

ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL



**FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICAS
DEPARTAMENTO DE INVESTIGACIÓN Y POSTGRADO**

PROYECTO DE TITULACIÓN

PREVIO A LA OBTENCIÓN DEL TÍTULO DE:

“MAGÍSTER EN ESTADÍSTICA APLICADA”

TEMA:

Análisis Multivariante del mantenimiento preventivo y correctivo de equipos médicos.

AUTOR:

MARCOS JAVIER GUAMÁN OCHOA

Guayaquil - Ecuador

2024

RESUMEN

Los países en desarrollo han realizado grandes inversiones en la compra de equipos médicos, a pesar de ello tienen grandes deficiencias en sus sistemas de salud y eso tiene en gran parte que ver la poca atención en cuanto al mantenimiento preventivo y correctivo de los mismos, dando como resultado que muchos de estos equipos estén fuera de servicio debido a una alta tasa de fallos. Al mejorar las políticas de mantenimientos preventivos no solo logramos mejorar el rendimiento de un equipo sino que también reducimos los costos relacionados al mismo. Es en este punto donde los modelos matemáticos juegan un rol importante en el entendimiento y resolución de problemas que se aplican al mundo real y en este caso al de los equipos médicos. Las técnicas Multivariantes y de Regresión Lineal son herramientas que han sido ampliamente aplicadas a varios ámbitos, sobre todo en los que se producen y almacenan grandes cantidades de datos. Aplicar un Modelo Multivariante nos puede llevar a un mejor entendimiento del comportamiento y correlación que tienen los distintos procedimientos que se realizan en los equipos médicos para su correcto mantenimiento. Pudiendo encontrar áreas en las que se tienen deficiencias así como fortalezas. Además de poder llegar a un mejor entendimiento del ciclo de consumo de los repuestos con mayor rotación de stock.

Palabras Clave: equipos médicos, mantenimiento preventivo y correctivo, técnicas multivariantes, regresión lineal.

ABSTRACT

Developing countries have made large investments in medical equipment, yet they continue to have major deficiencies in their health systems. This has largely to do with the lack of attention to the equipment's preventive and corrective maintenance, resulting in a high breakdown rate causing many of these machines to be out of service. By improving preventive maintenance policies, we not only improve the performance of equipment, but also reduce costs related to it. In this point, mathematical models have played an important role in understanding and solving real world problems, and in this case, we can apply them to medical equipment. Multivariate and Linear Regression techniques are tools that have been widely applied to various areas, especially those in which large amounts of data are produced and stored. Applying a Multivariate Model can lead us to a better understanding of the behavior and effectiveness of the different procedures carried out on medical equipment for its correct maintenance, including identifying. With this model we can identify deficiencies as well as strengths, and will ultimately reach a better understanding of the consumption cycle of the spare parts with the highest stock rotation.

Keywords: medical devices, preventive and corrective maintenance, multivariate techniques, linear regression.

DEDICATORIA

A mi esposa por ese apoyo incondicional en todo momento y lugar que convierte cada proyecto en una aventura sin el cual no podría cumplirla. A mis hijos que son la sonrisa y alimento del alma que necesitamos cada día. A mi familia por enseñarme a luchar por mis metas y no dejarme vencer para lograr mis objetivos.

Marcos Javier Guamán Ochoa.

AGRADECIMIENTO

A cada uno de los docentes que durante el proceso de aprendizaje supieron compartir sus conocimientos y despertar ese sentido de curiosidad e investigación en cada uno de nosotros.

A mis directores de trabajo Mat. John Ramírez PhD e Ing. Omar Ruiz PhD, quienes me guiaron y alentaron en todo momento para llegar a un buen término en mi trabajo de graduación.

A Fresenius Medical Care Ecuador, por permitirme ser parte de esta gran empresa y usar los datos generados en su sistema para el estudio realizado.

DECLARACIÓN EXPRESA

La responsabilidad por los hechos y doctrinas expuestas en este Proyecto de Titulación, me corresponde exclusivamente y ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría. El patrimonio intelectual del mismo, corresponde exclusivamente a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

Marcos Javier Guamán Ochoa

TRIBUNAL DE GRADUACIÓN

M.Sc. Heidi Roa López
PRESIDENTE

John Ramirez Figueroa, Ph.D
TUTOR

Omar Ruiz Barzola, Ph.D
DOCENTE EVALUADOR

ABREVIATURAS O SIGLAS

SVM	Máquinas de Soporte Vectorial.
IQ	Coefficiente Intelectual.
ARIMA	Autoregressive Integrated Moving Average.
ACF	Autocorrelación.

TABLA DE CONTENIDO

RESUMEN.....	II
ABSTRACT	III
DEDICATORIA	IV
AGRADECIMIENTO.....	V
DECLARACIÓN EXPRESA	VI
ABREVIATURAS O SIGLAS	VIII
1. INTRODUCCIÓN	1
1.1. Antecedentes.....	1
1.2. Descripción del problema.....	3
1.3. Objetivos	3
1.4. Alcance	4
2. MARCO TEÓRICO	5
2.1. Métodos Explicativos	8
2.1.1. Regresión Múltiple	8
2.1.2. Análisis Canónico (Correlación Canónica).....	8
2.1.3. Análisis Discriminante.....	9
2.1.4. Segmentación Jerárquica	9
2.2. Métodos Descriptivos.....	10
2.2.1. Análisis de Componentes Principales	10
2.2.2. Análisis de Conglomerados (Clúster).....	12
2.3. Biplot	14
2.4. Fases para el Análisis Multivariante de Datos	15
2.5. Series de Tiempo.....	18
2.5.1. Modelos Arima	18
2.5.2. Prueba de Saphiro-Wilk.....	19
2.5.3. Correlograma de Residuales	20
3. METOLOGÍA	21
3.1. Datos.....	21
3.2. Tipo de Investigación	21
3.3. Diseño de la Investigación	22
3.3.1. Localización del estudio.....	22
3.3.2. Población de estudio	22
3.3.3. Técnicas de recolección de datos	23
3.4. Tipo de Investigación	23
3.4.1. Definición de variables para Estadística Descriptiva	23
3.5. Modelos ARIMA.....	24
3.5.1. Pasos para el Análisis de Datos	24
3.5.2. Definición de las variables en las series de tiempo	25
3.6. Análisis de Componentes principales (ACP)	26
3.6.1. Definición de variables para Análisis de Componentes Principales	26
4. RESULTADOS	28
4.1. Estadística Descriptiva de las máquinas HD	28
4.2. Series de tiempo de los trabajos realizados	34
4.3. Representación Gráfica de las series de tiempo	35
4.3.1. Suavizado de las series de tiempo.....	49
4.3.2. Selección del modelo.....	52

4.4.	<i>Análisis Multivariante</i>	60
4.4.1.	<i>Análisis de Componentes Principales (PCA)</i>	60
5.	CONCLUSIONES Y RECOMENDACIONES	71
5.1.	<i>Conclusiones</i>	71
5.2.	<i>Recomendaciones</i>	73
6.	REFERENCIAS	75

LISTADO DE FIGURAS

Gráfico 1.- Técnicas de Análisis Estadístico en función de las variables que se manejan y su principal objetivo de su tratamiento. Fuente: “Técnicas de Análisis Multivariante de Datos, César Pérez López.2004	7
Gráfico 2.- Composición de un gráfico Biplot	15
Gráfico 3.- Pasos para el Análisis Multivariante de Datos.	17
Gráfico 4.- Metodología Box-Jenkins	26
Gráfico 5.- Porcentaje de reportes de servicio en relación al total que conforma la base de datos.	28
Gráfico 6.- Datos faltantes de las series de tiempo, cumplimiento debe ser mayor al 75%.	29
Gráfico 7.- Boxplot general de los códigos de Trabajo de reportes de Servicio.	31
Gráfico 8.- Boxplot de los reportes de servicio generados por tipo de trabajo en la tecnología HD2.	33
Gráfico 9.- Boxplot de Repuestos usados durante el período en Análisis.	34
Gráfico 10.- Serie de tiempo de los trabajos realizados en las máquinas.	35
Gráfico 11.- Tendencia de serie de Tiempo de la variable 0A (Visita)	36
Gráfico 12.- Estacionalidad de la serie de tiempo Visita Técnica.	37
Gráfico 13.- Residuales de la serie de tiempo Visita Técnica.	37
Gráfico 14.- Resultados de Test. Saphiro-Wilk para la serie Visita Técnica	38
Gráfico 15.- Correlograma de la serie Visita Técnica.	38
Gráfico 16.- Gráfico de Tendencia de la serie Mantenimiento Preventivo	39
Gráfico 17.- Estacionalidad de la serie Mantenimiento Preventivo	40
Gráfico 18.- Residuales de la serie Mantenimiento Preventivo.	40
Gráfico 19.- Resultados de la prueba Saphiro-Wilk de la serie Mantenimiento Preventivo.	41
Gráfico 20.- Gráfico de Residuales de la serie Mantenimiento Preventivo.	41
Gráfico 21.- Gráfico de tendencia de la serie Reparación Preventiva.	42
Gráfico 22.- Gráfico de la estacionalidad de la variable Reparación Preventiva	42
Gráfico 23.- Gráfico de residuales de la variable Reparación Preventiva.	43
Gráfico 24.- Resultados prueba Saphiro-Wilk de la variable Reparación Correctiva.	43
Gráfico 25.- Correlograma de la variable Reparación Preventiva.	44
Gráfico 26.- Gráfico de la tendencia de la variable Reparación.	44
Gráfico 27.- Estacionalidad de la serie de tiempo de Reparaciones.	45
Gráfico 28.- Residuales de la serie Reparación.	45
Gráfico 29.- Resultados de prueba Saphiro-Wilk en variable Reparación.	46
Gráfico 30.- Correlograma de Residuales de la variable Reparación.	46
Gráfico 31.- Serie de tiempo de la variable 6751421 repuesto.	47
Gráfico 32.- Estacionalidad de la serie de tiempo Repuesto.	48
Gráfico 33.- Residuales de series de variable Repuesto.	48
Gráfico 34.- Resultados de la prueba Saphiro-Wilk.	49
Gráfico 35.- Correlograma de la variable repuesto.	49
Gráfico 36.- Suavizado de la serie de tiempo de la variable Visita Técnica.	50
Gráfico 37.- Suavizado de la serie de tiempo de la variable Mantenimiento Preventivo.	51
Gráfico 38.- Suavizado de la serie de tiempo de la variable Reparación Preventiva.	51
Gráfico 39.- Suavizado de la serie de tiempo de la variable Reparación Correctiva.	52

Gráfico 40.- Resultados de modelado ARIMA a variable Visita Técnica.....	53
Gráfico 41.- Gráfica de predicción de la serie de tiempo de la variable Visita Técnica.....	53
Gráfico 42.- Suavizado exponencial de la variable Visita Técnica.	54
Gráfico 43.- Modelado ARIMA de la serie de tiempo suavizada.	54
Gráfico 44.- Gráfico de predicción de la serie de tiempo de la variable Visita Técnica.....	55
Gráfico 45.- Modelado ARIMA a serie de Mantenimiento Preventivo.	55
Gráfico 46.- Predicción de valores para la serie de tiempo Mantenimiento Preventivo.	56
Gráfico 47.- Resultado valores modelo ARIMA en serie Reparación Preventiva.	56
Gráfico 48.- Predicción con modelo ARIMA de la serie Reparación Preventiva.	57
Gráfico 49.- Resultado de modelado ARIMA en serie Reparacion.	57
Gráfico 50.- Predicción de la serie de tiempo Reparación con modelado ARIMA.....	58
Gráfico 51.- Suavizado de la serie de tiempo Repuesto.	58
Gráfico 52.- Resultado de ajuste del modelado de la serie Repuesto.	58
Gráfico 53.- Predcción de la serie de tiempo Repuesto.....	59
Gráfico 54.- Gráfico del codo para determinación de las componentes principales.	61
Gráfico 55.- Pesos de la componente PC1.....	62
Gráfico 56.- Pesos de la componente PC2.....	63
Gráfico 57.- Gráfico de dispersión de las dos primeras componentes.....	63
Gráfico 58.- Biplot del Análisis Multivariantes de nuestra base de datos y sus proyecciones	65
Gráfico 59.- Biplot compuesto por 3 componentes principales.....	66
Gráfico 60.- Prueba del codo de la serie Repuesto.	67
Gráfico 61.- Aportes de las variables para la construcción de la componente 1.	67
Gráfico 62.- Biplot de las componentes de Repuestos y códigos de Trabajo.....	69

LISTADO DE TABLAS

Tabla 1.- Resumen de valores generales de las variables.	30
Tabla 2.- Resumen de valores de las variables de trabajos realizados en las máquinas con categoría HF.	32
Tabla 3.- Valores obtenidos del Análisis PCA de la base de datos.	60
Tabla 4.- Resultados de componentes principales de repuestos.	66

CAPÍTULO 1

1. INTRODUCCIÓN

1.1. Antecedentes

Los países en desarrollo realizan importantes inversiones en la compra de equipos médicos, pero no prestan suficiente atención al mantenimiento y reparación de los mismos. Como resultado, muchos de estos equipos están fuera de servicio debido a una alta tasa de averías (Kwaku Kutor et al., 2017)

Las políticas de mantenimiento han sido estudiadas por mucho tiempo, teniendo en cuenta dos tipos principales de mantenimiento como son el predictivo y el correctivo. El mantenimiento correctivo se lleva a cabo cuando un equipo falla, e implica un mayor costo y tiempo de trabajo. El fallo de un equipo en servicio, es de particular preocupación debido a su posible uso en condiciones críticas para el paciente (Murtaza Noor et al., 2016).

El mantenimiento preventivo en cambio, es una serie de tareas en las que se incluye el reemplazo de partes, calibración, ajuste, lubricación, etc., para verificar la seguridad y operatividad del equipo, detectando fallos potenciales u ocultos y tomando las acciones apropiadas para su solución. Estos mantenimientos son programados y su periodicidad está dada por las recomendaciones de cada fabricante (Badnjević & Pokvić, 2020). Siempre está acompañado de un “chek list” o protocolo con el registro de las acciones realizadas.

Las fallas en los equipos ocurren por diversas causas entre las que podemos citar: almacenamiento y transporte inadecuado, falla temprana, errores de manejo, mantenimiento incorrecto, fallas aleatorias, técnicas de reparación inapropiada y fallos por desgaste (Kwaku Kutor et al., 2017)

Entre los problemas enumerados en las líneas anteriores podemos destacar la tasa de fallos tempranos que hace referencia a los equipos que luego de ser instalados presentan fallas que pueden deberse a un mal manejo o a problemas de manufacturación. La tasa de fallos, en cambio es la cantidad de equipos que fallan en relación a la cantidad de máquinas que se mantienen a nivel institucional, regional o país; y por último, la tasa de repetición de fallos, que hace referencia a equipos que luego de haber sido intervenidos (mantenimiento correctivo o preventivo) presentan una nueva falla dentro de un tiempo previamente determinado (3 ó 6 meses).

Tal como menciona (Mahfoud et al., 2016), todo esto supone un cambio valioso en la percepción del mantenimiento de un “mal necesario” a una “oportunidad de inversión”. La excelencia en el mantenimiento se puede lograr tomando decisiones que equilibren los costos y el rendimiento industrial.

Los modelos matemáticos juegan un rol importante en el entendimiento y resolución de problemas del mundo real (Murthy et al., 2015). Y en el contexto de los equipos médicos ayudan a la toma de decisiones para la contratación de servicios de mantenimiento.

Las técnicas de agrupamiento y regresión lineal múltiple son herramientas que han sido ampliamente aplicadas en varios ámbitos, como son el financiero, técnico y biomédico, donde se producen y almacenan grandes cantidades de datos. Estas técnicas son prometedoras para analizar el desempeño de los departamentos responsables y relacionados con el mantenimiento de equipos médicos, (Cruz et al., 2007).

Como ejemplo de lo dicho tenemos, hallazgos que indican que las Máquinas de Soporte Vectorial (SVM) funcionan mejor en mantenimiento preventivo y sistemas predictivos de reemplazo con la mayor precisión de 99,42% y 99,80%, respectivamente. Mientras que el método de K-vecinos obtuvieron la mayor precisión en los sistemas predictivos de mantenimiento correctivo con 98,93%, (Abulaish et al., 2021).

Finalmente, también debemos mencionar que estudios realizados (Astivia-Chávez & Ortiz-Posadas, 2023) han detectado una fuerte relación entre el total de mantenimientos correctivos y la antigüedad de los dispositivos médicos. Sin embargo, usar la técnica apropiada es crucial al momento de manejar grandes cantidades de datos (big data), que puedan proveer indicadores y asistir en la planificación de una estrategia adecuada para el mantenimiento de los equipos médicos.

Al aplicar un modelo multivariante queremos llegar a un mejor entendimiento del papel que tienen los distintos procedimientos que se planifican en los equipos para su correcto funcionamiento, pudiendo determinar las áreas en las cuales los técnicos muestran deficiencias y fortalezas. Así como también entender el ciclo de consumo de los distintos repuestos que se utilizan para la realización del mantenimiento.

1.2. Descripción del problema

Los equipos médicos son una parte fundamental dentro del tratamiento de enfermedades y en muchos casos incluso llegan a reemplazar el funcionamiento de ciertos órganos de pacientes que por varias circunstancias perdieron su funcionalidad.

Los avances tecnológicos en estos dispositivos ha mejorado significativamente los servicios de salud. Sin embargo los fallos en su confiabilidad, disponibilidad y seguridad afectan la calidad de los servicios y pueden mostrar un impacto significativo en los gastos de operación, (Abulaish et al., 2021).

Es por esto que la predicción de daños en este tipo de equipos es una parte fundamental para poder garantizar su correcto funcionamiento y por ende el tratamiento de los pacientes en las instituciones de salud.

1.3. Objetivos

El principal objetivo del presente trabajo, es optimizar el uso de recursos como personal, tiempo y repuestos mediante técnicas de clasificación Multivariante. Así mismo podemos mencionar los siguientes objetivos específicos:

- Revisar la bibliografía científica aplicada para la evaluación estadística que guíe a la selección de una metodología apropiada para el estudio.
- Realizar un Análisis Multivariante al conjunto general de datos en busca de las variables relevantes para el estudio tanto desde el punto de vista de máquinas como de repuestos.
- Interpretar de manera gráfica los resultados obtenidos.
- Evaluar los resultados obtenidos para la generación de modelos de predicción de mantenimientos preventivos y repuestos relacionados en la categorización realizada.

1.4. Alcance

El estudio propuesto será en base a los reportes obtenidos entre los años 2018 y 2023 a nivel nacional. Se realizará en primer lugar un análisis de componentes principales para determinar el orden y peso de incidencia de cada una de las variables. Teniendo en cuenta que debemos dar importancia a datos relevantes como son la serie de las máquinas, el tipo de categoría al que pertenecen, tipo de trabajos realizados, descripción del trabajo realizado, entre otros.

CAPÍTULO 2

2. MARCO TEÓRICO

Antes de iniciar con nuestro análisis debemos realizar una revisión general de los métodos estadísticos que existen y poder entender la razón por la cual se eligió uno u otro método.

La estadística ha pasado de ser solamente una representación de datos, gráficas elegantes o números; ha ser un auxiliar de casi todas las demás ciencias, como son la medicina, la ingeniería, las ciencias sociales, la investigación, etc. La estadística trata de la recolección de datos, presentación, análisis y aplicación, para el soporte en la toma de decisiones, solucionar problemas y diseñar productos y procesos. Además se ha convertido en una parte esencial en el incremento de la calidad de los productos y servicios, basado en la variabilidad (Teresa Carot Sanchez, 2014).

El Análisis Multivariante es un método que tiene ya mucha historia, incluso se puede decir que los primeros en usar esta técnica fueron Francis Galton y Karl Pearson a finales del siglo XIX cuantificando la relación entre la descendencia y las características de los padres y desarrollando el Coeficiente de Correlación. Luego en los primeros años del siglo XX Charles Spearman sentó las bases para el Análisis Factorial mientras investigaba la correlación de los Test de Coeficiente Intelectual (IQ). En estos tiempos no existía ningún tipo de ayuda para llevar a cabo todas las operaciones matemáticas que el método requería por lo que la aplicación del Método Multivariante estuvo limitada y muy poco desarrollada.

No fue sino hasta la segunda mitad del siglo 20 y con el auge de la computación que la aplicación práctica del método se incrementó, lo que derivó en un aumento significativo de la aplicación de los métodos existentes y se renueve el interés en la creación de nuevas técnicas.

El Análisis Multivariante se aplica sobre variables que pueden ser cualitativas, cuantitativas o de ambos tipos simultáneamente. Desde sus inicios ha probado su eficacia puesto que se trata de un método descriptivo

y a la vez multidimensional. Por ejemplo, en Data Mining es muy utilizado el método factorial que junto a los métodos de clasificación resultan un complemento necesario para facilitar la construcción de topologías de individuos según las variables en estudio (Pérez López, 2004a).

Dentro del Análisis Multivariante podemos encontrar varios objetivos o técnicas que podemos usar dependiendo del objetivo principal de nuestro estudio, entre los que tenemos la reducción de dimensión, en la cual partimos de una gran cantidad de datos a las que se les aplica distintos tipos de técnicas como el análisis de componentes principales, análisis factorial, correspondencias, etc., en la cual tratamos de eliminar la información que se encuentra redundante o excesiva; y la clasificación en grupos en la que se puede aplicar la técnica de análisis de conglomerados, árbol de decisión, entre otros, este método tiene por objetivo el analizar las relaciones entre variables. En ocasiones pueden existir una o más variables cuya dependencia de otra sea interesante detectar, o incluso se puede llegar a la predicción de una variable a partir de las que son conocidas.

Pero el Análisis Multivariante no solo se basa en el tipo de técnica que debo aplicar, también se debe tener en cuenta el tipo de variables con las que están conformadas las bases de datos que quiero tratar. Es así que si tengo variables cuantitativas, las técnicas que puedo aplicar son el Análisis de Componentes Principales y el Análisis Factorial, y si son variables cualitativas las técnicas recomendadas son el Análisis de Correspondencias y si se trata de variables cualitativas ordinales podemos usar el Escalamiento Multidimensional. Sin embargo, podemos encontrarnos con variables que no destacan en dependencia, lo cual conlleva a querer clasificarlos en grupos más o menos homogéneos, con lo cual llegaríamos a la aplicación del Análisis de Conglomerados.

Para poder tener una visión general de lo dicho anteriormente en la cual se pueda observar las distintas técnicas del análisis estadístico en función de las variables que se manejan y el objetivo principal de su tratamiento, (Pérez López, 2004a), propone el siguiente cuadro clasificatorio.

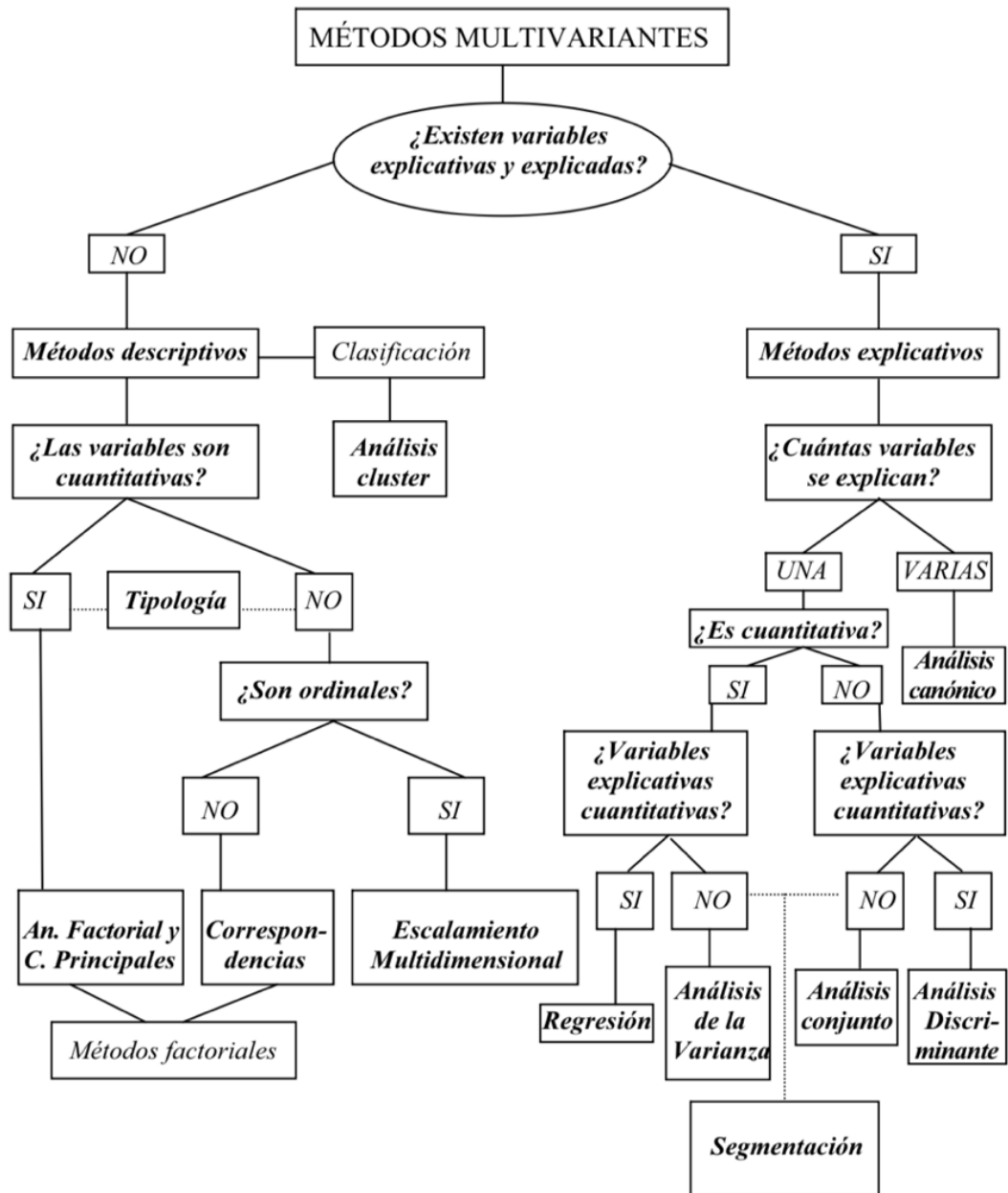


Gráfico 1.- Técnicas de Análisis Estadístico en función de las variables que se manejan y su principal objetivo de su tratamiento. Fuente: "Técnicas de Análisis Multivariante de Datos, César Pérez López.2004

2.1. Métodos Explicativos

Dentro del Análisis Multivariante podemos definir métodos que se usan para aplicar las distintas técnicas existentes y estas se discriminan a partir de la existencia o no de variables explicativas y explicadas. Una parte se desarrolla a partir de la existencia de la dependencia entre las variables explicadas y sus correspondientes variables explicativas, dando lugar a los denominados métodos explicativos (Pérez López, 2004). Entre los que podemos mencionar los siguientes:

2.1.1. Regresión Múltiple

Es una técnica estadística muy utilizada para obtener la relación entre una variable dependiente y varias variables independientes. Su objetivo principal es utilizar las variables independientes, para predecir la única variable criterio, cuyos valores son conocidos, para predecir la única variable criterio.

