

Análisis y optimización de dos sistemas de líneas de espera de empresas de logística y transporte de los Estados de Querétaro y Colima

Analysis and optimization of two queueing line systems of logistics and transportation companies of the States of Queretaro and Colima

Juan Manuel Izar Landeta¹

Carmen Berenice Ynzunza Cortés²

Jaime Garnica González³

¹ Juan Manuel Izar Landeta, Doctor en Administración, Profesor Investigador, Universidad del Centro de México, Departamento de Investigación, Líneas de Generación y Aplicación del Conocimiento: Ingeniería Industrial, Pymes y Pensiones y Jubilaciones, 444-123-4194, Código ORCID: 0000-0002-3187-6481, jmizar@hotmail.com

² Carmen Berenice Ynzunza Cortés, Doctora en Administración, Profesor Investigador de Tiempo Completo, Universidad Tecnológica de Querétaro, División Industrial, Líneas de Generación y Aplicación del Conocimiento: Estudios sobre los recursos y capacidades de las organizaciones, 442-209-6148 extensión 2180, Código ORCID: 0000-0003-1469-4763, ynzunzaber@hotmail.com

³ Jaime Garnica González, Doctor en Planeación Estratégica y Dirección de Tecnología, Profesor Investigador de Tiempo Completo, Universidad Autónoma de Hidalgo, Instituto de Ciencias Básicas e Ingeniería, 771-717-2000, jgarnicag@gmail.com

Autor de correspondencia: Juan Manuel Izar Landeta, jmizar@hotmail.com

Resumen:

Objetivo: El objetivo de este estudio es definir mediante el análisis de 2 casos empíricos: el primero un sistema M/G/1 o M/D/1, y el segundo M/M/1 o M/M/S, la tasa óptima de servicio y la conveniencia económica de utilizar o no los mismos, así como validar dos ecuaciones desarrolladas ex profeso para calcular el número óptimo de servidores.

Metodología: Primero se identificaron los modelos que aplican para cada organización, luego se estimaron los parámetros de cada sistema para obtener la tasa óptima de servicio, lo cual se comparó con lo calculado con las ecuaciones desarrolladas para ello.

Resultados: Para ambos casos, al incrementarse la razón de costos servicio/demoras, disminuye el número óptimo de servidores y aumenta el costo del sistema. Ambas variables han sido menores para el modelo M/D/1 que para el M/G/1, debido al fuerte impacto de la desviación estándar del tiempo de servicio. Asimismo, el costo y número de servidores han sido menores para el sistema M/M/S comparado con el M/M/1, lo que lo hace preferible en caso de poder llevar a cabo su implementación.

Limitaciones del Estudio: Se obtuvieron los parámetros de cada sistema con un número limitado de datos y sólo se estudiaron dos organizaciones.

Originalidad: Las ecuaciones desarrolladas para calcular el número óptimo de servidores.

Conclusiones: Hay diversos factores inciden en el comportamiento de los sistemas de servicio. Su estudio permite a las organizaciones tomar mejores decisiones y desarrollar modelos de optimización que incrementen el aprovechamiento de sus recursos, la calidad de su servicio y la rentabilidad.

Palabras clave:

Línea de espera, Tiempo de espera, Tasa de servicio, Tasa de llegadas, Número óptimo de servidores

Abstract:

Objective: The objective of this study is to define by means of the analysis of 2 empirical cases: the first an M/G/1 or M/D/1 system, and the second M/M/1 or M/M/S, the optimal rate of service and the

economic convenience of using them or not, as well as validating two equations developed specifically to calculate the optimal number of servers.

Methodology: First, the models that apply to each organization were identified, then the parameters of each system were estimated to obtain the optimal service rate, which was compared with that calculated with the equations developed for it.

Results: For both cases, as the service/delay costs ratio increases, the optimal number of servers decreases and the cost of the system increases. Both variables were lower for the M/D/1 model than for the M/G/1, due to the strong impact of the standard deviation of the service time. Also, the cost and number of servers have been lower for the M/M/S system compared to the M/M/1, which makes it preferable in case of being able to carry out its implementation.

Study Limitations: The parameters of each system were obtained with a limited number of data and only two organizations were studied.

Originality: The equations developed to calculate the optimal number of servers.

Conclusions: There are several factors that affect the behavior of the service systems. Its study allows organizations to make better decisions and develop optimization models that increase the use of their resources, the quality of their service and profitability.

Keywords:

Queueing line, Waiting time, Service rate, Arrivals rate, Optimal number of servers

“**Clasificación JEL:**” C44 Operations Research; Statistical Decision Theory

Introducción

La economía de los servicios es muy importante hoy día, pues genera la mayor parte del producto interno bruto (PIB) de las naciones. En la economía norteamericana a fines de la década de los 90, los servicios generaban aproximadamente 76% del PIB del país y ocupaban al 79% de la fuerza laboral (Zeithaml y Bitner, 2002). Además ha sido el sector de mayor crecimiento en la economía global en la última década.

Esto se debe a factores diversos que tienen que ver con las tendencias actuales, como son la globalización de las economías, el desarrollo tecnológico, el boom de los medios de comunicación, la mayor orientación al cliente de los negocios, la necesidad de cuidar el medio ambiente, el aumento de la población mundial, la mayor responsabilidad social de las empresas, la creciente participación de las mujeres en todas las actividades, el valor del conocimiento como parte del capital intelectual de las organizaciones, el envejecimiento de la población y otras.

Los servicios excepcionales no son frecuentes, pero definitivamente muy rentables, razón por la cual deben realizarse con altos estándares de calidad, la cual se define como cumplir con las expectativas del cliente, lo que no es tarea fácil dado que éstas pueden cambiar con cada cliente, o incluso en un mismo cliente en ocasiones distintas, lo que es muy diferente al concepto de calidad para los productos, que normalmente se define como la adecuación al uso pretendido del consumidor, o cumplir con las especificaciones de diseño del artículo.

Un servicio es un proceso o una acción, no un objeto tangible y puede definirse como un conjunto de actividades económicas cuyo resultado se produce y consume al mismo tiempo, por lo cual no puede almacenarse y el prestador del servicio se convierte a los ojos del cliente en la imagen misma de la organización, razón por la cual la lealtad del cliente y del prestador del servicio son esenciales para producir un servicio sobresaliente.

Dada la intangibilidad de los servicios, la lealtad de los consumidores es mayor que para los productos y la publicidad “boca a boca” es un medio muy efectivo para su promoción. Por otro lado, la inspección y prueba de un servicio es problemática, ya que su producción y consumo se lleva a cabo

simultáneamente, por lo cual se requiere hacer uso de diversas técnicas para su supervisión, siendo una medida usual la del “comprador misterioso”.

Para las buenas compañías de servicios las quejas constituyen una fuente muy valiosa de información y si éstas se repiten, indican la existencia de un problema que debe solucionarse lo antes posible.

Además están las expectativas del cliente, las cuales son mayores si éste conoce el servicio.

Una manera de medir la calidad de un servicio es mediante el tiempo de espera en la fila, o bien con el número de clientes en espera. Algunos académicos señalan que un buen servicio no debe rebasar un tiempo de espera de 5 minutos (Heskett, Sasser y Hart, 1993) y aun cuando un cliente quizás no sea un buen juez respecto al tiempo que ha pasado esperando en la fila, definitivamente sí puede decir con certeza si dicho tiempo le ha parecido excesivo.

En virtud de ello y ante las variaciones de la demanda de servicios, las organizaciones buscan contar con capacidad para atender la demanda durante periodos altos y bajos, lo que requiere la comprensión de la naturaleza de la demanda del servicio y los esfuerzos de coordinación de la gestión, tanto de los clientes como de las personas responsables de proveer el servicio.

De este modo, la incapacidad de atender una demanda variable puede ser un elemento muy grave para muchos negocios, siendo las consecuencias más adversas a largo plazo, la de operar unas instalaciones en las que la demanda excede permanentemente a la capacidad óptima, ya que en tales circunstancias, se incrementará el número de clientes insatisfechos.

Objetivo

El objetivo de este trabajo es determinar mediante dos casos empíricos, uno que opera como M/G y el otro como M/M, cuáles son los parámetros de las líneas de espera que permiten optimizar los costos operativos y maximizar la rentabilidad de ambas organizaciones. Para ello se toman en cuenta los costos del servicio y demoras, con el fin de alcanzar la tasa óptima de servicio, que redunde en una operación eficiente. No obstante que existen muchos modelos de líneas de espera, en este caso se han aplicado los que describen el comportamiento de las organizaciones estudiadas en este trabajo, siendo además los modelos más usuales en la práctica, como es el caso de instituciones financieras, tiendas de conveniencia, supermercados, estaciones de gasolina y otras.

Además se han desarrollado dos ecuaciones para calcular el número óptimo de servidores para los sistemas M/G/1 y M/M/1, las cuales buscan validarse, esto es, ver si concuerdan con el valor obtenido a prueba y error.

Marco Teórico

Lovelock, Reynoso, D´Andrea y Huete (2004) señalan que en los servicios las filas de espera son un mecanismo de almacenaje, de modo que la espera constituye un fenómeno universal. Al respecto afirman que los estadounidenses pasan esperando en filas un promedio de 150 horas al año, habiendo diferencias por cultura, ya que los latinos tienen mayor aceptación por la espera que los anglosajones, quienes son clientes más exigentes e incluso con la disposición de pagar una prima extra por una atención más rápida.

