



Universidade de Vigo

Trabajo Fin de Máster

---

# Análisis y modelización del tráfico marítimo.

---

Lady Diana Amador Ysabel

Máster en Técnicas Estadísticas

Curso 2020-2021



# Propuesta de Trabajo Fin de Máster

<b>Título en galego:</b> Análise e modelización do tráfico marítimo.
<b>Título en español:</b> Análisis y modelización del tráfico marítimo.
<b>English title:</b> Maritime Traffic Analysis and Modeling.
<b>Modalidad:</b> Modalidad A
<b>Autor/a:</b> Lady Diana Amador Ysabel, Universidad de La Coruña
<b>Director/a:</b> Salvador Naya Fernández y Javier Tarrío Saavedra , Universidad de La Coruña
<b>Breve resumen del trabajo:</b> En este trabajo se utilizan datos correspondientes a los tránsitos de distintos tipos de buques a través de rutas que utilizan el Canal de Panamá para el análisis y estimación de modelos estadísticos que permitan describir, predecir y controlar variables críticas para la calidad del transporte marítimo, mediante la aplicación de distintas técnicas estadísticas como modelos de regresión multivariante, gráficos de control y análisis de capacidad.



Don Salvador Naya Fernández, profesor de la Universidad de La Coruña, Don Javier Tarrío Saavedra , profesor de la Universidad de La Coruña, informan que el Trabajo Fin de Máster titulado

**Análisis y modelización del tráfico marítimo.**

fue realizado bajo su dirección por doña Lady Diana Amador Ysabel para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En A Coruña, a 20 de Junio de 2021.

El director:

Don Salvador Naya Fernández

El director:

Don Javier Tarrío Saavedra

La autora:



Doña Lady Diana Amador Ysabel



# Agradecimientos

En primer lugar quiero agradecer a Dios por darme esa fortaleza que siempre me da cuando he sentido que ya no puedo más.

A mis padres que son todo para mi y siempre me han apoyado incondicionalmente, especialmente a mi madre por sus palabras de aliento.

A mis compañeras de clase Jomayra Ramírez y Carolyn Bautista por su empatía, ayuda y paciencia para estudiar.

A mi amiga Evelyn Reyes por siempre darme ese empujón para seguir y no desistir de mi meta cada vez que me iba mal en alguna asignatura.

A mis tutores Javier Tarrío y Salvador Naya por hacer posible que culminara este trabajo la entrega del mismo.

Gracias a la coordinación del Máster por ser tolerante en cada etapa de este proceso.



# Índice general

Resumen	xI
<b>1. Introducción</b>	<b>1</b>
1.1. Canal de Panamá . . . . .	1
1.1.1. Antecedentes y evolución. . . . .	1
1.1.2. Canal de Panamá Ampliado. . . . .	2
1.1.3. Principales componentes del Canal de Panamá . . . . .	4
1.1.4. Operaciones de Tránsito . . . . .	5
1.1.5. El piloto en el Canal de Panamá Ampliado. . . . .	8
1.2. Descripción de los problemas a tratar . . . . .	9
1.2.1. Estimación de la curva de aprendizaje del proceso de tránsito de buques a través del Canal de Panamá Ampliado . . . . .	9
1.2.2. Control y análisis de capacidad del proceso de tránsito a través del Canal Ampliado . . . . .	10
1.3. Métodos de resolución . . . . .	10
1.4. Objetivos . . . . .	11
<b>2. Métodos estadísticos: Control estadístico de calidad</b>	<b>13</b>
2.1. Gráficos de Control . . . . .	14
2.1.1. Gráficos de control univariante . . . . .	16
2.1.2. Gráficos de control multivariante . . . . .	21
2.2. Análisis de capacidad de procesos . . . . .	25
2.2.1. Análisis de capacidad univariante . . . . .	25
2.2.2. Análisis de capacidad multivariante . . . . .	28
<b>3. Análisis exploratorio y estimación del efecto del aprendizaje</b>	<b>31</b>
3.1. Descripción de datos . . . . .	31
3.1.1. Base de datos correspondiente al buque Baroque . . . . .	31
3.1.2. Base de datos correspondiente a los datos reales . . . . .	32
3.2. Estimación de las curvas de aprendizaje . . . . .	33
3.2.1. Curva de aprendizaje para las instalaciones y pilotos maniobrando el buque Baroque . . . . .	33

3.2.2. Estimación de la curva de aprendizaje correspondiente al tiempo de operación real . .	36
3.2.3. Ajuste de modelos aditivos generalizados . . . . .	40
<b>4. Control estadístico de procesos y análisis de capacidad</b>	<b>47</b>
4.1. Técnicas de gráficos de control, Fase I . . . . .	47
4.1.1. Gráficos de control univariantes . . . . .	47
4.1.2. Gráficos de control multivariantes . . . . .	58
4.2. Técnicas de gráficos de control, Fase II . . . . .	63
4.2.1. Gráficos de control univariantes . . . . .	63
4.2.2. Gráficos de control multivariante . . . . .	67
4.3. Análisis de capacidad de las instalaciones del Canal . . . . .	71
<b>5. Conclusiones</b>	<b>77</b>
<b>Bibliografía</b>	<b>79</b>

# Resumen

## Resumen en español

El Canal de Panamá ha estado recientemente en el centro de atención mundial. En agosto de 2014, celebró 100 años de servicio ininterrumpido y, en junio de 2016, se inauguró el proyecto de expansión del canal. El proyecto final implicó la construcción de un tercer juego de esclusas. Una vez que el canal comenzó a operar, la forma en que los buques transitaban seguía siendo la misma. Sin embargo, las dimensiones de las esclusas y sus nuevos procedimientos operativos han tenido un efecto en el tamaño del buque y las maniobras para los buques más grandes. Después del primer tránsito, el 26 de junio de 2016, fue posible tener acceso a datos sobre los nuevos sistemas de esclusaje para buques Neopanamax.

Antes de la implementación del Canal Ampliado, un número reducido de pilotos con experiencia en el canal existente fueron certificados para el tránsito de buques Neopanamax mediante un aprendizaje individual planificado e innovador. Luego de esta capacitación y con la inauguración del Canal Ampliado, se inició el entrenamiento de rutina. De ahí que, al realizar estas maniobras, la curva de aprendizaje represente la creciente habilidad adquirida tanto por los pilotos como por la organización.

Dado que la curva de aprendizaje es medible, en este trabajo estudiaremos la curva de aprendizaje operativa de la organización y la curva de aprendizaje para pilotos que realizan maniobras de tránsito en las vías fluviales ampliadas del Canal de Panamá. Específicamente, se pretende identificar patrones de cambio en el tiempo de paso de los buques por las esclusas de Cocolí y Agua Clara, tomando en cuenta la dirección, el tipo de embarcación, dimensión de la manga y eslora.

Para el análisis se ha aplicado un estudio estadístico que incluye: análisis exploratorios, regresión multivariada, análisis de varianza (ANOVA), y técnicas del control estadístico de la calidad, tales como gráficos de control y análisis de capacidad. Los tiempos de maniobra en las esclusas serán estudiados continuamente desde junio 2016 a octubre 2019.

## English abstract

The Panama Canal has recently been in the world spotlight. In August 2014, it celebrated 100 years of uninterrupted service and, in June 2016, the canal expansion project was inaugurated. The final project involved the construction of a third set of locks. Once the canal started to operate, the way the ships transited remained the same. However, the dimensions of the locks and their new operating procedures have had an effect on ship size and maneuvering for the larger vessels. After the first transit, on June 26, 2016, it was possible to have access to data on the new lockage systems for Neopanamax ships.

Prior to the implementation of the expanded Canal, a small number of pilots with experience in the existing canal were certified for the transit of Neopanamax vessels through planned and innovative individual learning. After this training and with the inauguration of the expanded channel, the routine training began. Hence, when performing these maneuvers, the learning curve represents the growing skill acquired by both the pilots and the organization.

Given that the learning curve is measurable, in this paper we will study the organization operative learning curve and the routine learning curve model for pilots performing transit manoeuvres in the expanded waterways of the Panama Canal. Specifically, it is intended to identify patterns of change in the time the vessels pass through the Cocolí and Agua Clara locks, taking into account the direction, type of vessel, beam dimension and length.