La expresión funcional del análisis de la regresión múltiple es la siguiente:

$$y = F(x_1, x_2, \dots, x_n)$$

La regresión múltiple admite la posibilidad de trabajar con variables independientes, si se emplean variables ficticias para su transformación.

2.1.2. Análisis Canónico (Correlación Canónica)

Aunque esta técnica estadística se encuentra muy poco aplicada en este tiempo debemos indicar que es usada para analizar la relación entre variables dependientes e independientes. Su objetivo es usar las variables independientes para predecir las variables dependientes.

Su expresión funcional es:

$$G(y_1, y_2, \dots, y_n) = F(x_1, x_2, \dots, x_n)$$

Su principio es desarrollar una combinación lineal de cada conjunto de variables para minimizar la correlación entre ellos, obteniendo de esta manera las ponderaciones para las variables dependientes e independientes. Esto se debe a que los dos conjuntos de datos no tienen un patrón consistente y las correlaciones necesitan ser ajustadas de alguna manera por las correlaciones interiores de los conjuntos de datos.

2.1.3. Análisis Discriminante

El Análisis Discriminante se encarga de analizar la relación entre una variable dependiente no simétrica y varias variables independientes. Al utilizar las variables independientes para predecir la categoría de la variable dependiente a la que corresponde. Se puede utilizar en instituciones financieras en las que se puede categorizar a un cliente que solicita crédito (riesgo crediticio).

El análisis discriminante se puede expresar de la siguiente manera:

$$y = F(x_1, x_2, \dots, x_n)$$

Se puede decir que esta técnica permite agrupar los elementos de una muestra en una o más categorías. Este tipo de análisis es útil cuando la muestra total puede dividirse en grupos basándose en una variable dependiente que se encuentra caracterizada por varias clases conocidas. Su objetivo es entender la diferencia de grupos y predecir la verosimilitud de que un objeto pertenezca a una clase o grupo particular, partiendo de varias métricas independientes. Entre sus aplicaciones se incluyen la distinción entre usuarios ocasionales o habituales de un producto, el riesgo de crédito, entre otros.

2.1.4. Segmentación Jerárquica

Es una técnica que trata de distinguir grupos homogéneos en una población utilizando una variable dependiente y varias independientes que hacen las veces de variables predictoras.

La expresión funcional que se usa para representarla es:

$$y = F(x_1, x_2, \dots, x_n)$$

A esta técnica se lo conoce también con el nombre de técnica de árboles de decisión, ya que puede mostrarse como la partición de la población total en grupos sucesivos a partir de un valor tomado de una variable dependiente, el cual se obtiene de las variables explicativas. Son muy útiles en el análisis y aprendizaje automático, dada su capacidad de manejar variables categóricas y continuas. En este caso cada nodo del árbol representa una condición de la variable independiente y las ramas resultantes es el subconjunto de datos que cumple la condición del nodo.

La segmentación puede ser realizada por dos métodos o enfoques que son el aglomerativo y divisivo. El enfoque aglomerativo toma a cada objeto como un clúster y en cada paso combina los clúster más cercanos hasta que todos los objetos llegan a un solo clúster o hasta el criterio de corte predefinido. Por otro lado el enfoque divisivo, comienza con todos los objetos en un solo clúster y en cada paso éste se divide en dos clúster más disímiles, continuando con este proceso hasta que cada objeto está en su propio clúster o se alcanza el nivel predefinido.

Esta técnica se usa ampliamente en varios campos como son la medicina, marketing, finanzas y más, debido a su buena interpretabilidad y capacidad de revelar relaciones complejas entre las variables.

2.2. Métodos Descriptivos

Los Métodos Descriptivos se desarrollan cuando suponemos la no existencia de dependencia entre las variables explicativas y las variables explicadas.

2.2.1. Análisis de Componentes Principales

El Análisis de Componentes Principales es una técnica multivariante, que se usa para la reducción de dimensión y se aplica cuando el conjunto de datos dispone de un elevado número de variables con datos cuantitativos y cuyo objetivo final es obtener un menor número de variables. Su aplicación sobre cualquier conjunto de datos es de forma directa y se les considera en bloque, sin que se haya establecido jerarquías previas (Pérez López, 2004c).

El objetivo del análisis de componentes principales es transformar un conjunto de variables a las que se les denominará variables *originales interrelacionadas*, en un nuevo conjunto de variables, que es una combinación lineal de las originales que se denominan *componentes principales*, y que se caracterizan por estar incorrelacionadas entre sí y que pueden ordenarse de acuerdo con la información que llevan incorporada. En una componente principal podemos medir la cantidad de información incorporada gracias a la varianza que esta contiene, mientras mayor sea esta, mayor es la información que lleva incorporada la componente y por lo tanto será la primer componente, mientras que la última será la de menor varianza.

Inicialmente contamos con una muestra de tamaño n y con p variables que se pueden expresar como X_1, X_2, \dots, X_p inicialmente correlacionadas, para a partir de ellas obtener un número $k \leq p$ variables incorrelacionadas que se pueden expresar como C_1, C_2, \dots, C_p que son la combinación lineal de las variables iniciales y explican la mayor parte de la variabilidad del conjunto original, a las cuales llamamos componentes principales.

La primer componente principal, es una combinación lineal de las variables originales que se puede expresar como:

$$C_{1i} = u_{11}X_{1i} + u_{12}X_{2i} + \dots + u_{1p}X_{pi}$$

Dado que no contamos solo con una sino con n observaciones muestrales, la misma ecuación puede expresarse matricialmente

$$\begin{bmatrix} C_{11} \\ C_{12} \\ \vdots \\ C_{1n} \end{bmatrix} = \begin{bmatrix} X_{11} & X_{21} & \dots & X_{p1} \\ X_{12} & X_{22} & \dots & X_{p2} \\ \vdots & \vdots & \dots & \vdots \\ X_{1n} & X_{2n} & \dots & X_{pn} \end{bmatrix} \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1p} \end{bmatrix}$$

Si usamos notación abreviada tenemos que: $C_1 = Xu_1$

En cambio si expresamos las X_j con respecto a su media muestral, la media de C_1 es cero, pudiendo expresarlo con la ecuación:

$$E(C_1) = E(Xu_1) = E(X)u_1 = \mathbf{0}$$

En este caso la varianza de C_1 será:

$$V(C_1) = \frac{\sum_{i=1}^n C_{1i}^2}{n} = \frac{1}{n} C_1' C_1 = \frac{1}{n} u_1' X' X u_1 = u_1' \left[\frac{1}{n} X' X \right] u_1 = u_1' V u_1$$

Si expresamos las variables en desviaciones respecto a la media, la expresión $\frac{1}{n} X' X$ (matriz de inercia) es la matriz de covarianzas muestral a la que denominamos V y para las variables tipificadas $\frac{1}{n} X' X$ es la matriz de correlaciones R .

En este caso la primer componente C_1 se obtiene de forma que su varianza sera máxima y mantenga la condición de que la suma de sus pesos u_{1j} al cuadrado sea igual a la unidad, en otras palabras, los pesos o ponderaciones $(u_{11} u_{12} \dots u_{1p})'$ es normalizada.

Si minimizamos $V(C_1) = u_1' V u_1$ para hallar C_1 debemos basarnos en la restricción $\sum_{j=1}^p u_{1j}^2 = u_1' u_1 = 1$.

Para resolver este problema de optimización con restricciones podemos aplicar el método de multiplicadores de Lagrange considerando la función:

$$L = \mathbf{u}'_1 \mathbf{V} \mathbf{u}_1 - \lambda (\mathbf{u}'_1 \mathbf{u}_1 - 1)$$

Si derivamos con respecto a \mathbf{u}_1 e igualamos a cero la ecuación, tenemos:

$$\frac{\partial L}{\partial \mathbf{u}_1} = 2\mathbf{V}\mathbf{u}_1 - 2\lambda\mathbf{u}_1 = \mathbf{0} \Rightarrow (\mathbf{V} - \lambda\mathbf{I})\mathbf{u}_1 = \mathbf{0}$$

Este es un sistema homogéneo en \mathbf{u}_1 , que solo tiene una solución, cuando el determinante de la matriz de los coeficientes es $|\mathbf{V} - \lambda\mathbf{I}| = 0$, pero esta expresión equivale a decir que λ es un valor propio de la matriz \mathbf{V} .

Podemos decir que la ecuación $|\mathbf{V} - \lambda\mathbf{I}| = 0$ tiene n raíces $\lambda_1, \lambda_2, \dots, \lambda_n$, que pueden ser ordenadas de mayor a menor $\lambda_1 > \lambda_2 > \dots > \lambda_n$.

Si la ecuación $|\mathbf{V} - \lambda\mathbf{I}| = 0$ la multiplicamos por \mathbf{u}'_1 a la derecha, obtenemos $\mathbf{u}'_1 |\mathbf{V} - \lambda\mathbf{I}| = 0 \Rightarrow \mathbf{u}'_1 \mathbf{V} \mathbf{u}_1 = \lambda \Rightarrow V(\mathbf{C}_1) = \lambda$. Por lo tanto, para maximizar $V(\mathbf{C}_1)$ tomamos el mayor valor propio de λ de la matriz \mathbf{V} .

Si tomamos λ_1 como el mayor valor propio de \mathbf{V} y \mathbf{u}_1 como su vector propio podemos expresar $(\mathbf{u}'_1 \mathbf{u}_1 = 1)$, ya se define el vector de ponderaciones que aplica a las variables iniciales para obtener la primer componente principal, componente que viene definida, como:

$$\mathbf{Z}_1 = \mathbf{X}\mathbf{u}_1$$

Esta deducción podemos aplicarla para las demás componentes principales como lo expresa (Pérez López, 2004c).

2.2.2. Análisis de Conglomerados (Clúster)

Es una técnica de la Estadística Multivariante, que busca la clasificación automática de datos a partir de una tabla de casos – variables, tratando de situar todos los casos en grupos homogéneos (conglomerados o clústers), no conocidos anteriormente pero sugeridos por la propia naturalidad de los datos. Si tenemos una serie de variables o muestra de individuos de las cuales se dispone varias observaciones, el análisis de conglomerados sirve para clasificarlos en grupos lo más homogéneo posible en base a las variables observadas. Todos los individuos que queden clasificados dentro del mismo grupo deberán ser tan similares como sea posible.

El análisis de conglomerados usa las características o atributos de las variables para clasificarlas, basada en patrones. Las unidades que se clasifican de manera similar estarán geoméricamente mas próximas entre sí (alta homogeneidad interna) en comparación de unidades de diferentes grupos. El Análisis de conglomerados es un proceso que comprende varios pasos en los que se incluye:

- Selección del método de agrupación, se tiene disponible varios métodos de agrupación y cada uno cuenta con su propia parametrización. Los más comunes son, K-means, Jerárquico, DBSCAN.
- Distancia o Similitud, su métrica puede afectar significativamente a los resultados. Las más comunes son Euclidiana, Manhattan, Coseno.
- Inicialización, en métodos como el K-means la inicialización de los centroides puede influir en el resultado final.
- Validación, para evaluar la calidad de los conglomerados formados, se pueden usar varios índices como Índice de Silueta, índice de Dunn y el Coeficiente de Rand ajustado.

El análisis de conglomerados se puede dividir en dos grandes grupos: los no jerárquicos, que son aquellos en los que las variables se asignan a grupos diferenciados que se configuran dentro del mismo análisis, sin que dependan de otros grupos. Y los jerárquicos son aquellos que se configuran con una estructura arborescente; es decir, que los niveles más bajos van siendo englobados en otros de niveles superiores.

Los métodos no jerárquicos pueden además producir clústers distintos (cada caso pertenece a un solo clúster), o solapados (un caso puede pertenecer a más de un grupo a la vez), éstos resultan difíciles de interpretar, por lo que en la práctica son muy poco utilizados.

Una vez obtenido el clúster final sea en subconjuntos jerárquicos o no jerárquicos, se podrá aplicar técnicas estadísticas comparativas convencionales siempre que la relevancia práctica lo permita; así como, otras pruebas multivariantes, para que ya se contará con una variable dependiente (grupo), aunque haya sido creada “artificialmente”.

2.3. Biplot

Toda matriz con un rango de 2×2 puede ser dibujada como un Biplot el cual consiste en un vector para las filas y otro para cada una de las columnas de la matriz. El Biplot provee una herramienta completa para el análisis de los datos y permite tener un panorama visual de la estructura de matrices de datos de gran tamaño. Es muy usada en el Análisis de Componentes Principales, donde puede mostrar distancias y clúster de unidades, así como, la varianza y correlación de las variables (GABRIEL, 1971).

Los elementos de una matriz son el producto interno de los vectores representando las correspondientes filas y columnas, esto es conceptualmente útil para el entendimiento de las propiedades de la matriz.

Cualquier matriz Y de $n \times m$ de rango r puede ser factorizada como:

$$Y = GH'$$

donde, las dos matrices $G_{n \times r}$ y $H_{m \times r}$, ambas son necesariamente de un rango r .

En una matriz de rango dos, los efectos $g_1 \dots g_n$ y $h_1 \dots h_m$ son vectores de orden dos. Estos $n + m$ vectores pueden ser dibujados en un plano dando una representación de nm elementos de Y por las medias del producto interno con el correspondiente vector de efecto de la fila y la columna.

En otras palabras un Biplot es una representación de una matriz de datos obtenida por los valores propios y los vectores propios de la matriz de covarianza y obtenida como:

$$x_2 = (p_1, p_2) \begin{pmatrix} \sqrt{\lambda_1} & 0 \\ 0 & \sqrt{\lambda_2} \end{pmatrix} \begin{pmatrix} q_1^T \\ q_2^T \end{pmatrix}$$

donde, x_2 es el rango dos de aproximación de los datos de la matriz x , λ_1 y λ_2 son los dos primeros valores propios de la matriz, y q_1 y q_2 los vectores propios. Los vectores p_1 y p_2 son obtenidos como:

$$p_i = \frac{1}{\sqrt{\lambda_i}} X q_i; \quad i = 1, 2$$

La distancia entre los puntos representan la distancia entre las unidades y la longitud del vector desde el origen a la coordenada representa la varianza de

dicha variable. Por último la correlación entre las variables es el ángulo entre los vectores, mientras menor sea el ángulo mayor será la correlación entre las variables.

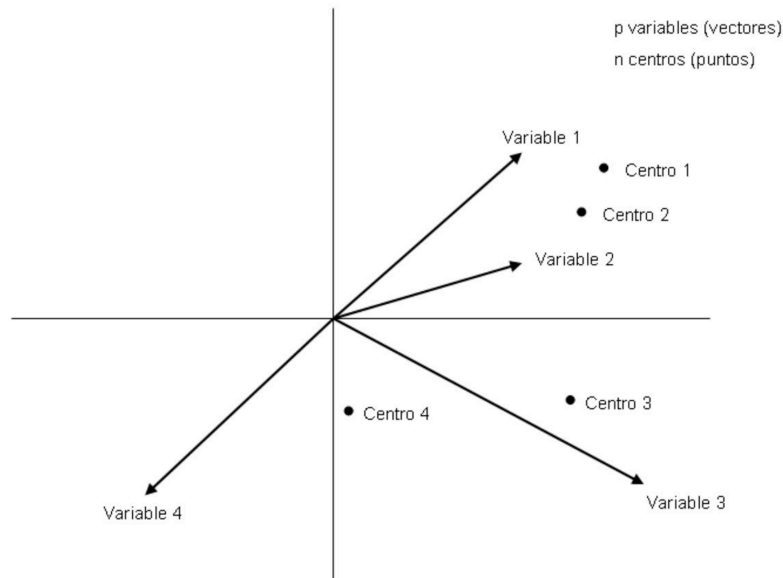


Gráfico 2.- Composición de un gráfico Biplot

2.4. Fases para el Análisis Multivariante de Datos

Para la correcta realización de un Análisis Multivariante debemos tomar en cuenta ciertos aspectos que nos ayuden a definir correctamente el modelo a aplicar hasta el diagnóstico crítico de los resultados. Debemos iniciar primero por definir un plan de investigación con un modelo conceptual que detalle las relaciones a examinar. Una vez definido el modelo, debemos seleccionar una técnica multivariante específica y ponerla en práctica.

Luego de obtener los resultados significativos debemos interpretar los mismos para finalmente validar el modelo para las muestras de datos y ver su generalización.

El Análisis Multivariante debe iniciarse con la definición del problema de investigación, objetivos y técnica conveniente para el estudio. Se debe definir los conceptos e identificar las relaciones fundamentales a estudiar.

A continuación se debe desarrollar el análisis aplicando la técnica multivariante elegida, se debe desarrollar para cada técnica un plan de análisis específico en los que se defina el tamaño de la muestra, los tipos de variables permitidas, métodos de estimación entre otros.

Una vez realizada la toma de datos se deberá evaluar los mismos para comprobar supuestos, dependiendo de las técnicas de dependencia o interdependencia de los datos como son normalidad, linealidad, homoscedaticidad, correlación, etc., antes de aplicar el modelo.

El siguiente paso es la estimación del modelo multivariante y la valoración del ajuste, una vez satisfechos todos los supuestos se procede a la estimación efectiva del modelo realizando una valoración global del mismo.

Luego de la valoración debemos interpretar los valores obtenidos. Esta interpretación puede llevar a la especificación del modelo y a una nueva estimación hasta que la interpretación de los resultados sea coherente a los valores teóricos. Mientras no se cumpla esta condición no existe evidencia empírica de que las relaciones multivariantes de los datos muestrales puedan generalizarse para toda la población.

Por último, debemos validar el modelo y aceptar los resultados mediante la aplicación de contrastes específicos de cada técnica.

El proceso descrito anteriormente se puede esquematizar de la siguiente manera:

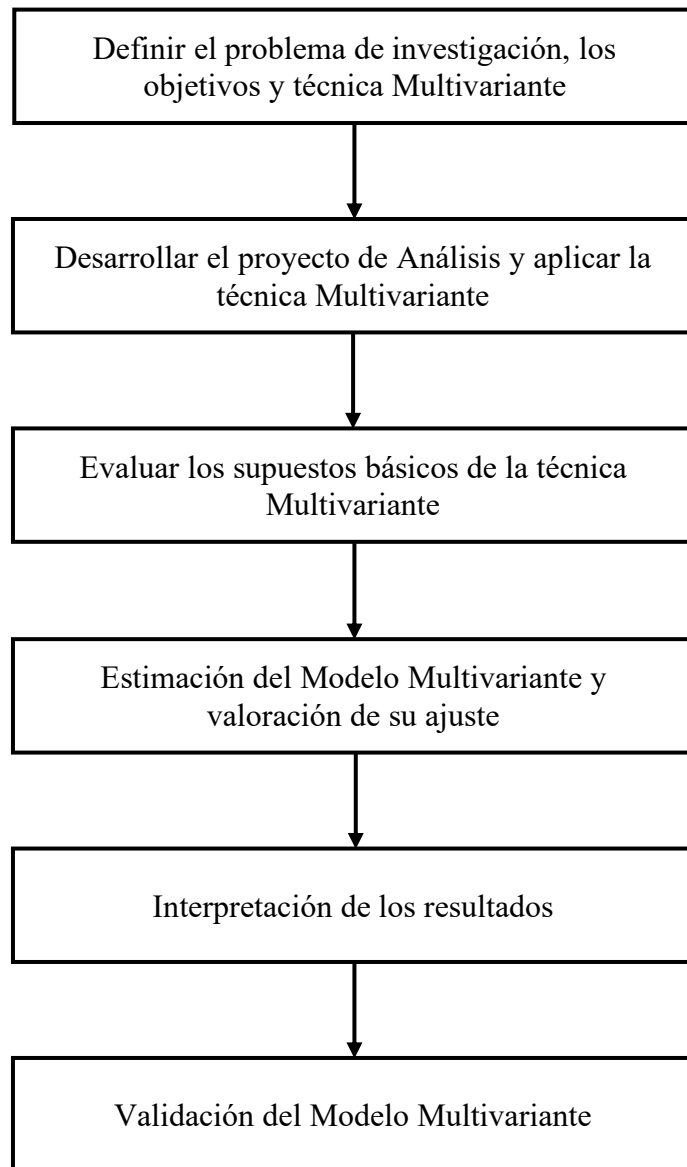


Gráfico 3.- Pasos para el Análisis Multivariante de Datos.

2.5. Series de Tiempo

Dentro del análisis de datos, también nos encontramos con un conjunto de datos que se denomina series temporales, este tipo de series se caracterizan por tener información que se prolonga a lo largo del tiempo y que nos sirve para encontrar tendencias, periodicidad y fenómenos aislados que pueden afectar de cierta manera los registros que se pueden obtener. Además podemos basarnos en los hechos pasados para realizar la predicción de hechos futuros. Como ejemplo podemos indicar la recolección de datos meteorológicos para la predicción del clima.

2.5.1. Modelos Arima

Para el modelado de las series de tiempo disponemos de los modelos clásicos y los modelos ARIMA.

La principal diferencia entre los modelos ARIMA (AutoRegresive Integrated Moving Average) y los modelos clásicos es el enfoque estocástico con el que se trata las series de tiempo, en lugar de tratarlas de forma determinística. Bajo este concepto se toma a las series de tiempo como un conjunto de valores aleatorios, que se generan a partir de un proceso desconocido. Es por esto que su objetivo es tratar de determinar el modelo probabilístico que represente el comportamiento de la serie.

De entre las distintas metodologías que se utilizan para el proceso de series de tiempo podemos destacar la Metodología Box-Jenkins, la cual se basa en tratar de determinar el modelo probabilístico que rige al comportamiento del proceso a lo largo del tiempo.

La metodología Box-Jenkins se resume en cuatro fases:

- Identificar el posible modelo ARIMA que sigue la serie.
- Estimación de los parámetros AR y MA por máxima verosimilitud, además se obtienen los errores estándar y los residuos del modelo.
- Diagnóstico o comprobación que los residuos y su estructura de dependencia y proceso de ruido blanco.
- Predicciones con el modelo obtenido.

2.5.2. Prueba de Saphiro-Wilk

La prueba de Saphiro-Wilk, es una de las más utilizadas para la comprobación de la normalidad de los datos. Su objetivo es determinar si una muestra proviene de una población con una distribución normal.

La prueba realiza el cálculo de una estadística W que mide cuánto difieren las observaciones ordenadas de la muestra de los valores esperados de una distribución normal.

En la prueba de Saphiro-Wilk nos planteamos la hipótesis de la normalidad de los datos, teniendo que:

$$H_0 = \text{Los datos siguen una distribución normal}$$

$$H_1 = \text{Los datos no siguen una distribución normal}$$

El valor de W , viene dado por

$$W = \frac{(\sum_{i=1}^n a_i x_{(i)})^2}{\sum_{i=1}^n (x_i - \tilde{x})^2}$$

donde:

- $x_{(i)}$ son los valores ordenados de los datos de la muestra.
- \tilde{x} es la media de los datos.
- a_i con coeficientes calculados a partir de la media y la varianza de los datos, y dependen directamente de la distribución normal.

Durante el test también obtenemos el valor p que es el coeficiente que se usa para rechazar o no la hipótesis nula. Si su valor es menor que el nivel de significancia (0.05 por defecto), se rechaza la hipótesis nula, lo que nos indica que los datos analizados no siguen una distribución normal.

Esta prueba tiene como ventaja que se puede utilizar para muestras pequeñas, en la que resulta de mucha utilidad. Además es una prueba exacta, lo que implica que la probabilidad de cometer un error es muy bajo igual al nivel de significancia.

Al contrario si tenemos muestras muy grandes puede ser menos efectiva y además es muy sensible a los valores atípicos.

2.5.3. Correlograma de Residuales

El correlograma de residuales es una herramienta gráfica utilizada en el análisis de series temporales, la misma sirve para evaluar la independencia de los residuos de un modelo. También es conocido como gráfico de autocorrelación.

Su objetivo es verificar si los residuos de un modelo de serie temporal son aleatorios, es decir, no están correlacionados. Esto es importante ya que uno de los supuestos fundamentales de muchos modelos de series de tiempo es que los residuos deben ser independientes y distribuidos de manera idéntica. La autocorrelación (ACF) es la correlación de una serie de tiempo con los rezagos de sí misma. Este representa el coeficiente de autocorrelación para diferentes rezagos.