A las personas no les gusta esperar, a veces es aburrido, incómodo e inclusive estresante. La cola se genera al momento que la demanda del servicio es mayor que la oferta, de modo que representa un problema de administración de la capacidad.

Ahora con la tecnología de la información, los clientes acuden a medios como el correo electrónico, el Internet, el WhatsApp y otros, que les permiten un contacto instantáneo, pero esto genera líneas de espera virtuales, que son por lo general más cómodas, ya que el cliente puede hacer un mejor uso de su tiempo o hacer la espera en casa.

Añadir capacidad puede ser una solución, pero incrementa el costo. No obstante hay otras estrategias que recomiendan algunos académicos, entre ellas: 1) utilizar la lógica operativa, eliminando ineficiencias del sistema que pueden llegar hasta el rediseño del proceso; 2) implementar reservaciones, lo que ayuda a hacer más uniforme la demanda; 3) establecer diferenciación en los clientes que esperan, lo que suele hacerse según su urgencia, su importancia, la duración del servicio y la aplicación de primas por proveer el servicio; y 4) hacer la espera más tolerable (Zeithml y Bitner, 2002; Lovelock et al., 2004).

Otras estrategias son la minimización de la longitud percibida de la cola, lo que puede hacerse mediante divisiones, filas en zigzag y moviendo las filas de prisa, tal como lo hace Disney en sus centros de diversiones, o bien la participación del cliente en la prestación del servicio, como es el caso de las estaciones de gasolina y los cajeros automáticos o algunas mueblerías, como Levitz Furniture, que incentivan a que el cliente transporte sus muebles otorgándole un descuento (Heskett et al., 1993).

Por otro lado, existen numerosos estudios sobre las aplicaciones de los modelos de líneas de espera a casos prácticos con el fin de determinar sus principales parámetros, como son los tiempos de espera y de servicio, así como el número promedio de clientes en espera y en el sistema.

Asimismo, otras investigaciones tratan el problema de optimización de los sistemas de espera, buscando minimizar su costo operativo, que se compone del costo de servicio y el de la demora, entre ellos Heyman (1977), quien afirma que para un sistema M/G/1 es mejor definir la política óptima con base en el tiempo del servidor y no en el número de clientes en el sistema.

Chang y Ke (2011) han estudiado el problema de manejo óptimo de un sistema de líneas de espera M/M/1/K con una política (p,E), donde todos los clientes que llegan desean el servicio esencial y algunos de ellos servicios adicionales. Cuando el número de clientes alcanza la capacidad del sistema K, no se permite la entrada a más clientes. Mediante esta técnica han derivado importantes medidas de desempeño y han desarrollado un modelo de costos para definir la política óptima considerando la sensibilidad del sistema.

También, Wang y Chang (2002) han investigado un sistema M/M/S con capacidad finita y clientes que rechazan entrar, otros que abandonan la fila y con interrupciones del servicio por parte de los servidores, encontrando que los tiempos de llegada de clientes, del servicio y de las interrupciones siguen una distribución exponencial negativa. Mediante una matriz geométrica han derivado las probabilidades de estado estable y con un modelo de costos han obtenido el número óptimo de servidores, además de analizar la sensibilidad del sistema.

Rodríguez-Jáuregui, González-Pérez, Hernández-González y Hernández-Ripalda (2017) han aplicado un modelo de líneas de espera al área de urgencias de un hospital para obtener con parámetros empíricos, el número mínimo de médicos para brindar buena atención a los pacientes, el cual en las condiciones analizadas no es suficiente.

De manera similar, Gorunescu, McClean y Millard (2002) han aplicado un modelo M/M/S para optimizar el uso de los recursos hospitalarios, concretamente el número de camas disponibles, para mejorar la atención al cliente, con la cual optimizan el número de camas que debe haber disponibles. Si el número de camas ocupadas está por encima del 80%, el rechazo de nuevos pacientes aumenta de manera significativa.

Por su parte, Hayel, Ouarranou y Tuffin (2007) han modelado un sistema como una línea de espera M/M/1, en la que diversas tareas son enviadas a la fila y se caracterizan por una función de demanda y un costo de demora por unidad de tiempo. La meta es diseñar un esquema óptimo de precios para la cola, donde el cargo total depende de la demora promedio de la fila y de la tasa de llegada de los clientes. Este modelo puede aplicarse al caso de redes de telecomunicaciones, en las que la fijación de precios puede usarse para controlar el congestionamiento y la red puede caracterizarse por una

cola de cuello de botella única, en la que el flujo se determina mediante mediciones pasivas y la demora con mediciones activas.

En otro estudio Hyytiä, Aalto, Penttinen y Virtamo (2012) han desarrollado funciones para políticas óptimas de envíos aplicables a sistemas de colas en paralelo $M/G/1$ con orden primero en llegar – primero en salir (FIFO) o último en llegar – primero en salir (LIFO), para escenarios multiclase de los que se dispone de estimaciones independientes del tiempo de servicio y el sistema está sujeto a costos de mantenimiento de los clientes por unidad de tiempo.

Asimismo, Barak y Fallahnezhad (2012) han analizado los sistemas de colas con datos difusos, que han aplicado tanto a las tasas de llegadas y servicio, como a los costos del sistema. Lo han aplicado para modelos $M/M/1$ y $M/E2/1$, para definir la política óptima de asignación de personas a estaciones de servicio, lo que constituye un enfoque novedoso.

Al igual que, Haridass y Arumuganathan (2012) quienes analizaron un sistema de colas con llegadas a granel bajo una distribución Poisson con vacaciones múltiples y restricción a la admisión de nuevas llegadas, con diferentes probabilidades de admisión si los servidores están ocupados u ociosos. El servidor se asigna para tareas secundarias (vacaciones) cuando el número de tareas en espera no es adecuado para el proceso. Desarrollaron la función generadora de probabilidad para el sistema, así como medidas de desempeño y un modelo de costos, a fin de brindar decisiones administrativas que optimicen su funcionamiento.

En otro estudio, Kumar, Jain y Som (2014) de un sistema de colas $M/M/1/N$ con retención de clientes impacientes, en el cual han analizado el tamaño del sistema con y sin retención de estos clientes y con un modelo de costos han obtenido los parámetros que optimizan el sistema en cuanto a su capacidad y la tasa de servicio promedio.

De manera similar, Srinivas, Rao y Kale (2011) han estimado los parámetros de un modelo $M/M/1$ mediante las metodologías de máxima verosimilitud y varianza mínima uniforme insesgada, encontrando mediante el criterio de deficiencia esperada asintótica que es mejor la segunda técnica para la estimación de tales parámetros.

Por su parte, Stadje (2009) ha estudiado mediante un proceso markoviano de dos dimensiones, un sistema de colas $M/M/2$ con la característica de tener clientes sensibles al costo y persuasivos, asumiendo que un nuevo cliente entra al sistema sólo si el costo de hacerlo no excede cierto valor límite. La función del costo es una combinación lineal de la longitud de la cola y una cuota de servicio.

Stidham (2005) ha investigado un sistema de colas con política de servicio completo y costos descontados, buscando la solución que minimice el costo esperado descontado en un horizonte infinito de tiempo, el cual ha aplicado para varios modelos de colas, extendiendo el análisis para el caso en que las funciones de costos sean no lineales.

Armony, Plambeck y Seshadri, (2009) han encontrado que cuando la capacidad en un sistema $M/M/S$ es costosa o la tasa de renuncia del cliente es alta, la capacidad óptima disminuye con cualquier aumento adicional de la tasa de renuncia, por lo cual los administradores deben comprender la impaciencia de los clientes para evitar una capacidad excesiva, para lo cual conviene prevenir las renunciaciones para optimizar el costo total del sistema y evitar la pérdida de ventas.

Finalmente, Gómez-Jiménez (2008), ha aplicado la teoría de colas para mejorar el funcionamiento de una entidad financiera colombiana, la cual funciona como un sistema $M/M/S$ y mediante un modelo de nivel de aceptación define el número óptimo de servidores, el cual está en función del patrón de llegadas de los clientes.

Respecto a los costos, algunos académicos argumentan que es inapropiado asumir que el costo de demora es una función lineal del número de clientes en el sistema.

Por lo que, en esta investigación se busca definir el número de servidores que optimice el costo del sistema para los modelos M/G/1 y M/D/1 del primer caso empírico, y de los modelos M/M/1 y M/M/S del segundo caso. Asimismo, se indaga lo que sucede al variar la razón de los costos de servicio y demoras, así como los impactos de la media y la desviación estándar del tiempo de servicio en el primer caso, para lo cual se desarrollaron las ecuaciones respectivas.

Metodología

Para hacer el análisis de los dos casos empíricos presentados, lo primero fue identificar con sus patrones de llegada y servicio, el modelo de espera que se alinea más al patrón de comportamiento de los datos de proceso de las organizaciones estudiadas. Una vez hecho lo anterior, se procedió a medir los parámetros de cada sistema, concretamente las tasas de llegada de los clientes y los tiempos de servicio, para con ello proceder a la aplicación de los modelos operativos para cada caso. En el primero de ellos (M/G) se requiere la media del tiempo de llegadas entre cada cliente y la media y desviación estándar del tiempo de servicio; mientras que para el segundo caso con la media del tiempo de llegadas y de servicio es suficiente.

