
Grupo1	14	Rsn
Grupo2	2	CxGn
Grupo3	6	Ginc
Grupo3	11	Ortp
Grupo3	15	Urlg

7.3.4 Prueba para cada uno de los métodos

Para la selección del procedimiento con el cual se efectuara la ejecución de cada uno de los Clusters, se evaluaran en su totalidad, y se realizara una prueba de hipótesis, por medio de a cual se pretende realizar una validación de los Clusters, partiendo de una prueba de medias para la cual la Hipótesis a priori parte de la premisa que todas las medias de cada uno de los grupos son iguales, de ser eso cierto no habrá diferencia entre ninguno de los grupos ejecutados para que esta

hipótesis sea nula el nivel de significancia de la prueba de medias debe ser menor al 5%. Para aquellos procedimientos cuyo valor supere dicho porcentaje, será descartado automáticamente, por otro lado, la selección se hará a partir del menor valor de significancia.

Dichas pruebas de medias están definidas de la siguiente manera:

- Para dos grupos → prueba t muestras
- Para 2 o mas grupos → prueba Chi cuadrado.
- Para mas de dos grupos → prueba ANOVA

Por medio de la evaluación de cada uno de los métodos se pretende seleccionar el método que tenga un menor nivel de significancia al evaluarlo dentro del clúster, el ejercicio desarrollado para este caso se presenta a continuación, donde, cada uno de los gráficos (dendográmas) representa los Clusters realizados y se presentara también posteriormente, el resultado obtenido a partir de la prueba de medias. En este caso se ha determinado que el corte del dendográma para determinar el número de grupos se realizara a partir del criterio de máxima distancia entre la correlación, es decir, partiendo del clúster más alejado se tomara esta referencia para realizar el corte y definir el número de grupos; en los 5 dendográmas generados se notó que el número de grupos es igual a 3 es decir 3 Clusters por método, lo que es bastante consistente. Dichos resultados se presentan a continuación:

Tabla 10 prueba ANOVA para método de amalgamiento simple

PRUEBA ANOVA PARA SINGLE						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
	Entre grupos	7899512,282	2	3949756,141	31,389	0%
SxM	Entre grupos	3566418,196	2	1783209,098	9,237	0%
SxA	Entre grupos	7,796	2	3,898	31,484	0%
2 - 91	Entre grupos	3606160,737	2	1803080,369	18,033	0%
92 - 181	Entre grupos	4234704,008	2	2117352,004	15,214	0%
182 - 271	Entre grupos	102820,149	2	51410,075	3,025	8%
272 - 360	Entre grupos	6981,067	2	3490,533	2,734	10%
361 - 450	Entre grupos	739,129	2	369,565	2,103	16%
451 - 540	Entre grupos	163,325	2	81,663	2,142	15%
541 - 629	Entre grupos	244,031	2	122,016	11,885	0%
630 - 719	Entre grupos	168,784	2	84,392	11,008	0%

720 - 809	Entre grupos	5,725	2	2,863	1,074	3.7%
810 - 899	Entre grupos	2,384	2	1,192	1,061	3.7%
900 - 988	Entre grupos	19,929	2	9,965	5,911	1%
989 - 1078	Entre grupos	44,384	2	22,192	83,221	0%

Fuente: Elaboración propia

Tabla 11 prueba ANOVA para método del promedio

PRUEBA ANOVA PARA AVERAGE						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
SxM	Entre grupos	4981550,094	2	2490775,047	27,085	0%
SxA	Entre grupos	7,940	2	3,970	34,960	0%
2 - 91	Entre grupos	4314826,112	2	2157413,056	43,701	0%
92 - 181	Entre grupos	5646470,582	2	2823235,291	73,661	0%
182 - 271	Entre grupos	235773,985	2	117886,992	15,727	0%
272 - 360	Entre grupos	11772,923	2	5886,462	6,298	1%
361 - 450	Entre grupos	801,940	2	400,970	2,341	13%
451 - 540	Entre grupos	173,828	2	86,914	2,326	13%
541 - 629	Entre grupos	270,329	2	135,164	16,113	0%
630 - 719	Entre grupos	172,374	2	86,187	11,631	0%
720 - 809	Entre grupos	5,315	2	2,658	0,986	4%
810 - 899	Entre grupos	2,220	2	1,110	0,978	4%
900 - 988	Entre grupos	19,632	2	9,816	5,751	2%
989 - 1078	Entre grupos	43,759	2	21,879	70,271	0%
1079 - 1168	Entre grupos	7,837	2	3,919	14,858	0%
1169 - 1257	Entre grupos	0,000	2	0,000		
1258 - 1347	Entre grupos	3,765	2	1,882		

Fuente: Elaboración propia

Tabla 12 Prueba ANOVA para método del centroide

PRUEBA ANOVA PARA CENTROIDE						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
	Entre grupos	8534798,190	2	4267399,095	53,040	0%

SxM	Entre grupos	4981550,094	2	2490775,047	27,085	0%
SxA	Entre grupos	7,940	2	3,970	34,960	0%
2 - 91	Entre grupos	4314826,112	2	2157413,056	43,701	0%
92 - 181	Entre grupos	5646470,582	2	2823235,291	73,661	0%
182 - 271	Entre grupos	235773,985	2	117886,992	15,727	0%
272 - 360	Entre grupos	11772,923	2	5886,462	6,298	1%
361 - 450	Entre grupos	801,940	2	400,970	2,341	13%
451 - 540	Entre grupos	173,828	2	86,914	2,326	13%
541 - 629	Entre grupos	270,329	2	135,164	16,113	0%
630 - 719	Entre grupos	172,374	2	86,187	11,631	0%
720 - 809	Entre grupos	5,315	2	2,658	0,986	4%
810 - 899	Entre grupos	2,220	2	1,110	0,978	4%
900 - 988	Entre grupos	19,632	2	9,816	5,751	2%
989 - 1078	Entre grupos	43,759	2	21,879	70,271	0%
1079 - 1168	Entre grupos	7,837	2	3,919	14,858	0%
1169 - 1257	Entre grupos	0,000	2	0,000		
1258 - 1347	Entre grupos	3,765	2	1,882		

Fuente: Elaboración propia

Tabla 13 Prueba ANOVA para método de WARD

PRUEBA ANOVA PARA MÉTODO DE WARD						
		Suma de cuadrados	gl	Media cuadrática	F	Sig.
	Entre grupos	8534798,190	2	4267399,095	53,040	0%
SxM	Entre grupos	4981550,094	2	2490775,047	27,085	0%
SxA	Entre grupos	7,940	2	3,970	34,960	0%
2 - 91	Entre grupos	4314826,112	2	2157413,056	43,701	0%
92 - 181	Entre grupos	5646470,582	2	2823235,291	73,661	0%
182 - 271	Entre grupos	235773,985	2	117886,992	15,727	0%
272 - 360	Entre grupos	11772,923	2	5886,462	6,298	1%
361 - 450	Entre grupos	801,940	2	400,970	2,341	3%
451 - 540	Entre grupos	173,828	2	86,914	2,326	3%
541 - 629	Entre grupos	270,329	2	135,164	16,113	0%
630 - 719	Entre grupos	172,374	2	86,187	11,631	0%
720 - 809	Entre grupos	5,315	2	2,658	0,986	4%
810 - 899	Entre grupos	2,220	2	1,110	0,978	4%

900 - 988	Entre grupos	19,632	2	9,816	5,751	2%
989 - 1078	Entre grupos	43,759	2	21,879	70,271	0%
1079 - 1168	Entre grupos	7,837	2	3,919	14,858	0%
1169 - 1257	Entre grupos	0,000	2	0,000		
1258 - 1347	Entre grupos	3,765	2	1,882		

Fuente: Elaboración propia

Prueba de hipótesis

Para comprobar la valides de cada uno de los Clusters será necesario realizar la comprobación de la hipótesis, cuyo punto de partida será:

H0: Parte de la premisa que todos los grupos generados poseen las mismas características, y de ser correcto el nivel de significancia será superior Al 5%.

H1: De no ser así, y su nivel de significancia sea menor a cero, querrá decir que el clúster generado tiene las características suficientes para diferenciar entre un grupo y el otro, por lo tanto, este será aceptable.

Como es posible observar de las pruebas realizadas para cada uno de los Clusters desarrollados, solo uno de ellos cumple con un nivel de significancia en cada una de las variables para cada uno de los grupos, de forma tal que la premisa inicial pase a ser hipótesis nula, por lo que se cumpliría mi H1, por ello el método seleccionado será el método de Ward.