La independencia de los residuos no debe mostrar autocorrelación, los cuales deberían ser cercanos a cero para todos los rezagos. En el correlograma esto se visualiza con barras que no exceden los intervalos de confianza, que usualmente se encuentran representados con líneas interpunktadas alrededor de cero.

CAPÍTULO 3

3. METOLOGÍA

3.1. Datos

Para el presente trabajo de investigación se ha tomado la base de datos generada a partir del sistema de gestión de trabajos realizados por los técnicos de una empresa encargada del mantenimiento de máquinas de diálisis a nivel nacional. Los técnicos a nivel nacional generan luego de cada trabajo un reporte el cual se registra en el sistema y se alimenta una base de datos la cual diariamente se sincroniza. El administrador del sistema puede revisar los trabajos generados por cada técnico y a la vez exportar los datos en formato Excel.

Entre las variables que podemos encontrar en dicha base de datos se encuentran aspectos como: fecha de los trabajos, serie del equipo, tipo de máquina, horas de uso, tipos de trabajo realizado, tiempos de trabajo, detalle de los trabajos, repuestos usados, etc. Teniendo un total de 242 variables. Se estima que la producción mensual de registros se encuentra en un promedio de alrededor de 450 registros, lo que da como resultado un promedio de 2000 registros mensuales.

Los datos son de propiedad de la empresa y contiene información de clientes y nombres de los técnicos que realizan los trabajos; por lo que, en este punto se procederá con la anonimización de los mismos para proteger su integridad y asegurar la privacidad de datos que la empresa mantiene como política interna.

3.2. Tipo de Investigación

La investigación propuesta por el método a aplicar es una investigación cuantitativa, puesto que vamos a analizar variables como la cantidad de mantenimientos preventivos, correctivos, tiempo de vida de las máquinas, cantidad de tipo de máquinas por categorías (modelo), entre otros.

Desde el punto de vista de profundización en la investigación, el presente estudio es analítico en series temporales; puesto que, nos vamos a basar en datos recopilados a lo largo del tiempo para determinar tendencias y comprender como ciertas variables o factores afectan a los resultados, usando estadística analítica, y análisis multivariante de componentes principales.

Por la manipulación de variables de este trabajo, estamos ante un trabajo no experimental ya que los datos de estudio se obtuvieron del sistema de registro de trabajos de mantenimiento de la empresa, y los mismos se registran de manera diaria.

Por último desde el punto de vista del período temporal, esta investigación es longitudinal, puesto que los datos se han ido registrando a lo largo del tiempo.

3.3. Diseño de la Investigación

El presente trabajo de investigación se clasifica como un diseño no experimental, debido a que la data es obtenida a partir de los registros del sistema de gestión de la empresa; es decir, no se realizan alteraciones en las variables, para observar el comportamiento de los trabajos de mantenimiento en los distintos equipos y regiones, obteniendo de esta manera datos longitudinales.

3.3.1. Localización del estudio

Para la localización del estudio, nos hemos referido al mismo como a todo el territorio ecuatoriano el cual cuenta con un aproximado de 400 clientes entre privados y públicos, que están divididos en las provincias de las distintas regiones del país. Dando como resultado un aproximado de 2500 máquinas a nivel nacional.

La base de datos contiene registros desde el 1 de enero de 2018 hasta el 31 de diciembre de 2023

3.3.2. Población de estudio

Para el desarrollo del estudio la data obtenida de la empresa tiene un total de 133807 registros que comprenden el período enero 2018 a diciembre 2023.

3.3.3. Técnicas de recolección de datos.

Los datos se obtienen a partir, de la función de exportación de datos del sistema que es propiedad de la empresa, la información se actualiza de manera diaria con los registros que ingresa cada técnico, quien en este caso tiene tiempos específicos en los cuales debe registrar las actividades.

3.4. Tipo de Investigación

El análisis de estadística descriptiva es donde vamos a recoger, almacenar, ordenar y obtener tablas o gráficos que describan el escenario ante el que nos encontramos, para calcular varios parámetros del conjunto de datos para el análisis propuesto, entre los valores que se puede evidenciar tenemos: media, mediana, moda, mínimo, máximo, etc. Para el desarrollo del Análisis Estadístico descriptivo se utilizó los datos de la empresa, en donde encontramos variables, entre las cuales podemos nombrar:

- Número de mantenimientos preventivos realizados de manera mensual por tipo de máquina. Estos mantenimientos debemos diferenciarlos en dos grupos que son Reparaciones preventivas y mantenimientos programados (se realizan cada dos años por recomendación del fabricante).
- Número de mantenimientos correctivos realizados de manera mensual por tipo de máquina.

3.4.1. Definición de variables para Estadística Descriptiva

Para la realización de la estadística descriptiva se ha realizado el cambio de nombres de las variables de la base de datos originales, de la siguiente manera:

- Tecnología
- Mes / Año
- Tiempo de Vida
- Región

3.5. Modelos ARIMA

3.5.1. Pasos para el Análisis de Datos

- **Recogida de Datos:** Se recomienda en este caso disponer de al menos 50 datos o más.
- **Representación gráfica:** Es de mucha utilidad usar gráficos, se utilizan medias y desviaciones típicas por subperíodo para juzgar la estacionariedad de la serie.
- **Transformación previa de la serie:** con series no estacionales se debe trabajar en base a una transformación logarítmica. Lo habitual es comprobar los resultados entre la serie original y la serie en logaritmos.
- **Eliminación de la tendencia:** el gráfico de la serie ayuda a visualizar la existencia o no de tendencia, sea esta lineal o no lineal.
- **Identificación del modelo:** refiere a la elección del modelo más adecuado, esto se logra en base a las funciones de autocorrelación (FAC) y autocorrelación parcial (FAC parcial), tanto en su parte regular como estacional. Se puede seleccionar varios modelos alternativos, desde las opciones más simples AR(1), AR(2), MA(1), MA(2), ARMA(1,1). Estos modelos serán estimados y luego contrastados para definir el modelo adoptado.
- **Estimación de los coeficientes del modelo:** Una vez seleccionado el modelo, se procede con la estimación de parámetros. Debido a que se trata de un proceso iterativo de cálculo, se puede sugerir valores iniciales.
- **Contraste de validez del modelo:** Se valora el modelo o modelos seleccionados con distintos procedimientos como contraste de significación de parámetros, covarianza entre estimadores, coeficiente de correlación, suma de cuadrados de errores, etc.
- **Análisis detallado de los errores:** Se toman en cuenta las diferencias entre los valores reales y estimados por el modelo. Se verifica el comportamiento no sistemático de los modelos, así como, la posible existencia de errores especialmente significativos.

- Selección del modelo: Luego de analizar los resultados de los pasos anteriores se decide sobre el modelo a adoptar.
- Predicción: Se utiliza el modelo seleccionado como fórmula inicial de la predicción.

3.5.2. Definición de las variables en las series de tiempo

Para realizar el análisis de una serie de tiempo se recomienda contar con al menos 50 registros, en nuestro caso disponemos de los datos mensuales de trabajos realizados desde enero 2018 a diciembre 2023 (5 años, 60 registros en total) que cumple con esta condición inicial. Esto nos permitirá ver el comportamiento de los distintos trabajos que se realizan en las máquinas.

Las variables en este caso no han sido renombradas debido a que las mismas ya se encuentran codificadas según el protocolo que la empresa tiene.

El método Box-Jenkins se usa también para realizar una predicción a corto plazo de la concentración de ozono troposférico en la zona urbana de la ciudad de Cali, Colombia, mediante el análisis Univariante de series de tiempo. El modelo propuesto puede ser usado para predecir y alertar a la población sobre concentraciones altas de ozono con hasta 8 horas de anticipación (Jaramillo Ayerbe et al., 2007).

Encontramos además el uso del método Box-Jenkins para identificar el modelo ARIMA, el cual se ajuste al comportamiento de los precios nominales del jitomate en México. De acuerdo a los resultados se concluye que las series de tiempo en consideración se ajustan a un modelo ARIMA (23,0,1) el cual tiene dos factores autoregresivos y una media móvil. Con este modelo se hizo una predicción para 12 meses, desde diciembre 2008 a noviembre 2009 (Marroquín Martínez & Chalita Tovar, 2018).

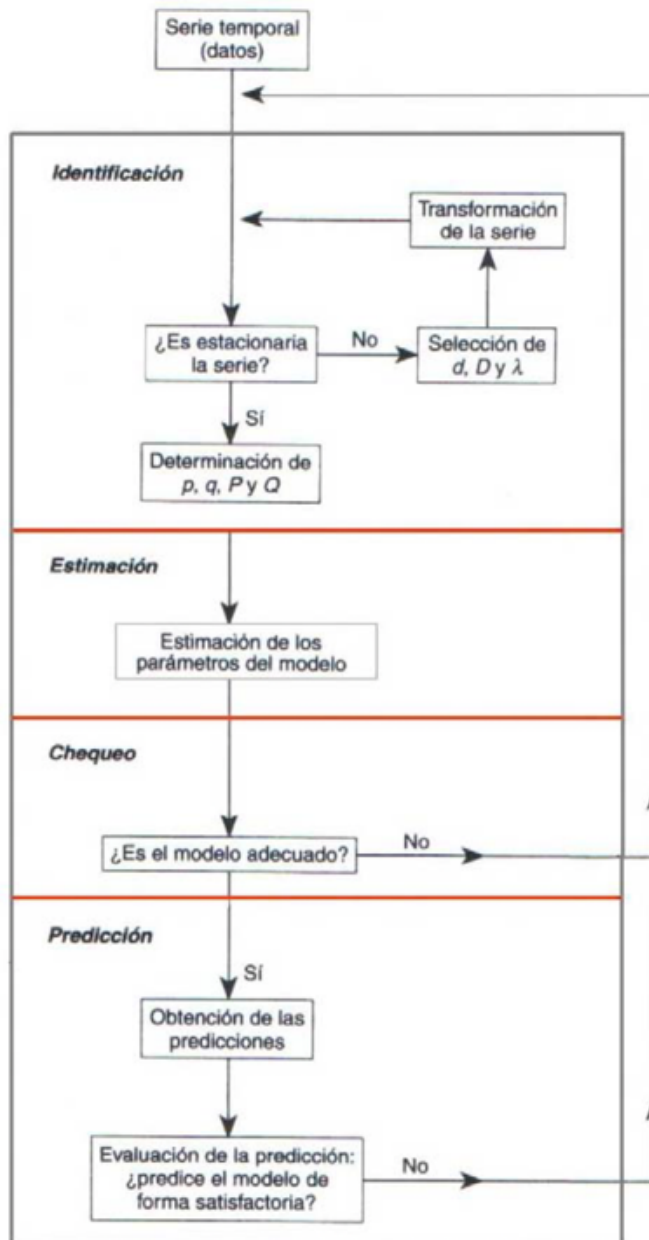


Gráfico 4.- Metodología Box-Jenkins

3.6. Análisis de Componentes principales (ACP)

3.6.1. Definición de variables para Análisis de Componentes Principales

Al realizar el análisis de la base de datos se puede identificar las variables que se pueden usar para la aplicación del Análisis de Componentes Principales, para lo que utilizaremos los códigos de trabajo que realizan los técnicos en las máquinas de HD (Hemodiálisis), los cuales se encuentran ya

filtrados y contabilizados de manera mensual. Entre las principales variables de trabajos realizados podemos mencionar:

0A: Visita Técnica, es un código muy usado ya que con este código se reportan visitas anuales para mantenimientos preventivos “cortos”, y además para justificar la visita de una máquina bajo contrato de mantenimiento. En muchos casos las visitas de contratos pueden ser mensuales, trimestrales, semestrales o anuales.

0B: Reparación preventiva, hace relación a trabajos de reparación o cambio de piezas en las cuales el equipo no presentó fallas, el problema fue detectado antes de que provoque “paras”.

1A: Mantenimiento TSC-MA (Technical Safety Check-Maintenance), por recomendación del fabricante se debe realizar de manera bi anual.

1B: Reparación, es todo trabajo que implica el arreglo del equipo que puede incluir o no el cambio de una parte o pieza. La condición inicial se da cuando el equipo ya no puede ser utilizado de manera correcta y se tuvo que llamar al técnico para su revisión.

1E: Instalación, se usa para realizar la puesta en marcha de un equipo nuevo, este código se usa una sola vez en el tiempo de vida útil de la máquina.

1F: Reinstalación, si el equipo es retirado de servicio e instalado en un nuevo cliente se utiliza el código de reinstalación.

1G: Check no presenta error, se usa en el caso que un usuario reporta un error y este no puede ser reproducido por el técnico.

1H: Modificación, son acciones generalmente mandatorias en las cuales se hacen modificaciones o actualizaciones a los equipos de manera mandatoria, por lo general solicitado desde fábrica.

CAPÍTULO 4

4. RESULTADOS

4.1. Estadística Descriptiva de las máquinas HD

Para tener una perspectiva del aporte de cada una de las tecnologías a nuestra base de datos en primer lugar se procedió con la categorización de los distintos modelos de máquinas en 3 tecnologías de la siguiente manera:

- HD1
- HD2
- HD3

Al realizar el análisis de la cantidad de reportes que aporta por cada tipo de tecnología, podemos observar que una tecnología prevalece claramente sobre las demás, teniendo un aporte del 68% de los datos obtenidos, mientras que las otras tecnologías tienen aportes del 11% y 21%.

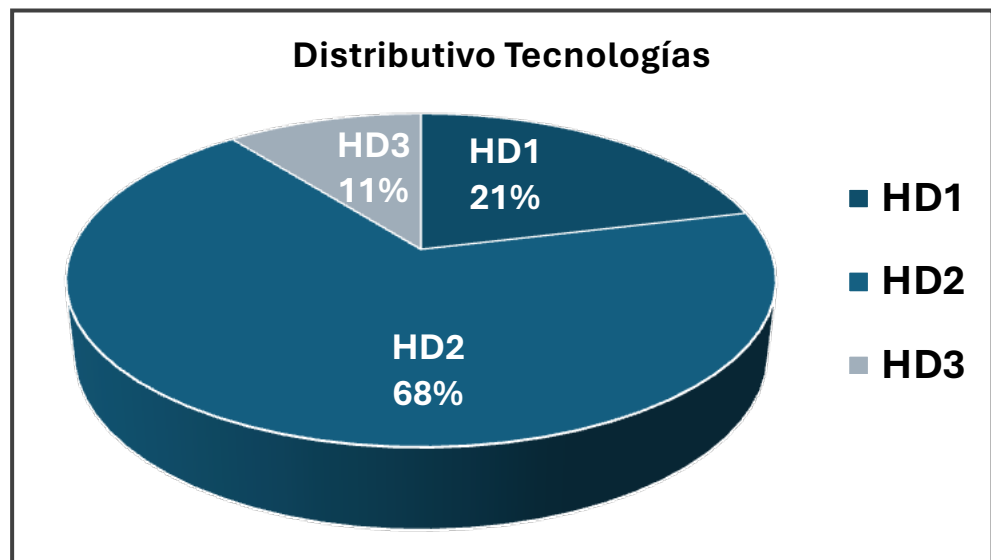


Gráfico 5.- Porcentaje de reportes de servicio en relación al total que conforma la base de datos.

Esto puede entenderse debido a que la tecnología HD2 es la que más penetración de mercado tiene, es usada en su gran mayoría en las clínicas de los prestadores privados. La tecnología HD1 en cambio es una tecnología que se encuentra ya con un tiempo de vida de entre 10 a 20 años, por lo que poco

a poco se han ido retirando del mercado siendo estas reemplazadas por las de tecnología HD2. Por el otro lado tenemos la tecnología HD3 que corresponde a una nueva generación de equipos con un tiempo de vida de no más de 5 años, es una versión moderna de las máquinas que se introduce al mercado como una oferta de un nuevo tratamiento más moderno y con mayores beneficios para el paciente, aunque con un costo más elevado; por lo que, los clientes son reacios a introducir la tecnología. El gobierno no reconoce estas diferencias de costos, pero estudios posteriores podrían relevar que con este nuevo tratamiento se podría mejorar la condición de salud de los pacientes y por ende disminuir los rubros que las clínicas invierten en medicamentos y control de los pacientes.

Es así que el aporte de reportes en base a cada tipo de tecnología se reparte de la siguiente manera:

- HD1: 11241 reportes
- HD2: 36070 reportes
- HD3: 5637 reportes

Antes de realizar el análisis de varianza de la base, realizamos un análisis de los datos faltantes y encontramos que varias variables tienen un alto porcentaje de datos en blanco (N/A). En el gráfico a continuación se observa claramente las variables que del total de la serie de tiempo tienen menos del 75% de datos, en este caso el proceso que debemos adoptar es el de descarte, esto debido al poco aporte que realizan a la variabilidad de los datos. Tomando en cuenta para nuestro estudio solamente las variables que tienen mas del 75% del total de variables esperadas.

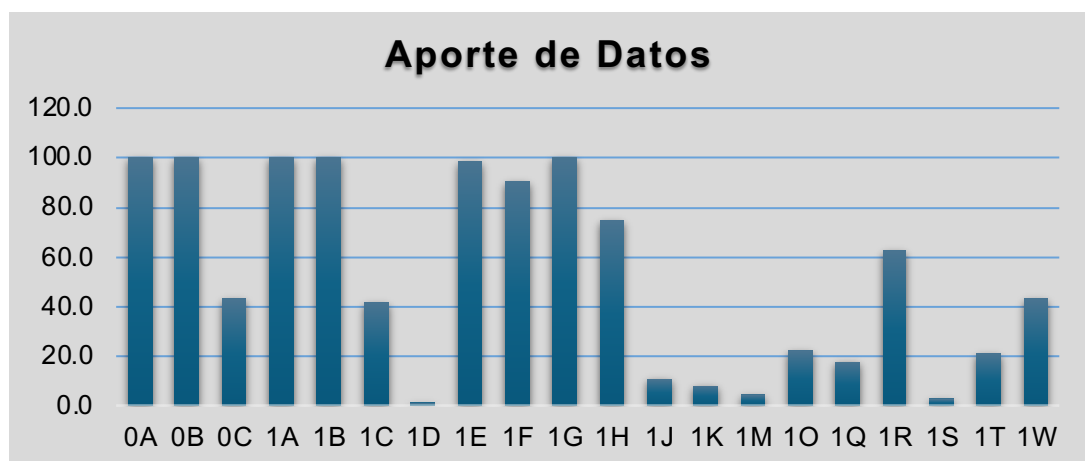


Gráfico 6.- Datos faltantes de las series de tiempo, cumplimiento debe ser mayor al 75%.

Bajo este criterio las variables a tener en cuenta para el estudio serán: 0A, 0B, 1A, 1B, 1E, 1F, 1G y 1H. Las demás variables tienen muy pocos datos como para aportar con información significativa a la comprensión de los datos.

Para poder determinar la influencia o cambio que tenemos al considerar la base de datos completa de datos versus la base tomando en cuenta solamente la tecnología HD2, vamos a realizar un comparativo de los resultados para determinar si es relevante o no el aislamiento de las tecnologías.

Para iniciar la estadística descriptiva vamos a realizar un boxplot con el total de trabajos realizados durante los años de muestreo. El cual nos da una visualización de la forma de distribución de los datos. Entre los cuales podemos observar los cuartiles, valores máximos y mínimos en incluso en el caso de tengamos dentro de nuestros datos valores atípicos (outliers).

En el primer caso podemos observar que las variables que contienen mayor cantidad de datos son los que corresponden a los trabajos de reparaciones preventivas y correctivas (0B y 1B), de igual manera las visitas sobresalen sobre los mantenimientos, esto debido a que las visitas en los clientes públicos se realizan en distintos ciclos que van desde mensuales hasta semestrales. En menor proporción encontramos los reportes de Instalación (1E), reinstalación (1F), Check no presenta error (1G) y Modificación (1H). En el caso de valores atípicos podemos encontrarlos en los códigos 1A, 1E, 1F, 1G y 1H. La distribución de valores, cuartiles, mínimos y máximos podemos observarlos claramente en la tabla 1.

Tabla 1.- Resumen de valores generales de las variables.

	0A	0B	1A	1B	1E	1F	1G	1H
Min	24.0	83.0	16.0	147.0	1.0	1.0	7.0	1.0
1st Qu	66.8	180.0	55.3	198.0	11.5	3.0	16.0	4.0
Median	89.5	244.5	69.5	223.5	20.0	7.0	20.0	10.5
Mean	91.1	249.2	70.9	231.9	23.2	10.8	21.8	23.8
3rd Qu	117.5	311.2	87.3	261.0	31.5	16.0	26.0	28.3
Max	177.0	446.0	148.0	347.0	82.0	61.0	51.0	111.0
N/A	0.0	0.0	0.0	0.0	1.0	7.0	0.0	18.0

El Boxplot, nos ayuda también a visualizar la distribución de los valores que tiene cada variable, en el caso de los mantenimientos (1A) y una distribución normal, esto se observa al ver que la mediana se encuentra prácticamente al centro de la caja al igual que sus “bigotes”. Por el contrario en las reparaciones (1B) y visitas (0A) y reparaciones preventivas (0B) tienen un leve sesgo debido a la ubicación de la mediana dentro de la caja y además es más notorio la diferencia de tamaño de los “bigotes”, en ciertos casos es más marcada la diferencias como es el caso de las reparaciones y las modificaciones (1H) que incluso contiene valores atípicos (outliers) que indica que en ciertos meses los trabajos realizados salen de la “tendencia” encontrada en los datos.

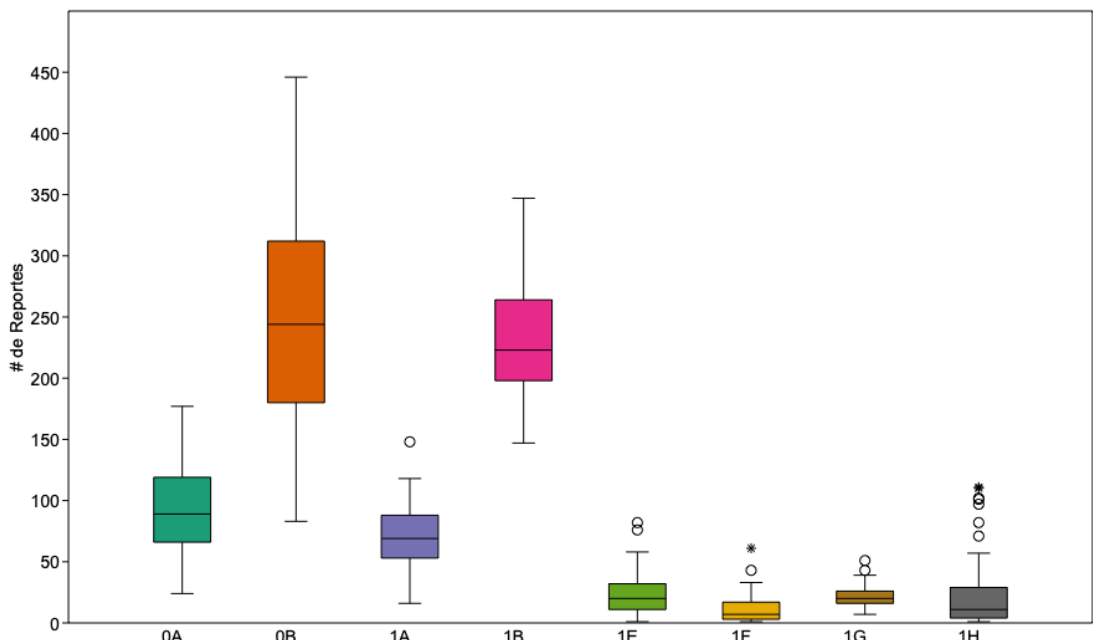


Gráfico 7.- Boxplot general de los códigos de Trabajo de reportes de Servicio.

Si observamos el resumen de datos obtenidos tanto con los datos generales como en el caso de HD2, las tendencias se mantienen tanto en el caso de distribución de los valores así como en su “normalidad”. Encontramos que los valores de mediana y media de la variable 0B son prácticamente los mismo (181 y 180.0 respectivamente). Esto lo podemos observar en la tabla de resumen de los datos. Podemos encontrar así mismo que las reparaciones tanto preventivas como correctivas son las que más se realizan, en comparación con los demás tipos de trabajo. Debemos también tener en cuenta que en muchos casos las reparaciones preventivas (0B) se realizan

junto con los trabajos de mantenimiento (0A y 1A), por lo que se espera que este valor sea alto, por el contrario las actividades preventivas se realizan de manera alterna cada año; es decir, si en una máquina en el 2018 se realizó una visita (0A) en el 2019 se debió haber realizado un mantenimiento preventivo (1A), mientras que en el mismo período a esos reportes se acompañó un reporte de reparación preventiva (0B), este es el principal motivo por el cual el valor de este tipo de trabajos tiene un valor alto.

Si analizamos el porcentaje de valores faltantes (N/A) vemos que las actividades que se encuentran fuera de las de mantenimiento y reparaciones tienen valores faltantes, unas más que otras, a excepción de los reportes 1G. Esto es lo esperado en estos casos puesto que no todos los meses se realizan actividades de instalación de equipos nuevos (1E) ó usados (1F) y menos aún los trabajos de modificación de un equipo (1H). Los reportes 1G nos indican que tenemos muchos casos en los que el cliente reporta un daño, pero el técnico no es capaz de reproducir el error o en su defecto el error fue provocado por el usuario (error de manejo). Podemos ver que este tipo de reportes se producen con bastante frecuencia, hasta el punto de no tener valores vacíos en el rango de datos en el cual estamos trabajando. Desde el punto de vista de la empresa, esta variable se podría estudiar más a fondo para determinar si debemos capacitar al usuario final para mejorar el manejo de la misma y mejorar la forma como se manejan las distintas alarmas que se pueden presentar durante el tratamiento.

Si comparamos las medias, los cuartiles y demás valores, en los dos casos la tendencia se mantiene; por lo que, para fines de estudio vamos a basarnos solamente en los valores obtenidos de la tabla general de datos.