Se desarrollaron además las ecuaciones para los modelos M/G/1 y M/M/1 para obtener el número de servidores que optimiza el costo total del sistema, colocando en primer lugar la ecuación del costo total en función del número de servidores, derivándola e igualándola a cero, lo cual permite en este trabajo su validación y obtener las tasas óptimas de servicio con un número menor de cálculos.

Para los casos en que se conoce el costo de la espera, el costo total del sistema viene dado por la ecuación siguiente:

$$\text{Costo del Sistema} = \text{Costo de Demoras} + \text{Costo del Servicio} \quad (1)$$

Siendo el costo de las demoras:

$$\text{Costo Demoras} = C_D L \quad (2)$$

Donde:

C_D = Costo unitario de la demora, \$/(día)(unidad)

L = Número promedio de clientes en el sistema, unidades

Mientras que el costo del servicio es:

$$\text{Costo Servicio} = C_S S \quad (3)$$

Siendo C_S el costo diario del servicio por cada servidor y S el número de servidores.

El número promedio de clientes en el sistema, L , viene dado por una relación matemática, que depende del modelo de la línea de espera bajo análisis. Para el caso del sistema M/G/1, esta expresión es (Izar-Landeta, 2012):

$$L = \rho + \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \quad (4)$$

Siendo λ la media de la llegada de clientes al sistema, ρ el factor de ocupación del sistema y σ la desviación estándar del tiempo de servicio.

Si la desviación estándar es cero, este modelo se convierte en el M/D/1, siendo L :

$$L = \rho + \frac{\rho^2}{2(1 - \rho)} \quad (5)$$

Por su parte ρ viene dada por:

$$\rho = \frac{\lambda}{\mu} \quad (6)$$

Siendo μ la tasa de servicio, la cual debe estar en las mismas unidades que λ .

Para el sistema M/M/1, L viene dada por la siguiente fórmula (Izar-Landeta, 2012):

$$L = \frac{\rho}{(1-\rho)} \quad (7)$$

Y para el modelo M/M/S (Izar-Landeta, 2012):

$$L = \rho S + \frac{(\rho S)^2}{S!} \frac{1}{(1-\rho)^2} P_0 \quad (8)$$

Siendo P_0 la probabilidad de que el sistema esté vacío, la cual puede estimarse mediante la siguiente relación matemática (Izar-Landeta, 2012):

$$P_0 = \frac{1}{\sum_{k=0}^S \frac{(\rho S)^k}{k!} + \frac{(\rho S)^S \rho}{S!(1-\rho)}} \quad (9)$$

Donde ya todos los términos han sido definidos previamente.

Con esto, el número de servidores S que minimiza el costo total para el modelo M/M/1 es:

$$S = \frac{\lambda + \sqrt{\frac{\lambda \mu C_D}{C_S}}}{\mu} \quad (10)$$

Y para el modelo M/G/1:

$$C_S + \frac{C_D \lambda \left[2\lambda - 2S\mu - \frac{\lambda^2}{S\mu} - \lambda^2 \sigma^2 S\mu \right]}{2S(S\mu - \lambda)^2} = 0 \quad (11)$$

Que es no lineal en S, por lo cual puede resolverse a prueba y error o mediante algún método numérico.

Esta misma ecuación puede aplicarse al modelo M/D/1, colocando la desviación estándar con un valor de cero.

Para el caso del modelo M/M/S no es posible derivar la expresión analítica para encontrar el valor de S que optimice el sistema, por lo cual esto debe hacerse a prueba y error, partiendo de un número mínimo de servidores con el cual μ supere a λ , hasta encontrar el valor óptimo.

Luego se analiza cómo cambia esta respuesta y el monto del costo mínimo al modificar la razón del costo de servicio al de demora C_S/C_D , así como el impacto de la desviación estándar y la media del tiempo de servicio en el número de servidores y el costo óptimo del modelo M/G/1 y un comparativo de costos de los sistemas M/M/1 con el M/M/S..

Discusión y Análisis de Resultados

Para el primer caso empírico, se observaron durante una semana del mes de septiembre de 2017, los movimientos de carga de materiales de una empresa mediana de 10,000 m² de logística y transporte ubicada en el Estado de Querétaro con sucursales en los Estados de Nuevo León, San Luis Potosí,

Coahuila y Guanajuato, que presta servicios para la gestión total de la cadena de suministro del sector automotriz, la cual cuenta con una estación de carga de camiones y por motivos de confidencialidad se omite su nombre. A la estación llegan los camiones para ser cargados con diferentes materiales y enviados a sus destinos para su entrega. La estación trabaja 14 horas diarias y se ha verificado que el patrón de llegadas sigue una distribución poissoniana con media de 34 camiones diarios y el tiempo de servicio sigue una distribución normal con tiempo promedio de 2 horas para su carga y desviación estándar de 1.96 horas, por lo cual el sistema es un modelo M/G/1. La carga de los camiones se hace con 8 montacargas, los que tienen un costo de 2,500 pesos diarios cada uno y la demora en la carga cuesta 4,200 pesos por día. Se busca determinar el número óptimo de montacargas, así como el costo mínimo del sistema y cómo cambian estas respuestas ante variaciones en la relación de costos y la desviación estándar y la media del tiempo de servicio.

Para este caso se tiene que λ es 34 camiones/día y la tasa de servicio μ es 7 camiones diarios por montacargas, dado que se trabajan 14 horas/día. La desviación estándar es 1.96 horas (0.14 días).

Si se aplica la ecuación (11) con tanteos, se llega a un valor S de 14.75 para obtener el costo mínimo, con el cual se da la siguiente igualdad:

$$2,500 + \frac{(4,200)(34) \left[2(34) - (2)(14.75)(7) - \frac{34^2}{(14.75)(7)} - (34^2)(0.14^2)(14.75)(7) \right]}{2(14.75)((14.75)(7) - 34)^2} = -1.54$$

Que es un valor muy próximo a cero.

Si se hace el procedimiento a prueba y error, el valor mínimo de S para que μ sea mayor que λ , es 5, en cuyo caso ρ es:

$$\rho = \frac{34}{(5)(7)} = 0.971$$

Por su parte al aplicar la ecuación (4) se calcula L:

$$L = 0.971 + \frac{(34)^2(0.14)^2 + 0.971^2}{2(1 - 0.971)} = 413.99$$

Siendo el costo del sistema:

$$\text{Costo sistema} = (4,200)(413.99) + (2,500)(5) = 1,751,274 \text{ \$/día}$$

Que es altísimo, ya que ρ es casi la unidad y por lo tanto el costo de demora es muy elevado. Si se repiten estos cálculos para un número mayor de servidores, se generan los resultados que se muestran en la tabla 1, en la que se aprecia que el costo mínimo sucede para 15 servidores con un monto de 109,552 \\$/día, lo que concuerda con el valor obtenido mediante la ecuación (11).

Tabla 1: Costos y parámetros del sistema para diferentes valores de S

S	P	L	Costo servicio	Costo demora	Costo total
5	0.971	413.99	12,500	1,738,774	1,751,274
6	0.810	62.01	15,000	260,425	275,425
7	0.694	38.49	17,500	161,648	179,148
8	0.607	29.91	20,000	125,636	145,636
9	0.540	25.47	22,500	106,961	129,461
10	0.486	22.74	25,000	95,522	120,522
11	0.442	20.90	27,500	87,791	115,291
12	0.405	19.58	30,000	82,214	112,214
13	0.374	18.57	32,500	78,000	110,500
14	0.347	17.79	35,000	74,703	109,703
15	0.324	17.16	37,500	72,052	109,552
16	0.304	16.64	40,000	69,874	109,874
17	0.286	16.20	42,500	68,053	110,553

18	0.270	15.84	45,000	66,508	111,508
19	0.256	15.52	47,500	65,180	112,680
20	0.243	15.24	50,000	64,026	114,026