7.4 PLANTEAMIENTO TEORICO PARA SU APLICACIÓN EN DE AGENDAMIENTO COMO PROBLEMA DE RUTEO

En este caso se ha planteado el problema de asignación, como un problema de ruteo, donde la agenda generada, dependerá directamente del orden resultante en la ruta, teniendo en cuenta que el manejo de los tiempos se ha dado por medio de los Clusters, los cuales como se verá seguidamente son fundamentales para el desarrollo del grafo. A continuación, se explica la forma en la cual esta configurados cada uno de los nodos y su respectiva disposición para aristas o arcos:

- Los Clusters componen una serie de grupos los cuales internamente están compuestos por una serie de etiquetas, dichas etiquetas (de especialidad) comparten varias similitudes dentro del grupo. La forma en la cual se disponen las especialidades esta estrechamente relacionado al clúster al cual pertenecen, de esta forma se determina a cuál de las intervenciones a realizarse contendrá.
- Cada procedimiento corresponde a una especialidad, puesto que, dicho procedimiento se encuentra caracterizado dentro del grupo de dicha especialidad; esta caracterización, le da cierto carácter a la arista que se proyecta sobre el nodo de especialidad, este carácter se encuentra definido por una clasificación que se muestra a continuación:

Tabla 14 Tabla de penalización por cluster

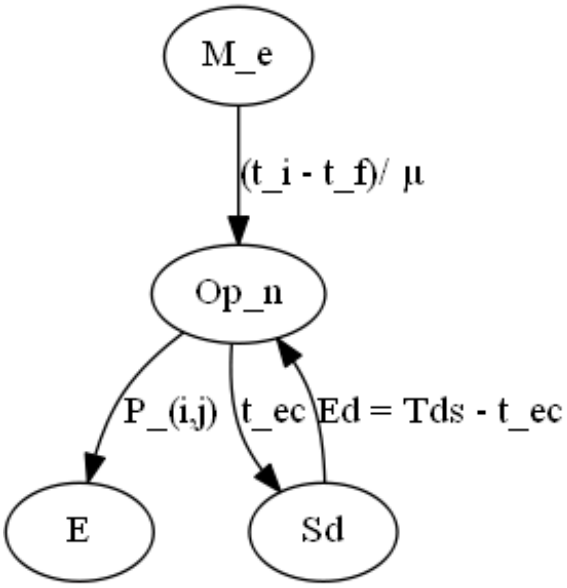
GRUPO	CARACTERISTICAS	PENALIZACIÓN
1	Tiempos medios	5
2	Tiempos bajos	1
3	Tiempos altos	10

Fuente: Elaboración propia

Cada uno de los grupos contenidos representan una serie de características las cuales fue posible determinar por medio de la prueba de comparación de medias, dichas características a su vez contienen una seria de penalizaciones las cuales están encargadas de asignar el valor que contendrá la arista.

- La parametrización de la arista de los nodos de operaciones a especialidades, se encuentran dados por el valor de la media del procedimiento, ya que este valor representa una escala de procedimientos para el grafo.
- La función del nodo sala esta dada por la capacidad instalada, la cual retroalimenta a la operación siguiente con los datos de la disponibilidad, si la disponibilidad de la sala en términos de tiempo a escala de minutos es mayor a el tiempo estimado por el clúster para esta operación entonces genera el agendamiento de no ser así, salta a una sala con mayor disponibilidad.

Figura 5 Grafo modelo teórico para agendamiento



Fuente: Elaboración propia

Notación

E (Especialidad): el numero de especialidades es de 17 especialidades en total, las cuales se encuentran agrupadas por el intervalo de tiempo y las características de procedimiento.

Sd (Sala de cirugía): el hospital San José cuenta con un total de 22 salas de cirugías de las cuales 7 son de uso exclusivo para determinadas especialidades.

Op_n (Operación o intervención): la operación o intervención, representa un paciente el cual se encuentra a la espera que se le sea asignada la agenda para su cirugía.

M_e (Médico especialista): Doctor que se encuentra a cargo de la intervencion.

7.4.1 Reglas generales

Para determinar la capacidad que tendrá cada sala y que esta sea asignada según la disponibilidad de tiempo, de forma tal que cada una de salas al ser agendadas resten la capacidad de tiempo, ya que estas se encuentran con una intervención la cual ocupa cierto tiempo, lo cual genera que su disponibilidad en horas disminuya; se ha determinado esta disponibilidad de la siguiente forma:

Ecuacion14: cálculo del tiempo disponible por sala

$$Ed = Tds - t_{ec}$$

Fuente: Elaboración propia

Donde t_{ec} corresponde al estimado por clúster, para el cual debe tenerse en cuenta la especialidad y el estimado de tiempo para esta especialidad:

Donde t_i corresponde al tiempo inicial de la hora en formato de 24h y t_f al tiempo de finalización de la intervención.

Las características de estas reglas generales se presentan a continuación:

- El agrupamiento inicial se realiza según el clúster, el cual determina el tiempo de duración, este tiempo se encuentra asignado por la distancia en el vértice que es el valor entre el nodo de E1, E2…En para la representación de cada especialidad.
- Las entidades se encuentran determinadas por el histórico de número de intervenciones que debe realizar cada médico.
- El clúster brinda las características al nodo de procedimientos el cual proyecta dichas características por medio de una arista o arco, estas características (numéricas) son de ayuda para la selección del ruteo de la agenda.
- Cada procedimiento es una entrada independiente y debe pasar por su asignación de sala.

- Cuando una sala se ocupa (su Ed es igual a cero) la intervención a agendarse ya no tendrá en cuenta dicha sala, por lo tanto, ya no será tenida en cuenta para la asignación futura
- La ocupación de la sala esta determinada por el tiempo que es asignado de forma secuencial, lo que quiere decir que, si la sala tiene un tiempo disponible de 18 horas para agendar intervenciones, a medida que son agendadas dichas intervenciones este tiempo se hará cada vez menor. Por ejemplo, suponga que para esta sala se ha agendado ya una intervención quirúrgica cuyo tiempo estimado por clúster se encuentra entre 2 a 4 horas, esto quiere decir que el espacio de tiempo disponible para agendar una intervención en esta sala ya no es de 18 horas si no de 14 horas.
  De esta forma el nodo es alimentado por la arista y a su vez este genera una retro alimentación al nodo de Op_n, de esta forma el algoritmo tendrá conocimiento de si le es posible o no agendar ciertas intervenciones con ciertos tiempos.
- Los procedimientos emergen del médico especialista y se irán asignando a la especialidad correspondiente por medio de la estimación del tiempo, es decir, dependerá del delta del procedimiento sobre el promedio de tiempo estimado para este rango de tiempo. Por ejemplo, si el tiempo estimado para una especialidad contenida en cierto clúster es de 2 a 91 minutos hay que ver cual es el promedio de tiempo dentro de este rango, para este caso es 55.8 minutos de intervención, este será el valor de la media que debe reemplazarse en la siguiente ecuación:

Ecuación15: calculo del estimado para el tiempo de operación

$$\Delta t_{op} = \frac{(t_f - t_i)}{\mu}$$

Fuente: Elaboración propia

Donde como ya se ha dicho μ representa la media del tiempo estimado para dicho rango y el delta del dividendo es determinado por el rango que esta siendo evaluado.

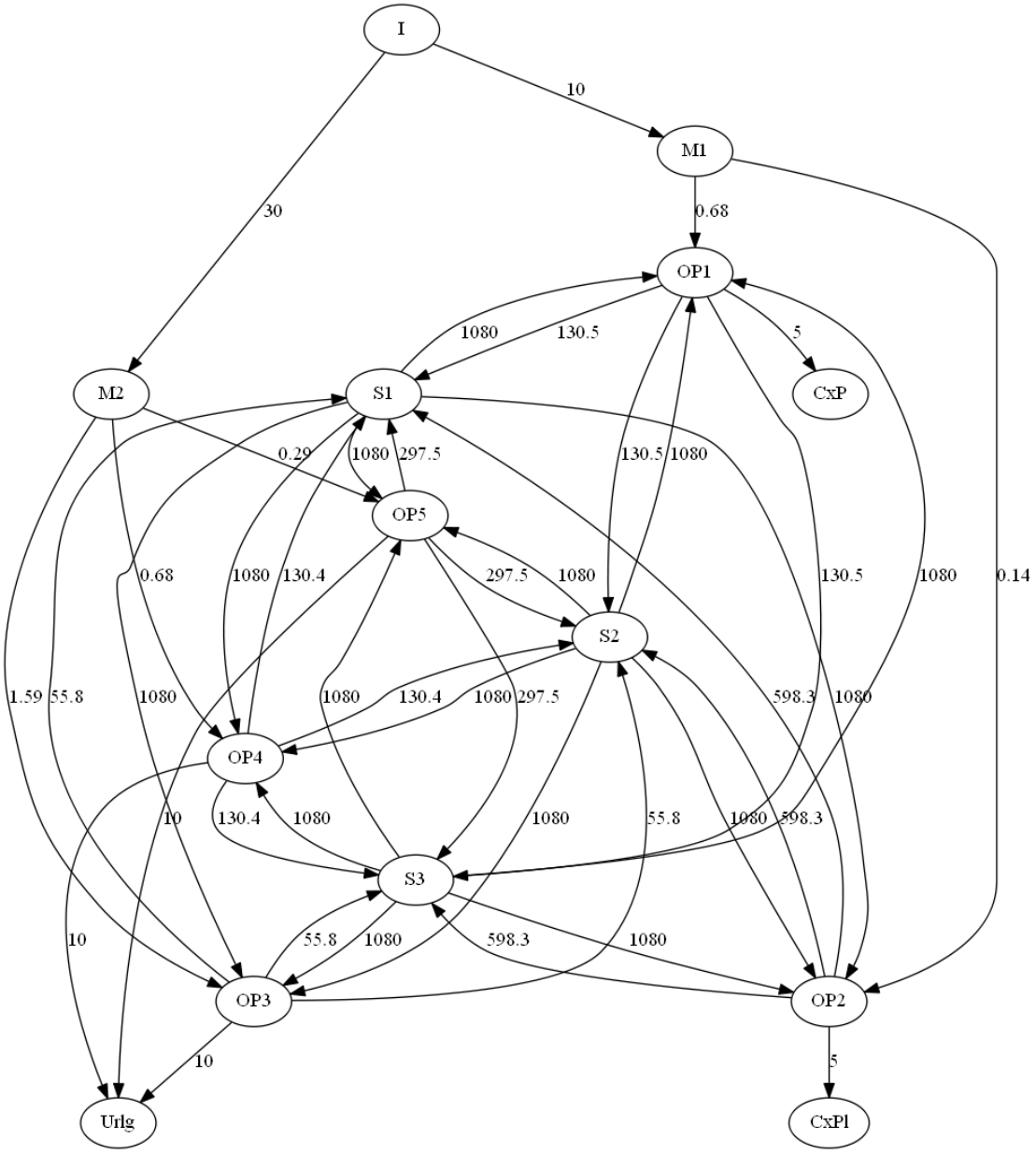
- La arista que compone la distancia entre el nodo del médico especialista y el nodo del procedimiento, se encuentra determinada por el estimado de tiempo arrojado por el clúster para

la realización de este calculo se requiere tener en cuenta el delta del tiempo estimado por clúster.