For the analysis, a statistical study has been applied that includes: exploratory analysis, multivariate regression, analysis of variance (ANOVA), and statistical quality control techniques, such as control charts and capacity analysis. The maneuvering times in the locks will be continuously studied from June 2016 to October 2019.

# Capítulo 1

## Introducción

En este primer capítulo trataremos de contextualizar la problemática sobre la identificación y modelación de los patrones de aprendizaje correspondientes al proceso de tránsito de embarcaciones por las esclusas Cocolí y Agua Clara, a través del nuevo Canal de Panamá Ampliado, así como en generar la estimación de la posición y variabilidad de las variables que caracterizan a este proceso, incluyendo la detección de patrones y anomalías, además del análisis de su capacidad de cumplir las especificaciones de gerencia, armadores o normativa. Este patrón de aprendizaje, también conocido como curva de aprendizaje, está relacionado con la experiencia de los pilotos y la mejora continua de los métodos e instalaciones del Canal de Panamá.

Comenzaremos por presentar brevemente la organización del Canal de Panamá, sus componentes afectando su tránsito y una breve explicación sobre la actividad de pilotaje. A continuación, se seguirá con una exposición del problema y los motivos que han generado la creación de este proyecto. Posteriormente se introducen los métodos de resolución utilizados y por último, se definen los objetivos.

### 1.1. Canal de Panamá

#### 1.1.1. Antecedentes y evolución.

El Canal de Panamá es una vía marítima de aproximadamente 80 kms(50 millas) que conecta los océanos Atlántico y Pacífico a través de uno de los puntos más estrechos del continente Americano. Utiliza un sistema de esclusas por gravedad y canales de navegación para permitir el tránsito seguro y eficiente de los buques de altamar a través del istmo de Panamá.

El Canal es considerado como una de las grandes obras de la ingeniería mundial del siglo XX. Dada su ubicación y al ofrecer un paso expedito, confiable y seguro entre ambos océanos, representa un punto clave para el comercio y tráfico marítimo internacional, reduciendo las distancias marítimas, los tiempos de tránsito y los costos de un buque al transportar bienes entre los principales centros de producción y consumo.

Huebner (1915) se refirió a las dos funciones principales del Canal de Panamá. Uno era su valor militar, ya que permitía a los Estados Unidos (EE.UU) mover embarcaciones navales entre los océanos Atlántico y Pacífico. El otro era su valor económico; ofrecía un atajo cuando se transportaba carga entre los dos océanos.

De acuerdo a Sabonge (2014) la construcción del Canal de Panamá a principios del siglo XX cambió los patrones comerciales al abrir nuevas rutas entre países y regiones que tradicionalmente no podían comercializar a precios competitivos debido a la gran distancia entre ellos. El Canal fue construido principalmente para fines militares, pero, con el tiempo, se convirtió en un facilitador del comercio al acortar el tiempo y la distancia entre los mercados de producción y consumo. Especifica que desde su inauguración hasta el año fiscal 2013, más de un millón de buques han transitado por el Canal con más de 9,4 mil millones de toneladas de carga. Estas estadísticas demuestran que el Canal de Panamá ha sido un catalizador muy efectivo del

comercio internacional, reduciendo el tiempo y la distancia entre países y que se ha traducido en una mayor competitividad y crecimiento económico para los países y regiones.

Con respecto a la evolución del canal, Carral et al. (2017) señalan que el sistema operativo del Canal de Panamá se mantuvo sin cambios en 102 años de operación. Sin embargo, hubo cambios importantes en la infraestructura y el equipo. Por ejemplo, el corte Culebra se amplió para que los buques Panamax pudieran transitar el canal al mismo tiempo. Además, las vías fluviales se iluminaron y se mejoró la señalización. Sin embargo, las dimensiones de las esclusas habían limitado el tamaño de los buques que las atravesaban Martin et al. (2015) hasta el punto de que se estableció una categoría específica de Panamax para graneleros, petroleros y portacontenedores.

Durante las primeras cinco décadas que funcionó (1914-1950), el aumento en el tamaño de la carga comercial condujo a un aumento en el tránsito, por ejemplo, en los años 30 ya se hablaba de la necesidad de ampliar el canal para solucionar la congestión del tráfico y permitir el paso de barcos más grandes. Sin embargo, durante la década de 1970, el número de tránsito se estabilizó, ya que el aumento en la cantidad de carga se mantuvo estable debido a que se construyeron buques progresivamente más grandes.

### 1.1.2. Canal de Panamá Ampliado.

El volumen comercial aumentó y, al mismo tiempo, los buques de tamaño Neopanamax se hicieron más populares para el transporte marítimo. Estos eran demasiado grandes para transitar las esclusas y, por esta razón, la Autoridad del Canal de Panamá (ACP) influye para el proceso de expansión del Canal.

Las obras de Ampliación del Canal de Panamá dieron inicio en 2007, mediante la construcción de un tercer juego de esclusas y la ampliación del cauce de navegación, y de esta forma permitir el paso de buques tamaño Post-Panamax, los cuales, dado que superan el tamaño Panamax, no podían navegar por la vía interoceánica.

En junio de 2016, se inauguró el proyecto de expansión del canal. El proyecto final implicó la construcción de un tercer conjunto de esclusas más grandes que las hechas en 1914. Un conjunto de esclusas está ubicado en el lado atlántico del canal, en la orilla este de las esclusas de Gatón. Otro conjunto de esclusas se encuentra en el lado del Pacífico, al suroeste de las esclusas de Miraflores. La expansión también incluyó el dragado de ambas entradas del Canal (Atlántico y Pacífico), así como la ampliación y profundización de los canales de navegación existentes en el Lago Gatún y la profundización en el Corte Culebra.

A diferencia de las esclusas de la configuración de 1914, las nuevas esclusas tienen dimensiones apropiadas para maniobrar los buques Neo-panamax entre 150,000 y 170,000 toneladas de desplazamiento, según se indica en Carral et al. (2019).

La accesibilidad que ofrece el Canal de Panamá une de manera competitiva a todos los mercados, principalmente a Asia, Europa, Norte y Sudamérica, como se visualiza en la Figura 1.1.

La vía interoceánica conecta 144 rutas marítimas que llegan a 1,700 puertos en 160 países, posicionando a Panamá como un centro de transporte, logística y servicios. Las principales rutas comerciales con tráfico en el Canal de Panamá son:

- Costa Este de E.U.A. y Asia (Lejano Oriente)
- Costa Este de E.U.A. y Costa Oeste de Sudamérica
- Europa y Costa Oeste de Suramérica
- Costa Este de E.U.A. y Costa Oeste Centroamérica
- Costa a Costa de Suramérica.

Estas principales rutas, en conjunto con todas las demás que utilizan la vía interoceánica, reportaron para el año fiscal 2019 un total de 468,779 toneladas netas CP/SUAB<sup>1</sup>, representando un incremento del 6.3% con respecto al año fiscal 2018.

<sup>1</sup>CP/SUAB es el Sistema Universal de Arqueo de Buques del Canal de Panamá.