Fuente: Elaboración propia

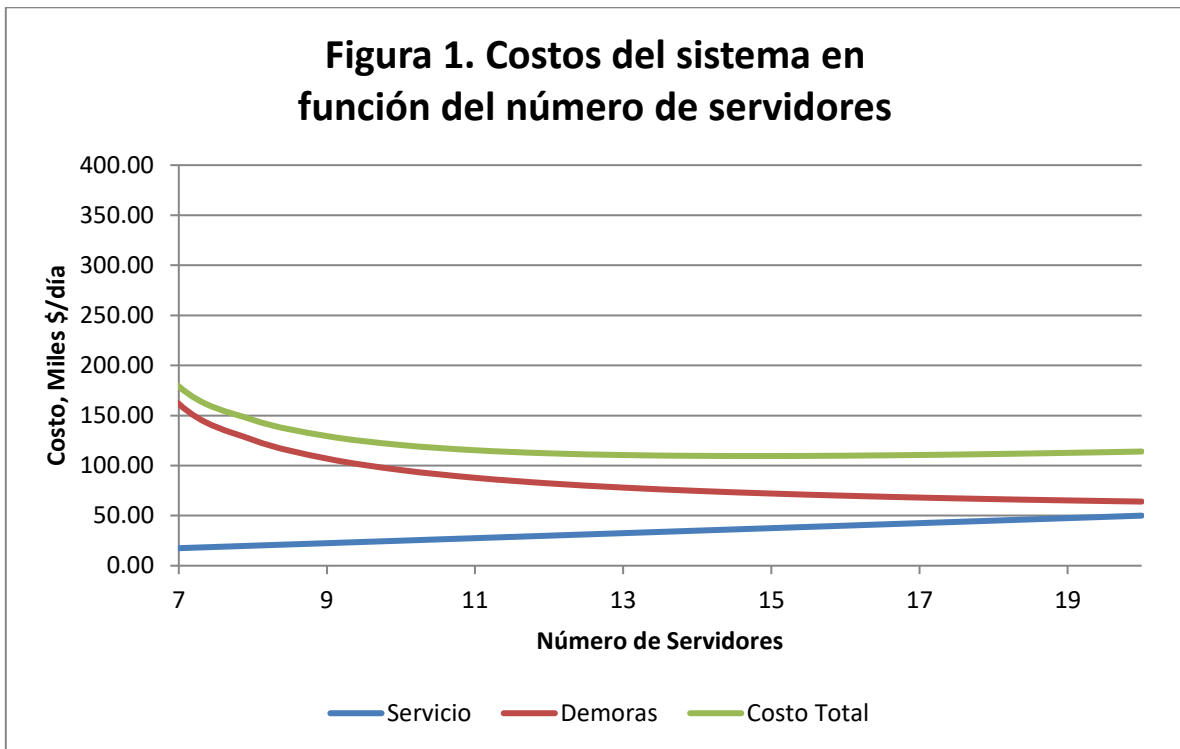
Al presentar estos costos de demora, servicios y total del sistema en un gráfico, se produce el de la figura 1, que es típico de estos sistemas, en el cual puede apreciarse que el costo de demora es superior al de servicio en todo el rango analizado, de modo que las líneas del costo total y de demoras aparecen muy próximas en todo el gráfico.

Si la desviación estándar fuese cero, lo que significa que el sistema es M/D/1, para un número de servidores de 5, se obtiene L aplicando la ecuación (5):

$$L = 0.971 + \frac{0.971^2}{2(1-0.971)} = 17.486$$

Siendo el costo del sistema:

$$\text{Costo sistema} = (4,200)(17.486) + (2,500)(5) = 85,940 \text{ \$/día}$$



Fuente: Elaboración propia

Si se repiten estos cálculos para otros valores de S, se producen los resultados de la tabla 2:

Tabla 2: Costos del sistema M/D/1 para valores de S

S	P	L	Costo servicio	Costo demora	Costo total
5	0.971	17.49	12,500	73,440	85,940
6	0.810	2.53	15,000	10,625	25,625
7	0.694	1.48	17,500	6,217	23,717
8	0.607	1.08	20,000	4,520	24,520
9	0.540	0.86	22,500	3,595	26,095
10	0.486	0.72	25,000	3,003	28,003
11	0.442	0.62	27,500	2,588	30,088
12	0.405	0.54	30,000	2,278	32,278

Fuente: Elaboración propia

El costo mínimo se ha dado con 7 servidores y es muy bajo en comparación con el obtenido con los datos del caso base.

Si se hubiese utilizado la ecuación (11) con σ igual a cero, produce un valor para S de 6.79, lo que concuerda con el esquema de cálculos que llevó a 7 servidores,

Es notable que con desviación estándar de cero, el número de servidores ha sido menos de la mitad en comparación con el caso base de desviación estándar de 0.14 días.

Si se hace variar la desviación estándar en un rango de 0 a 0.3 días para definir el número óptimo de servidores, se producen los resultados de la tabla 3:

Tabla 3: Variación del número de servidores y el costo mínimo con σ

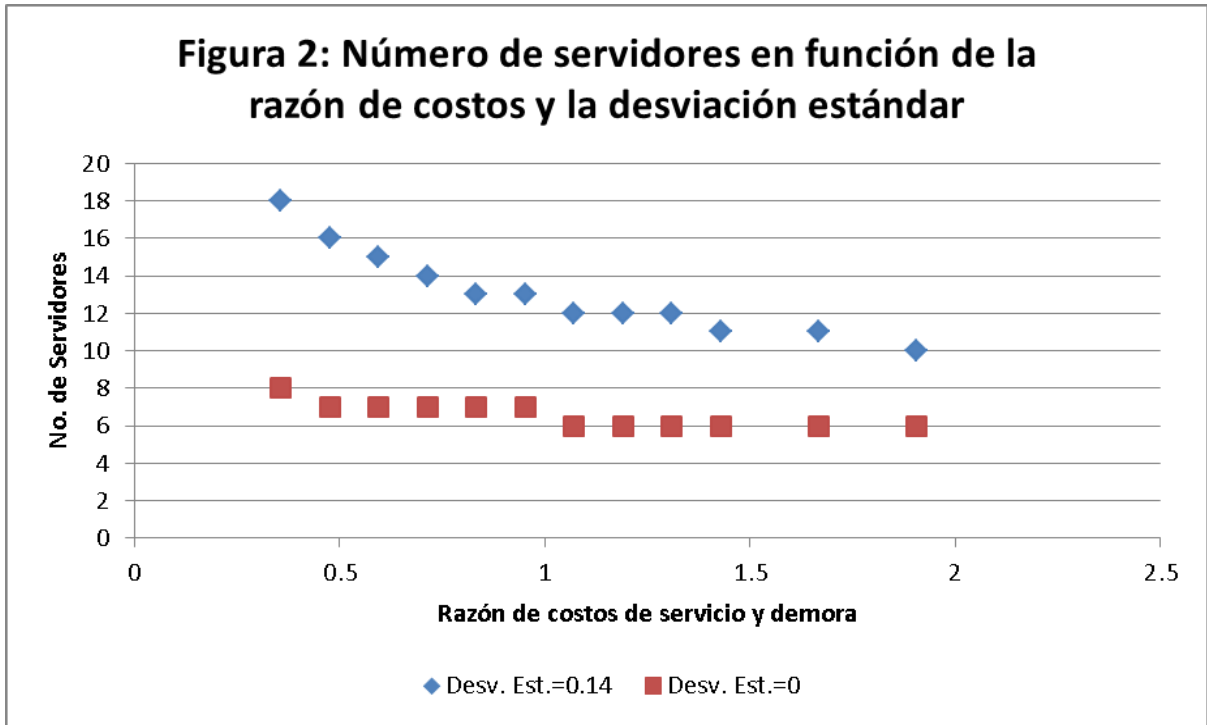
σ	S	Costo mínimo, \$/día
0	7	23,717
0.07	10	51,133
0.14	15	109,552
0.17	17	141,161
0.20	19	179,211
0.25	22	250,772
0.30	25	334,551

Fuente: Elaboración propia

Se observa que en todo el rango estudiado el número de servidores y el costo mínimo del sistema aumentan directamente con la desviación estándar del tiempo de servicio.

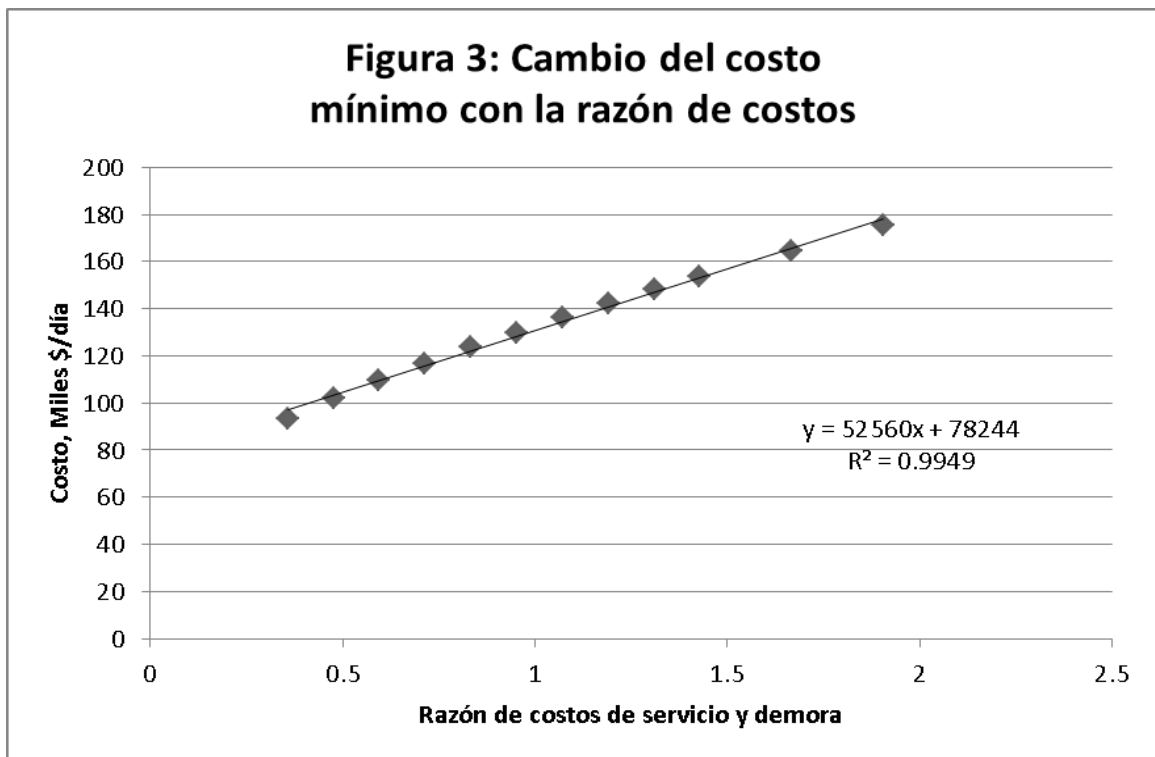
Si se modifica el costo del servicio, de modo que la razón de costos de servicio y demora, C_S/C_D cambie, a fin de ver su impacto en el número óptimo de servidores utilizando desviaciones estándar del tiempo de servicio de 0 y 0.14 días, se produce el gráfico de la figura 2, en el que se observa que en todo el rango estudiado el número óptimo de servidores es mayor con la desviación estándar del caso base comparada con la opción alterna de cero y dicha diferencia es mayor en el lado izquierdo del gráfico, para razones menores del costo de servicio y demoras.