Tabla 2.- Resumen de valores de las variables de trabajos realizados en las máquinas con categoría HF.

	0A	0B	1A	1B	1E	1F	1G	1H
Min	8.0	57.0	9.0	83.0	1.0	1.0	5.0	1.0
1st Qu	44.8	113.2	36.8	115.5	8.0	3.0	9.0	4.0
Median	64.0	181.0	53.0	141.5	13.0	6.0	11.5	13.0
Mean	65.3	180.8	51.9	145.7	17.0	9.6	13.6	25.9
3rd Qu	83.3	238.5	65.0	174.2	22.0	12.0	16.3	32.5
Max	132.0	320.0	114.0	226.0	81.0	60.0	42.0	106.0
N/A	0.0	0.0	0.0	0.0	3.0	18.0	0.0	33.0

A pesar de que la información que encontramos en los Boxplot y la tabla de resumen de datos son muy similares, se pueden usar cualquiera de las dos opciones para realizar el estudio exploratorio inicial de los datos. En el primer caso podremos visualizar los valores si lo que nos interesa es saber los valores exactos de la distribución, y el gráfico para tener una visión más general de los datos y su distribución, pero en este caso los indicadores del datos faltantes no se pueden obtener, dato que puede ser de mucha utilidad para justificar el descarte de variables que no aporten significativamente al análisis de los datos.

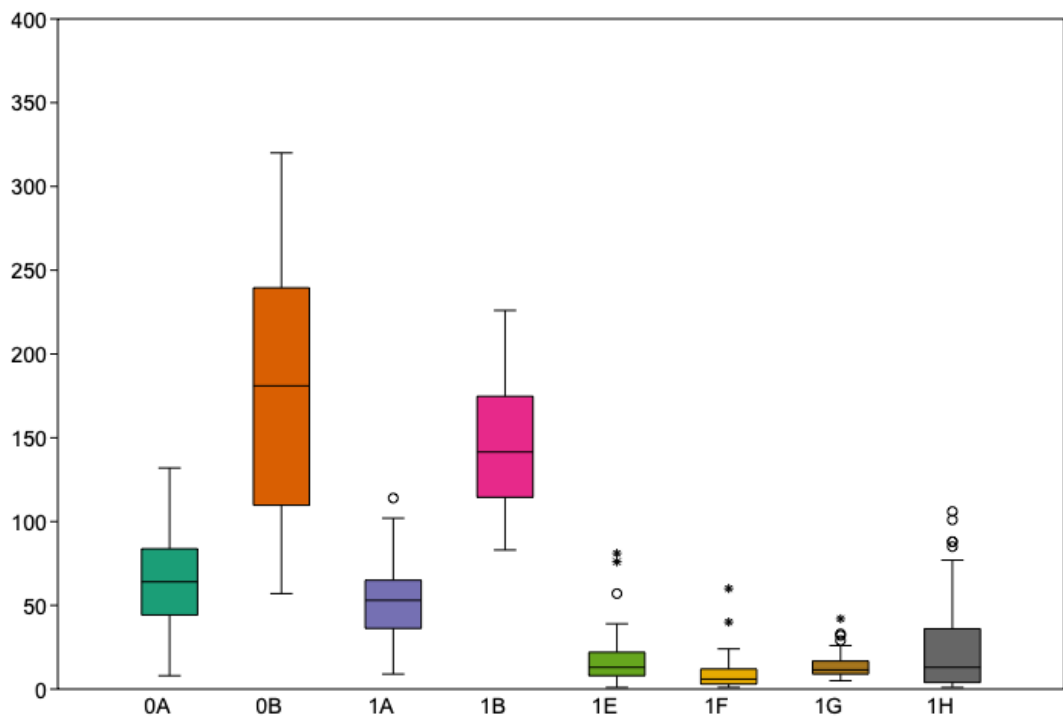


Gráfico 8.- Boxplot de los reportes de servicio generados por tipo de trabajo en la tecnología HD2.

Por otro lado si vemos el mismo gráfico desde el punto de vista de los repuestos que fueron utilizados durante el mismo período de tiempo vemos que un repuesto sobre sale sobre todos los demás. El mismo sabemos que ha provocado problemas debido a que tenemos reportes de ciertos clientes de cambios antes del tiempo de duración. Lo que provoca que no se ha podido predecir a ciencia cierta el consumo promedio de este repuesto, el cual se lo está cambiando en el mismo período de las Visitas o los Mantenimiento Preventivos, pero no se encuentra coincidencia entre las tablas que demuestren el consumo a la par de los trabajos realizados.

Incluso en el gráfico se aprecia que el repuesto 6751421 tiene una diferencia notoria entre los cuartiles, además muestra “outliers” superiores que claramente muestran que en el tiempo el repuesto ha presentado problemas de consumo excesivo y muchas veces aún más de lo esperado, así como un claro sesgo en sus datos.

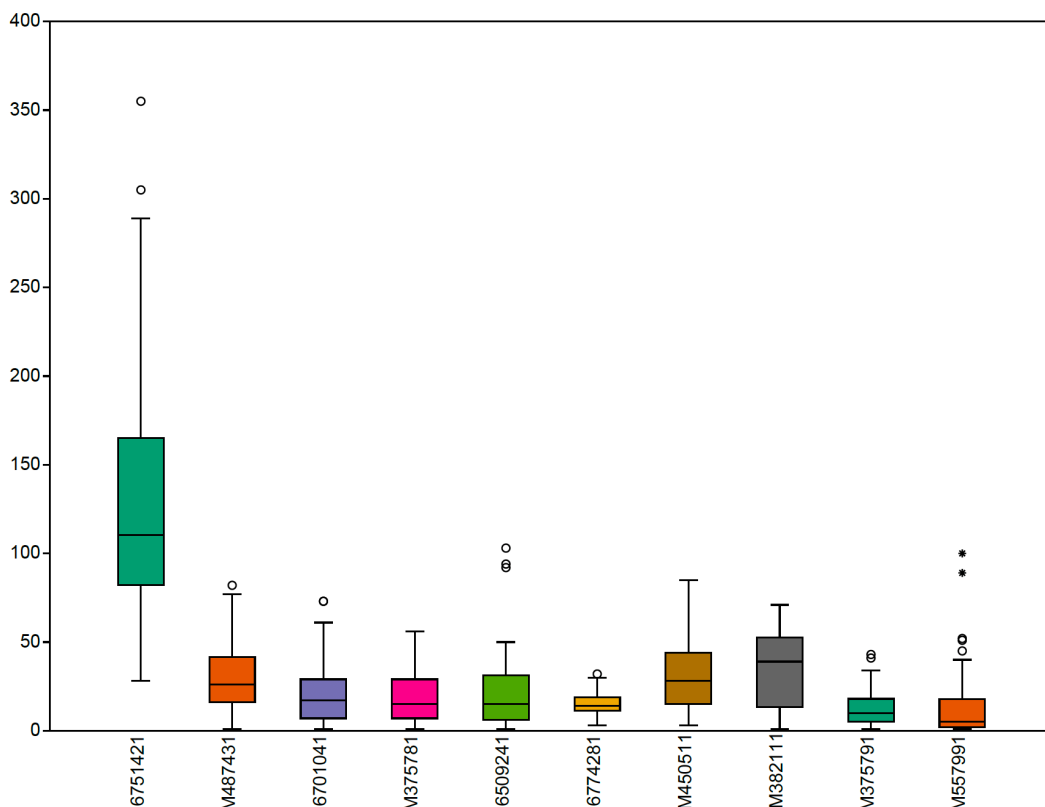


Gráfico 9.- Boxplot de Repuestos usados durante el período en Análisis.

4.2. Series de tiempo de los trabajos realizados

Si analizamos el comportamiento de los trabajos de mantenimiento y reparación tanto preventivos como correctivos podemos ver que no hay una periodicidad clara en los mismo. Se puede ver que los trabajos de mantenimiento preventivo (1A) y visita (0A), tratan de mantener cierta linealidad a través del tiempo, esto se explica ya que debido a que estas son actividades planificadas y los técnicos tratan de distribuir las de manera regular para que la carga mensual sea relativamente igual. Es por esto que notamos que los meses de noviembre y diciembre de todos los años estos valores descienden y nuevamente ascienden en enero del año siguiente. Incluso encontramos que durante los meses de marzo y abril y mayo de 2020 hay una caída de los trabajos de mantenimiento y esto expresa claramente el

confinamiento que sufrimos por el inicio de la pandemia (COVID-19). Y en los meses siguientes en cambio notamos un pico que de muestra que los técnicos trataron de realizar los trabajos que estaban atrasado y por ende igualar la planificación anual de cada uno de ellos. No se ve esta baja en los trabajos de reparación debido a que estos trabajos no dejaron de realizarse en ningún momento, se puede ver un descenso, pero no en la misma medida que los trabajos realizados. De igual manera los trabajos de reparación preventiva al realizarse junto con las visitas y mantenimientos preventivos, estos trabajos se duplican en cantidad a las actividades anteriores y los picos de caídas y aumento de trabajo es más notorio en este tipo de trabajos.

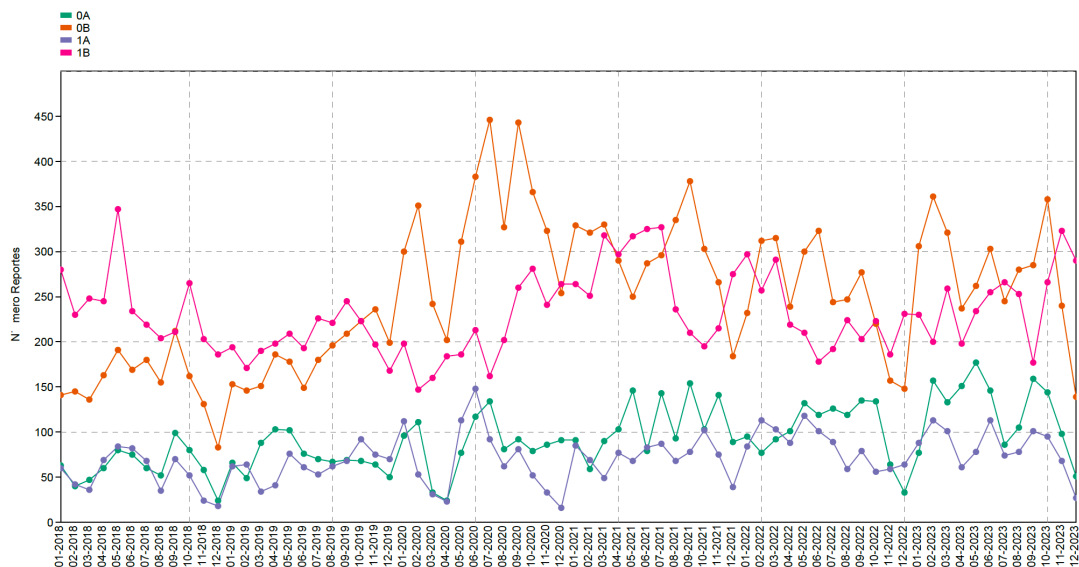


Gráfico 10.- Serie de tiempo de los trabajos realizados en las máquinas.

Una vez que tenemos una visión general del comportamiento de los datos. Vamos a aplicar el modelo ARIMA a cada una de las series entender y predecir la evolución en el tiempo de cada uno de los trabajos de mantenimiento que tenemos presente.

4.3. Representación Gráfica de las series de tiempo

Para cada una de las series de tiempo debemos analizar y entender el comportamiento de cada uno de los trabajos para posteriormente aplicar los pasos para finalmente proceder con el suavizado de las series y analizarlas para saber cual es el mejor método que podemos aplicar para poder llegar a

la predicción del comportamiento de las series de tiempo en los próximos 12 meses.

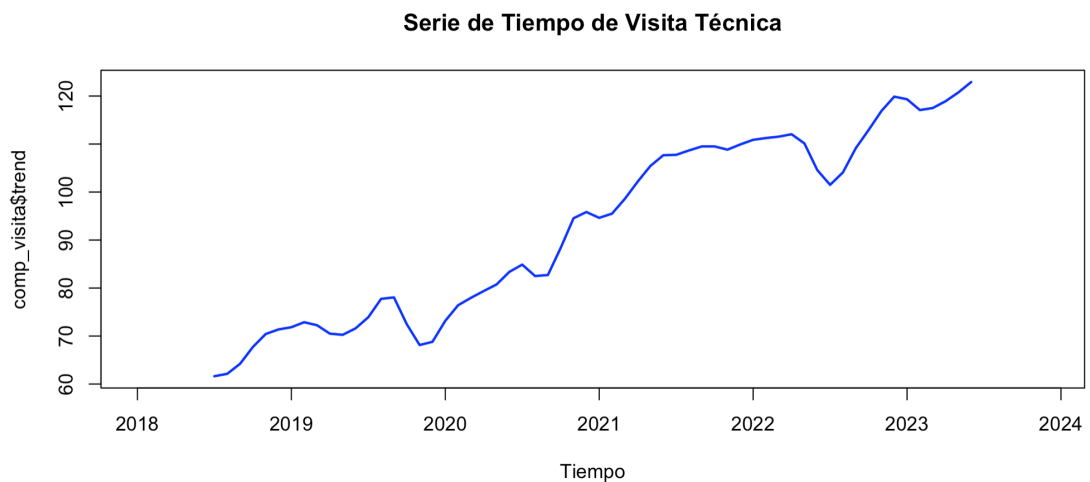


Gráfico 11.- Tendencia de serie de Tiempo de la variable 0A (Visita)

En el caso de los trabajos de Visita podemos observar una tendencia aditiva positiva en la que con el paso de los años en la que la primera caída está a finales del año 2019, de igual manera se puede ver una caída en el segundo semestre del 2020 que puede ser efectos del confinamiento por motivos de la pandemia COVID-19, de la misma manera encontramos la misma caída en el segundo trimestre del año 2022, las demás variaciones en la tendencia del aumento de la cantidad de visitas de deben a vencimientos de los contratos de mantenimientos que se tiene tanto con entidades públicas y privadas las que por lo general tienen equipos propios.

Desde el punto de vista de la estacionalidad podemos ver que la serie se comporta con un ciclo anual de actividades que tiene una caída pronunciada a finales del año y sube en dos pasos para conseguir picos máximos en el primer trimestre del año posterior. Esto se puede explicar debido a que los técnicos planifican este tipo de actividades en un calendario anual y en el mismo se tiene presente que a partir de noviembre se debe bajar la carga de trabajo por cierres contables en las áreas de logística.

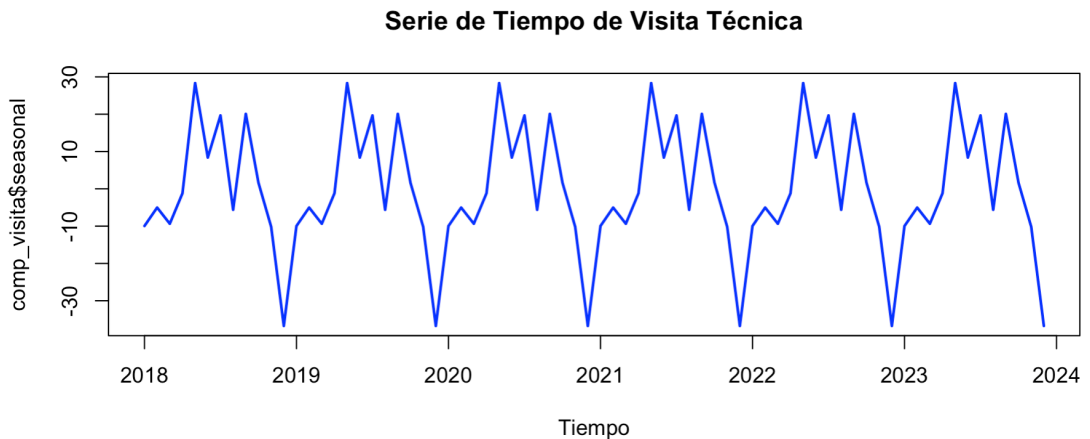


Gráfico 12.- Estacionalidad de la serie de tiempo Visita Técnica.

El siguiente gráfico que se utiliza para analizar la serie de datos son la cantidad de datos que no pueden ser explicados ni por la tendencia ni la estacionalidad a estos los llamamos los residuales. En este caso los residuales no muestran un patrón evidente, ya que, se mueven alrededor del punto cero de manera aleatoria.

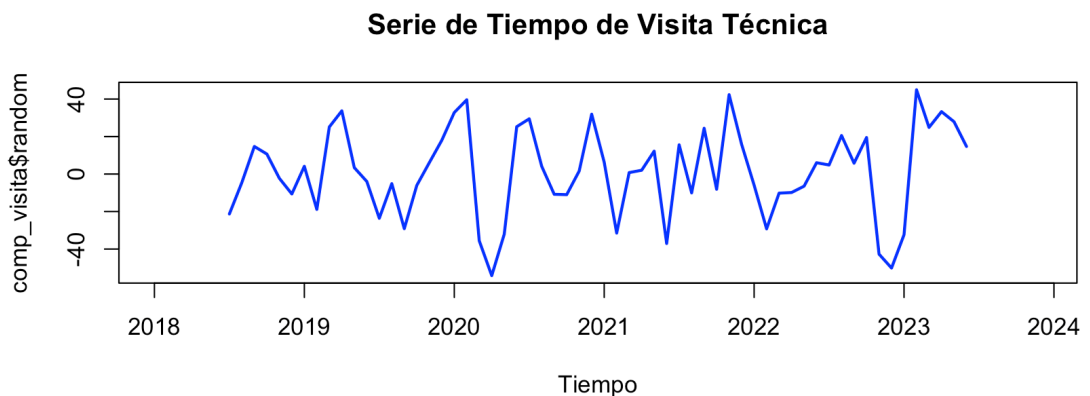


Gráfico 13.- Residuales de la serie de tiempo Visita Técnica.

Finalmente a los residuales debemos realizarles la prueba de “Saphiro-Wilk” para comprobar si los residuales de la serie de tiempo siguen una distribución normal. El Resultado del Análisis nos dará un valor ρ , el cual es un valor de significancia y su valor referente es de 0.05 con lo cual tenemos dos opciones de hipótesis que se componen de

- Hipótesis Nula (H_0): Los residuales siguen una distribución normal.
- Hipótesis Alternativa (H_1): Los residuales no siguen una distribución normal.

En el caso de la serie de tiempo Visita, tenemos que el valor resultante de $\rho = 0.4009$, el cual es mayor al de significancia, lo que nos lleva a la aceptación

del la hipótesis nula indicando que residuales tienen una distribución normal y podemos continuar con el análisis de la serie de tiempo.

Shapiro-Wilk normality test

```
data: residuales_visita  
W = 0.97933, p-value = 0.4009
```

Gráfico 14.- Resultados de Test. Saphiro-Wilk para la serie Visita Técnica

Como punto final procedemos con el análisis de las autocorrelaciones de los residuales, para poder determinar si tenemos patrones o dependencia temporal en el modelo. Las autocorrelaciones deberán encontrarse dentro de los intervalos de confianza que por lo general se grafican con líneas interpuntadas en donde debemos considerar que las autocorrelaciones que se encuentran dentro de los intervalos debe ser mayor a $\pm 1.96/\sqrt{n}$, donde n es el número de observaciones del modelo. Si se cumple esta condición podemos decir que los residuales no tienen dependencia temporal significativa, y por lo tanto es adecuado para continuar con el análisis.

Correlograma de los Residuales

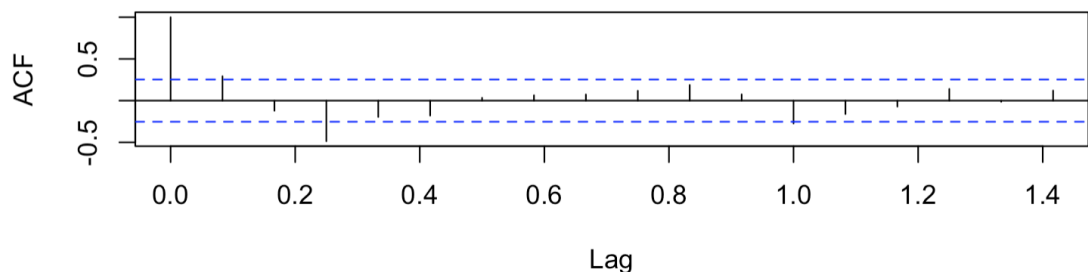


Gráfico 15.- Correlograma de la serie Visita Técnica.

Para la serie de tiempo de los Mantenimientos Preventivos debemos tomar el mismo flujo de análisis gráfico para determinar todas sus variables. En el caso de la tendencia de los Mantenimientos Preventivos tenemos una tendencia aditiva positiva que nos indica que las máquinas bajo responsabilidad de la empresa han ido en aumento desde el 2018, en este caso podemos ver que a finales de los años 2019, 2021 y 2023 tenemos un poco de caída de los mantenimientos y eso se puede explicar por la misma planificación de los técnicos. Ellos debido a cierres contables y de logística tienen previsto terminar su trabajo de Mantenimientos anuales hasta Noviembre. Por el

contrario podemos observar que durante los años 2020 y 2022 en cambio a inicios de cada período observamos un pico de trabajos que nos ayuda a explicar el hecho de que al inicio del año los técnicos empiezan con un trabajo bajo presión que va disminuyendo con el paso de los meses.

Serie de Tiempo de Mantenimiento Preventivo

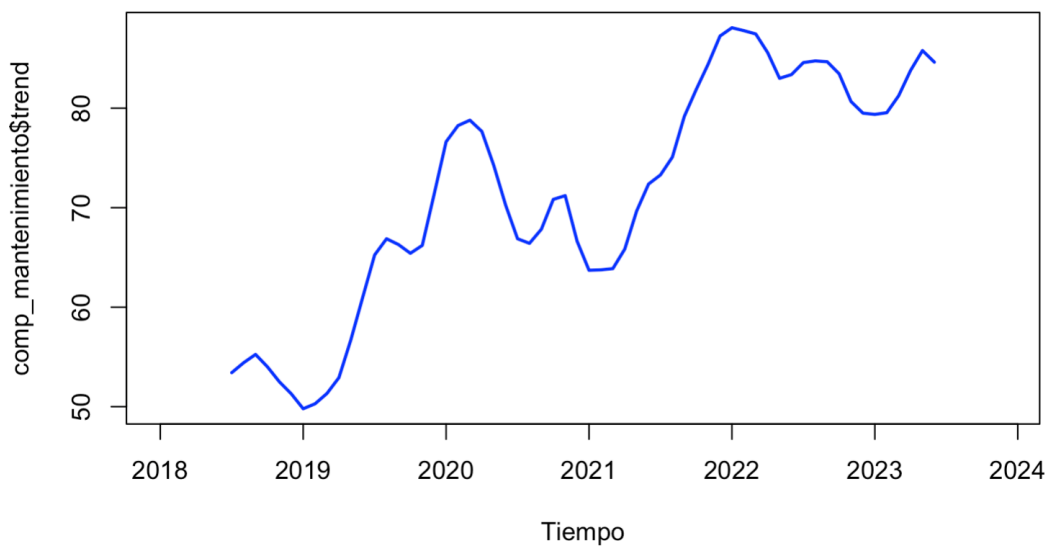


Gráfico 16.- Gráfico de Tendencia de la serie Mantenimiento Preventivo

Desde el punto de vista de la estacionalidad de la variable en el tiempo, podemos observar de una manera más marcada que todos los años el comportamiento de la serie es muy similar, en el que los dos primeros meses del año inicia con un pico de trabajos, para luego caer y subir nuevamente a mediados luego de lo cual cae con un tercer pico de subida y finalmente a en el último mes del año encuentra su pico más bajo. Estos picos de caída se pueden explicar debido las temporadas vacacionales escolares en las regiones en las cuales se encuentran los técnicos. En la costa se tiene como meses de vacación entre Marzo y Abril y para la sierra se marca entre mediados de Junio a mediados de Agosto. Es así que los técnicos apresuran sus actividades antes de salir de vacaciones, en lo que encontramos los picos en los meses previos al período de vacaciones de cada una de las temporadas.

Serie de Tiempo de Mantenimiento Preventivo

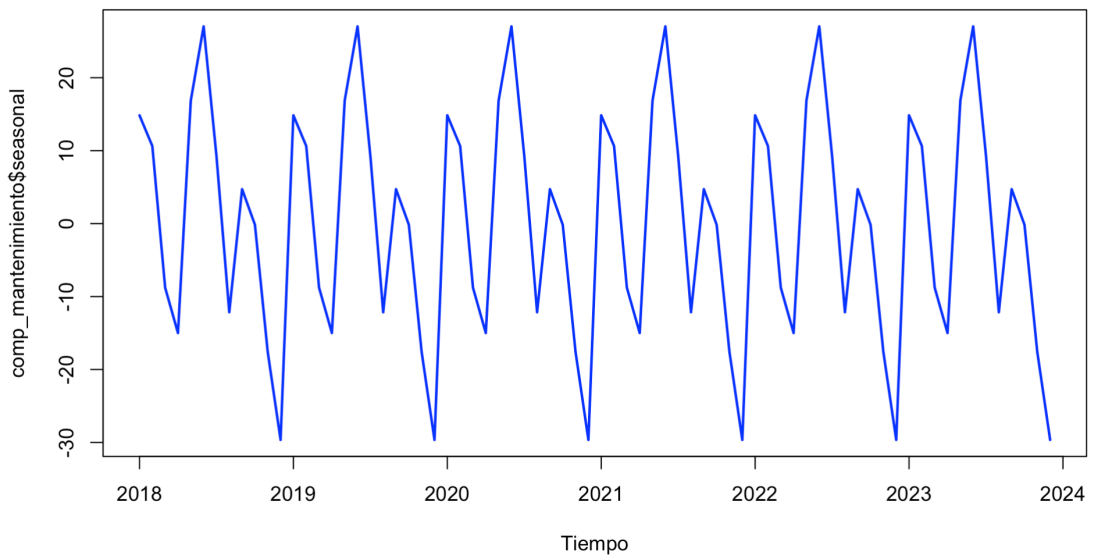


Gráfico 17.- Estacionalidad de la serie Mantenimiento Preventivo

Si observamos en cambio los residuales vemos que solamente durante el año 2020 en los primeros meses hay valores que muestran un evento que afecta a la serie de tiempo y aquí es donde se reflejan los efectos de la pandemia COVID-19, por la cual no se pudieron realizar los trabajos de manera habitual y aunque no marcó el ritmo de trabajo, trae valores que salen de lo que normalmente se encuentra.