7.4.2 Ejercicio aplicado

para este caso suponga que se encuentra un hospital en el cual se dispone de dos médicos, solo tres salas de cirugías y se pretende por medio del algoritmo y técnicas de clusterizacion demostradas anteriormente realizar la programación de 5 intervenciones de cirugía para la próxima semana, por lo que tener una aproximación al tiempo de intervención es imprescindible y conocer la forma de programar a través de este es fundamental. En este caso el grafo generado para un problema tan sencillo es el siguiente:

Gráfico 11 Grafo ejercicio aplicado



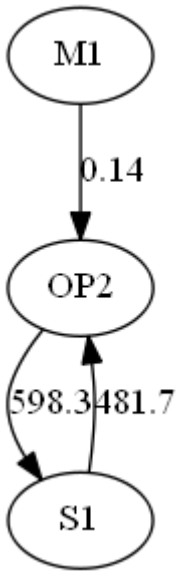
Fuente: Elaboración propia

Como es posible observar, existe un nodo de inicio, a partir del cual se van tomando las decisiones de por que medico iniciara el agendamiento, este nodo inicial toma el valor de la sumatoria de las penalizaciones determinadas para cada uno de los Clusters por intervención a realizar. Por ejemplo, en este caso el medico dos (M2) tiene que operar tres pacientes (OP3, OP4, OP5) estas intervenciones a realizar, hacen parte del clúster numero 3 o Grupo3 para el cual se ha determinado como se

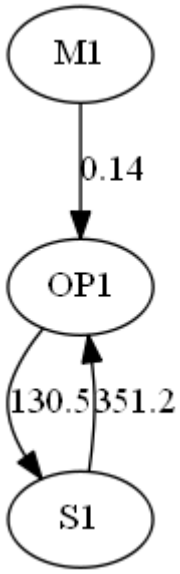
observa en la **tabla de penalizaciones** un valor de 10 por intervención a realizar, esto quiere decir que si debe realizar 3 intervenciones el valor del vértice que comunica el nodo inicial con el medico será de 30, de igual forma es realizado el calculo para el vértice entre el nodo inicio y el médico, el cual da como resultado 10 siendo la penalización para este de 5.

El valor 0.68 que se encuentra en el vértice que se encuentra uniendo el medico 1 (M1) con la intervención 1 (OP1) es calculado por medio de la *Ecuación15: cálculo del estimado para el tiempo de operación* en donde los valores son reemplazados por el rango de tiempo que esta siendo evaluado, contra el promedio del tiempo para dichos datos este proceso es desarrollado para todas las aristas.

Uno de los proceso mas importantes que ocurren dentro de la interacción de nodos y aristas, es el proceso de alimentación y retroalimentación de los nodos de salas de cirugías (S), con los nodos de operaciones (OP), como es posible observar, la capacidad inicial para todas las salas es de 1080 minutos lo que es equivalente a 18 horas de disponibilidad de sala, a la cual debe restarse el tiempo estimado, hay que tener en cuenta que cualquier algoritmo tomara el camino por la arista que tenga un menor valor, por ello, mientras exista capacidad de tiempo dentro de la sala, el algoritmo seguirá generando la agenda dentro de la misma; el proceso para la primera iteración de este segmento seria (por tura más corta):



Como se puede observar la primera ruta arrojada devuelve por medio de la arista que comunica el nodo de S1 con OP2 el nuevo valor de la capacidad disponible para esa sala, este proceso de redistribución ocurre para cada una de las operaciones, de esta forma el cargue de la información le permite reevaluar el punto mas optimo para hacer la distribución de la agenda.



De esta forma se van actualizando constantemente cada una de las aristas que componen el grafo y una vez se haya ocupado la capacidad de tiempo de la sala para el día, continuara agendando de igual manera cada una de las intervenciones

8. CONCLUSIONES

El trabajo que se ha desarrollado para la reducción de las incertidumbres en los tiempos de intervención quirúrgica, con el fin de mejorar el proceso de asignación de quirófanos presenta grandes oportunidades de aplicación para la realización de un proceso para la adaptación a diferentes métodos, el método que se ha aplicado a este problema particular presenta la facilidad en la aplicación e implementación, desarrollando un coste menor en las primeras etapas ejecutadas para el proceso de construcción del agendamiento o asignación de salas, este proceso, genera una forma mucho mas flexible para desarrollar la asignación del bloque quirúrgico teniendo en cuenta una serie de etapas que son fundamentales en el desarrollo de la misma.

En primera instancia se encuentra e factor de inputs, este factor determina la estructura que tendrán los datos para ser ingresados en el clúster, lo primero que se ha observado durante este proceso, es que fundamentalmente se cuentan con técnicas que no simplifican el proceso de homogeneización de las componentes de las variables, esto, para casos en los cuales no se tiene una clara estructura de datos predefinida, este proceso de alistamiento o ingeniería de datos es uno de los mas tardados en desarrollarse, si las organizaciones cuentan con DataSets que cuenten con información estructurada y homogénea cualquier proceso de implementación, será mucho más fácil de desarrollar. Respecto a este punto también fue claro que en el proceso de ingeniería de datos existen herramientas que, aunque muy potentes quedan cortas para generar un relacionamiento profundo de algunas componentes, como por ejemplo, los casos particulares en los cuales existía mas de una forma de estructurar cierta información que contenía una misma clasificación, o pertenecía a un mismo grupo de datos.

En segundo lugar los complejos algoritmos para el desarrollo del agendamiento del bloque quirúrgico existentes dentro de la literatura aunque poco flexibles, presentan grandes ventajas, por ejemplo, existen algunos algoritmos que presentan ventaja en la solución para el agendamiento simple, pero no tienen en cuenta algunas variables que son fundamentales para la evaluación completa del entorno, los algoritmos y métodos desarrollados no solo deben ser flexibles en su aplicación si no también en sí mismos, tener la facilidad de disgregarse de tal forma que la aplicación de su estructura más fuerte sea sencilla de aplicar; es por ello que el método de clusterizacion, el cual presenta una gran fortaleza en la solución de problemas fundamentales, aplicado en otras estructuras, o unido con otros métodos de agendamiento para la asignación de quirófanos puede presentar una forma

novedosa de interactuar con el problema, interactuando no solo con el problema si no también con diferentes algoritmos para el agendamiento.

No todos los métodos para generar los clúster pueden ser los mas adecuados a un problema en particular, es por ello que, aunque en apariencia la clusterizacion presenta un método sencillo y de bajo costo para disminuir la incertidumbre del tiempo de operación, también tiende a ser un método bastante complejo, el cual requiere de una excelente y perfecta preparación de datos, tal como se presenta en el presente documento, en cuanto a algunos problemas que tenga el desarrollo de los Clusters en su mayoría todos se encuentran relacionados a estructuras de datos.

Finalmente se encuentra el desarrollo e implementación teórica del método de clusterizacion a un modelo en particular, en este caso se ha desarrollado una adaptación de un problema en particular, para visibilizar la efectividad en agendamiento. Tomando como punto de partida los tiempos con los cuales se cuenta y su estimación dentro de las componentes para cada una de las variables, en cuando a la forma en la cual se comportan estas clases, es necesario tener en cuenta que como ya se ha dicho existen diversos factores que pueden incidir en el tiempo de intervención quirúrgica y una de las facilidades que presenta la clusterizacion es que estas pueden ser incluidas dentro de un modelo, teniéndolas en cuenta solo como una variable mas para generar el clúster; es por ello que este método presenta flexibilidad y una estructura dimensional de datos que se puede adaptar a una necesidad en particular o adaptarse propiamente al cambio ambiental o adaptándose y evolucionando a medida que lo hacen los factores externos.

9. ANEXOS

8.1 ECUACIONES DE BÚSQUEDA

Anexo 8.1.xlsx

8.2 HERRAMIENTAS DE INGENIERÍA APLICADAS

Anexo 8.2.xlsx

8.3 INFORME TECNICO PROCESAMIENTO

ANDRÉS FELIPE VEGA SIERRA

Estudiante Ingeniería industrial

Universidad Santo Tomás

EQUIPO DEL PROYECTO

Helien Parra Riveros

Sebastián Gustavo Moreno Barón

Investigador Principal

INFORME PROCESAMIENTO DE INFORMACIÓN

INTRODUCCIÓN

En el siguiente informe presenta la metodología utilizada para el proceso de preparación de la información y la selección de las variables, con las cuales se trabajará para el desarrollo del algoritmo. Este proceso normalmente se conoce como ingeniería de datos, en el cual el tratamiento que se realiza a los mismos se enfoca en la estructura que tiene la información según su procedencia y las variables que se han definido previamente para el procesamiento que se desea realizar, es decir, para este caso, se ha notado que algunos formatos de las variables no corresponden al análisis que se desea hacer, algunos de los datos que componen la información no tienen la estructura adecuada para su procesamiento, por lo tanto, las componentes que corresponden a formas no binarias o datos cualitativos requieren de un profundo análisis para una adecuada transformación. A continuación, se presenta detalladamente el proceso para cada una de las variables.

OBJETIVO DEL INFORME

Presentar la metodología para el preprocesamiento de los datos de intervenciones quirúrgicas del hospital San José, desarrollando la solución aplicada a cada una de las principales problemáticas par la homogeneización de las componentes de las variables.