Además, al incrementarse la razón de costos, el número óptimo de servidores disminuye, lo que es obvio, ya que con una razón más alta, el costo del servicio es mayor en comparación al de demora, lo que lleva el número de servidores a la baja.



Fuente: Elaboración propia

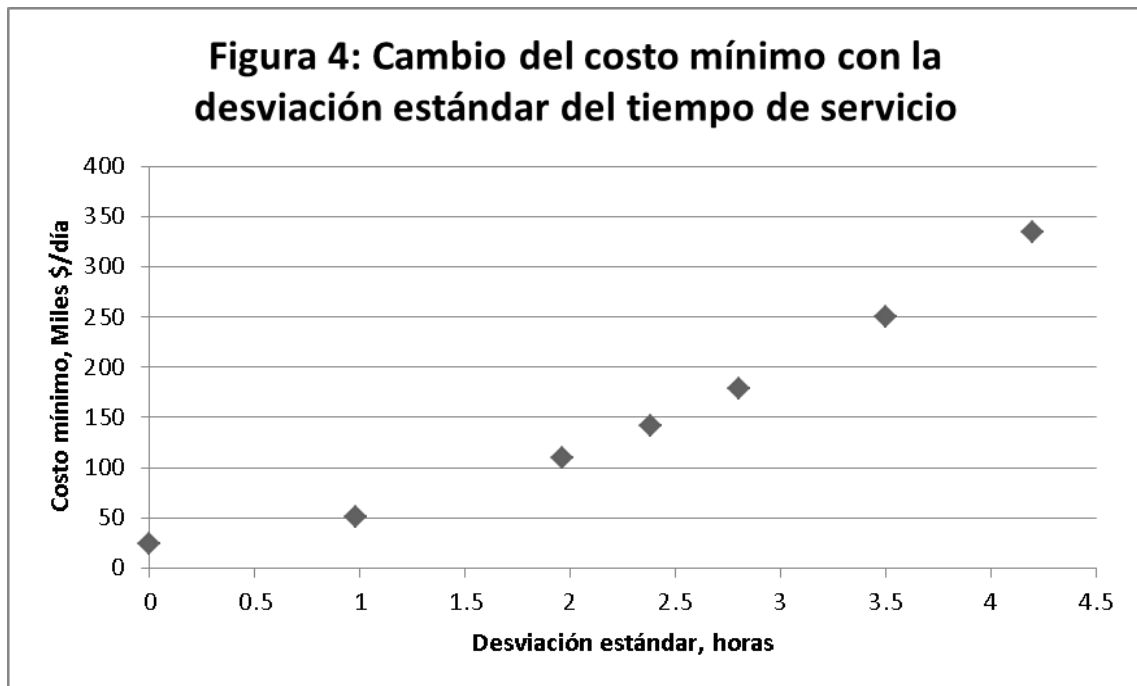
Para la desviación estándar del caso base (0.14), se analizó cómo cambia el costo mínimo del sistema con la razón de costos C_s/C_D , lo que se muestra en la figura 3, donde se incluye la ecuación del ajuste lineal de mínimos cuadrados, que es excelente, puesto que el coeficiente de determinación es 0.9949, lo que indica un ajuste casi perfecto y hace ver que por cada unidad de incremento en la razón de costos, el costo mínimo del sistema se eleva en \$52,560 diarios.



Fuente: Elaboración propia

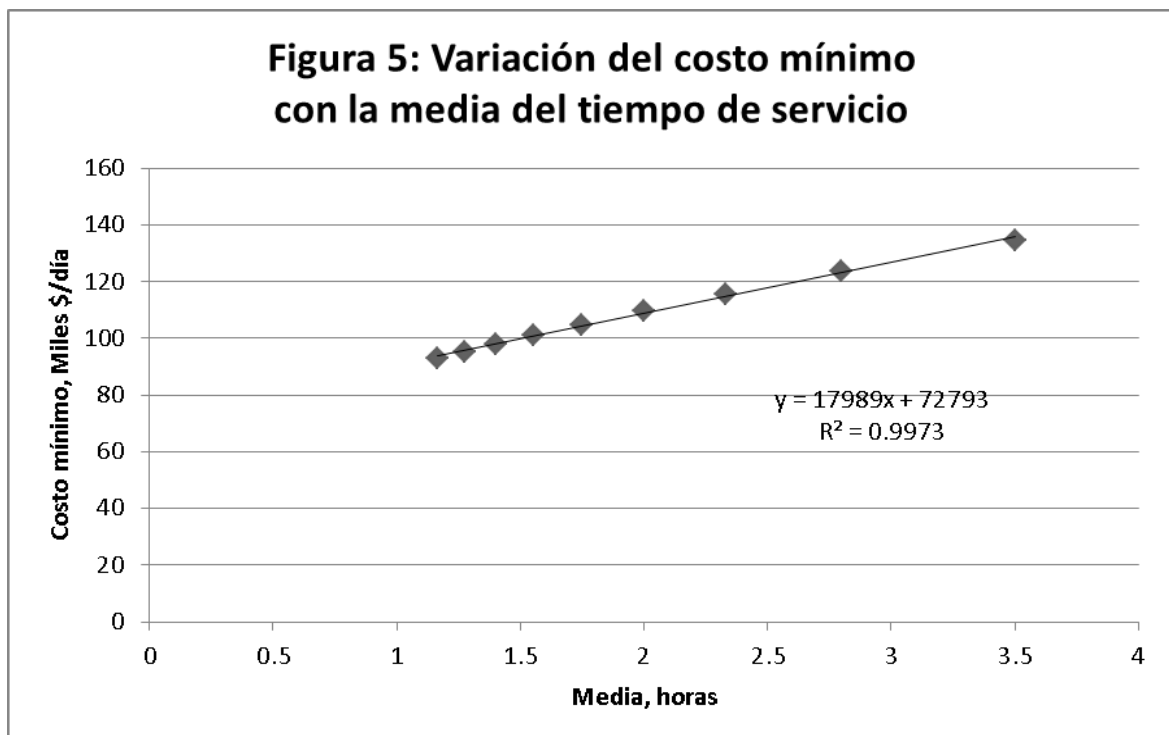
Finalmente se ha realizado un análisis de sensibilidad comparativo del sistema, para definir qué variable impacta en mayor medida al costo, si la media o la desviación estándar del tiempo de servicio.

Para la media del caso base de 2 horas ($\mu=7$ camiones/día), el gráfico que muestra la variación del costo mínimo del sistema con la desviación estándar, expresada en horas es el de la figura 4, en la cual puede observarse que la curva resultante no es lineal y el cambio en el costo es más pronunciado en el lado derecho del gráfico, donde la pendiente con los 2 puntos máximos es de casi 120,000 pesos por día, mientras que en el lado izquierdo de la figura, tomando los 2 puntos de desviación estándar mínima, la pendiente es un poco menor a 28,000 pesos diarios.



Fuente: Elaboración propia

Por su parte, si se hace algo similar para los datos del caso base, variando ahora la media del tiempo de servicio y estimando el costo mínimo del sistema, se produce el gráfico de la figura 5, en el cual la línea resultante es recta, con un ajuste muy bueno ($R^2=0.9973$) y una pendiente que indica que el costo del sistema se incrementa en casi 18,000 pesos diarios por cada hora que aumente el tiempo promedio de servicio.



Fuente: Elaboración propia

Esto da una clara idea que el impacto de la desviación estándar del tiempo de servicio es notoriamente mayor que el de su valor promedio en el costo del sistema.