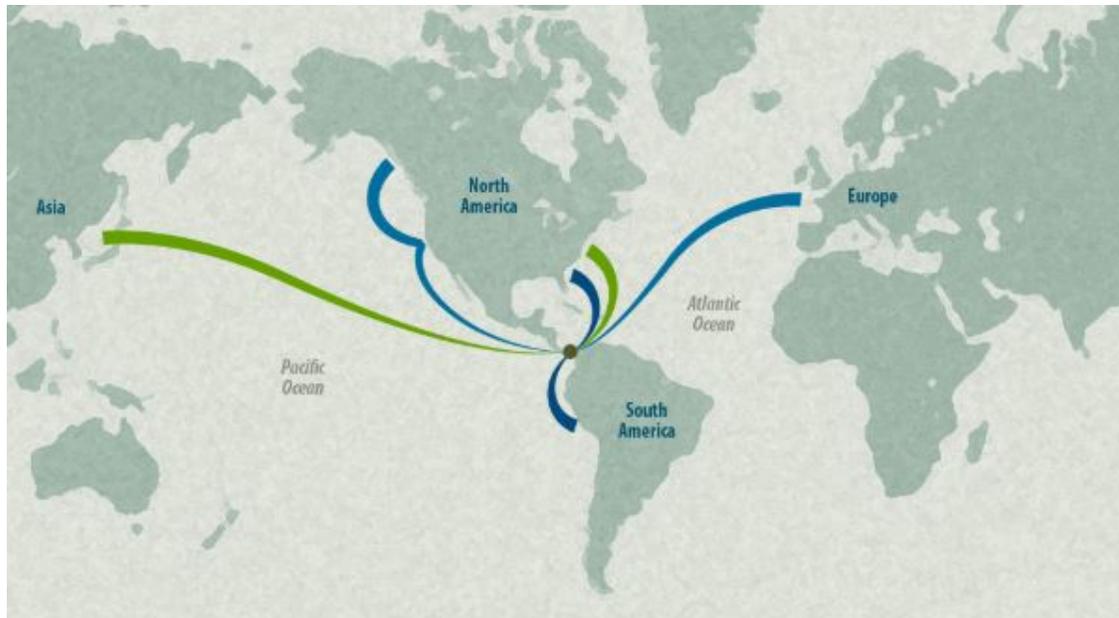


Figura 1.1: Flujo tránsito. Fuente: Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá.

El Canal representa aproximadamente el 5 % del comercio marítimo mundial, y unos 14,000 buques transitan esta importante vía fluvial cada año.

Al analizar la estadística sobre el tránsito acumulado del Canal de Panamá por segmento de mercado y tipo de esclusa, al año fiscal 2020, se verifica que casi el 75 % del total de los tránsitos son buques tipo Panamax, mientras que el resto son buques Neopanamax, los más grandes que el Canal puede acomodar. Los buques Neopanamax transitan por el tercer juego de esclusas, que incluye dos esclusas de tres niveles y la ampliación de los canales de acceso en el Atlántico y el Pacífico.

Además, los buques principales que transitan el Canal de Panamá son: Graneleros, Portacontenedores, Químicos, de transporte de Gas Licuado de Petróleo (GLP) y Petroleros. Específicamente, los buques Portacontenedores, los GLP y aquellos dedicados al Gas Licuado Natural (GNL), que son lo que estudiaremos en este trabajo, representan casi el 80 % de los tránsitos de los buques Neopanamax, con 42.9 %, 24.6 % y 12.4 %, respectivamente.

A partir de la información institucional existente en la página web de la Autoridad del Canal de Panamá (2020), desde su inauguración en 1914, se han realizado alrededor de 1.1 millones a través del Canal de Panamá, esto al cierre del año fiscal 2016.

EL 10 de agosto de 2020 se alcanzaron 10 mil tránsitos por el Canal Ampliado desde su apertura en junio de 2016.

La respuesta de expertos del Canal al preguntarles cómo ha afectado el Canal Ampliado a los tránsitos en los últimos cuatro años es la siguiente: “En 2016, el Canal Ampliado representaba el 8 % de los tránsitos en la vía interoceánica. Desde entonces, los navieros han podido capitalizar el ahorro de tiempo y costos que ofrece el Canal Ampliado, trayendo nuevos segmentos, mayor cantidad de carga para los segmentos existentes y redirigiendo sus líneas de tránsito a la vía interoceánica. Cuatro años más tarde, el Canal Ampliado representa actualmente no solo el 27 % de los tránsitos, sino también el 50 % del tonelaje en el Canal de Panamá”, es decir, la apertura del Canal de Panamá Ampliado ha aumentado significativamente las posibilidades y capacidad del transporte marítimo de grandes dimensiones, llegando a ser tan importante, en términos de carga transportada, como la ruta tradicional o Canal antiguo, y todo ello en un periodo de apenas 4 años.

### 1.1.3. Principales componentes del Canal de Panamá

El Canal posee varios componentes físicos importantes que trabajan de manera integrada para proporcionar los servicios de tránsito a toda clase de naves: desde pequeñas unidades flotantes hasta barcos de altamar. Estos componentes son las esclusas, los canales de navegación, los lagos, las áreas de anclaje y las estaciones de amarre (ver Figura 1.2).



Figura 1.2: Componentes del Canal Ampliado. Fuente: Autoridad del Canal de Panamá.

#### Esclusas

Las esclusas del Canal de Panamá funcionan como elevadores de agua que suben y bajan a los buques desde los océanos Atlántico y Pacífico hasta el Lago Gatún a 26 metros (85 pies) sobre el nivel del mar. Cada esclusa cuenta con dos carriles para el tránsito de los buques. Cada carril puede ser usado en ambas vías.

El Canal anterior tenía tres juegos de esclusas: Gatún (en el lado Atlántico), Pedro Miguel y Miraflores (en el lado Pacífico). Las esclusas de dos vías permiten a los buques transitar simultáneamente en dirección opuesta si es requerido. Los tres juegos de esclusas tienen diferentes niveles: Miraflores tiene dos niveles de levantamiento de los buques hacia el lago Miraflores; la esclusa Pedro Miguel tiene un solo nivel hasta el lago Gatún; mientras que Gatún, que es la esclusa más grande, tiene tres niveles para descender el buque hacia el lado Atlántico.

Las cámaras de las esclusas tienen las siguientes medidas: 33.5 mts (110 pies) de ancho, 304.8 mts (1,000 pies) de largo y 12.8 mts (42 pies) de profundidad. Estas dimensiones restringen a los barcos a un máximo de 32.3 mts (106 pies) en manga; 294.1 mts (965 pies) en eslora; y 12.04 mts (39.5 pies) en calado de agua dulce.

Con la expansión del Canal de Panamá, se construyeron dos esclusas adicionales, la esclusa “Cocolí” en el Pacífico y “Agua Clara” en el Atlántico. Las esclusas tienen 55 mts (180 pies) de ancho, 427 mts (1,400 pies) de largo y 18.3 mts (60 pies) de profundidad con el fin de permitir el tránsito de naves con mayor capacidad de carga. Cada complejo de esclusas cuenta con tres niveles.

#### Canales de navegación

Los canales de navegación del Canal de Panamá se extienden por 72 km (45 millas) entre las áreas de anclaje del Atlántico y del Pacífico. Estos canales de navegación varían en lo ancho, su profundidad y las características del barco. Estos canales se dividen en los siguientes, ordenados del sur al norte:

- Entrada del Pacífico: Entrada al canal por el océano pacífico y conecta con el final sur de la esclusa Cocolí, eludiendo el lago miraflores.

- Lago Miraflores: El lago miraflores se creó luego de la construcción de una pequeña represa para contener los Ríos Grande y Cocolí, los cuales fluían hacia el Océano Pacífico. Se ubica entre las esclusas de Pedro Miguel y Miraflores. Tiene una extensión de 3.94 km cuadrados (1.52 millas cuadradas) con una elevación media de 16.5 metros (54') sobre el nivel del mar. El almacenaje de agua disponible equivale a 2.46 millones de metros cúbicos (650 millones de galones).
- Corte Culebra: El Corte Culebra mide 13.7 kilómetros de largo. Se extiende desde el Lago Gatún hasta las Esclusas de Pedro Miguel. Es la sección más estrecha del sistema de canales de navegación del Canal. Sus espacios para maniobrar son limitados. Varias curvas y la incidencia de la neblina restringen la visibilidad durante horas de la noche, se considera el carril más difícil de navegar, de ahí la asignación de restricciones especiales a naves de alto calado que transitan por esta área.
- Lago Gatún: El lago Gatún es un elemento clave del canal y la razón principal por la que es posible. El Lago Gatún tiene un área de 163.38 millas cuadradas y existe gracias a una presa que se construyó en el estuario del río Chagres en la entrada del Océano Atlántico. Estas aguas se utilizan para llenar los canales para que los buques puedan transitar y las esclusas puedan ser operadas. Como consecuencia de este uso dual del agua, el volumen de suministro de agua de las cerraduras es limitado.
- Entrada del Atlántico: Va desde final norte de las esclusas Agua Clara hasta la entrada del Atlántico, eludiendo las esclusas de Gatún.