ALCANCE DEL INFORME

El presente informe contendrá el desarrollo que se ha ejecutado para el procesamiento de cada una de las variables que se incluirán para alimentar la tesis, se explicara a fondo la metodología de forma general, con el fin de que esta pueda

ser replicada en cualquier tipo de bases de datos, con errores de tipeo comunes, clasificación para géneros y anonimización de información, para codificación de diferentes clases. Junto con este análisis de metodología se anexaran los resultados encontrados, en cuento a la información resultante, como cantidad o frecuencia de algunos datos, clasificación de los tiempos, entre otros, por medio de la visualización de los mismos.

ANONIMIDAD Y PREPARACIÓN DE LA INFORMACIÓN

El proceso de anonimizacion de la data corresponde al proceso de tomar cada una de los componentes que puedan poner en evidencia datos particulares que atenten contra la privacidad de las personas que se encuentren registradas en dicha base, tales como: médico y paciente. En este caso es necesario hacer una diferenciación entre cada una de las componentes de estas variables sin que las mismas pongan en evidencia de forma discriminatoria un caso en particular.

En segunda estancia para la utilización de los datos es indiferente componentes de las variables tales como, nombre completo del paciente, en este caso será de utilidad una discriminación que haga referencia al sexo del paciente, ya que este si puede ser un dato de gran utilidad. Para el caso de médico tratante, por lo anteriormente ya descrito en cuanto al tratamiento de la información y el respeto a la privacidad existente en los registros, fue necesario realizar una codificación que no hace alusión al nombre de un médico en particular si no simplemente discrimina que el paciente “A” fue tratado por el médico especialista x, y o z, que es una diferenciación para los estadísticos y resultados, con el cual será posible observar que existen N número de médicos especialistas diferentes.

PREPARACIÓN DE DATOS DEL PACIENTE

Como ya se ha expresado anteriormente por cuestiones de tratamiento de la información y de la sensibilidad de la misma, fue necesario transformar cada uno de los nombres de los pacientes y discriminarlos solo por sexo; puesto que cuando se desea alimentar el algoritmo, tener una variable que está compuesta por tantas características particulares no representa valor; puesto que dentro de la base suministrada por el hospital no estaba discriminado este dato del paciente, para ello el proceso seguido consta de:

4. **Alimentación por librerías:** Se tomo una librería la cual contenía más de 20.000 nombres ordenados de forma alfabética, esta librería estaba compuesta por la variable nombre y sexo, esto con el objetivo de realizar una comparativa entre el nombre del paciente y el sexo al cual el mismo podía corresponder.
5. **Comparación por composición:** Puesto que cada nombre tiene una composición particular de tres tipos
    a. **Nombre M & F:** Esta composición hacer referencia a los nombres que tienen según la librería, un primer nombre de tipo masculino, y en su segundo nombre uno de tipo femenino, en este caso el nombre será tomado como masculino, por ejemplo: José María = M.
    b. **Nombre F & M:** Esta segunda composición hace referencia a los nombres compuestos por dos tipos, el primero un nombre de tipo femenino y el segundo de tipo masculino, en cuyo caso el nombre será clasificado como tipo femenino, por ejemplo: María José = F
    c. **Nombre simple:** Este caso es el más simple de clasificar puesto que solo toma el nombre del paciente y comparándolo con la librería que se encuentra clasificada según género, determina si se trata de una persona de sexo masculino o femenino.
    d. **Nombre igual:** El caso de nombre igual es aquel en que los nombres que componen la identificación del paciente corresponden al mismo tipo, es decir, M & M o F &F; en cuyo caso la clasificación se hará basado en el sexo que defina.
6. **Generación de reglas:** Finalmente las reglas se encuentran resumidas a continuación en forma de código en lenguaje Python, aunque el procesamiento finamente se encuentre expresado en un archivo xlsx; por medio de condicionales se realizó la parametrización para cada uno de los casos descritos anteriormente.

```
paciente['PRIMER']==paciente['PRIMER'].str.lower()
paciente['SEGUNDO']==paciente['SEGUNDO'].str.lower()
datos['NOMBRE']==datos['NOMBRE'].str.lower()

print("_____PRIMER NOMBRE_____")
paciente['GENERO1']=paciente.PRIMER.isin(datos.NOMBRE)
datos['GENERO3']=paciente.PRIMER.isin(datos.NOMBRE)
print(paciente['GENERO1'].value_counts())
print(datos['GENERO3'].value_counts())
print(paciente.loc[paciente.GENERO1==True, 'PRIMER'])
print(datos.loc[datos.GENERO3==True,'GENERO'])
print("_____SEGUNDO NOMBRE_____")
paciente['GENERO2']=paciente.SEGUNDO.isin(datos.NOMBRE)
datos['GENERO4']=paciente.SEGUNDO.isin(datos.NOMBRE)
print(paciente['GENERO2'].value_counts())
print(datos['GENERO4'].value_counts())
print(paciente.loc[paciente.GENERO2==True, 'SEGUNDO'])
print(datos.loc[datos.GENERO4==True,'GENERO'])
```

En primera instancia se puede observar cómo se ha generado el cruce entre la Liberia que contiene los nombres juntos con los nombres de los pacientes, para que de este modo determine si dicho nombre corresponde a femenino o masculino.

Lo siguiente será aplicar las reglas establecidas para cada uno de los casos realizando una comparación con condicionales dentro de Excel donde se definirá de la siguiente manera:

if (and(a=”F”; b=”F”); “F”; if(and(a=”M”; b=”M”); “M”))

donde la anterior será la representación del cuarto caso, para el tercer caso el cruce inicial mostrado en el código anterior será suficiente. Los casos dos y tres se presentan a continuación:

if(and(a=”F”; b=”M”); “F”; if(and(a=”M”; b=”F”); “M”))

Para cualquiera de los casos los condicionales bastan para cubrir la primera parte de los casos; en aquellas ocasiones en las que el condicional aplicado no puede cubrir una particularidad, con el simple algoritmo mostrado anteriormente bastará

PREPARACIÓN DE DATOS POR MÉDICO

Uno de los procesos que fue necesario realizar de forma manual fu el tratamiento de los datos de médico, como ya se ha dicho cada uno de los médicos se encuentran clasificados por un código que diferencia uno entre otro, dicho código se compone de tres variables, las cuales se explicaran más adelante. Puesto que dentro de la base de datos se encuentran diferentes errores de digitación que hace muy complejo el proceso de identificación de forma automática se hace necesario hacerlo manualmente, aplicando diferentes filtros y encontrando equivalencias entre cada una de las expresiones que hagan referencia a la identificación de cada uno de los individuos.

Durante este proceso de identificación para llevar a cabo el cifrado de cada uno de los códigos que identificarían un doctor de otro, se encontraron diferentes tipos de errores que se encuentran clasificados en:

1. **Error por sexo:** Este hace referencia a los casos en los cuales:
    a. La enfermera o digitadora pone la sigla Dr. que hace referencia al masculino en los casos en los que la doctora se encuentra identificada con el género femenino lo que hace más complicado su identificación si esta se encuentra mezclada con el error por nombre incompleto.
    b. La o el digitador(a) antepone la sigla contraria a la identificación por sexo mal escrita, por ejemplo *"Srs"*, "*Drc".* Etc.
2. **Error por letras:** Uno de los errores más comunes cuando se requiere llenar un campo sin ningún tipo de plantilla es que la persona encargada de digitar escriba con errores ortográficos o de forma incompleta el nombre del sujeto, por ejemplo: en lugar de escribir José Luis Forero escriba "Jose Luis Frero" lo que también complica la identificación del sujeto.
3. **Error por espacios contenidos:** Los espacios contenidos en la escritura del nombre pueden también generar problemas en la identificación del médico, por ejemplo, en lugar de escribir José Luis Forero, escriba: "*José Lu is For*

ero" este también es uno de los errores más comunes a la hora de realizar la identificación.

4. **Error por puntuación:** El error de puntuación o generar diferencias en la forma en la cual esta puntuado un nombre se n el número de veces que se escribe el nombre y N el número de formas en las cuales se puede puntuar el mismo, genera confusiones para identificar si el nombre "a" corresponde realmente al mismo doctor, o hace referencia a un médico que pueda tener características similares en su nombre.
5. **Error por nombre incompleto:** Es muy común que dentro de la digitación se acostumbre llamar al médico por su apellido, en este caso, de tener dos doctores con el mismo apellido será complicado identificar a cuál de los dos sujetos se hace referencia; para ello fue necesario tener en cuenta la especialidad puesto que existen n formas de digitar el nombre del sujeto y x_1 las veces que este puede coincidir con un apellido o nombre incompleto similar para ello se compara con las especialidades y se cruza de la siguiente forma

$$\boldsymbol{n} = aX_1$$
$$\boldsymbol{donde}, E_1 = a'$$
$$aX_1|a = E_1 \rightarrow X_1 = 1$$
$$aX_1 = a'$$

Ecuación 1 calculo para nombre incompleto

Donde E_n representa la especialidad; cómo se puede observar en la expresión anterior muestra la comparación en la cual se hace la identificación del nombre al que se hace referencia por medio de un cruce de especialidades para ciertos nombres repetidos imposibles de identificar, esto reduce el número de formas en las cuales se puede escribir el nombre al resultado de aX_1 donde el numero de coincidencias que tendrá X_1 corresponde a 1 debido al cruce realizado.