Para el segundo caso empírico, el estudio se llevó a cabo durante una semana del mes de octubre de 2017 y se trata de una empresa comercializadora mediana de más de 30,000 m², especializada en el manejo de contenedores, graneles agrícolas y carga en general en un muelle de carga de barcos ubicado en el puerto de Manzanillo, Colima. Por motivos de confidencialidad se omite su nombre. El análisis de los datos así observados y recopilados permitieron verificar en primera instancia que el sistema tiene un patrón de llegadas con distribución poissoniana con media de 45 barcos diarios. La carga se realiza con grúas, que tienen un tiempo de servicio exponencial con una tasa promedio de 12 barcos diarios por cada grúa, de modo que el sistema es tipo M/M. El costo de servicio es de 1,100 dólares diarios por grúa y la demora en la carga le cuesta al muelle 6,000 dólares/día. La empresa utiliza actualmente 6 grúas y se busca definir su número óptimo, así como el costo mínimo del sistema y cómo afecta el cambio en la relación de costos de servicio y demoras para ambos sistemas.

En este caso λ es 45 barcos/día y la tasa de servicio μ es 12 barcos diarios por grúa.

Por tanto el número mínimo de grúas a utilizar para que la tasa de servicio sea mayor a la de llegadas es 4.

Si se aplica la ecuación (10) para el caso del sistema M/M/1, se obtiene el número óptimo de servidores:

$$S = \frac{45 + \sqrt{(45)(12)(6,000)}}{12} = 8.27$$

Si se hace un primer cálculo con 8 servidores, el sistema se conformaría de 8 subsistemas M/M/1, por lo cual a cada servidor llegarían $45/8=5.625$ barcos, con lo cual la tasa de ocupación, ρ es:

$$\rho = \frac{5.625}{12} = 0.469$$

Entonces puede estimarse L mediante la ecuación (7):

$$L = \frac{0.469}{1 - 0.469} = 0.882$$

Y el costo del sistema es:

$$\text{Costo sistema} = (6,000)(0.882) + (1,100)(8) = 14,094 \text{ Dólares/día}$$

Si se repiten estos cálculos para diferentes valores de S, se obtienen los resultados de la tabla 4:

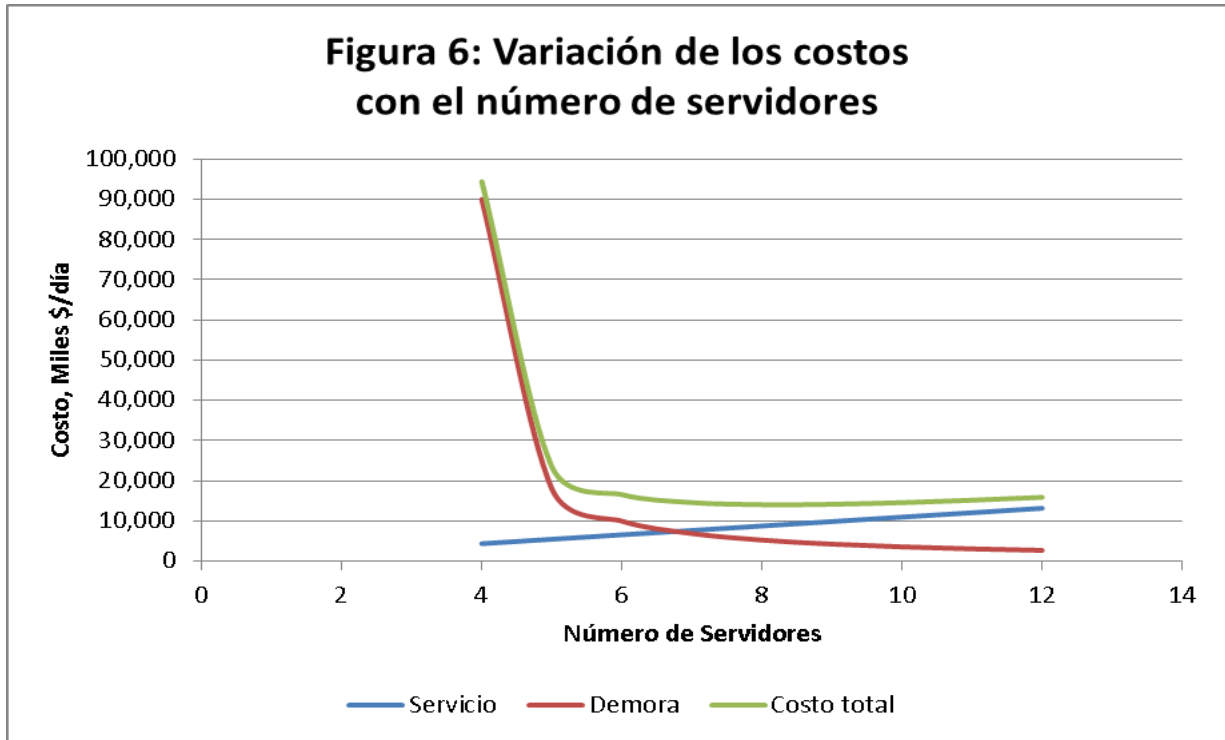
Tabla 4: Costos del sistema M/M/1 para valores de S

S	P	L	Costo servicio	Costo demora	Costo total
4	0.938	15.00	4,400	90,000	94,400
5	0.750	3.000	5,500	18,000	23,500
6	0.625	1.667	6,600	10,000	16,600
7	0.536	1.154	7,700	6,923	14,623
8	0.469	0.882	8,800	5,294	14,094
9	0.417	0.714	9,900	4,286	14,186
10	0.375	0.600	11,000	3,600	14,600
11	0.341	0.517	12,100	3,103	15,203
12	0.313	0.455	13,200	2,727	15,927

Fuente: Elaboración propia

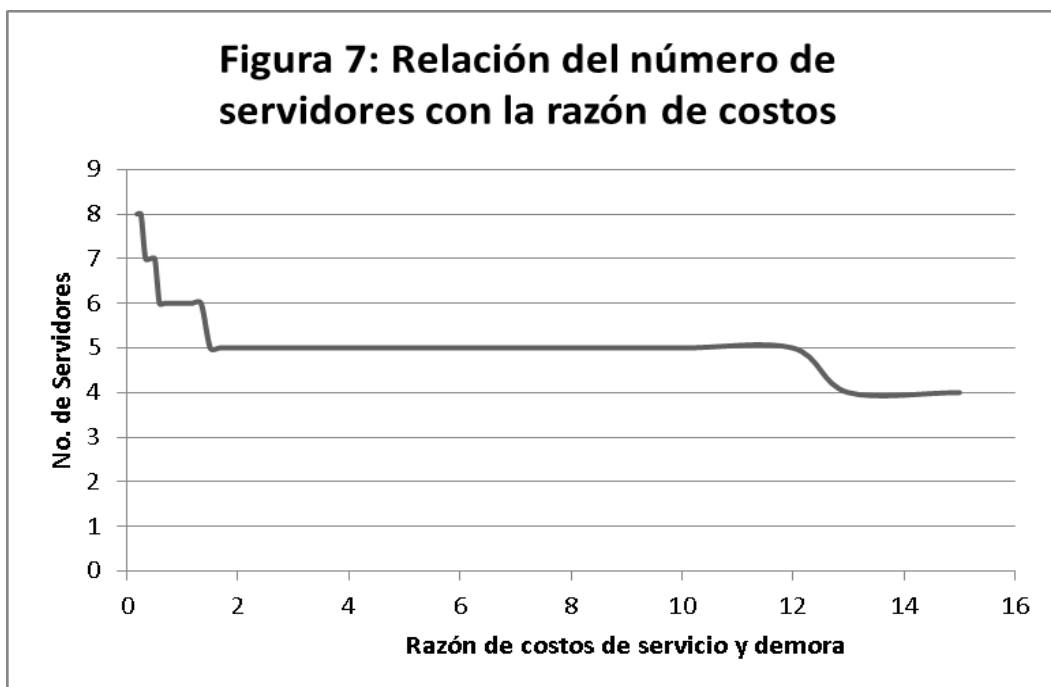
El costo mínimo se ha dado con 8 servidores, lo que concuerda con el valor calculado mediante la ecuación (10).

Esta variación de los costos se muestra en la figura 6 en la que puede verse que con pocos servidores el costo de la demora es mucho mayor que el de servicio, pero en un nivel un poco menor a siete, ambos costos se igualan y este punto no coincide con el del número óptimo de servidores.