Debido a las características propias de cada uno de estos canales, existen diferentes restricciones operacionales (que detallaremos más adelante) para asegurar el tránsito seguro de las naves por sus aguas. Estas restricciones tienen un impacto directo en la flexibilidad operacional, su mantenimiento y costos de mejoramiento.

### **Lagos**

Las operaciones de tránsito del Canal de Panamá dependen de la disponibilidad de agua fresca. El diseño original del Canal incluye la construcción de represas en los ríos Chagres y otros afluentes menores para formar los lagos que almacenan el agua necesaria para las operaciones de tránsito. Estas represas forman el Lago Gatún, el Lago Miraflores y el Lago Alajuela.

La capacidad de almacenamiento de agua de estos lagos está determinada por los niveles físicos y operacionales del Lago Gatún, el cual provee del calado necesario en las vías de acceso. El Canal administra estos lagos con un sistema hídrico integrado que permite el uso óptimo de este importante recurso.

### **Anclajes y estaciones de amarre**

Los barcos que arriban a las aguas del Canal para su tránsito deben esperar en las áreas de anclaje designadas en las entradas del Pacífico o Atlántico. Por razones operacionales, existen estaciones de amarre y áreas de anclaje internas (en los lagos o canales de navegación) en las cuales los barcos pueden esperar temporalmente sin interrumpir el paso de otras naves. Estas ubicaciones mejoran la flexibilidad de programación del barco incrementando el número de tránsitos.

#### **1.1.4. Operaciones de Tránsito**

El tráfico a través del canal se organiza de dos maneras diferentes:

1- Por orden de llegada de los buques, en este caso, son naves que no poseen reservaciones. El tráfico es administrado por medio un sistema de arribo (primero que llega, primero que sale) y de acuerdo a las restricciones individuales de cada nave.

2-La otra manera es mediante un sistema de reserva. En este caso los buques pueden comprar una reserva para transitar en una fecha específica, y mediante el cual el canal se compromete a realizar el tránsito en esa fecha dentro de las 18 horas posteriores a su inicio.

Bajo condiciones normales, el Canal ofrece cada día 17 cupos de reserva para naves con manga superior o igual a los 91 pies (conocidos como barcos Super), 8 cupos de reserva para naves con manga menor a los 91

pies (conocidos como barcos Regulares) y un cupo que es ofrecido por medio de una subasta vía internet para barcos de cualquier tamaño.

El Canal de Panamá ofrece una capacidad diaria del 70%. Las naves que posean la reservación de un cupo recibirán un servicio de tránsito expedito.

### Restricciones de navegación

La programación de tránsitos por el Canal toma en consideración las restricciones específicas que aplican para cada tipo de nave. Estas restricciones a la navegación por el Canal tienen como objetivo minimizar los riesgos de accidentes y se aplican a los buques sobre la base de dos criterios generales: su dimensión y la carga que transporta. Básicamente, las restricciones determinan si la embarcación puede transitar el Corte Culebra o las esclusas durante la noche.

Durante el día se aplican restricciones en el tránsito de buques con mangas mayores de 30.5 metros (100') y largos que exceden 244 metros (800'). A su vez, los buques con mangas mayores de 30.5 metros (100') y menos de 244 metros (800') de largo pueden transitar las esclusas de noche y, tras una inspección previa, pueden también transitar el Corte Culebra de noche.

Las restricciones más comunes consideradas para el tránsito por el Canal son:

- Corte Despejado (CD): Esto indica que una nave no puede encontrar a otra en el Corte Culebra. Esta restricción aplica generalmente a los buques con mangas mayores o igual a 91 pies.
- Corte Despejado con Luz de Día (CDLD): Una nave no puede encontrar a otra en el Corte Culebra y solamente puede transitar durante horas del día. Algunos barcos CDLD pueden ser programados para el Corte en horas nocturnas. Esto se decide según cada caso individual por el Capitán de Puerto del Canal.
- Tránsito con Luz de Día (LD): Una nave no puede transitar las esclusas ni el Corte Culebra en horas nocturnas.
- Tránsito completo con Luz de Día: Esta restricción indica que un barco debe realizar su tránsito completo en horas diurnas desde un anclaje externo hasta el otro anclaje externo sin demoras. Si se llega a estar dentro del Canal en horas nocturnas, la nave deberá ser amarrada en una estación o fondeada en un área designada de anclaje.
- Restricciones del canal de acceso: Algunos barcos no están permitidos para acercarse a otros en ciertas áreas del canal de acceso si sus mangas combinadas exceden entre 51.8 o 54.9 metros (170 o 180 pies), dependiendo de sus diseños y de la sección del canal de acceso.

### Programación del Tráfico del Canal: Sistema de Semiconvoy

Cuando las embarcaciones tienen un procedimiento operativo restringido transitan bajo condiciones específicas. El horario de tránsito del Canal se puede hacer de dos maneras: los buques pueden viajar como un convoy completo, donde todos los buques se envían en una dirección en un día y en la dirección opuesta al día siguiente. Sin embargo, la fórmula más común se conoce como semi-convoy, esto con el objetivo de maximizar la utilización de la vía acuática

Este sistema permite que se alcancen un máximo número de tránsitos cada día manteniendo la seguridad de las operaciones y minimizando el impacto de las restricciones como el corte despejado y tránsito de día. Para el proceso de programación, muchos elementos son tomados en consideración: las dimensiones de los barcos, el tipo de barco, su carga, así como otros requerimientos técnicos especiales.

Generalmente, la estrategia de semiconvoy opera de la siguiente manera, como se muestra en la Figura 1.3:

- Dos grupos de convoys son coordinados para zarpar en dos programaciones diferentes: a inicios de la madrugada para barcos Super y a inicios de la tarde para barcos Regulares. Ambos turnos son organizados con el mismo sistema de programación. Los tránsitos más críticos son los barcos Super debido a sus dimensiones y restricciones a lo largo de la ruta.