Serie de Tiempo de Mantenimiento Preventivo

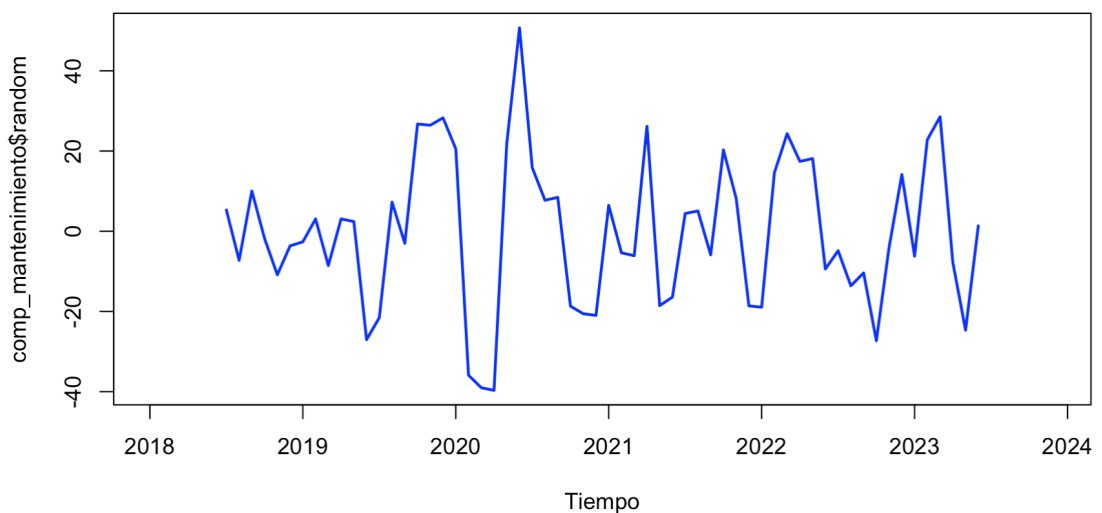


Gráfico 18.- Residuales de la serie Mantenimiento Preventivo.

En el caso de la variable la prueba de Saphiro-Wilk nos muestra que el valor de $\rho = 0.7566$ que es mayor al umbral y por lo tanto aceptamos que los residuales siguen una distribución normal.

Shapiro-Wilk normality test

data: residuales_mantenimiento
W = 0.98668, p-value = 0.7566

Gráfico 19.- Resultados de la prueba Saphiro-Wilk de la serie Mantenimiento Preventivo.

En el caso de las correlaciones encontramos ciertos puntos que se encuentran fuera de rango, pero esto se puede explicar desde el punto de vista de la pandemia y hechos aislados que no afectan a la serie al punto de indicarnos que debamos hacer algún procedimiento adicional para poder trabajar con la serie de tiempo.

Correlograma de los Residuales

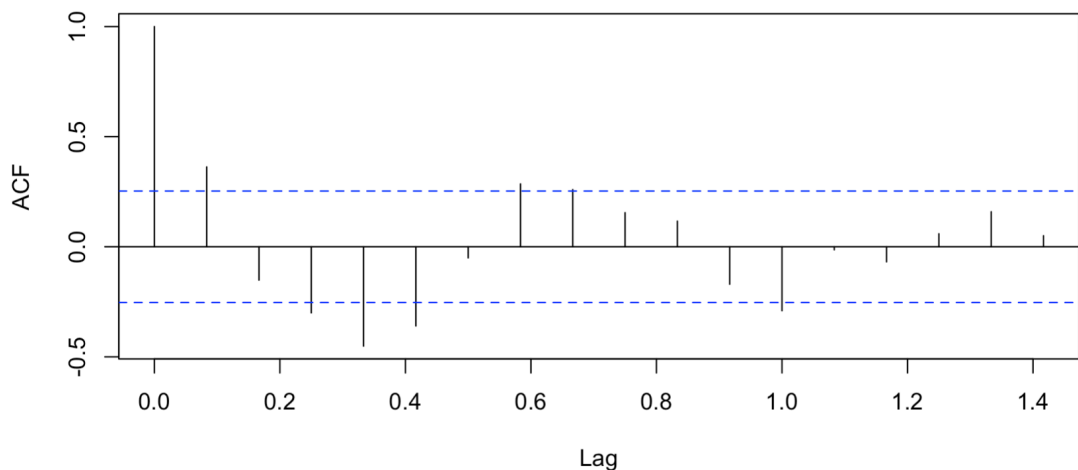


Gráfico 20.- Gráfico de Residuales de la serie Mantenimiento Preventivo

Para el caso de las Reparaciones Correctivas podemos ver que la tendencia es aditiva positiva hasta finales del año 2020 luego del cual tiene una caída estabilizándose a mediados del 2022 y luego de lo cual tiene una tendencia a mantenerse estable. En este período se inicia ciclos de capacitaciones para mejorar la calidad de datos y uso de los códigos de trabajo con los técnicos. Y esto puede ser en parte la caída que tenemos. Esto nos indica que los datos no solo estuvieron influenciados por la cantidad de trabajos que se realizaron sino también en cierto porcentaje por el “ruido” provocado por los mismos técnicos al no realizar un uso correcto de los códigos. A partir del cambio obtenido se implantó en la empresa el tener al menos un ciclo de capacitaciones a los técnicos durante el año.

Serie de Tiempo de Reparación Preventiva

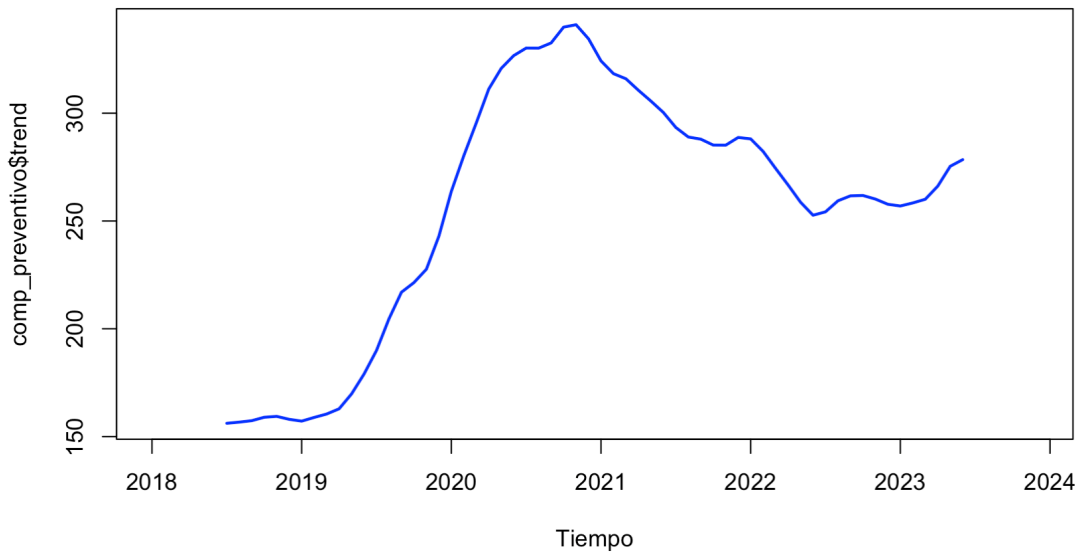


Gráfico 21.- Gráfico de tendencia de la serie Reparación Preventiva.

En cuanto a la estacionalidad de los datos podemos ver que guardan una relación bastante cercana al comportamiento de los mantenimientos preventivos, y esto es en gran parte debido a que los trabajos se realizan en conjunto, debido a que durante un mantenimiento se hace una revisión completa de la máquina, esto ayuda a encontrar problemas (principalmente fugas) internos que hasta ese momento no llegaron a provocar un daño que presente alarmas en la máquina pero que si no se soluciona más adelante puede caer en fallos que provoquen que el equipo deje de funcionar en espera de reparación.

Serie de Tiempo de Reparación Preventiva

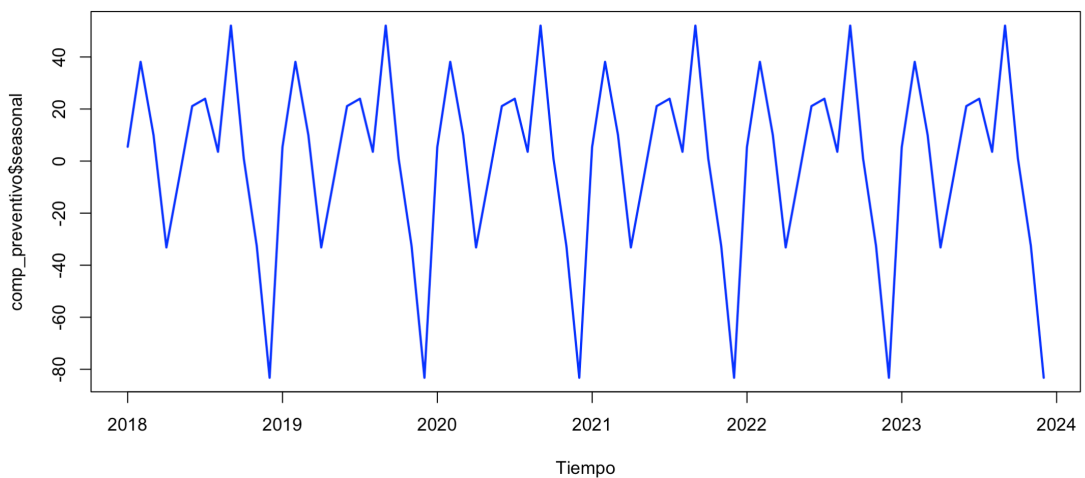


Gráfico 22.- Gráfico de la estacionalidad de la variable Reparación Preventiva

En el gráfico de los residuales podemos observar un comportamiento que se asemeja mucho al de las dos variables previas y esto justamente sustenta el hecho de que son trabajos que se realizan a la par y no muestran un patrón que indique que no se trate de ruido blanco.

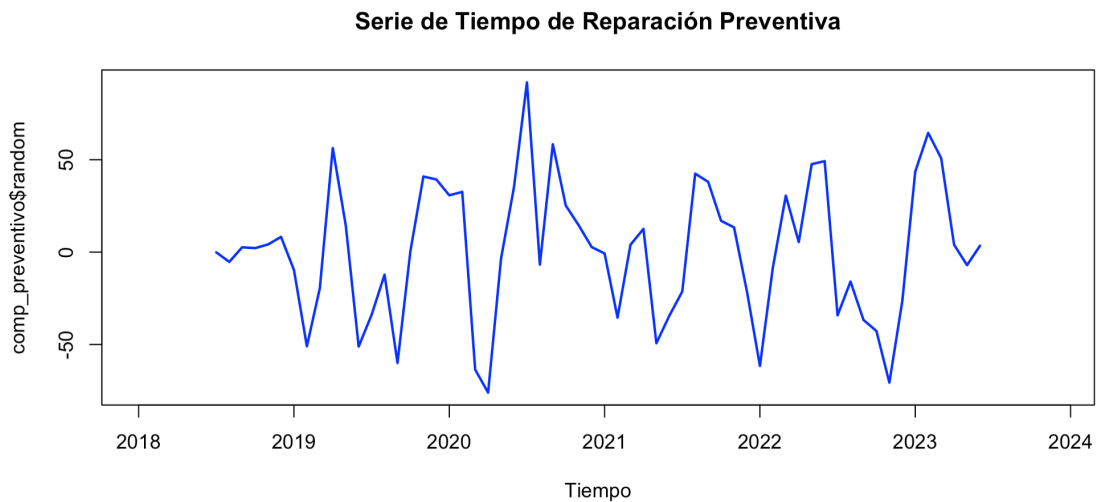


Gráfico 23.- Gráfico de residuales de la variable Reparación Preventiva.

Para poder aseverar lo dicho en el párrafo anterior de que el ruido es un ruido blanco debemos realizar en primer lugar la prueba de Saphiro-Wilk, la cual nos da como resultado un valor $\rho = 0.7185$ el cual es un valor por encima del umbral de 0.05 y nos da paso a la aceptación de la hipótesis nula que indica que los residuos tienen una distribución normal.

Shapiro-Wilk normality test

```
data: residuales_preventivo
W = 0.98593, p-value = 0.7185
```

Gráfico 24.- Resultados prueba Saphiro-Wilk de la variable Reparación Correctiva.

En este caso en el gráfico de correlación de los residuales podemos ver que de continuar con ese problema que presentaba el registro errado de las actividades de los técnicos, la serie no hubiera tenido datos válidos para el estudio, en cambio una vez corregido el problema vemos que los residuos se mantienen dentro del umbral por lo que determinamos que se trata de ruido blanco y el modelo se puede trabajar sin trabajos previos.

Correlograma de los Residuales

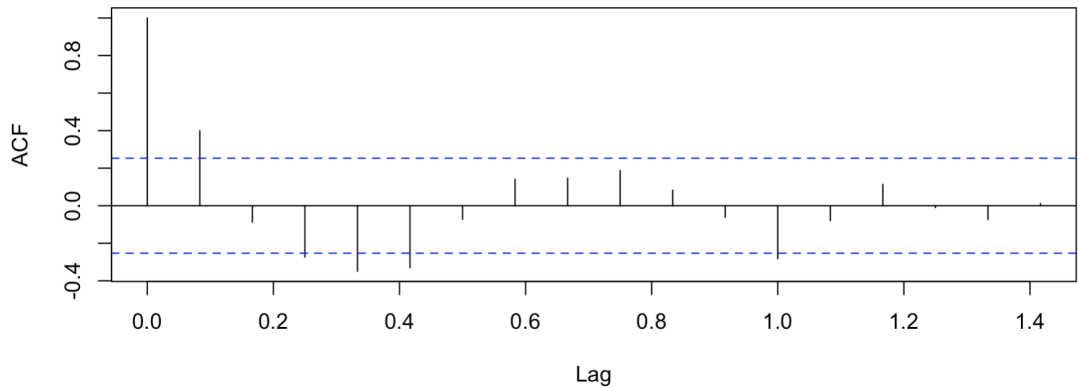


Gráfico 25.- Correlograma de la variable Reparación Preventiva.

A continuación tenemos a nuestra variable Reparación, la cual tiene una tendencia aditiva positiva con un registro de actividades que caen en los primeros meses del año 2021, esto puede explicarse debido al proceso de renovación de tecnología que inició la empresa en ese año. Dicha renovación tomó en cuenta a las máquinas que cumplían ciertas condiciones como año de fabricación, horas de uso, estado general para ser retiradas de funcionamiento (dadas de baja) y renovadas por máquinas nuevas. Esto tuvo un punto de vista contable, debido a que las máquinas estaban depreciadas desde el punto de vista de costos y un punto de vista técnico de mostraba un alto índice de daños y repetición de daños. Esta mejora dura hasta el año 2022, a que en el año 2023 se empieza a notar un nuevo crecimiento en la tendencia.

Serie de Tiempo de Reparación

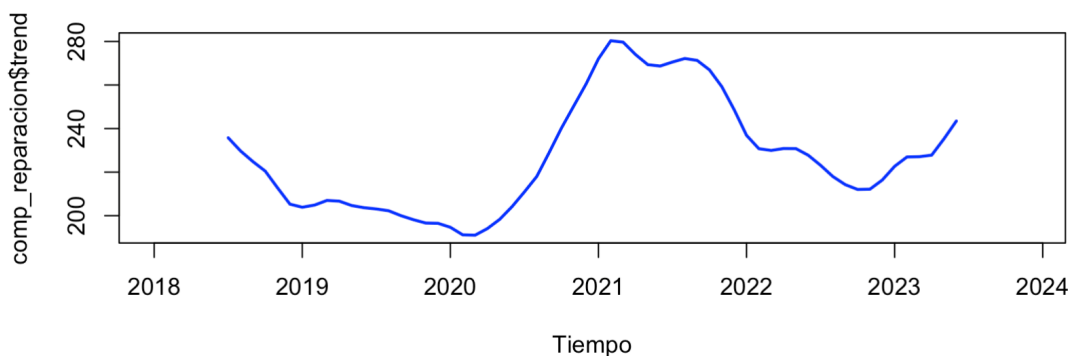


Gráfico 26.- Gráfico de la tendencia de la variable Reparación.

Desde el punto de vista de la estacionalidad del modelo, podemos encontrar de igual manera ciertos eventos que se muestran de manera cíclica que se repiten en la variable, sobre todo a inicios de cada año, esto puede de cierta

manera explicarse que durante el fin de año la parte de logística reduce su trabajo en cuanto a distribución de repuestos por cierres y cuadros de inventarios.

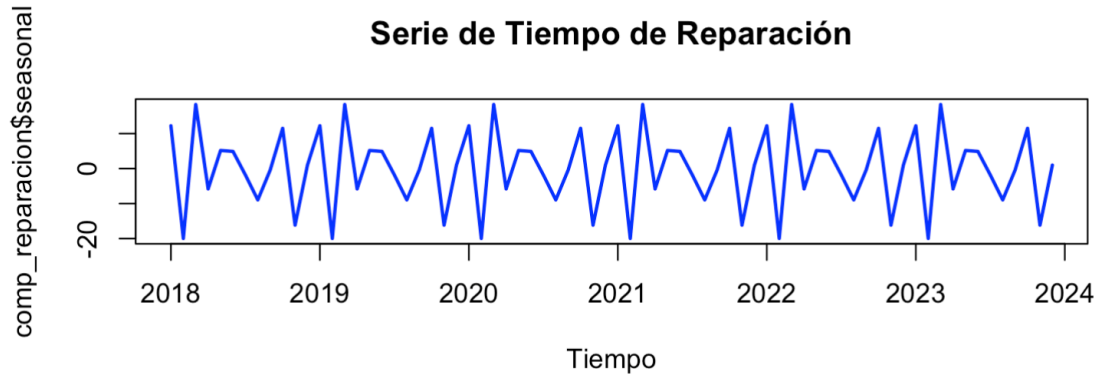


Gráfico 27.- Estacionalidad de la serie de tiempo de Reparaciones.

Desde el punto de vista de los residuales encontramos que la variabilidad de los mismos es menor que los casos anteriores encontrando un solo evento en el segundo semestre del año 2021, el cual principalmente puede estar relacionado con la renovación de máquinas.

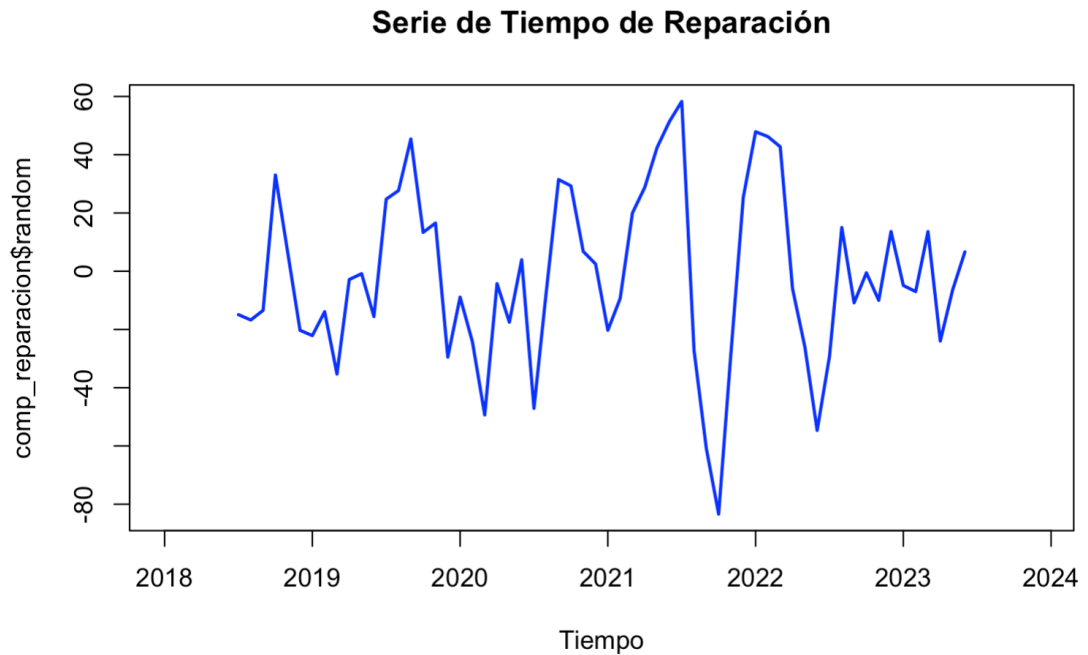


Gráfico 28.- Residuales de la serie Reparación.

No podemos tomar decisiones sobre el tratamiento de la serie solamente con la interpretación de los residuales, por lo que, debemos realizar el test de Saphiro-Wilk, del cual obtuvimos un valor $\rho = 0.5878$, que nos indica que los residuales tienen una distribución normal.

Shapiro-Wilk normality test

data: residuales_correctivo
W = 0.9834, p-value = 0.5878

Gráfico 29.- Resultados de prueba Saphiro-Wilk en variable Reparación.

En cuanto a la correlación que tenemos en los residuales se comporta de manera similar a las Reparaciones Preventivas y al inicio muestra resultados no favorables pero luego se estabiliza quedando dentro del rango de tolerancia para aceptación del modelo de la serie de tiempo, pudiéndose la misma trabajar y continuar con el análisis.

Correlograma de los Residuales

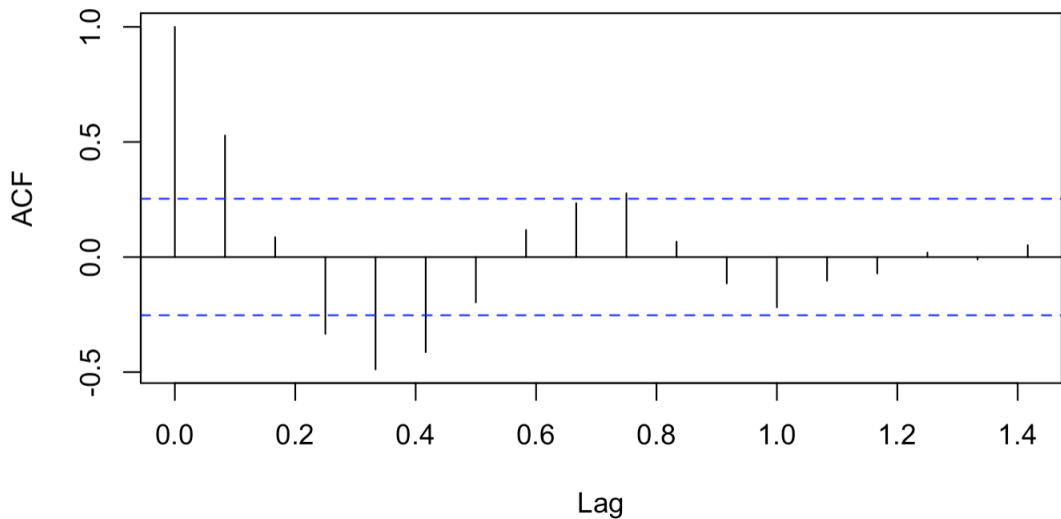


Gráfico 30.- Correlograma de Residuales de la variable Reparación.

En la serie de tiempo del repuesto 6751421 podemos ver claramente una serie aditiva positiva, que desde el 2018 hasta el 2022, muestra una tendencia aditiva positiva, pero a partir de ese año tiene una baja importante que puede estar relacionada con el reemplazo de tecnología en el mercado y la mejora también en la forma como se esta cambiando el repuesto así como la frecuencia con la que se pide, entrega y cambio en caso de ser necesario. Todo esto como indica la teoría, se está reportando como reparación preventiva.

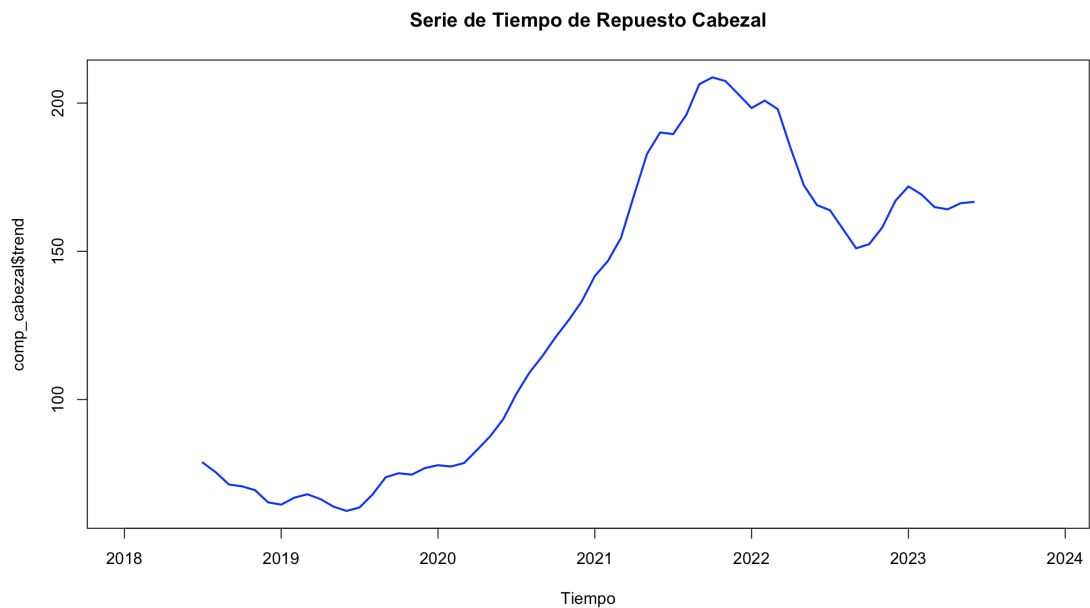


Gráfico 31.- Serie de tiempo de la variable 6751421 repuesto.