Teniendo en cuenta lo anterior se ha generado una expresión para el numero de formas en las cuales puede generarse la combinatoria de los errores de tal forma que se explique como por medio de la simple observación sea posible identificar los errores y traducirlos a un formato simple para su codificación

$n = número\ de\ formas\ de\ escritura$

$a = número\ de\ combinaciones\ de\ la\ palabra$

$X_1 = número\ de\ coincidencias\ por\ busqueda\ en\ la\ base$

$C = conteo\ de\ formas\ de\ combinación\ para\ una\ misma\ palabra\ (frecuencia)$

$E_n = numero\ de\ especialidades$

$\delta = desviación\ de\ numero\ existente\ segun\ historico\ para\ escritura$

$I = número\ de\ casos\ resumido\ posibles\ a\ observar$

$$I = \frac{((C * E_n)\delta) - ((aX_1 * E_n)\delta)}{n}$$

$Donde, n = aX_1\ Entonces$:

$$I = \frac{((C * E_n)\delta) - ((n * E_n)\delta)}{n}$$

Ecuación 2 resumen de número de casos observados

Ejemplo:

Para el siguiente caso se desea determinar si Jose corresponde a un determinado nombre, para ello se resumirá el número de posibles visualizaciones con la

expresión anterior. Se encuentran dos diferentes doctores con el nombre de pila José:

$Sujeto\ 1 = José\ bohorquez$

$Sujeto\ 2 = Daniel\ Jose$

$Sujeto\ a\ comparar = Jose$

También se encuentra que el número de formas escrito el nombre son 4, teniendo en cuenta los errores anteriormente descritos; su frecuencia encontrada, es decir, el número de veces que aparecen escritos se presenta frente a cada uno:

$Forma\ 1 = José \rightarrow 7$

$Forma\ 2 = \ Jo\ se\ \rightarrow 5$

$Forma\ 3 = Jose \rightarrow 8$

$Forma\ 4 = Jóse \rightarrow 10$

La desviación de estos datos teniendo en cuenta la frecuencia que aparece cada uno de ellos es: $\delta = 2.08$ y finalmente la frecuencia de datos encontrados corresponde a 30 que es igual a:

$$C = \sum_{i}^{1} n_i$$

$C = Fforma\ 1 + Fforma2 + Fforma3 + Fforma4$

$C = 7 + 5 + 8 + 10 = 30$

Resolviendo:

$$I = \frac{((30 * 5)2.08) - ((6 * 5)2.08)}{6}$$

$$I = 38$$

El número de casos observados tomando en cuenta todos los posibles errores para este ejemplo es de 38

Codificación por médico especialista: Cada uno de los médicos especialistas tienen una codificación la cual está compuesta por tres variables, cada una de ellas en el orden que se presentan separadas por un punto“.”:

4. **Sexo:** Esta es la primera letra que se encuentra dentro de la codificación y hace referencia al género que identifica al médico especialista (M para masculino y F para femenino)
5. **Especialidad:** Las especialidades han sido resumidas de la siguiente manera, y es el segundo digito que se encuentra en la codificación:
    a. Cdva: Cirugía Cardio Vascular
    b. ClDl: Clínica del dolor.
    c. CxGn: Cirugía general.
    d. CxPl: Cirugía plástica.
    e. Dmt: Dermatología.
    f. Gast: Gastroenterología.
    g. Ginc: Ginecología.
    h. Neum: Neumología.
    i. Neur: Neurología.
    j. Odnt: Odontología.
    k. Oftm: Oftalmología.
    l. Ortp: Ortopedia.
    m. Otgia: Otorrinolaringología.

n. Rdg: Radiología.
o. Rsn: Resonancia.
p. Urlg: Urología.
q. Hemt: Hematología.

6. **Nombre:** El nombre completo del médico especialista se ha resumido por medio de un consecutivo que va desde el 1 al 180, esto no quiere decir que existan exactamente 180 doctores dentro de la base de datos, puesto que los mismo tuvieron que ser modificados debido a la variabilidad de cada uno de los casos.

Finalmente, la combinación de cada una de estas variables da como resultado una codificación compuesta de la siguiente forma: Sexo/Especialidad/Cd nombre, en cada uno de los casos se encontrará la misma de la siguiente manera, por ejemplo:

Cirujano 1	sexo	especialidad	serie
MARQUEZ AZUERO	M	Cdva	33
ALVARO GOMEZ	M	Otgia	63

Dando como resultado para los casos mostrados en la tabla la siguiente codificación respectivamente:

M.Cdva.33

M.Otgia.63

Existencia de más de una especialidad para Cx: Dentro de los casos se encontró que para un mismo procedimiento se encontraba más de un médico especialista en la sala, para ello la codificación se vio combinada, pero no mezclada, separándola por medio de una barra "/" quedando las codificaciones de la siguiente forma:

M.Cdva.33/M.Otgia.63

TRATAMIENTO DE TIEMPOS

Para el tratamiento de los tiempos se definió en primera instancia la escala en la cual se iban a tratar los mismo, que para este caso se ha tomado la decisión de que sea en minutos, seguidamente se discriminaron las particularidades respecto a los tiempos y a partir de ahí se definió el proceso correspondiente a cada uno de estos casos para realizar la transformación de horas en formato hh:mm:ss a minutos como entero.

1. **Tiempo de inicio > a tiempo de finalización:** Este caso depende directamente del formato en el cual se ha estado trabajando la data, es decir, los tiempos al ser trabajos en formatos de 24 hora los casos para los cuales la hora de inicio superaba la hora de finalización como entero. Por ejemplo:

   Hora de inicio Cx: 23:30

   Hora fin Cx: 1:30

   Para este caso como es posible notar, la hora de inicio es las 11 de la noche y la intervención concluye a la 1:30 de la madrugada lo que da como resultado dos horas de cirugía.
2. **Tiempo de inicio < a tiempo de finalización:** Como en el caso anterior depende directamente del formato, aunque este caso dentro es de los más comunes dentro de la base. Por ejemplo:

   Hora de inicio Cx: 13:30

   Hora fin Cx: 15:30

   La hora de inicio para este caso es la 1 de la tarde y su hora de finalización son las 3:30 de la tarde lo que da como resultado 2 horas de cirugía.
3. **Tiempo de inicio = tiempo de finalización:** Existen casos dentro de la base en los cuales por cuestiones de digitación la hora de inicio es igual a la hora de finalización en estos casos al hacerse imposible obtener un tiempo de intervención lógico, por lo tanto, estos casos fueron eliminados.

4. **Formato no correspondiente imposible de transformar:** Durante el proceso de digitación y de compilación de todos los años dentro de una misma base, algunos de los datos perdieron su formato de origen, y al tratar de regresar los mismos a un formato en el cual se pueda realizar la transformación no fue posible, estos datos también se eliminaron.

Los dos últimos casos no representan un porcentaje considerable de la base, aunque es importante nombrar la particularidad de los mismos.

Teniendo en cuenta los casos anteriormente nombrados se han definido las siguientes expresiones:

$$T_f > T_i \rightarrow T_f - T_i = \Delta T$$

$$T_f < T_i \rightarrow T_i - T_f = \Delta T$$

$$T_f = T_i \rightarrow no\ operable$$

Donde ΔT es el delta del tiempo de cirugía, que representa el tiempo real de intervención, el cual inicia desde que el paciente ingresa a la sala de Cx y finaliza con el momento en el cual se inicia la limpieza de la sala.

ORGANIZACIÓN DE LAS VARIABLES

Para concluir se tomaron cada una de las variables organizadas y se ordenaron dentro de una tabla que contiene 5 diferentes variables:

6. **Fecha de registro:** Corresponde a un dato de control el cual indica la fecha en la cual se tomó dicho registro.

7. **Genero del paciente:** Teniendo en cuenta el tratamiento que se realizó anteriormente para tratar el género del paciente por medio de librerías y haciendo uso de condicionales.
8. **Código de especialista:** Este hace referencia a la clasificación que se le ha dado al médico especialista por medio de una codificación la cual se ha expuesto anteriormente.
9. **Tiempo de intervención:** Tiempo real que dura el procedimiento o intervención quirúrgica en minutos.
10. **Especialidad intervenida:** la especialidad intervenida hace referencia a la especialidad que prevalece sobre la intervención, esta se ha decidido dejar para aquellos casos en los cuales hay más de un médico especialista registrado para dicha intervención.

Las anteriores han sido hasta el momento las variables que se ha considerado que son de alta importancia para alimentar el algoritmo que se desarrollará, se ha efectuado todo este proceso puesto que es necesario tener un espectro un poco más cerrado y claro de las variables que tienen más peso y el efecto de las mismas sobre el desarrollo de la programación,