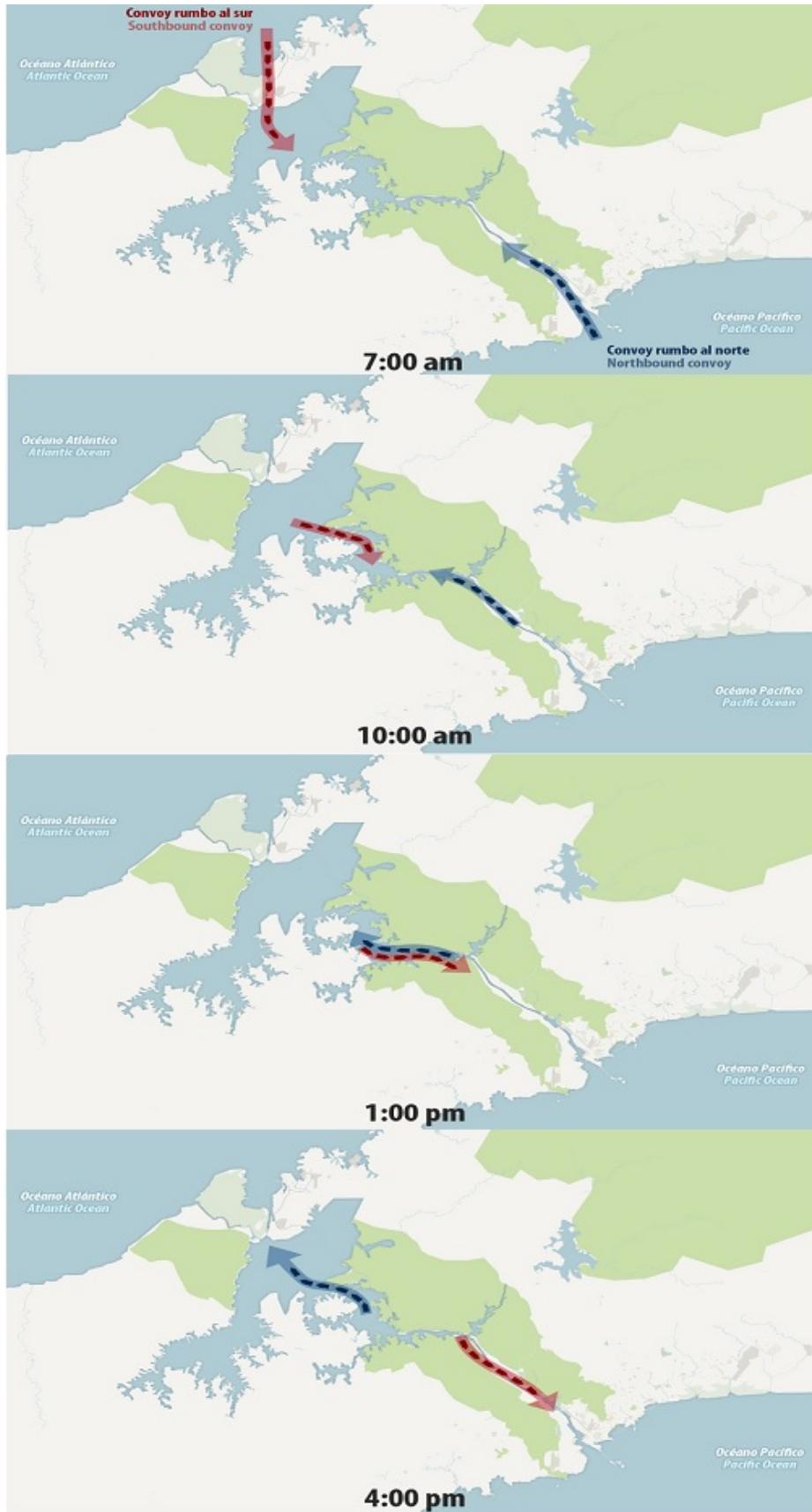


Figura 1.3: Sistema de Semiconvoy. Fuente: Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá.

- El convoy con dirección norte inicia su tránsito a tempranas horas de la mañana, según las restricciones de luz del día de los buques en el Corte y las esclusas. Al mismo tiempo, un convoy similar de barcos transitan hacia el sur por las esclusas de Gatún y esperan en el área de anclaje de Gatún hasta que el convoy norte despeja el Corte Culebra. El Corte Culebra es una de las áreas de mayor restricción, por ende la mayoría de las barcos de grandes dimensiones necesitan transitar esta sección en una condición especial denominada “corte despejado con luz de día”. La finalidad de esta programación intenta facilitar que los buques que no pueden realizar esclusajes de noche lleguen a las esclusas antes de la puesta del sol.
- El convoy sur está programado para entrar al Corte Culebra cuando la última nave en dirección norte lo haya despejado.
- Mientras el convoy sur transita por el Corte Culebra y alcanza las esclusas de Pedro Miguel y Miraflores, el convoy norte ya se encuentra completando su tránsito por las esclusas de Gatún en dirección al Océano Atlántico.
- El próximo turno de barcos Regulares (pequeños) inicia la misma rutina.

Los patrones de tránsito y programaciones son altamente susceptibles a las condiciones del clima a lo largo del día (por ejemplo, las restricciones de visibilidad por cambios en el clima) y el número de barcos en dirección norte y sur con restricciones en el corte y tránsito de día. Diariamente, el Canal de Panamá organiza y completa un detallado proceso de programación que puede diferir de los días anteriores.

### 1.1.5. El piloto en el Canal de Panamá Ampliado.

Cuando el Canal de Panamá comienza sus operaciones, en el año 1914, se implementa un sistema de control combinado en el cual el piloto del puerto asumió el control total de la embarcación durante su paso por las esclusas, y el capitán fue responsable de la navegación a lo largo de los canales bajo el asesoramiento del primero. Esta forma de operar cambia cuando la gerencia estadounidense, en 1953, decide que el pilotaje a través del Canal será diferente del resto de las vías navegables: los pilotos del puerto del Canal asumirán el control total de los movimientos y la navegación de los buques en tránsito, reemplazando al capitán en estas funciones.

La actividad que realizan los pilotos es clave para garantizar la eficacia y seguridad del Canal y de los buques que lo cruzan, así como también garantizar el control y la ejecución del horario de tránsito.

Con el desarrollo de otras alternativas de transporte logístico a través del Istmo de Panamá, se construyen puertos terminales en los extremos del Canal de Panamá. Con esto, los servicios portuarios se han agregado a las complejas operaciones de tránsito del Canal dentro de las responsabilidades de los pilotos portuarios. Este hecho, junto con el concepto de pilotaje mencionado anteriormente, ha provocado que, a diferencia de los sistemas convencionales VTS (Vessel Traffic Service), los servicios de control de tráfico se gestionen en el Canal de Panamá a través del Centro de control de tráfico marítimo (MTCC, por sus siglas en inglés).

El MTCC está a cargo de la generación del cronograma y los recursos asignados a cada buque para su tránsito, el cronograma de los pilotos del puerto que guiarán dichos buques y el seguimiento del cumplimiento del cronograma y de la disponibilidad de recursos para que los tránsitos se realicen de manera segura y eficiente.

En cuanto al proceso de certificación de los piloto, en Carral et al. (2021) especifican que la ACP recluta a sus pilotos y capitanes de remolcadores, principalmente, en medio del oficial de navegación de Panamá, y los entrena a través de programas de una duración promedio de 2 años. Una vez que se ha obtenido la certificación, en el caso específico de los pilotos de operación, estos progresan del nivel 1 al nivel 9 a lo largo de aproximadamente 9 años, y finalmente se convierten en pilotos calificados capaces de transitar cualquier embarcación a través del Canal existente. Para convertirse en pilotos certificados para desempeñarse en el Canal de Panamá Ampliado, se requiere un proceso doble de 4 años.

La capacitación de pilotos se imparte íntegramente en las instalaciones del Canal, principalmente a través del Centro de Simulación y Desarrollo Marítimo del Canal (SIDMAR), que está certificado por la Autoridad Marítima de Panamá bajo los esquemas del Acuerdo Internacional sobre Normas de Capacitación, Certificación y Guardias de la gente de mar, 1978 (Convenio STCW).

## 1.2. Descripción de los problemas a tratar

### 1.2.1. Estimación de la curva de aprendizaje del proceso de tránsito de buques a través del Canal de Panamá Ampliado

Con la construcción de un tercer conjunto de esclusas, el sistema operativo general de tránsito se mantendría igual. Sin embargo, por la forma en que operarían estas nuevas esclusas conllevaría cambios en el sistema de maniobras, así como en los tiempos de cierre y tránsito, según se indica en Carral et al. (2017).

De acuerdo a informaciones obtenida de la página del Canal de Panamá, una nave demora un promedio de 8 a 10 horas en transitar el Canal. Sin embargo, según Carral et al. (2019), este tiempo promedio se refiere a los buques estándar, con una manga de menos de 27.7m, y/o embarcaciones Panamax. El tiempo debe modificarse para las embarcaciones Neopanamax, ya que su tiempo de tránsito se ve afectado por factores clave: sus componentes, los principios operativos del esclusaje, las restricciones a la circulación de los buques y los principios operativos de los canales.

El tránsito a través del Canal de Panamá Ampliado consta una sucesión de procesos repetitivos. A medida que aumenta el número de repeticiones y la experiencia del personal y la organización, se podría esperar una reducción del tiempo y el esfuerzo requerido para efectuar el tránsito de los buques a través del Canal.

Esta reducción en el tiempo se debería a la realización del trabajo de una manera cada vez más eficiente, debido fundamentalmente a un fenómeno definido como “aprendizaje”, cuya representación gráfica se denomina, a su vez, “curva de aprendizaje”. Por lo que la visualización, modelización estadística y estimación de esta curva define el proceso de aprendizaje en un proceso, circunscribiéndose en este caso al pilotaje y maniobras de la Autoridad del Canal para efectuar los tránsitos de buques.