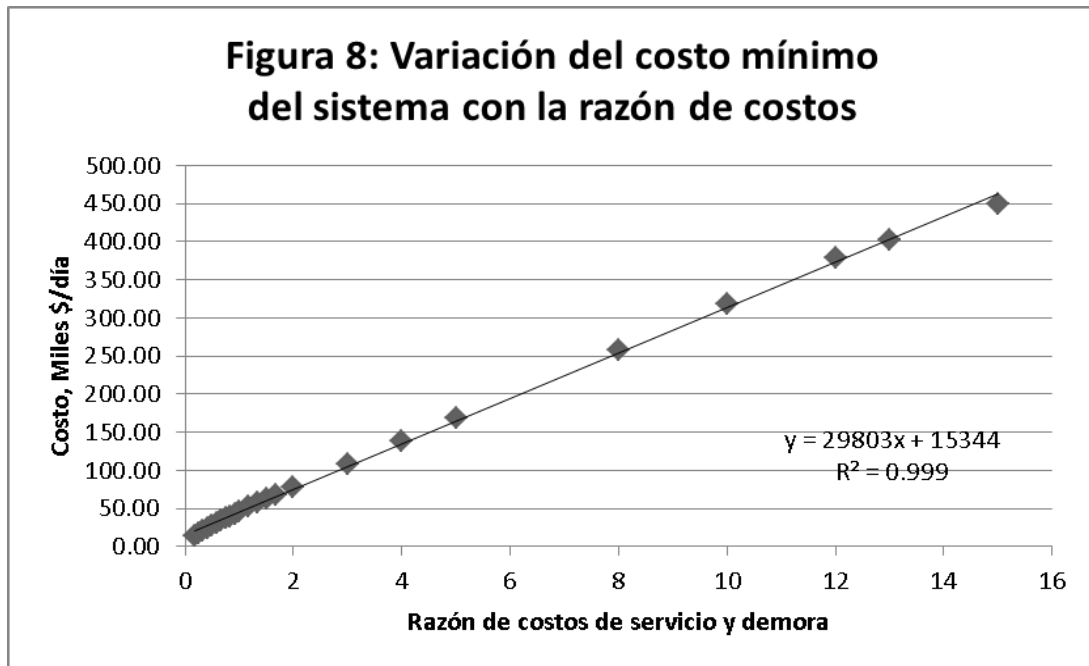
Fuente: Elaboración propia

Si ahora se varía la razón de costos de servicio y demoras C_S/C_D , para ver cómo afecta al número de servidores y al costo óptimo, se producen los gráficos de las figuras 7 y 8 respectivamente.



Fuente: Elaboración propia

Como era de esperarse, al aumentar la razón de costos, el número óptimo de servidores disminuye, siendo este efecto mayor en el lado izquierdo del gráfico, para valores menores de la razón de costos.



Fuente: Elaboración propia

La relación del costo mínimo con la razón de costos es lineal y por cada unidad que se incremente esta razón, el costo mínimo aumenta en casi 30,000 dólares diarios.

Si ahora se hace un análisis similar pero considerando el sistema M/M/S, que significaría una sola fila de barcos y todas las grúas brindando servicio a la misma, el factor de ocupación del sistema ρ no cambia.

Para obtener el valor de L, hay que ir por pasos, calculando primero el término de la sumatoria que aparece en la ecuación (9) para un número de servidores de 4 y un valor constante ρS de 3.75:

$$\sum_{k=0}^4 \frac{(\rho S)^k}{k!} = \frac{3.75^0}{0!} + \frac{3.75^1}{1!} + \frac{3.75^2}{2!} + \frac{3.75^3}{3!} + \frac{3.75^4}{4!} = 28.81$$

Se procede entonces a obtener la probabilidad de que el sistema esté vacío con la ecuación (9):

$$P_0 = \frac{1}{28.81 + \frac{(3.75)^4 (0.9375)}{4!(1-0.9375)}} = 0.00656$$

Ahora se estima L con la ecuación (8):

$$L = 3.75 + \frac{(3.75)^2}{4!} \frac{1}{(1-0.9375)^2} (0.00656) = 17.59$$

Y el costo del sistema es:

$$\text{Costo sistema} = (6,000)(17.59) + (1,100)(4) = 109,943 \text{ Dólares/día}$$

Si se repiten estos cálculos para otros valores de S, se obtienen los resultados que se sintetizan en la tabla siguiente:

Tabla 5: Costos del sistema M/M/S para valores de S

S	P	L	Costo servicio	Costo demora	Costo total
4	0.938	17.59	4,400	105,543	109,943
5	0.750	5.60	5,500	33,583	39,083
6	0.625	4.36	6,600	26,139	32,739
7	0.536	3.97	7,700	23,830	31,530

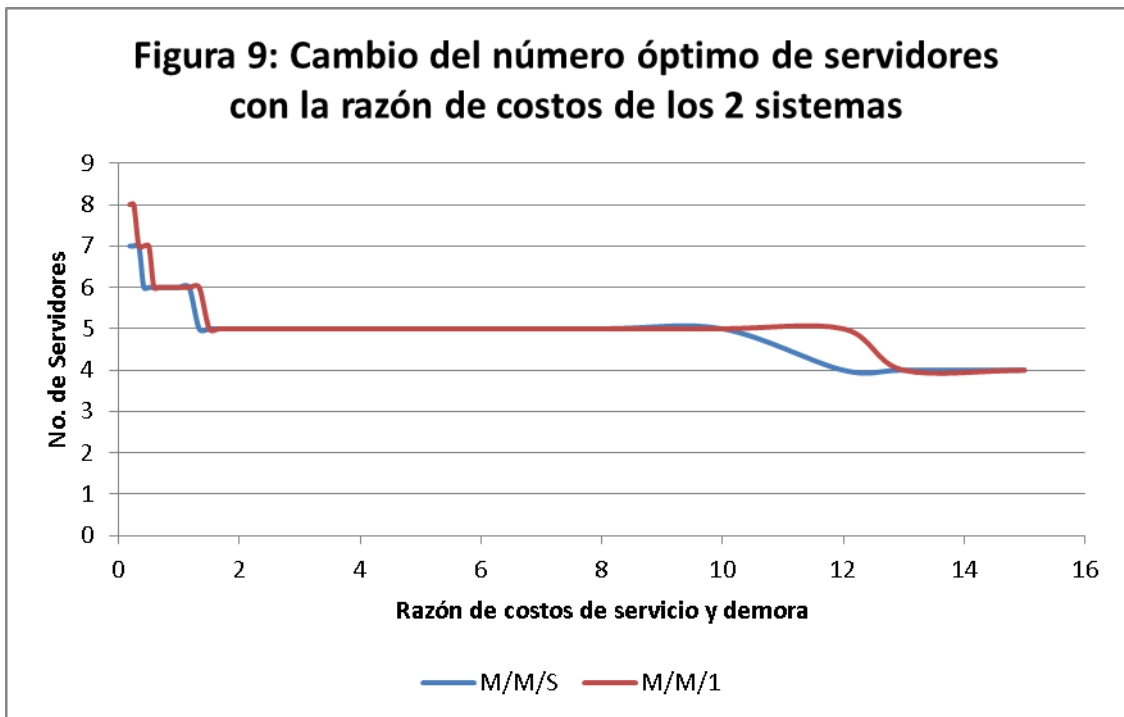
8	0.469	3.83	8,800	22,982	31,782
9	0.417	3.78	9,900	22,667	32,567
10	0.375	3.76	11,000	22,555	33,555
11	0.341	3.75	12,100	22,517	34,617
12	0.313	3.75	13,200	22,505	35,705

Fuente: Elaboración propia

El mínimo ha sucedido para un número de servidores de 7, valor que no coincide con el del sistema M/M/1.

Las variaciones del costo mínimo y el número de servidores de este sistema son muy parecidas a las del M/M/1.

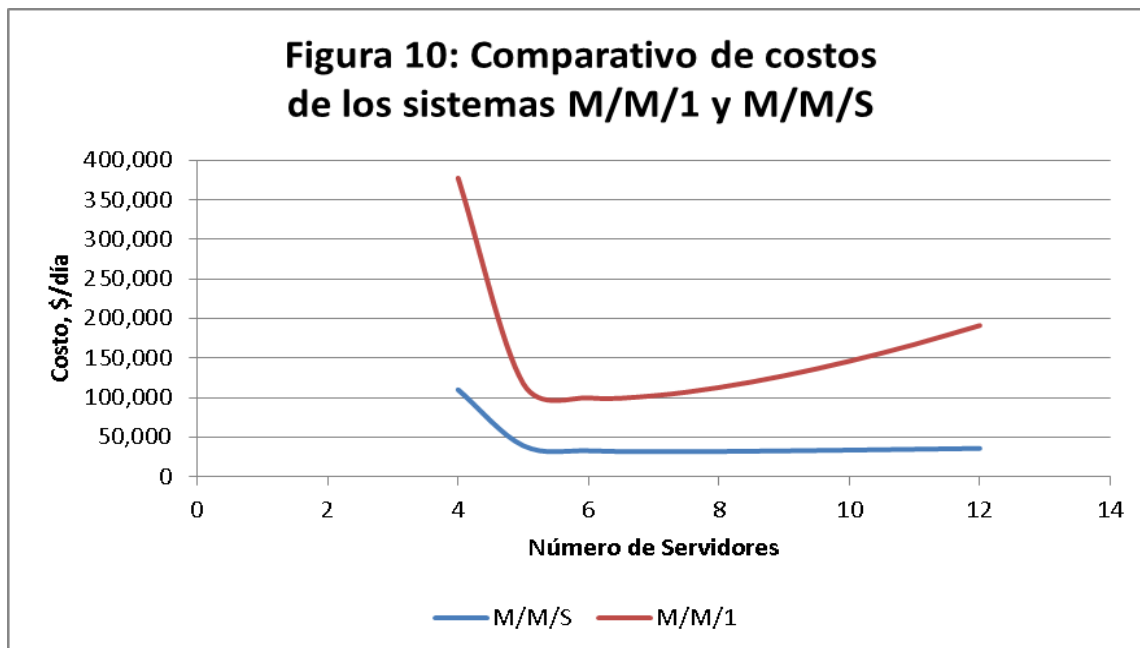
La figura 9 muestra gráficamente la variación del número óptimo de servidores en función de la razón de costos de servicio y demora para ambos sistemas, en la cual se observa que el número de servidores es similar para ambos modelos, siendo siempre mayor o igual para el modelo M/M/1. De hecho en el rango de la razón de costos de 2 a 10, el número óptimo de servidores coincide en 5 para ambos sistemas.