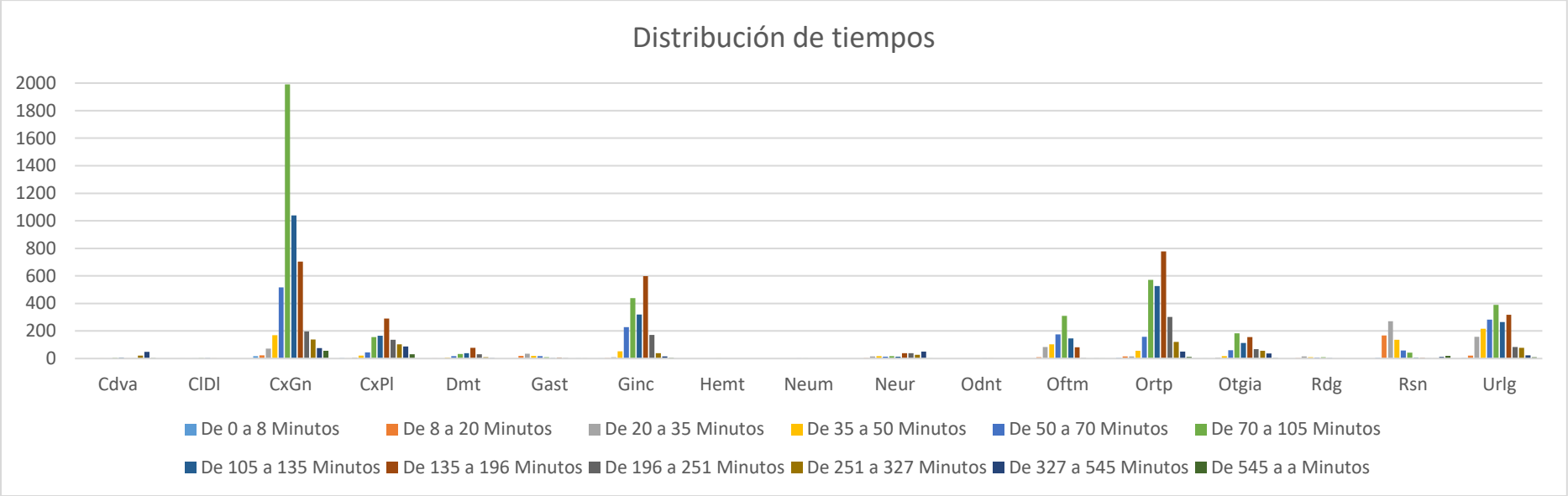
RESUMEN DE RESULTADOS

A continuación, se presentan los resultados obtenidos a partir de una tabla resumen que se realizó con cada una de las variables ordenadas. Los resultados arrojados presentan una alta frecuencia de intervenciones quirúrgicas para la especialidad de medicina general con una alta frecuencia de entre 70 a 105 minutos de intervención, es decir, que el tiempo de intervención que mas se repite para intervenciones de cirugía general el intervalo ya mencionado, las siguientes intervenciones con mas frecuencia son ortopedia y ginecología con un intervalo para ambos casos de tiempo de intervención, entre 135 y 196 minutos para ambos casos. El comportamiento de las variables especialidad, se ha encontrado que solo algunas como la especialidad de oftalmología, ortopedia, otorrinolaringología cuentan con dicha distribución, lo que hace ocasiona que definir la probabilidad de ocurrencia de cada uno de los tiempos que la componen sea aún más fácil.

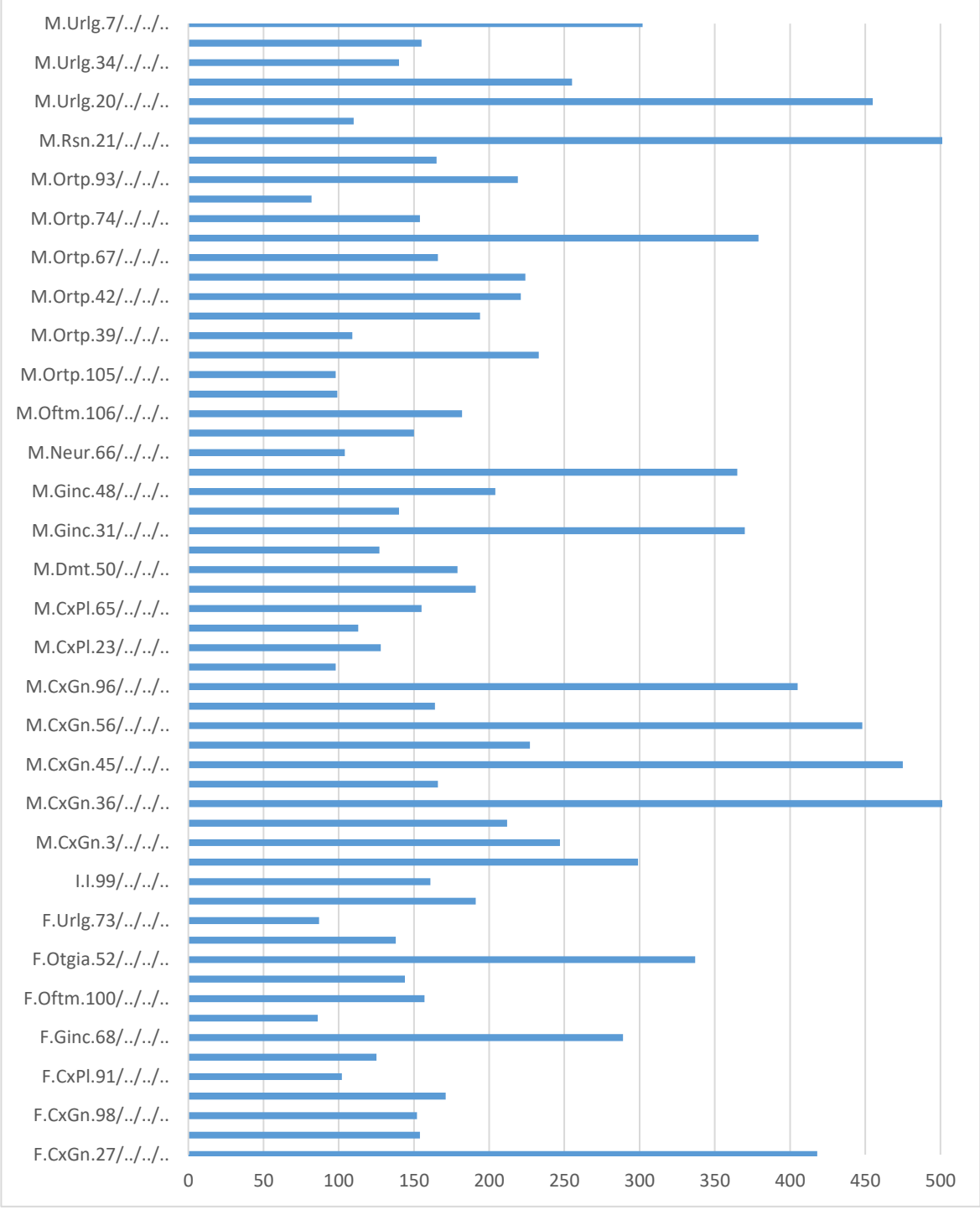
Existen otra serie de particularidades que se pueden inferir de la visualización de los datos, por ejemplo, es posible a partir de la simple observación clasificar aquellas intervenciones que tienen una serie de similaridades entre los intervalos para los tiempos de intervención, como cirugía plástica, ginecología y ortopedia, que aunque no tienen nada que ver una con la otra en cuanto a sus procedimientos, tienen una gran similaridad en la forma en la cual se presenta el porcentaje de sus frecuencias, de dichos o intervalos de tiempos.

Posteriormente se pasará a visualizar uno por uno los casos que menor frecuencia tienen, su comportamiento y el comportamiento por sexo por código de doctor, entre otros datos que serán de ayuda para entender la forma en la cual se comportan las intervenciones.

TABLA DE FRECUENCIA DE TIEMPOS



FRECUENCIA POR ESPECIALISTA

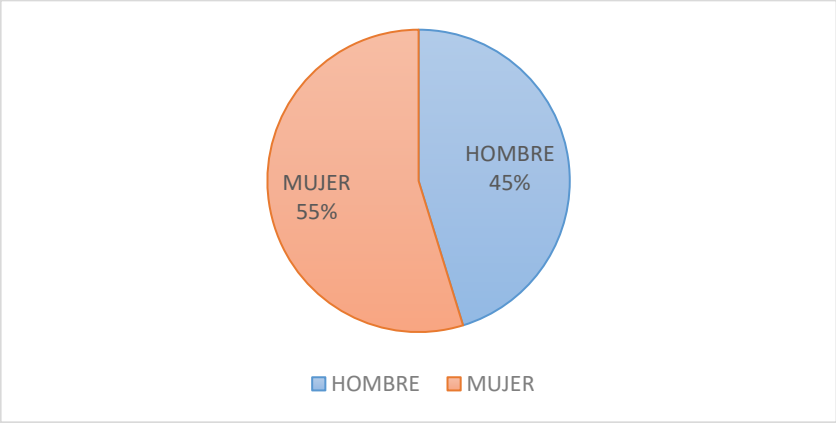


La tabla de frecuencia de especialista representa la frecuencia, es decir, el numero de veces que cada uno de los especialistas ha estado en una intervención

quirúrgica; teniendo en cuenta que la forma en la cual se presentan cada uno de los especialistas corresponde a la codificación mostrada anteriormente para la anonimización de los datos del médico especialista.

Como es posible observar, existen una serie de médicos y de que sobresalen, como ya se ha dicho antes, por la cantidad de intervenciones realizadas, dentro de estos encontramos al medico identificado con el código M.CxGn.36 el cual es quien mas sobresale seguido por M.CxGn.45, M.CxGn.56 y M.CxGn.96, como es posible observar estos cuatro primeros especialistas pertenecen a una misma especialidad, cirugía general, por lo cual, el dato de los especialistas, respalda la visualización del grafico anterior, a partir de aquí se podría partir de la premisa que estos doctores se encuentran en una media de tiempo de intervención quirúrgica de entre 70 a 105 minutos, ya que esta es respaldada por el DashBoard presentando anteriormente , en donde se visualizan los intervalos de tiempos para cada una de las especialidades. Esta visualización de las frecuencias, también será útil para determinar la relación existente entre cada uno de los datos, los grupos con los cuales cada variable se puede relacionar y complementar la caracterización para el análisis de los Clusters.