Las curvas de aprendizaje o, también llamadas, las curvas de experiencia, se basan en la premisa de que las organizaciones, lo mismo que las personas, hacen mejor sus trabajos a medida que estos se van repitiendo. Una gráfica de curva de aprendizaje, de horas de mano de obra por unidad versus el número de unidades producidas, normalmente tiene la forma de la distribución exponencial negativa. Las curvas de aprendizaje pueden aplicarse tanto a individuos como a organizaciones. El aprendizaje individual debe considerarse vinculado a la mejora obtenida cuando los individuos repiten un proceso y adquieren habilidad, eficiencia o practicidad a partir de su propia experiencia. Al mismo tiempo, podemos considerar el aprendizaje organizacional como resultado de la práctica, pero proveniente de cambios en la administración, el equipo y el diseño de productos y procesos (Lefcovich, 2003).

Lefcovich (2003) indica que en el caso de los empleados de una empresa, ambos tipos de aprendizaje ocurren al mismo tiempo, y que muy a menudo el efecto combinado se describe en una sola curva de aprendizaje. Al considerar el proceso de tránsito de embarcaciones a través del Canal Ampliado, el aprendizaje individual se refiere a la habilidad, eficiencia o práctica adquirida, a partir de su propia experiencia, por los pilotos como individuos responsables del desempeño de la operación, mientras que el aprendizaje de la organización resulta de la experiencia de un grupo de personas que trabajan en el Centro de Control de Tráfico Marítimo, en el Departamento de Hidrología, remolcadores y esclusas.

Dicho todo lo anterior entendemos que al identificar patrones de aprendizaje y determinar los factores que influyen en los valores del tiempo de tránsito por las nuevas esclusas ayudará a mejorar la gestión del tránsito, facilitando también la adopción de medidas efectivas para la reducción del tiempo de tránsito, siempre y cuando se requiera.

### 1.2.2. Control y análisis de capacidad del proceso de tránsito a través del Canal Ampliado

La posición y variabilidad del tiempo de tránsito a través de las esclusas del Canal Ampliado, Cocolí y Agua Clara, debido a su relativamente nueva inauguración (en 2016), todavía no ha sido completamente estimada, al menos no en la medida en la que se ha hecho en el antiguo Canal. Es por ello que todavía no hay una normativa interna definitiva ni una política de incentivos laborales basada en unos tiempos de tránsito de referencia a través de las nuevas esclusas.

Por tanto, uno de los retos a abordar en este trabajo es la estimación de la variabilidad y posición del tiempo de tránsito a través de las esclusas del Canal Ampliado, incluyendo también la puesta en marcha o aplicación de herramientas de control de los procesos que definen el tránsito, que además permitan evaluar la capacidad de la organización de cumplir unas determinadas especificaciones de gerencia y/o clientes. En este caso particular, se propone abordar el problema desde el marco del control estadístico de la calidad. De este modo, se estimará la posición y dispersión de las variables críticas para la calidad (CTQ) de los procesos de tránsito de buques a través del Canal Ampliado, que son los tiempos de tránsito a través de las esclusas Cocolí y Agua Clara, a través de la aplicación de gráficos de control univariantes y multivariantes. Mediante este tipo de herramientas se estimará la variabilidad natural de las características CTQ cuando el proceso está bajo control (sin otra causa de variación aparte de la propia aleatoriedad característica del proceso), permitiendo la detección de cambios, ya sean positivos como el aprendizaje o negativos, como son las alarmas, averías u otro tipo de anomalías.

Por otro lado, gracias a la aplicación de los gráficos de control, se pueden determinar una serie de valores para el tiempo de tránsito que sirvan de base para establecer políticas de incentivos entre los pilotos y demás trabajadores del Canal. Por último, la determinación de la capacidad del Canal para cumplir especificaciones definidas por un límite en el tiempo de tránsito puede ser muy útil para la Autoridad del Canal a la hora de evaluar el desempeño de su actividad y la necesidad o no de implementar un proceso de mejora. Esta tarea se puede llevar a cabo mediante la aplicación de técnicas de análisis de capacidad de procesos univariantes y multivariantes, que se probarán por primera vez en este ámbito.

## 1.3. Métodos de resolución

Para el análisis se tomarán datos correspondientes a los tiempos de tránsito de distintos tipos de buque (los más comunes en el Canal Ampliado), véase Contenedores, GNL y GLP, a través de las rutas marítimas que utilizan el Canal de Panamá Ampliado, en particular, a través de las esclusas Cololí y Agua Clara, en las direcciones de Atlántico a Pacífico (Sur) o de Pacífico a Atlántico (Norte). Las operaciones en cada esclusa están tomadas en orden cronológico, en el intervalo comprendido entre 2016 y 2019.

En este proyecto se aplican herramientas del análisis exploratorio de datos, se ajustan modelos de regresión no paramétricos y paramétricos, simples y multivariantes, además de utilizarse técnicas del control estadístico de la calidad como son los gráficos de control y los índices de capacidad de procesos. Estas últimas técnicas son aplicadas tanto desde un enfoque univariante, para controlar y analizar el comportamiento de cada una de las variables de forma individual, como multivariante, teniendo en cuenta que el tránsito por el Canal Ampliado se puede definir por más de una variable, en concreto por los tiempos de tránsito correspondientes a cada una de las dos esclusas. Además, se evaluará el efecto de otros factores influyentes en el tiempo de tránsito como es la propia dirección de trayecto (Norte o Sur) y el tipo de buque que transita el Canal.

Todos los cálculos se implementan utilizando el software estadístico R, en concreto mediante el uso de, principalmente, los paquetes `ggplot2` (Wickham et al., 2016), `qcc` (Scrucca et al., 2017), `qcr` (Flores, 2020), `MVN` (Korkmaz et al., 2019), `MPCI` (Santos y Scagliarini, 2015).

## 1.4. Objetivos

Se pretende detectar patrones de aprendizaje correspondientes al proceso de tránsito de embarcaciones que transitan por las esclusas Cocolí y Agua Clara a través del nuevo Canal de Panamá, debido a la formación continua y a la acumulación de horas de navegación de los pilotos que guían las operaciones.

Se tienen los siguientes objetivos específicos que permiten lograr el objetivo general:

- Explicar brevemente sobre el Canal de Panamá Ampliado y las principales componentes que afectan el tránsito.
- Estudiar, describir y estimar el tiempo de tránsito de buques a través de las esclusas de las que se compone el Canal Ampliado, Cocolí y Agua Clara.
- Identificar los factores que determinan el tiempo de tránsito de buques a través de las nuevas esclusas, de qué manera y en qué medida.
- Aplicar y comparar distintas técnicas de control estadístico de la calidad para indentificar reducciones en el tiempo de tránsito en las dos esclusas.

En este capítulo se utilizaron principalmente las referencias Canal de Panamá (2006), Carral et al. (2021), Carral et al. (2019), Sabonge y Sánchez (2009), las páginas webs del Canal de Panamá <https://micanaldepanama.com/> y del Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá. <https://logistics.gatech.pa/es/>.



## Capítulo 2

# Métodos estadísticos: Control estadístico de calidad

En este trabajo se han aplicado diferentes técnicas estadísticas, desde el análisis exploratorio a la regresión paramétrica y no paramétrica. Estas técnicas son bien conocidas, sin embargo, son menos comunes aquellas relacionadas con el control estadístico de la calidad, por ser mucho más específicas de esta rama de la estadística. Por ello, para facilitar la comprensión del documento, se proporciona en este capítulo una introducción al concepto del control estadístico de la calidad, la metodología Seis Sigma para la mejora continua y todas aquellas técnicas que se utilizarán en este trabajo para el control y análisis de la capacidad del proceso de tránsito de los buques Neopanamax a través del Canal Ampliado.

La calidad se ha convertido en uno de los factores de decisión del consumidor más importantes en la selección entre productos y servicios competidores. En consecuencia, comprender y mejorar la calidad son factores claves que conducen al éxito empresarial, el crecimiento y una mayor competitividad. En este sentido, la Estadística juega un papel importante en todas las actividades empresariales relacionadas con la calidad ya que proporciona técnicas que se pueden aplicar en el control de la misma.

De acuerdo a Montgomery (2009) calidad es inversamente proporcional a la variabilidad, es decir, que si la variabilidad en las características importantes de un producto disminuye, la calidad del producto aumenta. Refiriéndose a una variabilidad no deseada.