Desde el punto de vista de la estacionalidad de la serie de tiempo, podemos notar un patrón que se repite de manera anual, pero tiene demasiada variabilidad y a la final esto puede no ser del todo bueno al intentar acoplar nuestros datos a un modelo en específico.

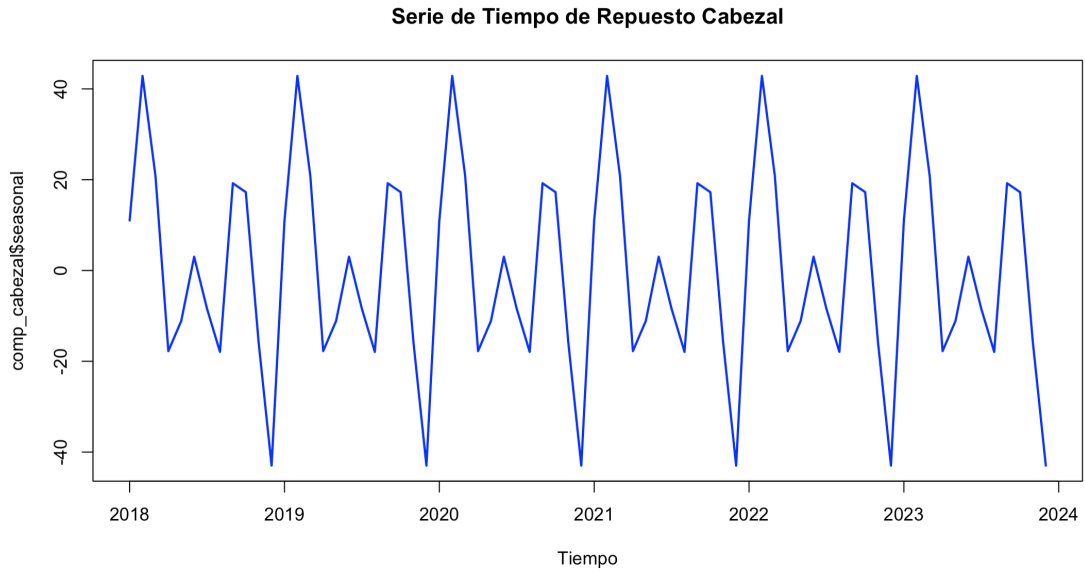


Gráfico 32.- Estacionalidad de la serie de tiempo Repuesto.

Desde el punto de vista de los residuales podemos seguir encontrando varios picos que demuestran que el alto consumo de ester repuesto no es hecho cíclico, sino que hay factores externos que están afectando de manera directa a su consumo.

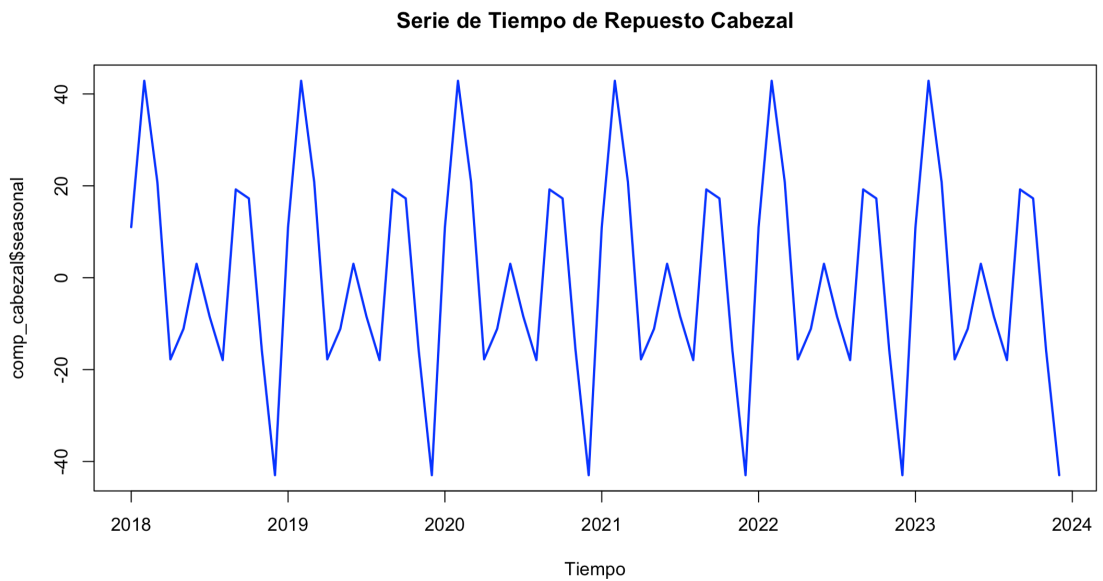


Gráfico 33.- Residuales de series de variable Repuesto.

Al realizar la prueba Saphiro-Wilk para tener una primera impresión de los resultados del modelado podemos encontrar que nos da resultados bastante alentadores puesto estamos dentro de un rango bastante aceptable.

Shapiro-Wilk normality test

```
data: residuales_cabecal  
W = 0.96869, p-value = 0.1259
```

Gráfico 34.- Resultados de la prueba Saphiro-Wilk.

Dado este resultado con un $p > 0.05$ dando como aceptada la no estacionalidad de la serie. Realizamos nuestro correlograma de residuales y no tenemos valores que digan o muestren que los datos disponibles, se traten de ruido blanco o de algún otro parámetro que pueda afectar a nuestro modelado y predicción de valores.

Correlograma de los Residuales

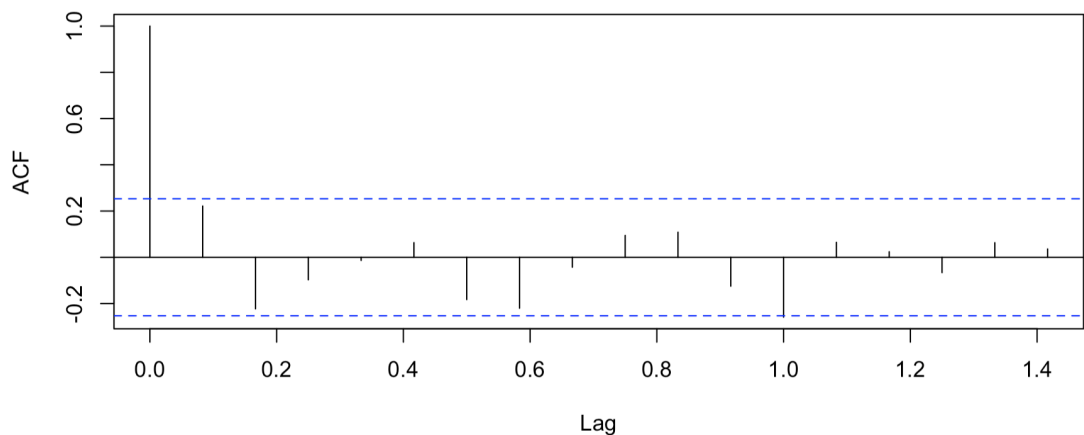


Gráfico 35.- Correlograma de la variable repuesto.

4.3.1. Suavizado de las series de tiempo

Una vez realizado el análisis de la serie de tiempo podemos continuar con el tratamiento de las mismas, para ello el siguiente paso es el suavizado de la serie con el fin de reducir el ruido y resaltar las tendencias en los datos. El suavizado de las series de tiempo se puede hacer por varios métodos entre los que podemos mencionar, el de medias móviles, el exponencial, y el suavizado con splines. En nuestro caso vamos a utilizar el suavizado de medias móviles. El cual se calcula un promedio de una ventana de tiempo deslizante. Vamos a trabajar con las medias móviles centradas para poder representar la tendencia de los datos. Este tipo de medias móviles se

caracterizan por utilizar un número impar de de observaciones, en el caso aplicado utilizaremos 3 observaciones para realizar el suavizado. Es así que una media móvil centrada en un tiempo t de longitud $2(n + 1)$, en este caso la formula que vamos a utilizar es:

$$\hat{y}_t = \frac{y_{t-1} + y_t + y_{t+1}}{3}$$

Siendo en este caso t el tiempo en el que se está suavizando y 3 el número de puntos que utilizamos para calcular el promedio.

De este modo aplicamos el método de suavizado de medias móviles a cada una de las variables de nuestra serie de tiempo.

Si realizamos la comparación de las gráficas de la serie de tiempo inicial con el modelo suavizado podemos notar que no perdemos la tendencia de la serie pero si eliminamos los picos que pueden producirnos ruido al momento de aplicar el modelo para la predicción.

Serie de Tiempo de Visita Técnica

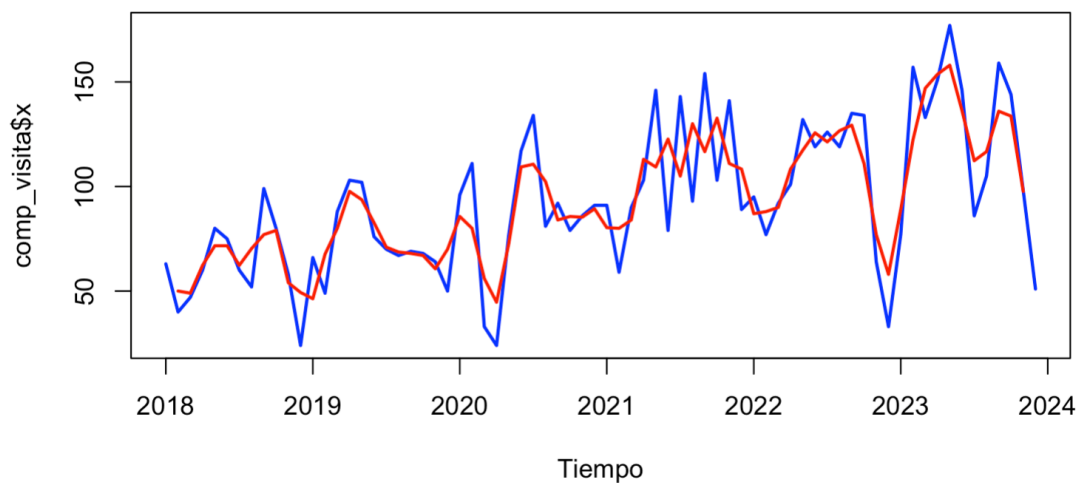


Gráfico 36.- Suavizado de la serie de tiempo de la variable Visita Técnica.

De la misma manera procedemos con el suavizado de la serie de la variable Mantenimiento, de igual manera reducimos la varianza y mejoramos la tendencia, debemos recordar que al aplicar este método estamos conservando la tendencia de los datos.

Serie de Tiempo de Mantenimiento Preventivo

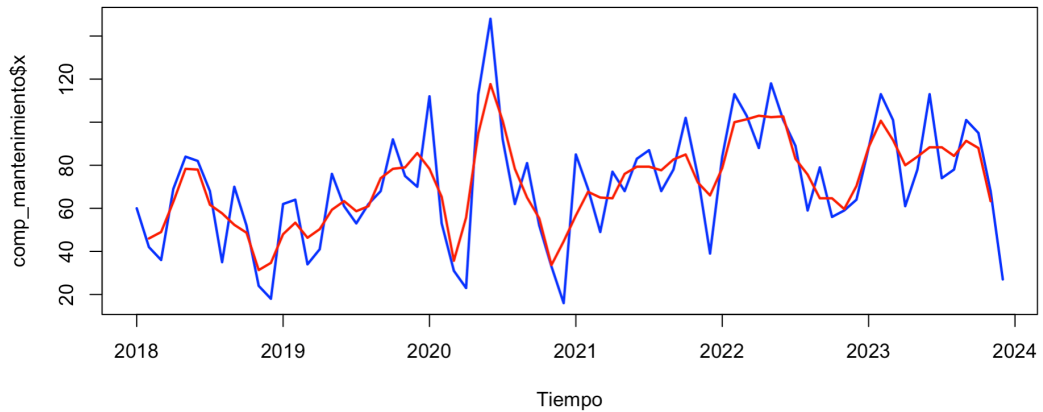


Gráfico 37.- Suavizado de la serie de tiempo de la variable Mantenimiento Preventivo.

Al aplicar el suavizado a la serie de Reparación Preventiva vemos que mantiene una alta varianza en los puntos más extremos de la serie, por lo que podríamos aumentar el factor de suavizado de 3 a 5, lo que implicaría que el lugar de tomar un valor antes y después del punto a suavizar tomaría dos valores, esto ayuda a mejorar la curva pero en cambio podríamos perder información. Esta es la razón por la cual nos mantenemos con un factor de 3.

Serie de Tiempo de Reparación Preventiva

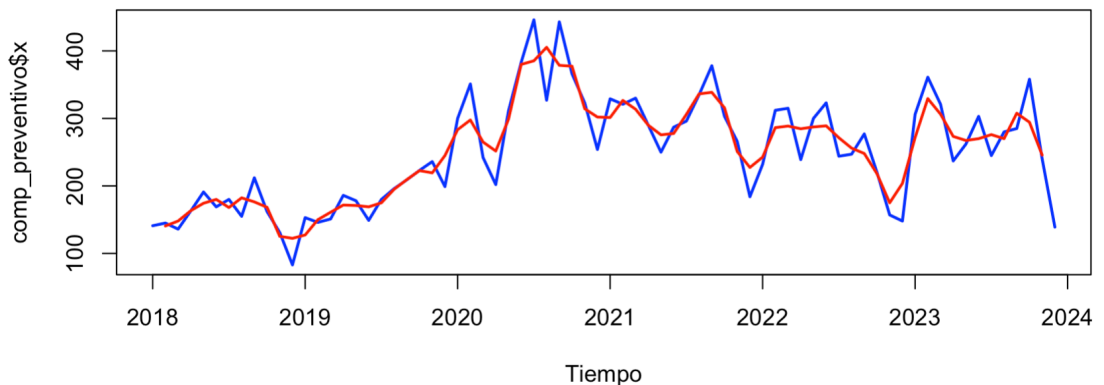


Gráfico 38.- Suavizado de la serie de tiempo de la variable Reparación Preventiva.

En el caso de las Reparaciones Correctivas encontramos que tenemos varios “picos” en el transcurso del tiempo y el suavizado se adapta muy bien a la tendencia que marcan los datos convirtiendo a la serie más manejable para el modelo que vayamos a aplicar.

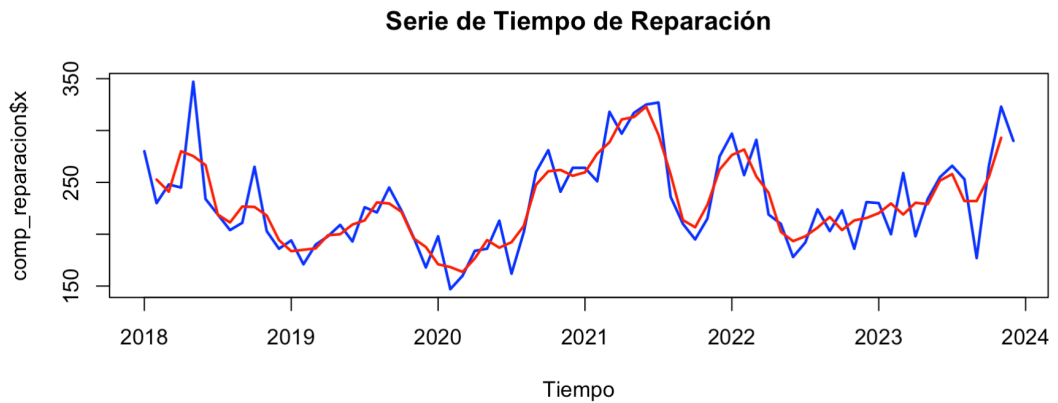


Gráfico 39.- Suavizado de la serie de tiempo de la variable Reparación Correctiva.

4.3.2. Selección del modelo

Una vez que hemos preparado adecuadamente nuestra serie de tiempo, podemos aplicar el test para validar el modelo que mejor se adapte a la serie de tiempo para poder realizar la predicción de valores del mismo. En este caso vamos a utilizar el paquete de R “forecast”, el cual posee funciones para la modelización de las series de tiempo y por ende funciones que nos ayuden a encontrar el modelo adecuado y la predicción de cada una de las series.

El modelado que vamos a aplicar en este caso es el ARIMA (Autoregressive Integrated Moving Average) para series temporales. La Función en R va a realizar la selección automática del modelo y parámetros óptimos para cada una de las series. Basándose en varios criterios de información y pruebas para asegurar que el modelo se ajuste correctamente.

Antes debemos recalcar que el ARIMA es una combinación de tres componentes que son:

- AR (Autoregressive), se trata de la parte del modelo que usa las dependencias entre una observación y un número de rezagos o lags.
- I (Integrated), es la parte de la serie temporal que hace que la serie sea estacionaria.
- MA (Movil Average), Esta parte de la función se encarga de modelar los errores de la serie como una combinación lineal de los errores pasados.

Además en el modelado ARIMA vamos a obtener unos coeficientes denominados p,d,q que en el modelo resultante vamos a verlo como ARIMA(p,d,q), donde:

- p, número de términos autoregresivos.

- d, número de diferencias estacionales necesarias para hacer que la serie se convierta en estacionaria.
- q, número de términos de la media móvil.

Dados estos antecedentes procedemos con la aplicación del modelado a nuestras series de tiempo.

En primer lugar realizamos el modelado de la variable Visita Técnica de la cual obtenemos los siguientes resultados

```
Series: smooth_visita
ARIMA(0,1,0)

sigma^2 = 250.7: log likelihood = -288.49
AIC=578.99  AICc=579.05  BIC=581.22

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.6816667 15.72012 12.03405 2.71186 14.27494 0.5597231
              ACF1
Training set 0.2179107
```

Gráfico 40.- Resultados de modelado ARIMA a variable Visita Técnica

En este caso podemos ver que los coeficientes del modelo son (0,1,0), lo que nos indica en el orden de los coeficientes los siguientes parámetros, en el caso de los términos autoregresivos es igual a 0, el modelo tuvo que realizar una sola diferencia para hacer la serie estacionaria y el término de media móvil también es 0. Estos parámetros no ayudan a un buen resultado, a lo que debemos adicionar la varianza del error residual que es alto (250.7). Y los valores de AIC, AIC_c y BIC son demasiado altos los mismos que indican un ajuste malo. Si procediéramos con la predicción de este modelo vamos a tener un mal resultado, por todos los resultados previos. Para prueba vamos a proceder con la gráfica de la predicción para 12 meses.

Forecasts from ARIMA(0,1,0)

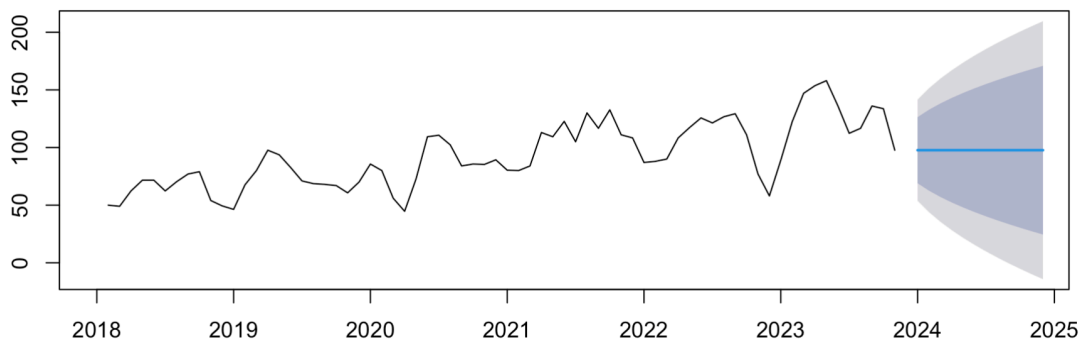


Gráfico 41.- Gráfica de predicción de la serie de tiempo de la variable Visita Técnica.

Esto nos demuestra que al revisar los valores resumidos resultantes del modelo son una gran fuente de información que indican un buen ajuste o no del mismo y por ende una correcta previsión de datos. En este caso al no ajustarse el modelo de manera correcta vemos que la predicción sale sin valores (línea continua), esto puede deberse a varios factores entre los que se encuentran un sobre ajuste del modelo al aplicar el suavizado de medias móviles. En este Caso vamos a proceder cambiando el método de suavizado por el de logaritmos y realizaremos un nueva análisis.

Al realizar el suavizado por medias móviles obtenemos una nueva serie temporal suavizada que da el siguiente modelo.

Serie de Tiempo de Visita Técnica

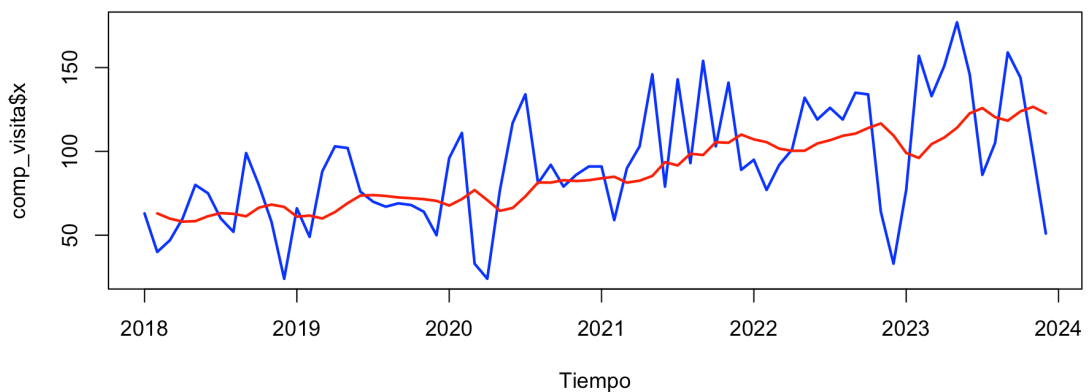


Gráfico 42.- Suavizado exponencial de la variable Visita Técnica.

El método de suavizado mejora la calidad de los datos, mantenimiento de tanto la tendencia como la estacionalidad. A continuación procedemos con la aplicación del modelo y predicción de la serie.

```
Series: exp_smooth_values[, "xhat"]
ARIMA(0,1,3) with drift

Coefficients:
      ma1      ma2      ma3      drift
      0.2757 -0.1880 -0.6877  0.9066
s.e.  0.1009  0.1251  0.0972  0.1676

sigma^2 = 10.72: log likelihood = -181.3
AIC=372.6  AICc=373.53  BIC=383.84

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE
Training set -0.09011336  3.157092  2.392441 -0.3555817  2.808473
              MASE      ACF1
Training set  0.2028133  0.05025716
```

Gráfico 43.- Modelado ARIMA de la serie de tiempo suavizada.

En este caso la primer gran diferencia con el modelado anterior es que el valor q del modelo es 3 que indica 3 términos de medias móviles y por ende ahora

contamos con coeficientes “ma” que son los términos de media móvil, así mismo el valor de σ^2 que es la varianza del error residual es un número pequeño, a pesar de que los valores de AIC, AIC_c, BIC son grandes tenemos un buen ajuste. Ahora procedemos con la gráfica de la predicción.

Forecasts from ARIMA(0,1,3) with drift

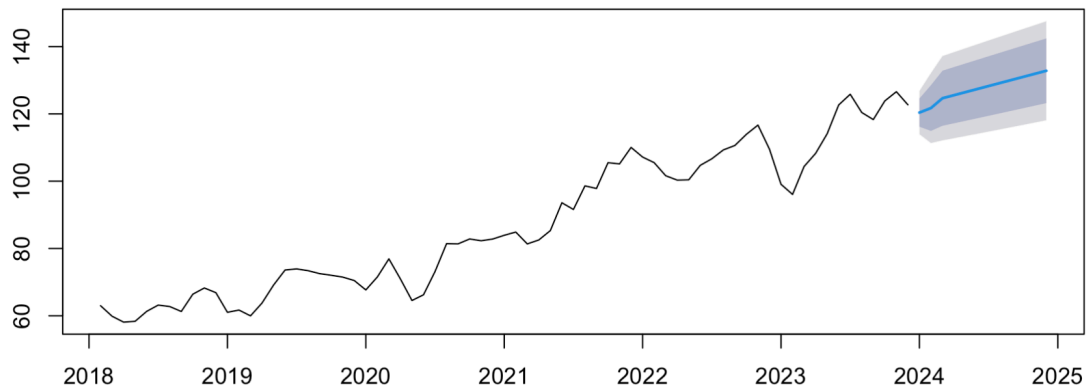


Gráfico 44.- Gráfico de predicción de la serie de tiempo de la variable Visita Técnica.

En este caso tenemos una tendencia a aditiva para el año 2024 con una tolerancia aceptable para la serie trabajada.

Para el caso de las series de Mantenimiento Preventivo el caso es distinto ya que al aplicar el modelado ARIMA a la serie con suavizado de medias móviles tuvimos los siguientes resultados.

```
Series: smooth_mantenimiento
ARIMA(0,1,1)(1,0,0)[12]

Coefficients:
      ma1      sar1
      0.4644  0.2554
s.e.  0.0856  0.1211

sigma^2 = 114.4:  log likelihood = -260.93
AIC=527.86  AICc=528.23  BIC=534.56

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.04453799 10.46302  8.27099  2.344703 13.30069  0.4102486
ACF1
Training set 0.0734438
```

Gráfico 45.- Modelado ARIMA a serie de Mantenimiento Preventivo.