PACIENTES ATENDIDOS SEGÚN SEXO

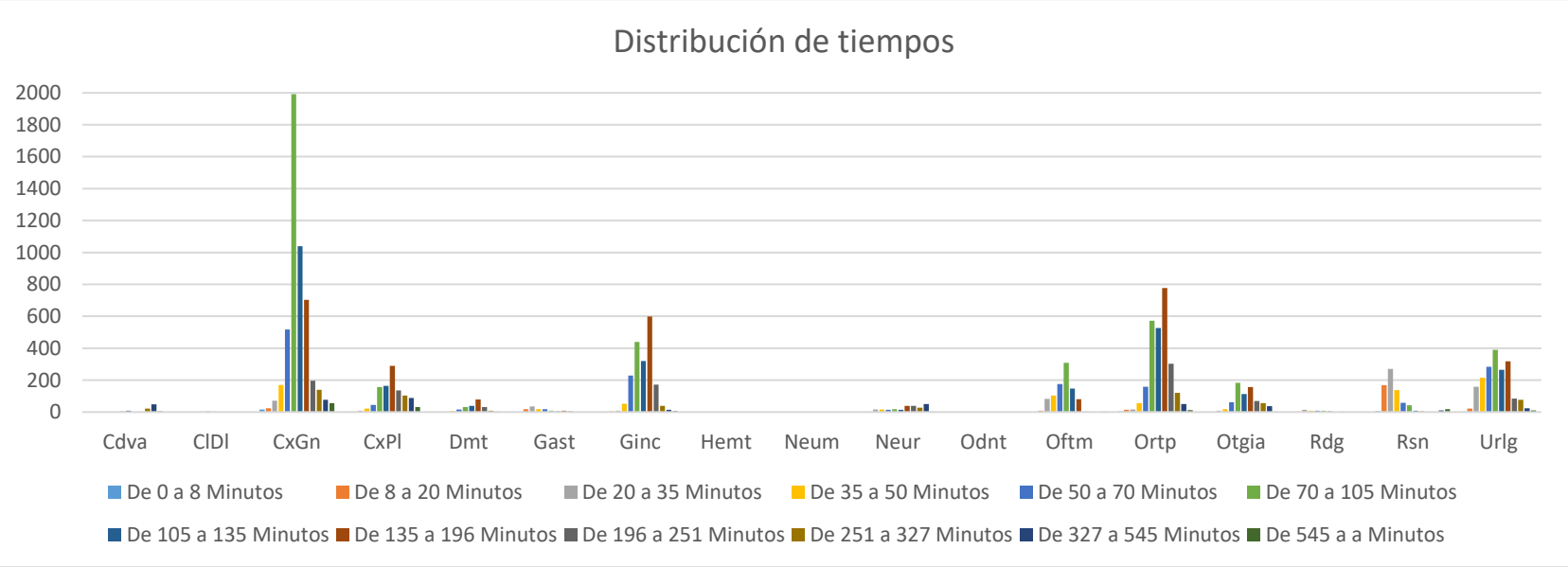


El numero de pacientes atendidos en su mayoría corresponden a hombres (sin caracterización de edad) quienes en su mayoría se someten a una intervención quirúrgica correspondiente a la especialidad de medicina general. Existen casos en los cuales no se especifica el campo de nombre, lo cual imposibilita. Que se realice la identificación del sexo.

8.5 PROCESAMIENTO DE DATOS.EXCE

PROCESAMIENTO DE DATOS.xlsx

8.6 TABLA DE FRECUENCIA DE TIEMPOS



8.7 ALGORITMO PYTHON PARA CLUSTERES

IMPORTE DE LIBRERIAS

```
mport numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import scipy.cluster.hierarchy as sch
from scipy.cluster.hierarchy import dendrogram, linkage, fcluster
%matplotlib inline
```

DESARROLLO DE MATRIZ

```
variables_array = datos.iloc[:,[1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16
,17,18]].values
variables_array
```

PRUEBA DE CADA UNO DE LOS METODOS AGLOMERATIVOS

```
M_mediana = linkage(variables_array,'median')
M_unico = linkage(variables_array,'single')
M_promedio = linkage(variables_array,'average')
M_centroide = linkage(variables_array,'centroid')
M_ward = linkage(variables_array, 'ward')
```

GRAFICA DE CADA UNO DE LOS DENDOGRAMAS

```
DISTANCIAM_1 = 7.08
plt.figure(figsize=(10, 5))
plt.title('DENDOGRAMA MEDIAN')
plt.xlabel('Etiquetas')
plt.ylabel('Distancia')
sch.dendrogram(M_mediana, truncate_mode='lastp', p=150, leaf_rotation=
90., leaf_font_size=8.,)
plt.axhline(y=DISTANCIAM_1, c='k')
plt.show()
DISTANCIAM_2 = 7.08
plt.figure(figsize=(10, 5))
plt.title('DENDOGRAMA SINGLE')
plt.xlabel('Etiquetas')
```

```
plt.ylabel('Distancia')
sch.dendrogram(M_unico, truncate_mode='lastp', p=150, leaf_rotation=90
., leaf_font_size=8.,)
plt.axhline(y=DISTANCIAM_2, c='k')
plt.show()
DISTANCIAM_3 = 7.08
plt.figure(figsize=(10, 5))
plt.title('DENDOGRAMA AVERAGE')
plt.xlabel('Etiquetas')
plt.ylabel('Distancia')
sch.dendrogram(M_promedio, truncate_mode='lastp', p=150, leaf_rotation
=90., leaf_font_size=8.,)
plt.axhline(y=DISTANCIAM_3, c='k')
plt.show()
DISTANCIAM_4 = 7.08
plt.figure(figsize=(10, 5))
plt.title('DENDOGRAMA CENTROID')
plt.xlabel('Etiquetas')
plt.ylabel('Distancia')
sch.dendrogram(M_centroide, truncate_mode='lastp', p=150, leaf_rotatio
n=90., leaf_font_size=8.,)
plt.axhline(y=DISTANCIAM_4, c='k')
plt.show()
DISTANCIAM_5 = 7.08
plt.figure(figsize=(10, 5))
plt.title('DENDOGRAMA WARD')
plt.xlabel('Etiquetas')
plt.ylabel('Distancia')
sch.dendrogram(M_ward, truncate_mode='lastp', p=150, leaf_rotation=90.
, leaf_font_size=8.,)
plt.axhline(y=DISTANCIAM_5, c='k')
plt.show()
```

SELECCIÓN Y ETIQUETADO DE LAS COMPONENTES

```
from sklearn.cluster import AgglomerativeClustering
hc = AgglomerativeClustering(n_clusters = 3, affinity = 'euclidean', l
inkage = 'ward')
y_hc = hc.fit_predict(variables_array)
y_hc
```

GRAFICA DE LOS GRUPOS EJECUTADOS

```
plt.scatter(variables_array[y_hc == 0, 0], variables_array[y_hc == 0,
1], c='red', label='Grupo 1')
plt.scatter(variables_array[y_hc == 1, 0], variables_array[y_hc == 1,
1], c='green', label='Grupo 2')
plt.scatter(variables_array[y_hc == 2, 0], variables_array[y_hc == 2,
1], c='blue', label='Grupo 3')
plt.title('CLUSTERIN PARA METODO DE WARD')
plt.xlabel('Etiquetas')
plt.ylabel('Dintacia')
plt.legend()
plt.show()
```

10. REFERENCIAS

Abedini, A., Li, W., & Ye, H. (2017a). An optimization model for operating room scheduling to reduce blocking across the perioperative process. *Procedia Manufacturing, 10*, 60-70. doi:10.1016/j.promfg.2017.07.022

Abedini, A., Li, W., & Ye, H. (2017b). An optimization model for operating room scheduling to reduce blocking across the perioperative process. *Procedia Manufacturing, 10*, 60-70. doi:10.1016/j.promfg.2017.07.022

Al Hasan, H., Guéret, C., Lemoine, D., & Rivreau, D. (2019a). Surgical case scheduling with sterilising activity constraints. *International Journal of Production Research, 57*(10), 2984-3002. doi:10.1080/00207543.2018.1521015

Al Hasan, H., Guéret, C., Lemoine, D., & Rivreau, D. (2019b). Surgical case scheduling with sterilising activity constraints. *International Journal of Production Research, 57*(10), 2984-3002. doi:10.1080/00207543.2018.1521015

Albareda, J., Clavel, D., Mahulea, C., Blanco, N., Ezquerra, L., Gómez, J., & Silva, J. M. (2017a). ¿Realizamos bien la programación quirúrgica? ¿Cómo podemos mejorarla? *Revista Española De Cirugía Ortopédica Y Traumatología, 61*(6), 375-382. doi://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/j.recot.2017.07.006

Albareda, J., Clavel, D., Mahulea, C., Blanco, N., Ezquerra, L., Gómez, J., & Silva, J. M. (2017b). Do we perform surgical programming well? how can we improve it? [¿Realizamos bien la programación quirúrgica? ¿Cómo podemos mejorarla?] *Revista Espanola De Cirugia Ortopedica Y Traumatologia, 61*(6), 375-382. doi:10.1016/j.recot.2017.07.006

Bam, M., Denton, B. T., Van Oyen, M. P., & Cowen, M. E. (2017). Surgery scheduling with recovery resources. *IISE Transactions, 49*(10), 942-955. doi:10.1080/24725854.2017.1325027

Benchoff, B., Yano, C. A., & Newman, A. (2017). Kaiser permanente oakland medical center optimizes operating room block schedule for new hospital. *Interfaces, 47*(3), 214-229. doi:10.1287/inte.2017.0885

Bravo Bastidas, J. J. (2013). Comentarios sobre la solución de problemas de optimización lineal, enteros mixtos y no lineales de gran escala.

C. West Churchman. (1967). Guest editorial: Wicked problems. *Management Science, 14*(4), B14-B142. Retrieved from https://www.jstor.org/stable/2628678

Calderón, L. A. T. (1996). El sistema de salud de colombia después de la ley 100. *Colombia Médica, 27*(1), 44-47.