El control de calidad es un conjunto de técnicas y actividades cuyo objetivo es asegurar que se mantiene la calidad de un producto o de un servicio y se reducen y eliminan posibles errores.

Por su parte, el control estadístico de la calidad (CEC) es el conjunto de herramientas y técnicas estadísticas utilizadas para medir la calidad y compararla con las especificaciones dadas por el cliente, con el fin de llevar a cabo acciones correctivas para su mejora (normalmente dirigidas a reducir la variabilidad). Se fundamenta en la propuesta realizada por el Dr. Walter Shewhart que los empleó por primera vez cuando trabajaba en la empresa Bell Telephone, en la década de los 20 del siglo pasado. Shewhart utilizaba los gráficos de control para identificar los problemas ocurridos durante el proceso que pueden ser por causas asignables o causas no asignables; su identificación daba como resultados la reducción de la variabilidad y la mejora del proceso.

Las causas de variabilidad no asignables, también llamadas aleatorias o comunes, son aquellas inherentes a la incertidumbre del proceso productivo, mientras que las causas asignables o especiales son las que producen efectos predecibles que aumentan la variabilidad y no son parte del propio proceso.

Es importante señalar la relación de las herramientas de control estadístico de calidad con la metodología Seis Sigma. Es una metodología de mejora de procesos que persigue la reducción de la variación, los defectos y otros tipos de errores para alcanzar objetivos como aumentar la cuota de mercado, minimizar los costes, incrementar la ganancia, la eficiencia, la imagen de marca, etc.

Promueve la utilización de herramientas y métodos estadísticos para la mejora de procesos, centrada en la reducción de la variabilidad. La meta de Seis Sigma es llegar a un máximo de 3.4 defectos por millón de eventos u oportunidades (DPMO) cumpliendo los requisitos del cliente.

Los elementos claves que soportan la filosofía Seis Sigma son los siguientes:

- Conocimiento de los requerimientos del cliente.
- Dirección basada en datos y hechos.
- Mejora de procesos.
- Implicación de la dirección.

La aplicación de la metodología Seis Sigma a un proceso consta de cinco etapas: Definir, Medir, Analizar, Mejorar y Controlar. Capacitando así a las empresas a diseñar productos y procesos que satisfagan las expectativas de los consumidores y que puedan ser producidas con los niveles de calidad seis sigma. Hoy en día, puede verse como un conjunto de métodos y procedimientos de aplicación de la estadística para la mejora y control de procesos productivos, de servicios o de I+D+I, desde un enfoque de CEC.

En la fase de control deben monitorizarse o someter a vigilancia los logros conseguidos en etapas anteriores, así como documentar las nuevas condiciones o especificaciones del proceso de estudio. Esta etapa está caracterizada por el empleo de herramientas basadas en la detección de errores y su corrección: las más representativas son los gráficos de control.

Luego de desarrollar algunos elementos básicos sobre el control estadístico de la calidad, se introducirá en este capítulo las técnicas estadísticas que se van a emplear para el análisis de los datos. Entre estas técnicas encontramos los gráficos de control del tipo univariante y multivariante, así como el análisis capacidad de procesos desde un punto de vista tanto univariante como multivariante.

## 2.1. Gráficos de Control

Los gráficos de control representan la herramienta más importante en el análisis de las variaciones de los procesos de producción y servicios.

Dentro del control estadístico de procesos la herramienta clave son los gráficos de control, en las que se representa la evolución del valor de un estadístico calculado sobre las mediciones de una característica de calidad efectuadas sobre los elementos de muestras sucesivas. Mediante los gráficos de control se estima la posición y variabilidad natural de la(s) variable(s) crítica(s) para la calidad del proceso, permitiendo su control y monitorización, además de la detección de anomalías y patrones que signifiquen un cambio en el mismo.

En el gráfico se van representando de forma correlativa los valores muestrales con objeto de poder monitorizar el proceso. Se acostumbra a unir los valores representados mediante líneas poligonales para poder identificar patrones indicadores de cambios significativos en el rendimiento del proceso. Gráficamente, se caracterizan por:

- Límite de control inferior (LCL), por debajo del cual es muy poco probable que se produzcan realizaciones de la variable.
- Límite de control superior (UCL). El LCL y UCL son simétricos si la distribución de probabilidad de la variable es simétrica (se suele utilizar la normal).
- Una línea central (CL). Media de las variables incluidas en la muestra. Los valores monitorizados varían en torno a esta media.

Conviene señalar que aún cuando todos los valores muestrales se encuentren en los límites de control pueden presentarse secuencias de valores por encima o por debajo de la línea central que indiquen claramente una

alteración en el proceso, para ello serán útiles realizar distintos contrastes que permitan detectar estas salidas de control.

Los gráficos de control se construyen en dos fases:

- En la fase I, los límites de control se estiman utilizando una muestra preliminar. Se utiliza un número  $m$  de muestras (normalmente  $m = 20$  o  $25$ ) cada una de un determinado tamaño,  $n$ , para establecer los límites de control y determinar de este modo si el proceso se encuentra bajo control en el momento en el que se recogen los datos. Los puntos que en esta fase no estén dentro de dichos límites deben ser examinados cuidadosamente y, si se encuentran causas asignables, eliminarlos del proceso. Una vez se eliminen deben ser recalculados los límites de control.
- En la fase II, se utiliza el gráfico de control para monitorizar y supervisar el proceso, comparando las muestras tomadas posteriormente con los límites de control ya calculados. Cuando las observaciones individuales de la variable de estudio,  $X$ , están dentro de los límites de control, se dice que el proceso está estadísticamente bajo control.

La mecánica de los gráficos de control puede ser vista como un contraste de hipótesis, donde la hipótesis nula,  $H_0$ , en CEC es que el proceso está bajo control. De lo contrario, si se rechaza esta hipótesis se aceptaría la alternativa,  $H_1$ , indicando que el proceso está fuera de control. Si se quiere verificar que el proceso está bajo control se verificaría para cada submuestra el correspondiente contraste (el establecido en  $H_0$ ).

En un contraste de hipótesis, rechazar la hipótesis nula cuando ésta es cierta se la denomina error de tipo I ( $\alpha$ ) o riesgo del vendedor. La decisión de no rechazar la hipótesis nula cuando es falsa se la denomina error de tipo II ( $\beta$ ) o riesgo del comprador.

Para medir la potencia del test se puede hacer uso de la curva característica de operación (OC, por sus siglas en inglés), ésta proporciona información acerca de la probabilidad de no detectar un cambio en el proceso cuando este ocurre de verdad. Esto se conoce generalmente como el error de tipo II, es decir, la probabilidad de aceptar erróneamente un proceso como en control.

Con respecto a la selección del tamaño de la muestra, suele hacerse uso de las curvas OC. A partir de estas curvas se obtiene las ARL (Average Run Length, por sus siglas en inglés) o longitud media de racha, que es el promedio de puntos que hacen falta representar en el gráfico (muestras a tomar) hasta que el primer punto muestre señal de fuera de control.

$$ARL = \frac{1}{1 - \beta}$$

Este corresponde a  $H_1$  que supone aceptar la asunción de que el proceso está bajo control cuando no lo está, por lo que se encuentra asociado a  $\beta$ .

El número de conforme hasta el primer fuera de control cuando el proceso está bajo control (dentro de la  $H_0$ ) viene dado por

$$ARL_0 = \frac{1}{\alpha}$$

Recordando que la probabilidad de rechazar dicha hipótesis cuando el proceso si está bajo control está asociada a  $\alpha$ .

El conocimiento de los valores de ARL para proceso bajo control y fuera de control es fundamental en el diseño de los gráficos de control de calidad. Pues debemos asegurarnos de tener una baja tasa de falsas alarmas (ARL bajo control alto) y, al mismo tiempo, rapidez suficiente para detectar los cambios (ARL fuera de control bajo).

Para detectar procesos fuera de control se pueden utilizar las reglas de la Western Electric (Western Electric Handbook, 1956) que señalan que un el proceso está fuera de control si:

1. Un punto cae fuera de los límites de control 3 sigma.
2. Dos de tres puntos consecutivos caen más allá de los límites 2 sigma.
3. Cuatro de cinco puntos caen a una distancia de 1 sigma o más allá de la línea central.
4. Ocho puntos consecutivos se encuentran en un mismo lado de la línea central.