En este caso obtuvimos dos casos de ajuste para el modelo ARIMA en el que el mejor resultados tenemos coeficientes “ma”, el valor σ^2 tiene un buen valor sobre los 100, en este caso tenemos que los valores AIC, AIC_c, BIC, son valores altos, lo que indica un ajuste del modelo no tan bueno como el que desearíamos.

Si revisamos el gráfico de la predicción tenemos que si hay predicción en el modelo los valores de tolerancia tienen un rango demasiado alto.

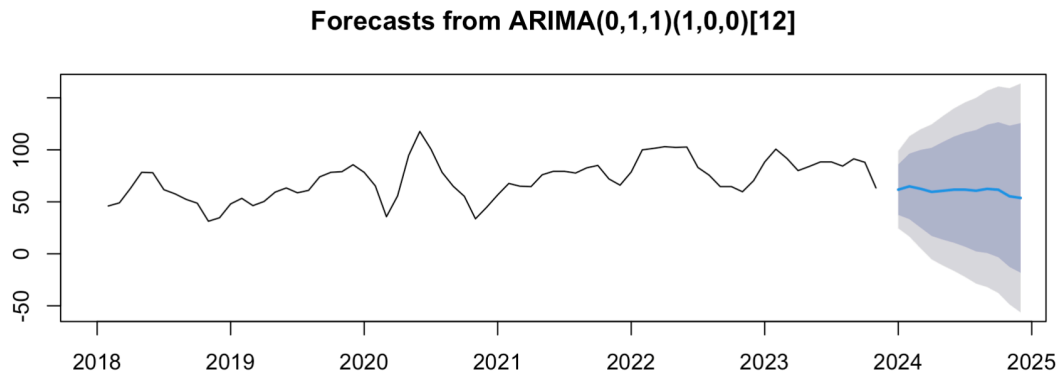


Gráfico 46.- Predicción de valores para la serie de tiempo Mantenimiento Preventivo.

Procedemos con el procedimiento de ajuste para las reparaciones preventivas. Cuyo ajuste nos da los siguientes resultados.

```

Series: smooth_preventivo
ARIMA(2,1,0)(1,0,1)[12]

Coefficients:
      ar1      ar2      sar1      sma1
    0.5488  -0.2917  0.8787  -0.6001
s.e.  0.1169   0.1170  0.1377   0.2463

sigma^2 = 445.9: log likelihood = -309.04
AIC=628.08  AICc=629.04  BIC=639.25

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE      MASE
Training set 0.3193977  20.34749  15.7451  0.8320544  6.419728  0.2495353
              ACF1
Training set -0.08586922

```

Gráfico 47.- Resultado valores modelo ARIMA en serie Reparación Preventiva.

Para este caso encontramos una mejor relación en los coeficientes del modelos en el uno nos muestra estacionalidad y en el otro no, con el coeficiente “d”. En esta serie de tiempo tenemos un mejor modelado ya que tenemos coeficientes de varios tipos (ar1, ar2, sar1, sma1) que son los coeficientes autoregresivos, atuoregresivos estacionales, y media móvil estacional, en conjunto estos coeficientes juegan un papel importante para la predicción ya que capturan de mejor manera las dependencias temporales en los datos. Dando una gráfica de predicción de muy buen ajuste.

Como se puede apreciar en el gráfico la predicción se ajusta a la tendencia que tenemos en en año 2023 sobre todo. Las sombras de ajuste son buenas lo que indica que el modelo se ajustó de muy buena manera a los datos. Dando como resultado que todo el proceso dio un resultado de predicción bastante satisfactorio para nuestro estudio.

Forecasts from ARIMA(2,1,0)(1,0,1)[12]

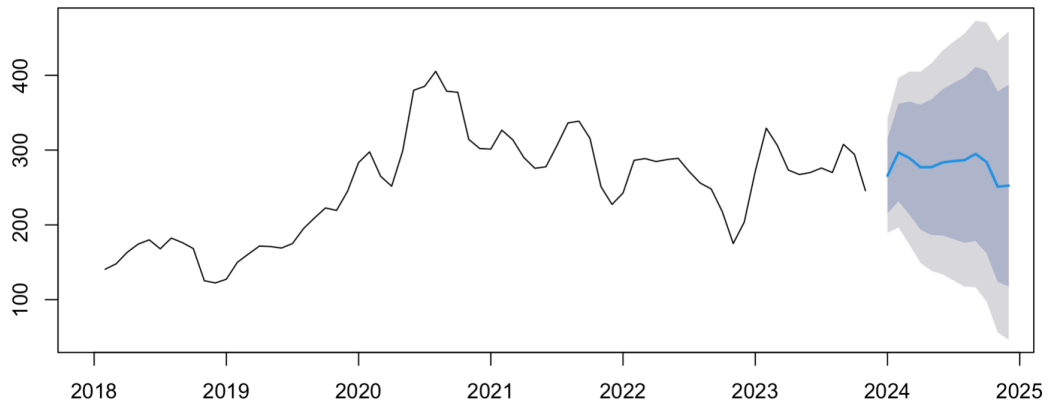


Gráfico 48.- Predicción con modelo ARIMA de la serie Reparación Preventiva.

Tenemos a continuación el análisis de la serie en cuanto a Reparaciones.

```

Series: smooth_correctivo
ARIMA(2,0,0) with non-zero mean

Coefficients:
      ar1      ar2      mean
      1.3071  -0.4733  234.8287
s.e.   0.1082   0.1081  11.0272

sigma^2 = 254.9: log likelihood = -292.75
AIC=593.49  AICc=594.11  BIC=602.49

Training set error measures:
              ME      RMSE      MAE      MPE      MAPE
Training set -0.1986742  15.61913  12.07959  -0.1644928  5.228839
              MASE      ACF1
Training set 0.2519901  0.01756554
    
```

Gráfico 49.- Resultado de modelado ARIMA en serie Reparacion.

En este caso podemos observar que el modelo arima me da un indicativo de no estacionalidad y no medias pero si contiene términos autorregresivos lo que deriva en valores de coeficientes ar1, ar2 y mean que contienen información valiosa para la construcción del modelo. Los demás valores como σ^2 , AIC, AIC_C, BIC y los valores de entrenamiento del modelo indican un buen ajuste, por lo que procedemos con la gráfica de la predicción resultante. La gráfica de ajuste en este caso presenta un modelo aditivo negativo para luego estabilizarse y mantenerse en el transcurso del tiempo. Los intervalos de confianza de la predicción por lo general siempre se encuentran en el 80% y 95% en todos los modelos realizados, por lo que podemos considerarlo un estandar que manejamos a nivel de programación. Y se interpretan como la incertidumbre que se asocia a las predicciones realizadas.

Forecasts from ARIMA(2,0,0) with non-zero mean

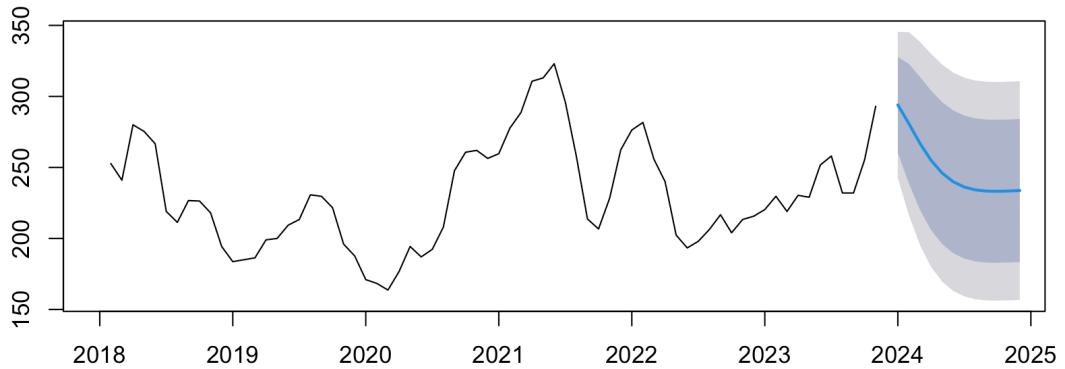


Gráfico 50.- Predicción de la serie de tiempo Reparación con modelado ARIMA.

Dados los resultados procedemos con el suavizado y comparación de la serie resultando con la serie original. En la misma podemos ver que la serie resultante disminuye considerablemente los picos y variaciones, viendo que a pesar de que se aplica la técnica de medias móviles ayuda de muy buena manera a mantener la tendencia de la serie.

Serie de Tiempo de Cabezal

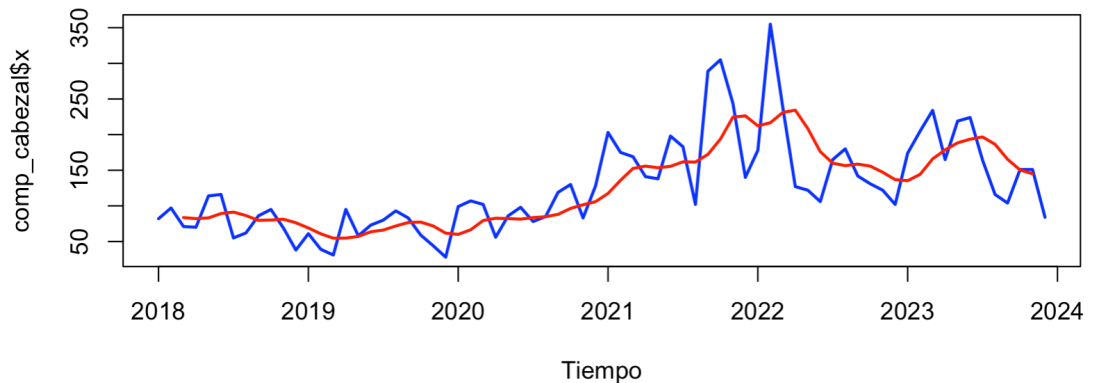


Gráfico 51.- Suavizado de la serie de tiempo Repuesto.

Los resultados de la prueba de ajuste del modelo si bien es cierto no da los mejores resultados que ya no me da más de 3 coeficientes y parece estar bastante bien adaptado a los datos

```

Series: exp_smooth_values_cabezal[, "xhat"]
ARIMA(2,1,0)

Coefficients:
      ar1      ar2
    0.3077  -0.4097
s.e.  0.1082   0.1067

sigma^2 = 245.4; log likelihood = -291.12
AIC=-588.24  AICc=-588.6  BIC=-594.98

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE
Training set 1.002307 15.33055 11.12715 0.1595148 9.202815 0.2522878
ACF1
Training set 0.02743848
    
```

Gráfico 52.- Resultado de ajuste del modelado de la serie Repuesto.

A pesar de intentar con distintos procesos de suavizado para intentar ajustar nuestros datos al modelo, no se logra obtener unos resultados de predicción que muestren una variación en el resultado, esto no muestra más sino que el modelo por mejor que muestre buenos resultados en los test de ajuste. Al final al momento de intentar resultados de predicción esto no es posible, viendo que en este caso el modelado y ajuste para predicción no se ajusta de manera correcta al modelo. Incluso probando con varios tipos de suavizado no se logra obtener una predicción. La varianza es excesiva y vemos que tiene muchas variables adicionales que terminan en este tipo de modelos. En este caso el estudio univariante no nos es de ayuda y podemos encontrar las diferencias del modelo Multivariante, en el cual a pesar de estos contra tiempos podemos utilizarlo para analizarlo y tener una mejor interpretabilidad de los datos.

Forecasts from ARIMA(2,1,0)

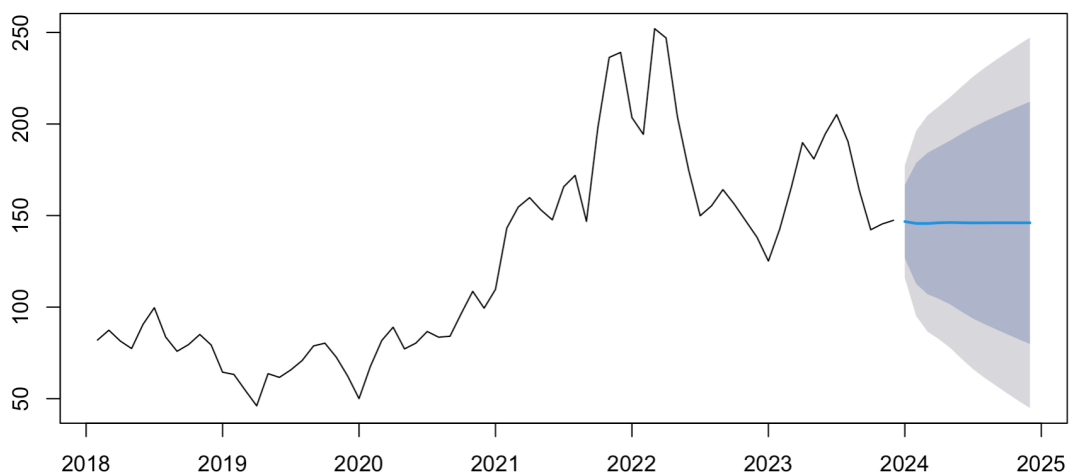


Gráfico 53.- Predcción de la serie de tiempo Repuesto.

4.4. Análisis Multivariante

Todo el análisis que hemos realizado de las series de tiempo hasta el momento es desde el punto de vista Univariante; es decir, considerando exclusivamente a cada serie sola, como la responsable de toda la variabilidad de los datos obtenidos.

Pero debemos tener en cuenta que en general nada se produce como un hecho aislado, desde el punto de vista de los equipos médicos debemos verlos que el resultado de daños, pueden ser la consecuencia de varios factores en los que se incluyen los mantenimientos y otros trabajos adicionales. Es por esto, que podemos utilizar este tipo de análisis para poder obtener otro punto de vista de como nuestras variables se interrelacionan.

4.4.1. Análisis de Componentes Principales (PCA)

El análisis de componentes principales busca variables hipotéticas (componentes) para explicar la mayor parte de la varianza de los datos multivariantes. Estas nuevas variables son combinaciones lineales de las variables originales. El análisis de componentes principales puede ser usado para la reducción de variables al mínimo (dos componentes) para fines de representación gráfica.

El PCA encuentra los valores y vectores propios de la matriz de covarianza o de correlaciones, para lo cual las variables deben ser medidas en las mismas unidades. Para este caso no vamos a encontrar las variables más significativas sino que el programa se encarga de evaluar todas las variables y a partir de ellos obtenemos los componentes principales. Como se muestra en la tabla a continuación.

Tabla 3.- Valores obtenidos del Análisis PCA de la base de datos.

PC	Eigenvalue	% variance
1	7228.16	62.016
2	2060.05	17.675
3	989.089	8.4861
4	563.64	4.8359
5	435.849	3.7395
6	208.258	1.7868
7	116.796	1.0021
8	53.5566	0.4595

Como se puede ver en la tabla con los tres primeros componentes de nuestra tabla logramos ya cubrir el 88% de la varianza de los datos, si vamos aumentando la cantidad de componentes elegidas aumenta la cantidad de variabilidad explicada pero el aumento de variabilidad es en pasos cada vez más pequeños, por el contrario si decimos que con 3 componentes explicamos ya el 88% de la variabilidad, con dos componentes tenemos el 80% de explicación de la variabilidad y una sola componente cubre el 62% de la variabilidad. Es decir que queda a elección de quien es responsable del estudio elegir la cantidad de componentes que desea usar que puede ser entre 2 ó 3 componentes que van a explicar la mayor parte de toda la variabilidad de los datos.

Una gráfica que se utiliza para poder determinar el número recomendado de componentes que debemos usar para el análisis de los datos es el gráfico simple de los valores propios el cual indica el número de componentes significantes. En esta gráfica se aplica la llamada “regla del codo” que lo que hace es encontrar el punto de quiebre de los valores propios en el cual se produce un cambio brusco de los mismos. Este es el punto que indica el número de variables que debemos elegir para nuestro estudio.

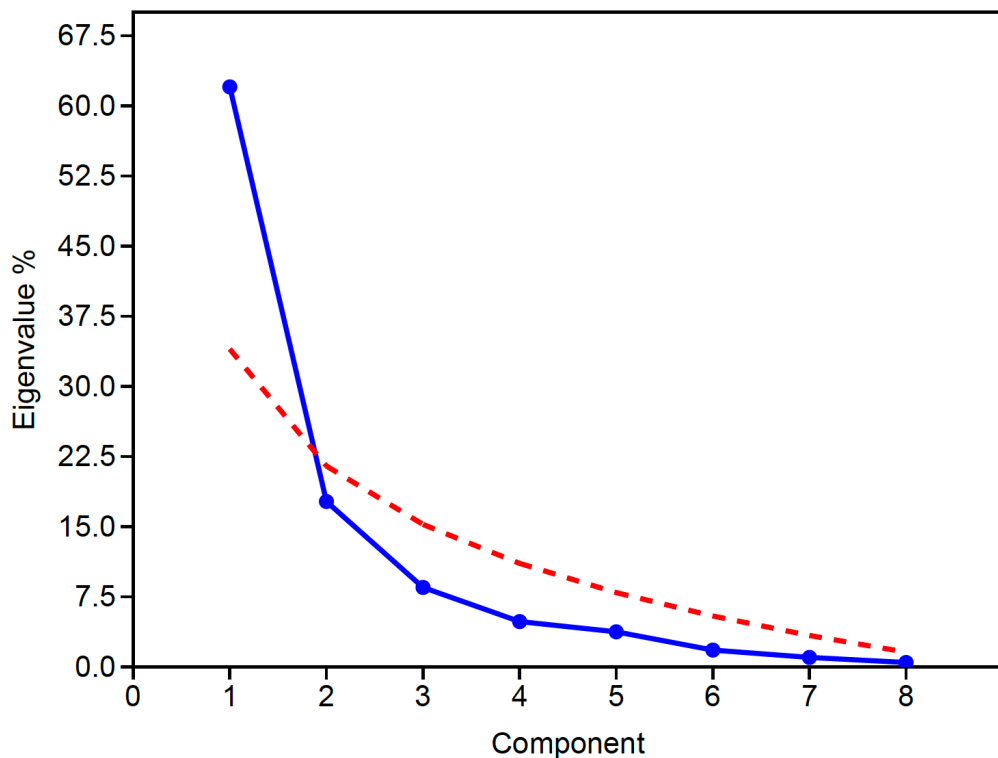


Gráfico 54.- Gráfico del codo para determinación de las componentes principales.

Para nuestro caso el punto de quiebre está en la segunda componente, por lo que este gráfico nos demuestra que al usar dos componentes estamos cubriendo satisfactoriamente la mayor parte de variabilidad de los datos. Ahora también podemos analizar como contribuye cada variable para las varianzas de las componentes, esto se puede observar con la gráfica de los pesos de los componentes, para el caso de la componente 1 tenemos que, está compuesta por la varianza de casi la totalidad de las variables pero principalmente por los trabajos 0B, 0A y 1A.

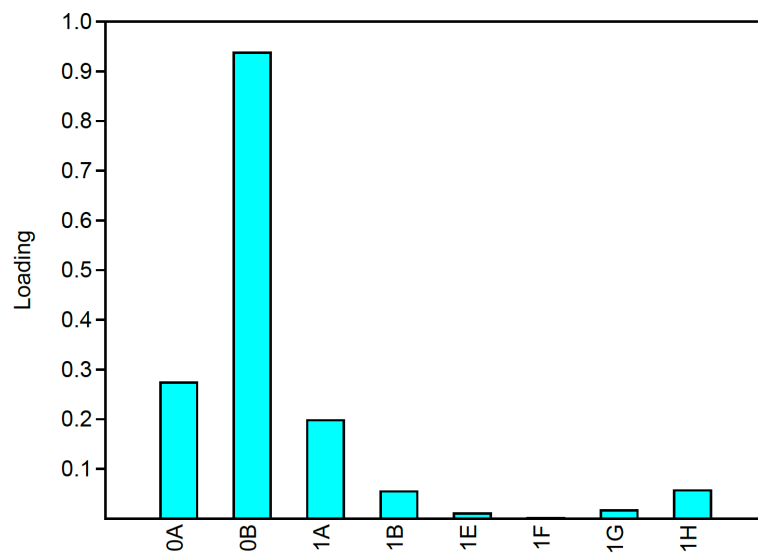


Gráfico 55.- Pesos de la componente PC1.

De la misma manera podemos observar como se encuentran los pesos de PC2. En este caso podemos observar que el principal aportante de la componente es la variable 1B y las demás variables aportan de manera muy poco significativa

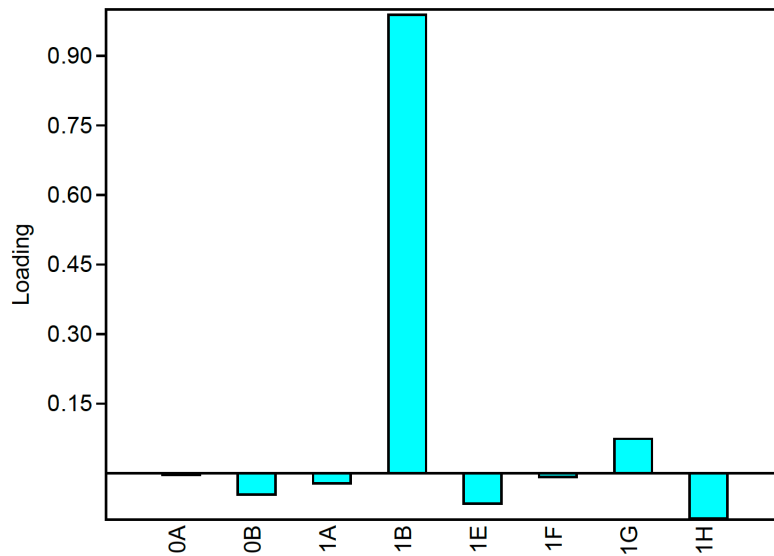


Gráfico 56.- Pesos de la componente PC2.

Vamos a realizar la gráfica de dispersión de los datos para las dos primeras componentes dando el siguiente resultado.

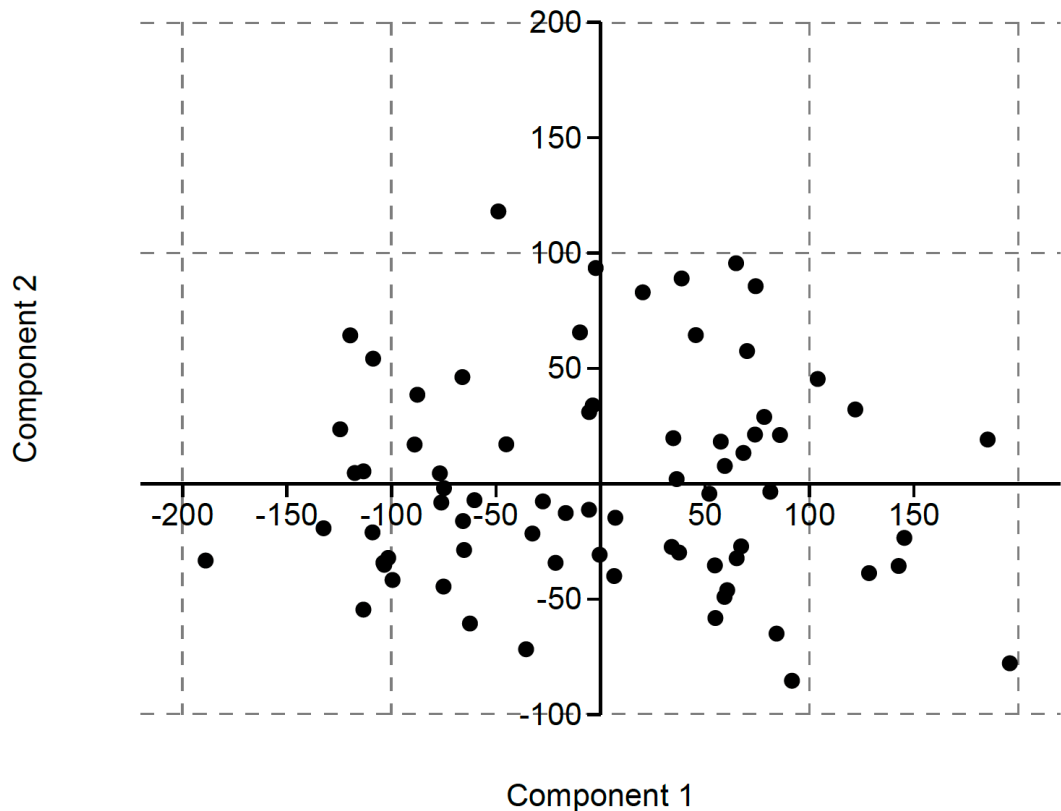


Gráfico 57.- Gráfico de dispersión de las dos primeras componentes.

Debemos tener en cuenta que la variabilidad explicada esta prácticamente en 80%. En cada uno de los cuadrantes de nuestro gráfico se encuentran puntos de las componentes, esto se debe al peso de cada componente tiene sobre cada punto que se representa.

Con todos los datos obtenidos hasta el momento podemos proceder con la creación del Biplot que es una combinación de dos gráficos, el de dispersión que representa el valor de cada una de las observaciones de la base de datos con el de vectores que representa a las variables.

En este caso podemos ver que el mayor peso se encuentran en las variables "1B" y "0B" y no se encuentran correlacionadas entre ellas, puesto que el ángulo que se forma entre las dos variables es de prácticamente 90 grados, lo que indica que no hay influencia de la una sobre la otra y son las que más información entregan a la base de datos por la distancia que tiene el vector al centro del eje.