Cappanera, P., Visintin, F., & Banditori, C. (2018a). Addressing conflicting stakeholders' priorities in surgical scheduling by goal programming. *Flexible Services and Manufacturing Journal, 30*(1-2), 252-271. doi:10.1007/s10696-016-9255-5

Cappanera, P., Visintin, F., & Banditori, C. (2018b). Addressing conflicting stakeholders' priorities in surgical scheduling by goal programming. *Flexible Services and Manufacturing Journal, 30*(1-2), 252-271. doi:10.1007/s10696-016-9255-5

Castro, P. M., & Marques, I. (2015). Operating room scheduling with generalized disjunctive programming. *Computers and Operations Research, 64*, 262-273. doi:10.1016/j.cor.2015.06.002

Chavez Quisbert, N. (2011). Modelos de programación lineal y no lineal con multiobjetivos. *Revista Varianza,* , 19.

Choque López, J. F. (2011). Tiempos quirúrgicos. *Revista De Actualización Clínica Investiga, 15*, 851.

Ley 1438, (2011).

Ley estatutaria 1751 de 2015, (2015).

Ley 100 de 1993,

Domínguez González, M. N., López-Pardo Pardo, M. E., Rey Liste, M. T., & García Sixto, M. M. (2011). Intervención para reducir la variabilidad de las indicaciones quirúrgicas y la lista de espera de pacientes con prioridad 1. una experiencia en galicia. *Gaceta Sanitaria, 25*(6), 545-548. doi://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/j.gaceta.2011.04.008

Durán, G., Rey, P. A., & Wolff, P. (2017a). Solving the operating room scheduling problem with prioritized lists of patients. *Annals of Operations Research, 258*(2), 395-414. doi:10.1007/s10479-016-2172-x

Durán, G., Rey, P. A., & Wolff, P. (2017b). Solving the operating room scheduling problem with prioritized lists of patients. *Annals of Operations Research, 258*(2), 395-414. doi:10.1007/s10479-016-2172-x

Gauthier, J. B., & Legrain, A. (2016). Operating room management under uncertainty. *Constraints, 21*(4), 577-596. doi:10.1007/s10601-015-9236-4

Guido, R., & Conforti, D. (2017). A hybrid genetic approach for solving an integrated multi-objective operating room planning and scheduling problem. *Computers and Operations Research, 87*, 270-282. doi:10.1016/j.cor.2016.11.009

Gür, S., Eren, T., & Alakas, H. M. (2019). Surgical operation scheduling with goal programming and constraint programming: A case study. *Mathematics, 7*(3) doi:10.3390/math7030251

Guzmán, E. L. (2016). Métricas para la validación de clustering. *Elizabeth León Guzmán,*

Hillier, F. S., Lieberman, G. J., & Osuna, M. A. G. (1997). *Introducción a la investigación de operaciones* McGraw-Hill.

INEGI. (2011). Diseño de la muestra en proyectos de encuesta.

Kuri, Á, & Galaviz, J. (2002). *Algoritmos genéticos* IPN.

Latorre-Núñez, G., Lüer-Villagra, A., Marianov, V., Obreque, C., Ramis, F., & Neriz, L. (2016). Scheduling operating rooms with consideration of all resources, post anesthesia beds and emergency surgeries. *Computers and Industrial Engineering, 97*, 248-257. doi:10.1016/j.cie.2016.05.016

Li, F., Gupta, D., & Potthoff, S. (2016). Improving operating room schedules. *Health Care Management Science, 19*(3), 261-278. doi:10.1007/s10729-015-9318-2

Li, X., Rafaliya, N., Baki, M. F., & Chaouch, B. A. (2017). Scheduling elective surgeries: The tradeoff among bed capacity, waiting patients and operating room utilization using goal programming. *Health Care Management Science, 20*(1), 33-54. doi:10.1007/s10729-015-9334-2

Marques, I., Captivo, M. E., & Barros, N. (2019). Optimizing the master surgery schedule in a private hospital. *Operations Research for Health Care, 20*, 11-24. doi:10.1016/j.orhc.2018.11.002

Martínez, E. (2011). Análisis factorial. *Técnicas De,*

Martí-Valls, J., Ballesta, E., González, R., Solé, M., & Torras, G. (2006). Resultados de un plan de gestión de listas de espera quirúrgica de prótesis articulares. *Gaceta Sanitaria, 20*(3), 248-250. doi://doi-org.crai-ustadigital.usantotomas.edu.co/10.1157/13088858

Marty, J. (2019). Organización del bloque quirúrgico. *EMC - Anestesia-Reanimación, 45*(3), 1-11. doi://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/S1280-4703(19)42458-4

Mateus, C., Marques, I., & Captivo, M. E. (2018). Local search heuristics for a surgical case assignment problem. *Operations Research for Health Care, 17*, 71-81. doi:10.1016/j.orhc.2017.04.001

M'Hallah, R., & Visintin, F. (2019a). A stochastic model for scheduling elective surgeries in a cyclic master surgical schedule. *Computers and Industrial Engineering, 129*, 156-168. doi:10.1016/j.cie.2019.01.030

M'Hallah, R., & Visintin, F. (2019b). A stochastic model for scheduling elective surgeries in a cyclic master surgical schedule. *Computers and Industrial Engineering, 129*, 156-168. doi:10.1016/j.cie.2019.01.030

Millan, M. (2010). Segmentación o clustering. *Universidad Del Valle, Cali, Colombia.(43),* , 2-43.

Min. salud y protección social. (2017). RESOLUCIÓN 5269

PLAN DECENAL DE SALUD PÚBLICA (2012).

Moreno Millán, E., García Torrecillas, J. M., & Carmen Lea Pereira, M. (2008). Variación de la estancia preoperatoria en españa según grupos de edad, sexo y modo de acceso hospitalario (urgente o programado). *Revista De Calidad Asistencial, 23*(5), 222-229. doi://doi-org.crai-ustadigital.usantotomas.edu.co/10.1016/S1134-282X(08)72611-3

Morillo, D., Moreno, L., & Díaz, J. (2014). Metodologías analíticas y heurísticas para la solución del problema de programación de tareas con recursos restringidos (RCPSP): Una revisión parte 1. *Ingeniería Y Ciencia, 10*(19), 247-271.

Novo, M., Arce, R., Fariña, F., & a de Vega, S. (2003). El heurístico: Perspectiva histórica, concepto y tipología. *Jueces: Formación De Juicios Y Sentencias,* , 39-66.

Olaguíbel, R. A., & GOERLICH, Y JOSÉ MANUEL TAMARIT. (1989). Algoritmos heurísticos deterministas y aleatorios en secuenciación de proyectos con recursos limitados. *Qüestiió, 13*(1, 2, 3), 173-191.

Parra Riveros, H. (2014). *Ingeniería de sistemas complejos y aplicación en la gestión en salud*

Sagnol, G., Barner, C., Borndörfer, R., Grima, M., Seeling, M., Spies, C., & Wernecke, K. (2018). Robust allocation of operating rooms: A cutting plane approach to handle lognormal case durations. *European Journal of Operational Research, 271*(2), 420-435. doi:10.1016/j.ejor.2018.05.022

La salud en Colombia: logros, retos y recomendaciones (2012). . Bogotá: Universidad de los Andes. Retrieved from http://ebookcentral.proquest.com/lib/bibliotecaustasp/detail.action?docID=3211755

Santoso, L. W., Sudiarso, A., Masruroh, N. A., & Herliansyah, M. K. (2017). Development of mathematical model for operating room scheduling. *Journal of*

Engineering and Applied Sciences, 12(21), 5413-5417. doi:10.3923/jeasci.2017.5413.5417

Shafaei, R., & Mozdgir, A. (2019). Master surgical scheduling problem with multiple criteria and robust estimation. *Scientia Iranica, 26*(1E), 486-502. doi:10.24200/sci.2018.20416

Soudi, A., & Heydari, M. (2019). Generating a stable primary schedule for an integrated surgical suite. *International Journal of Medical Engineering and Informatics, 11*(2), 174-203. doi:10.1504/IJMEI.2019.098755

Spratt, B., & Kozan, E. (2016). Waiting list management through master surgical schedules: A case study. *Operations Research for Health Care, 10*, 49-64. doi:10.1016/j.orhc.2016.07.002

Villar, L. A. (2004). La ley 100: El fracaso estatal en la salud pública. *Revista Deslinde, 36*, 28-48.

Wang, T., Meskens, N., & Duvivier, D. (2015). Scheduling operating theatres: Mixed integer programming vs. constraint programming. *European Journal of Operational Research, 247*(2), 401-413. doi:10.1016/j.ejor.2015.06.008

Yin, M., Zhou, B. -., & Lu, Z. -. (2016a). An improved lagrangian relaxation heuristic for the scheduling problem of operating theatres. *Computers and Industrial Engineering, 101*, 490-503. doi:10.1016/j.cie.2016.09.003

Yin, M., Zhou, B. -., & Lu, Z. -. (2016b). An improved lagrangian relaxation heuristic for the scheduling problem of operating theatres. *Computers and Industrial Engineering, 101*, 490-503. doi:10.1016/j.cie.2016.09.003

Zhang, J., Dridi, M., & El Moudni, A. (2019a). A two-level optimization model for elective surgery scheduling with downstream capacity constraints. *European Journal of Operational Research, 276*(2), 602-613. doi:10.1016/j.ejor.2019.01.036

Zhang, J., Dridi, M., & El Moudni, A. (2019b). A two-level optimization model for elective surgery scheduling with downstream capacity constraints. *European Journal of Operational Research, 276*(2), 602-613. doi:10.1016/j.ejor.2019.01.036

Zhou, B., & Yin, M. (2016). Novel operating theatre scheduling method based on estimation of distribution algorithm. *Journal of Southeast University (English Edition), 32*(1), 112-118. doi:10.3969/j.issn.1003-7985.2016.01.019