Fuente: Elaboración propia

Finalmente la figura 10 presenta un comparativo del costo de los sistemas M/M/1 y M/M/S en función del número de servidores, en el cual se aprecia que el costo del M/M/S es siempre menor que el del M/M/1, puesto que el costo calculado con este último modelo debe multiplicarse por el número de servidores, ya que funcionan en paralelo. Esto indica que será preferible el sistema M/M/S en caso que sea posible su implementación, ya que pudiera haber problemas de espacio en el muelle para hacer una sola fila de barcos.

Si se observa el gráfico, puede verse que con el mínimo de servidores, que es 4, hay una diferencia significativa en el costo de ambos sistemas (\$110,000 para el M/M/S y más de \$377,000 para el M/M/1), esta diferencia disminuye en el caso de 5 servidores, donde es mínima (\$39,000 para el M/M/S y \$117,500 para el M/M/1) y luego vuelve a incrementarse para un número mayor de servidores.



Fuente: Elaboración propia

Conclusiones

Los hallazgos encontrados validan las diferencias existentes en los modelos de líneas de espera analizados, así como la conveniencia de aplicar o no los mismos a casos reales, en función del patrón de comportamiento de los datos del proceso de servicio relacionado con el negocio.

Para el sistema M/G/1, al tener el tiempo de servicio una mayor desviación estándar, se incrementa el número de servidores y el costo mínimo del sistema, lo que hace un efecto devastador para éste, por lo cual una recomendación es la de tratar a como dé lugar de disminuir la variabilidad del tiempo de servicio, más aún que su valor promedio, ya que el impacto de la desviación estándar es mucho mayor en el costo del sistema que el de su media.

Para esto una recomendación es la de automatizar lo más que se pueda el proceso del servicio, lo cual permitiría disminuir la varianza de su tiempo y con ello el costo del sistema.

Al seguir la recomendación de manejar el número óptimo de montacargas, que resultó 15, la empresa obtendrá un ahorro de más de \$36,000 diarios.

Al incrementarse la razón de costos de servicio y demoras, como es de esperarse, disminuye el número de servidores y aumenta el costo mínimo del sistema, lo que aplica para todos los modelos analizados en este trabajo.

Para cualquier valor de la razón de costos de servicio y demoras, el número óptimo de servidores y el costo del sistema disminuyen con una menor desviación estándar del tiempo de servicio, lo que representa un beneficio económico para la empresa.

Por esto el costo del modelo M/D/1 será siempre menor al del M/G/1, ya que al ser el tiempo de servicio determinístico, su desviación estándar es cero.

Para el caso de la empresa que carga los barcos, bajo un modelo M/M, se ha dado un cruce en donde se igualan los costos de servicio y demoras, el cual no ha coincidido con el costo óptimo del sistema.

Al manejar 8 grúas en vez de las 6 que actualmente se utilizan, la empresa alcanzará un ahorro de más de 2,100 dólares diarios en la operación del muelle.

Para el caso de los modelos M/M/1 y M/M/S, el número óptimo de servidores ha sido menor o igual para el M/M/S en comparación con el M/M/1 para todo el rango estudiado de la razón de costos. De

hecho en un buen rango de valores de éste –de 2 a 10-, se ha dado una igualdad en el número de servidores –grúas en este caso- de ambos modelos.

El costo del sistema M/M/S es notoriamente menor que el del M/M/1, independientemente del número de servidores, lo que lo hace recomendable, en caso de ser factible su implementación, ya que requiere mayor espacio para formar una sola fila de espera, lo que no siempre es posible, como en el caso de las grúas, ya que al ser los clientes los barcos que van a ser cargados, el espacio que requerirían para formar una sola fila de espera en el muelle es enorme.

Con esto queda clara la gran utilidad de aplicar el análisis de líneas de espera a los casos empíricos presentados, lo cual permite optimizar los costos operativos de ambos sistemas, sin menoscabo de la calidad del servicio brindado a los clientes.

Además se ha validado la aplicación de las ecuaciones (10) y (11) para estimar el número óptimo de servidores para los sistemas M/M/1 y M/G/1 respectivamente, lo que evita tener que hacer los cálculos a prueba y error en ambos casos, con el consiguiente ahorro en el número de estimaciones y el tiempo dedicado a ello.

Como limitaciones de este estudio está el periodo de observación para la recopilación de la información, el cual podría ampliarse para determinar con un mayor nivel de precisión el patrón de comportamiento del sistema. De igual forma, ampliar el tipo y número de empresas estudiadas. Para trabajos posteriores, se sugiere buscar otro tipo de organizaciones cuyos patrones de llegadas y servicio sean diferentes, a fin de analizar el sistema y encontrar el valor de sus parámetros que optimice su costo operativo. Esto podría incluir empresas de manufactura con volúmenes altos de manejo de materiales. Asimismo, podría replicarse este estudio en estas mismas empresas, una vez determinado el nivel óptimo de servidores con fines de evaluación.

Bibliografía

Armony, M., Plambeck, E., & Seshadri, S. (2009). Sensitivity of Optimal Capacity to Customer Impatience in an Unobservable M/M/S Queue (Why you Shouldn't Shout at the DMV). *Manufacturing & Service Operations Management*, 11(1), 19-32.

Barak, S., & Fallahnezhad, M. S. (2012). Cost Analysis of Fuzzy Queueing Systems. *International Journal of Applied Operational Research*, 2(2), 25-36.

Chang, F. M., & Ke, J. C. (2011). Cost analysis of a two-phase Queue System with Randomized Control Policy. *Procedia-Social and Behavioral Sciences*, 25, 137-146.

Gómez-Jiménez, F. A. (2008). Aplicación de teoría de colas en una entidad financiera: herramienta para el mejoramiento de los procesos de atención al cliente. *Revista Universidad EAFIT*, 44(150), 51-63.

Gorunescu, F., McClean, S. I., & Millard, P. H. (2002). A queueing model for bed-occupancy management and planning of hospitals. *Journal of Operational Research Society*, 53, 19-24.

Haridass, M., & Arumuganathan, R. (2012). Optimum Cost Analysis of a Bulk Queueing System with Multiple Vacations and Restricted Admissibility of Arriving Batches. *International Journal of Operations Research*, 9(1), 27-43.

Hayel, Y., Ouarranou, M., & Tuffin, B. (2007). Optimal Measurement-based Pricing for an M/M/1 Queue. *Networks & Spatial Economics*, 7, 177-195.

Heskett, J. L., Sasser, W. E., & Hart C. W. L. (1993). *Cambios Creativos en los Servicios*. México: Editorial Díaz de Santos.

Heyman, D. P. (1977). The T-policy for the M/G/1 queue. *Management Science*, 23(7), 775-778.

- Hyttiä, E., Aalto, S., Penttinen, A., & Virtamo, J. (2012). On the Value Function of the M/G/1 FCFS and LCFS Queues. *Journal of Applied Probability*, 49, 1052-1071.
- Izar-Landeta, J. M. (2012). *Investigación de Operaciones*, 2ª edición. México: Editorial Trillas.
- Kumar, R., Jain, N. K., & Som, B. K. (2014). Optimization of an M/M/1/N Feedback Queue with Retention of Reneged Customers. *Operations Research and Decisions*, 3, 45-58.
- Lovelock, C., Reynoso, J., D'Andrea, G., & Huete, L. (2004). *Administración de Servicios*. México: Pearson Prentice Hall.
- Rodríguez-Jáuregui, H. G. R., González-Pérez, A. K., Hernández-González, S., & Hernández-Ripalda, M. D. (2017). Análisis del servicio del área de urgencias aplicando teoría de líneas de espera. *Contaduría y Administración*, próxima publicación.
- Srinivas, V., Rao, S. S., & Kale, K. (2011). Estimation of Measures in M/M/1 Queue. *Communications in Statistics-Theory and Methods*, 40, 3327-3336.
- Stadje, W. A. (2009). Queueing System with Two Parallel Lines, Cost-Conscious Customers, and Jockeying. *Communications in Statistics-Theory and Methods*, 38, 3158-3169.
- Stidham, S. (2005). On the optimality of a full-service policy for a queueing system with discounted costs. *Mathematical Methods of Operations Research*, 62, 485-497.
- Van Meighem, J. A. (1995). Dynamic scheduling with convex delay costs: the generalized $c\mu$ -rule. *Annals of Applied Probability*, 5, 809-833.
- Wang, K. H., & Chang, Y. C. (2002). Cost analysis of a finite M/M/R queueing system with balking, reneging and server breakdowns. *Mathematical Method of Operations Research*, 56(2), 169-180.
- Weber, R. R. (1988). Stochastic scheduling on parallel processors and minimization of concave functions of completion times. "Stochastic differential systems, stochastic control theory and applications", *IMA Volumes Mathematics and its Applications*, 10, 601-609.
- Zeithaml, V. A., & Bitner, M. J. (2002). *Marketing de Servicios*. Colombia: Mc Graw Hill.