Así mismo vemos que las variables "0A" y "1A" que hemos venido analizando desde el punto de vista Univariado, se encuentran correlacionadas y están muy cercanas la una de la otra y a su vez se encuentran en la misma dirección de la variable "0B" pero en menor proporción. Esto demuestra desde el punto de vista gráfico que como hemos explicado a lo largo de todo el análisis, estos trabajos se hacen en conjunto y por consiguiente están correlacionados, los trabajos de mantenimiento se hacen de manera alterna cada año repitiéndose el ciclo cada 2 años y eso también explica el motivo por el cual la longitud del vector "0B" es aproximadamente el doble de los vectores "0A" y "1A". Por los demás valores vemos que los trabajos que se podrían hacer a la par o de manera conjunta igualmente se muestran altamente correlacionados como son los trabajos "1E" y "1H".

Por otro lado si tomamos en cuenta los puntos de las observaciones que se encuentran distribuidos en todos los cuadrantes podemos obtener la proyección de las observaciones sobre los vectores que indica que tan bien representada se encuentra esa observación por el vector. Basados en su distancia podemos ver que mientras menor sea su valor, esto nos indica que la misma se encuentra mejor representada por la variable con la cual estamos realizando la medición.

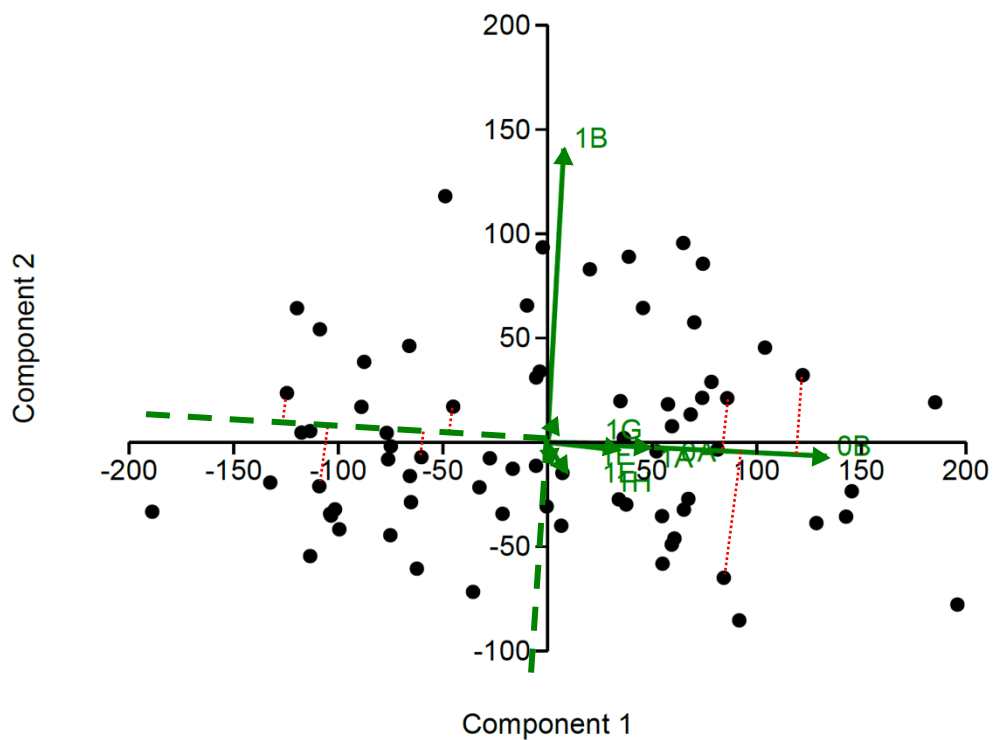


Gráfico 58.- Biplot del Análisis Multivariantes de nuestra base de datos y sus proyecciones

En el caso de intentar agregar una componente más a nuestro análisis debemos considerar que esto dificulta el análisis debido a que la proyección que vamos a tener de las variables y de cada una de las observaciones ya no se realizan sobre un plano sino sobre un gráfico de 3 dimensiones.

A pesar de ello notamos que se aprecia de una mejor manera la correlación entre variables, se mantiene la prácticamente nula correlación entre la variable “1B” y “0B”. Las variables “1A” y “0A” mantienen cierta correlación entre las dos y se encuentran hacia el vector “0B” esta correlación se debe a los tipos de trabajos que se realizan en conjunto. Por otro lado podemos observar que la “1H” mantiene nula correlación con las variables de reparaciones y cierta correlación con las de mantenimientos, esto debido a que ciertas modificaciones son mandatorias pero se las realiza junto con las visitas planificadas a los equipos. Siendo la variable “1E” la que tiene como vector adyacente a “0B” y esto puede reflejarse en una tasa de fallos tempranos (equipos nuevos) que es la que deberíamos evitar y sugerir seguir estudios más a fondo.

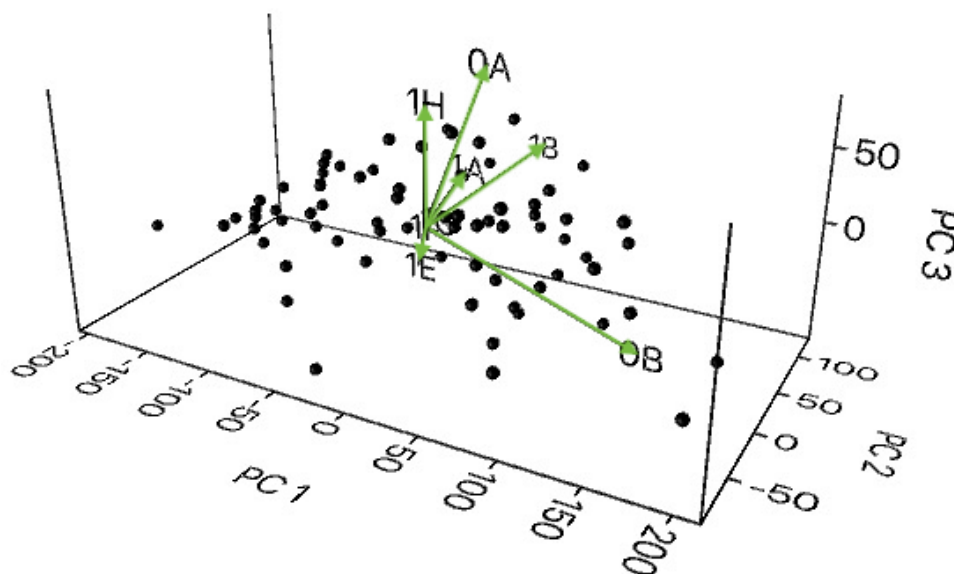


Gráfico 59.- Biplot compuesto por 3 componentes principales.

Para el análisis multivariante de los repuestos, hemos separado los que en cantidad total de consumo a lo largo del período de análisis son los 10 más representativos. En este caso podemos ver que en una sola componente encontramos una explicación del 95% de la variabilidad de los datos.

Tabla 4.- Resultados de componentes principales de repuestos.

PC	Eigenvalue	% variance
1	2.16E+06	95.435
2	44417.9	1.9621
3	36539.6	1.6141
4	9216.63	0.40712
5	5376.56	0.2375
6	4140.04	0.18288
7	2199.68	0.097166
8	1147.27	0.050678
9	317.3	0.014016

A diferencia del análisis de los trabajos realizados podemos ver que en la prueba del codo nos muestra como un repuesto es el que sobresale sobre los 10 elegidos para el análisis, pero para poder realizar el gráfico de dispersión debemos usar el menos 2 componentes, por lo que vamos a usar las dos variables, basándonos principalmente en la componente 1.

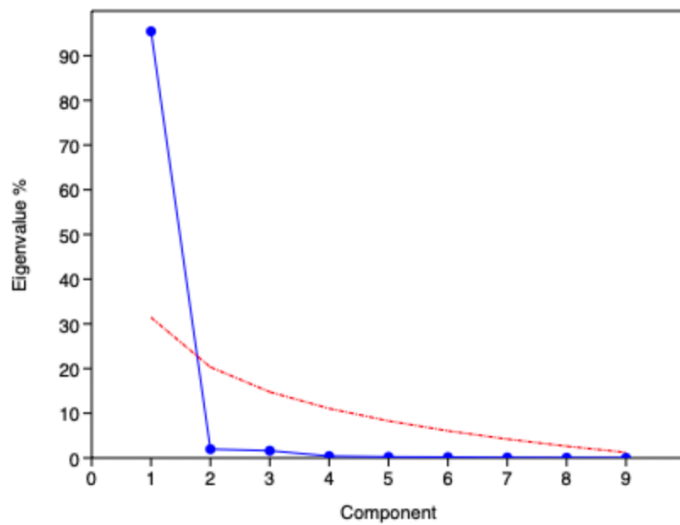


Gráfico 60.- Prueba del codo de la serie Repuesto.

Podemos en este caso revisar el aporte de las variables de la primera componente, donde la variable 6751421 es la de mayor peso, aunque la componente tiene información de la variabilidad del total de las variables incluidas en el estudio. Dicho repuesto es el que sobresale sobre los demás.

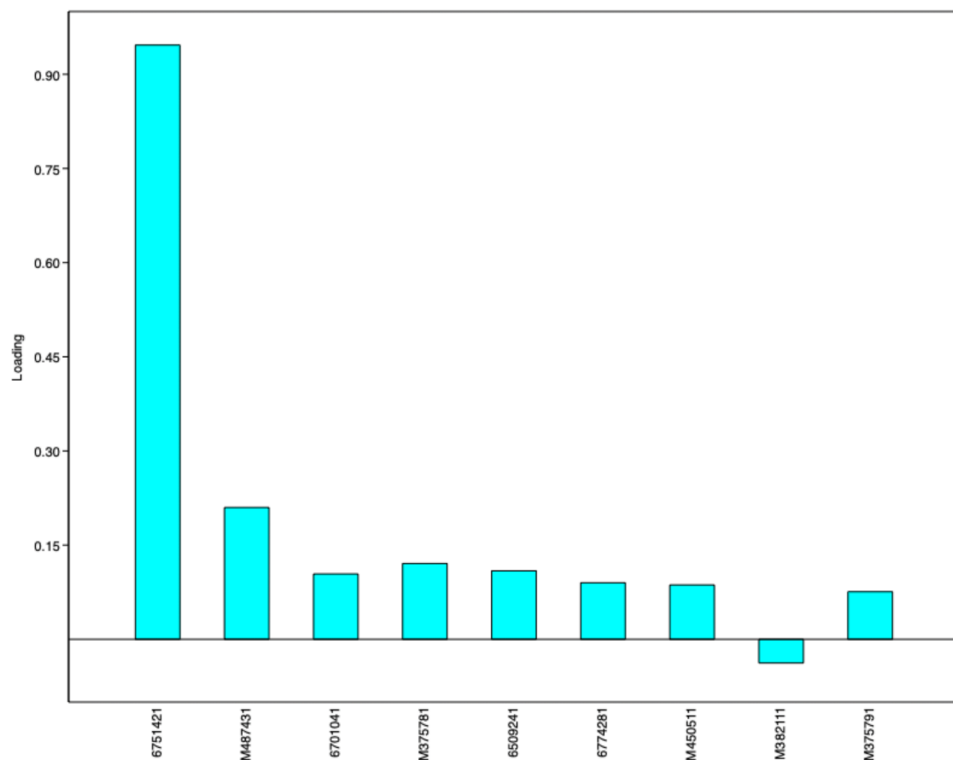


Gráfico 61.- Aportes de las variables para la construcción de la componente 1.

Dado el análisis anterior de la conformación de las componentes principales y el aporte de cada una de las variables, procedemos con la conformación del

Biplot de las 2 componentes principales, en las que en un eje vamos a tener los repuestos en análisis y la distribución de cada uno de los trabajos en los cuales se utilizan dichos repuestos, con esto podemos encontrar la relación entre ellos. Podemos ver que los repuestos 6751421 se encuentra relacionado con el repuesto 6774281, de igual manera los dos repuestos no tienen correlación alguna con los demás repuestos como son M487431, M375781. Incluso podemos ver que los repuestos M359841, se encuentran inversamente correlacionados con el repuesto 6774281. Si miramos desde el punto de vista de los trabajos realizados se distribuyen por todos los cuadrantes. Incluso encontramos que los códigos de trabajo que analizamos en los puntos anteriores se encuentran separados en cada cuadrante, lo que nos muestra que no se encuentran relacionadas y tienen un peso alto sobre todo en los trabajos 0B y 1B.

Si vemos la relación o representación de los repuestos con los trabajos vemos directamente que el vector para la variable 6751421 se encuentran bien representados en este caso los trabajos 1B, 0A, y 0B se encuentra completamente alejada, es decir no se encuentra representada por ninguno de los vectores.

Desde el punto de vista de clústers, podemos ver que podemos formar dos grupos de repuestos, los cuales se encuentran correlacionados y las distancias de sus vectores tienen mucha similitud. Este es el caso del clúster formado por 6751421 y 6774281. Además de ver la formación de otro cluster que está con M487431, M375781, M375791, entre otros. Esto nos indica que tenemos un consumo bastante similar entre ellos. Se consumen a la par o tienen una incidencia de uso en reparaciones del mismo nivel.

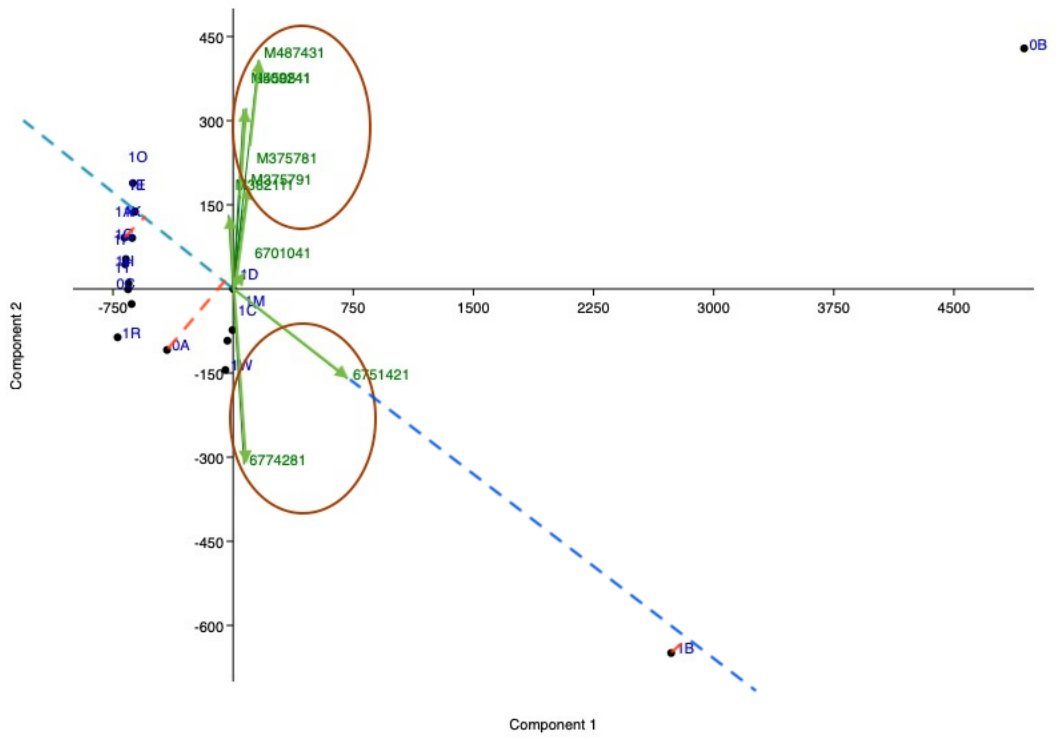


Gráfico 62.- Biplot de las componentes de Repuestos y códigos de Trabajo.

CAPÍTULO 5

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Luego de los distintos análisis que hemos realizado utilizando las distintas técnicas tanto Univariantes como Multivariantes para el conjunto de datos, podemos llegar a varias conclusiones entre las que podemos mencionar:

- Al tener un gran conjunto de datos lo primero que debemos lograr es recolectar la información de la manera correcta para analizarla de una manera correcta y que nos lleve a los resultados esperados. En muchos casos no podemos consolidar los datos en base a ciertas variables o condiciones si a la final lo que estamos queriendo es analizar en base al tiempo y la mejor forma para ello sería en primer lugar formar la serie de tiempo.
- Los datos a ser tratados sobre todo en series de tiempo deben tener una cantidad de variables mínimas correspondientes a un período significativo en el cual podamos obtener una tendencia y estacionalidad válidos que nos aporten la información suficiente para que al aplicar el modelado los test de los residuales tengan niveles aceptables de tal manera que el modelo se ajuste de una manera correcta y en consecuencia las predicciones que se puedan hacer tengan buenos resultados.
- Si bien poder tener predicciones, que se basan en el análisis de la información recolectada en el pasado, para intentar predecir lo que va a suceder en el futuro, esto puede no ser del todo aplicable a todos los campos de la ciencia, porque en este caso se puede ver que no hay factores de estacionalidad totalmente definidos. A pesar de usar variables que sabemos con toda certeza que corresponden a trabajos que se realizan de manera bi anual en las máquinas y que se distribuyen a lo largo del año.
- En el caso de nuestras series de tiempo vemos que presenta mucho ruido, el cual no solamente es ruido blanco ya que un suavizado de medias móviles, que es el más sencillo que podemos aplicar no fue suficiente en ciertos casos como para poder llegar a obtener un buen modelo ARIMA, esto no puede ser visto como un obstáculo, ya que tenemos otros métodos los cuales también pueden ser aplicados para lograr un buen modelado.
- Desde el punto de vista práctico el hecho de no lograr fácilmente un suavizado con un método sencillo, nos demuestra la versatilidad de los métodos de los que se dispone para suavizar las series de manera eficiente y a la vez sean

series que se acoplan de una mejor manera a cualquiera de los modelos para las predicciones.

- Desde el punto de vista Multivariante, vemos que a pesar de trabajar con todas las variables que extrajimos de la base de datos, el proceso trabaja de una manera muy eficiente e intuitiva. Y cumple de muy buena manera su principal objetivo que es la reducción de variables manteniendo la variabilidad de los datos.
- Podemos ver que en muchos casos con el solo uso de dos componentes obtenidas del análisis podemos explicar sobre el 80% de las variables. Y en las cuales se puede analizar como cada una de las variables aporta para la conformación de las componentes.
- Cuando realizamos el análisis gráfico de los aportes y correlaciones de las variables, es donde el modelo se complica, puesto que con dos variables es fácil desde el punto de vista de interpretabilidad el obtener los resultados. Pero al tomar en cuenta una componente más convertimos el gráfico en uno de tres dimensiones que ya complica la interpretabilidad de los resultados. Pero sigue siendo hasta cierto punto manejable. En el caso de querer seguir aumentando la cantidad de componentes debemos considerar que de igual manera la representación gráfica se va tornando cada vez más complicada.
- Dado el bajo aporte que realizan las tecnologías comparada con la que es más significativa, no se encontró objetivo en trabajar con cada categoría por separado. El cambio no era significativo como para que haya cambios relevantes en la variabilidad de los datos.
- Si bien es cierto el Análisis Multivariante no lo utilizamos para llegar a obtener una predicción de las variables, nos da una mejor idea de como se relacionan con los trabajos que realizamos, incluso podemos deducir que al reducir la cantidad de trabajos debería afectar de manera directa a ciertos repuestos. Con esto mejoramos los trabajos tanto preventivos como correctivos. De igual manera nos da una mejor visión de como se relacionan distintos tipos de repuestos. Incluso podríamos separar los repuestos por tipo de aplicación como pueden ser mecánicos, eléctricos, placas electrónicas y así mismo categorizarlos para asociarlos a una tecnología en particular, con eso podríamos ver como se correlaciona cada uno con la tecnología en las que se los utiliza.

5.2. Recomendaciones

- Durante el Análisis de las series de tiempo pudimos ver que las tendencias en ciertas variables llegaron a picos que luego bajaron y hasta cierto punto llegaron a estabilizarse. Esto se concluyó que no era el efecto de una sola acción.
- Las tendencias en las series de tiempo son útiles desde el punto de vista de poder mostrar como una variable va evolucionando con el paso del tiempo. Esto ayuda en su interpretación y análisis, ya que al ver la evolución de varios años (observaciones) podemos interpretar de una mejor manera lo que está pasando.
- Es de suma importancia el mantener una buena calidad de datos, y sobre todo que quienes son los encargados del ingreso de información sean conscientes de que la información que se produce día a día es de mucha importancia para poder llegar a tomar decisiones. La capacitación constante y revisión de proceso de ingreso de información se vuelve también una parte crucial en el análisis.
- En el Análisis Univariante y aplicación de modelos ARIMA, vemos que tenemos varias rutas para llegar al resultado final. El punto de vista importante en este caso es el poder tener la pericia necesaria para saber el punto desde el cual podemos ver que los datos que estamos trabajando o preparando se pueden ajustar o no a un modelo. Esto es parte de experiencia y pericia, pero también de constancia y probar varios modelos hasta llegar a la aplicación del adecuado.
- Para el análisis Multivariante vemos que las reglas indicaban que con dos componentes tenemos lo necesario para analizar nuestros datos. Y al final lo que se trata es reducir la dimensionalidad de los datos pero siempre y cuando mantengamos la variabilidad de los mismos.
- El aumento de las componentes que se utilizan para el estudio complica su interpretabilidad pero nos puede dar otras perspectivas adicionales que a lo mejor se eran del todo claras en un análisis de solo dos componentes principales.

- El desarrollo de nuevos softwares es siempre de utilidad desde el punto de vista gráfico, ya que pueden ayudarnos a mejorar la visualización de los resultados y por ende su interpretabilidad.
- Como mencionamos en la parte introductoria, debemos considerar siempre el acompañamiento de un experto en el área en la cual estamos realizando un estudio, puesto que esa visión nos ayuda a mejorar la interpretación de los datos, además de darnos una dirección de como realizar categorizaciones y discriminar los datos para mejorar el modelado e interpretación de los resultados.

6. REFERENCIAS

- Abulaish, A., Hasikin, K., Mokhzaini Azizan, M., Hilmi Zamzam, A., Khallel Ibrahim Al-Ani, A., Khairi Abdul Wahab, A., Wee Lai, K., Chandra Satapathy, S., & Khalil, A. (2021). Prioritisation Assessment and Robust Predictive System for Medical Equipment: A Comprehensive Strategic Maintenance Management. *Frontiers in Public Health* | www.frontiersin.org, 9, 782203. <https://doi.org/10.3389/fpubh.2021.782203>
- Astivia-Chávez, D. N., & Ortiz-Posadas, M. R. (2023). *Exploratory Data Analysis for Preventive and Corrective Maintenance for Medical Equipment in a General Hospital from the Health Institute of the State of Mexico* (pp. 805–815). https://doi.org/10.1007/978-3-031-18256-3_84
- Badnjević, A., & Pokvić, L. G. (2020). Medical devices maintenance. *Clinical Engineering Handbook, Second Edition*, 520–526. <https://doi.org/10.1016/B978-0-12-813467-2.00080-8>
- C. M. Cuadras. (2014). *Nuevos Métodos de Análisis Multivariante*. CMC Editions.
- Cruz, A., Barr, C., & Pozo Punaes, E. (2007). Improving corrective maintenance efficiency in clinical engineering departments - Multiple Linear Regression and Clustering Techniques for Analyzing Quality and Effectiveness of Technical Services. *IEEE Engineering in Medicine and Biology Magazine*, 26(3), 60–65. <https://doi.org/10.1109/MEMB.2007.364931>
- GABRIEL, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3), 453–467. <https://doi.org/10.1093/biomet/58.3.453>
- Jaramillo Ayerbe, M., González Gómez, D. E., Núñez Cabrera, M. E., Portilla, G. E., & Lucio García, J. H. (2007). *Análisis de series de tiempo univariante aplicando metodología de Box-Jenkins para la predicción de ozono en la ciudad de Cali, Colombia*. *Revista Facultad de Ingeniería Universidad de Antioquia*, No. 39. Marzo.
- Kwaku Kutor, J., Agede, P., & Haruna Ali, R. (2017). Maintenance Practice, Causes of Failure and Risk Assessment of Diagnostic Medical Equipment. *Journal of Biomedical Engineering and Medical Devices*, 02(01). <https://doi.org/10.4172/2475-7586.1000123>
- Mahfoud, H., El Barkany, A., & El Biyaali, A. (2016). Preventive Maintenance Optimization in Healthcare Domain: Status of Research and Perspective. *Journal of Quality and Reliability Engineering*, 2016, 1–10. <https://doi.org/10.1155/2016/5314312>
- Marroquín Martínez, G., & Chalita Tovar, L. E. (2018). APLICACIÓN DE LA METODOLOGÍA BOX-JENKINS PARA PRONÓSTICO DE PRECIOS EN JITOMATE. *Revista Mexicana de Ciencias Agrícolas*, 2(4), 573–577. <https://doi.org/10.29312/remexca.v2i4.1643>
- Murtaza Noor, M., Ram, D. C., Magray, I. A., & Chawla, E. S. (2016). *Maintenance and Modelling Strategies of Biomedical Equipment's in Hospitals: A Review* (Vol. 40). Online. www.iiste.org
- Murthy, D. N. P., Karim, M. R., & Ahmadi, A. (2015). Data management in maintenance outsourcing. *Reliability Engineering and System Safety*, 142, 100–110. <https://doi.org/10.1016/j.res.2015.05.002>
- Pérez López, C. (2004a). *Técnicas de análisis multivariante de datos : aplicaciones con SPSS®*. Pearson Educación.

- Pérez López, C. (2004b). *Técnicas de análisis multivariante de datos : aplicaciones con SPSS®*. Pearson Educación.
- Pérez López, C. (2004c). *Técnicas de análisis multivariante de datos : aplicaciones con SPSS* [Book]. Pearson Prentice Hall.
- Teresa Carot Sanchez. (2014). *Introducción a la Estadística y a las Probabilidades*.