No se suelen utilizar más de 3 reglas a la vez para evitar incurrir en falsos rechazos de la  $H_0$  (proceso bajo control).

### 2.1.1. Gráficos de control univariante

#### Gráficos de Control por variables

En el control estadístico de procesos, para el caso univariante, cuando la característica de interés es medible y representada por una variable aleatoria  $X$ , existen dos tipos de gráficos muy habitualmente utilizados, propuestos por Shewhart, por un lado los gráficos para monitorizar el promedio y controlar así la posición (los gráficos de media o de la mediana) y, por otro lado, los gráficos que controlan la dispersión de la variable (gráficos de desviaciones o de rango).

A continuación, se presentan gráficos de control para monitorizar la posición (tendencia centralidad) y variabilidad del proceso, donde se considerarán mejores estimaciones de la varianza (variabilidad).

Es importante señalar, para construir un gráfico de control de la media primero hay que observar los diagramas de rangos o desviaciones típicas (antes que los diagramas de la media). Es debido a que los UCL y LCL en el gráfico de control de la media se construyen a partir de la estimación de la desviación típica de la variable o característica estudiada:  $\hat{\sigma} = \frac{\bar{R}}{d_2}$ , obtenida usando el rango medio o  $\hat{\sigma} = \frac{\bar{s}}{c_4}$ , a partir del gráfico de desviaciones típicas muestrales.

Así pues, primero se establecerán los límites de control del gráfico de rangos (R) y, solo después, se hará lo mismo con el gráfico de medias.

#### ■ Gráficos de control $(\bar{x}, R)$

Uno de los gráficos de control de calidad para controlar la variabilidad del proceso es el basado en los rangos. El rango o recorrido es la diferencia entre la mayor y menor de las observaciones de una muestra. Sea  $x_1, x_2, \dots, x_n$  una muestra de  $n$  observaciones que sigue una distribución normal con media  $\mu$  y varianza  $\sigma^2$ . El rango será:

$$R = \max(x_i) - \min(x_i)$$

Tomando como línea central el valor esperado del rango muestral, que es  $d_2\sigma$ , donde  $d_2$  es una constante que depende del tamaño muestral,  $n$ , y que está tabulada.

Para construir los gráficos de control para la media se puede realizar una estimación de la varianza a través del rango puesto que  $\frac{R_i}{d_2}$  es un estimador insesgado de  $\sigma$  y, por tanto, su media (a lo largo de todas las muestras) también lo será. Definiendo

$$\bar{R} = \frac{\sum_i R_i}{k},$$

Se tiene entonces que podemos estimar  $\sigma$  mediante  $\frac{\bar{R}}{d_2}$  y construir los límites de control (llamados ahora gráficos  $(\bar{x}, R)$  mediante:

$$\bar{\bar{x}} \pm \frac{3\bar{R}}{\sqrt{nd_2}}$$

En donde  $\bar{\bar{x}}$  es la media global de todos los datos o, lo que es igual, la media de las medias de cada submuestra. De esta forma tendremos los límites de control:

$$LCL = \bar{\bar{x}} - A_2 \bar{R}$$

$$CL = \bar{\bar{x}}$$

$$UCL = \bar{\bar{x}} + A_2 \bar{R}$$

donde los valores de la constante  $A_2$  están tabulados para distintos valores de  $n$ .

■ **Gráficos de control ( $\bar{x}, \hat{s}$ )**

En este caso, en vez de utilizar el rango para la estimación de la desviación típica del proceso se utiliza directamente la desviación típica de cada submuestra.

Para la construcción del gráfico de medias se calculan la media y la desviación típica de cada muestra,  $\bar{x}_i$  y  $s_i$ , respectivamente ( $i = 1, \dots, k$ ). Seguidamente se calcula tanto la media como la desviación típica global.

La media global es la media de las medias muestrales, que puede calcularse promediando las medias de cada submuestra o promediando el total de los datos ( $kn$ ).

$$\bar{\bar{x}} = \frac{\sum_{i=1}^k \bar{x}_i}{k} = \frac{\sum_{i=1}^k \sum_{j=1}^k x_{ij}}{nk}$$

La media global es un estimador insesgado de la media teórica,  $\mu$ . La desviación típica muestral es un estimador sesgado (aunque asintóticamente insesgado) de la desviación típica teórica, ya que  $E(s_i) = c_4 \sigma$ , donde  $c_4$  es una constante dependiente de  $n$ .

$$\hat{\sigma} = \frac{\bar{s}}{c_4} = \frac{\sum_i s_i}{c_4 k}$$

como estimador (centrado) de la desviación típica teórica.

Sin embargo en los paquetes estadísticos que se utilizarán se implementan las cuasidesviaciones típicas para la estimación de la varianza,  $\hat{s}_i$ , en lugar de las desviaciones típicas muestrales.

$$\hat{\sigma} = \frac{\bar{\hat{s}}}{c_4} = \frac{\sum_i \hat{s}_i}{c_4 k}$$

El valor  $c_4$  viene dado por la expresión:

$$c_4 = \sqrt{\frac{2}{n-1} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}}$$

Usando cuasidesviaciones típicas, los límites de control de  $3\sigma$  para la media son:

$$LCL = \bar{\bar{x}} - A_3 \bar{\hat{s}}, \quad CL = \bar{\bar{x}}, \quad UCL = \bar{\bar{x}} + A_3 \bar{\hat{s}},$$

en donde

$$A_3 = \frac{3}{\sqrt{nc_4}}$$

Se pueden definir los parámetros del gráfico de desviaciones típicas mediante los estadísticos

$$LCL = B_3\bar{\hat{s}}, \quad CL = \bar{\hat{s}}, \quad LCL = B_4\bar{\hat{s}}$$

Lo que se hace es multiplicar la media de las cuasidesviación típica por  $B_3$  y  $B_4$ , las cuales están tabuladas, y cuyas expresiones son  $B_3 = 1 - \frac{3}{c_4}\sqrt{1 - c_4^2}$  y  $B_4 = 1 + \frac{3}{c_4}\sqrt{1 - c_4^2}$

### ■ Gráficos de medidas individuales

Además de los casos mencionados anteriormente, donde se tiene  $m$  muestras de tamaño  $n$  cada una, se utilizan también los gráficos de medidas individuales, en el supuesto que las submuestras utilizadas sean de un solo valor, es decir, el tamaño muestral  $n$  es 1.

En este trabajo se aplicarán fundamentalmente estos gráficos (junto con los gráficos EWMA y CUSUM), ya que se trabajará con datos tomados cada cierto tiempo, formando una única observación por muestra, de hecho, es complejo agrupar los diferentes buques en submuestras racionales, pues aquellos que pasan por el Canal en un mismo día, ni son un número constante ni por supuesto tienen que ser de la misma tipología. El problema en este tipo de estudio consiste en estimar la varianza, para ello pueden emplearse dos métodos: el de rangos móviles o la agrupación por bloques.

El método de rangos móviles consiste en agrupar  $n$  medias individuales consecutivas con intención de formar una submuestras que permita estimar los rangos. Una vez creado el primer grupo, se irán agrupando sucesivamente las muestras incorporando la nueva observación y eliminando la primera de cada muestra.

Por otra parte, cuando exista alguna razón que pueda justificar la agrupación de los datos en bloques puede recurrirse a formar los grupos con estos datos.

Para estimar los límites de control en este caso se procede como con el gráfico de la media pero tomando el tamaño de submuestras igual a 1. La estimación de  $\sigma$  puede hacerse utilizando la media de las desviaciones típicas  $\bar{s}_n$ , la media de los rangos  $\bar{R}$ , o la media de las cuasidesviaciones  $\bar{S}_{n-1}$

$$\bar{\bar{x}} \pm Z_{\alpha/2} \frac{\bar{s}_n}{c_{2,n}}$$

$$\bar{\bar{x}} \pm Z_{\alpha/2} \frac{\bar{R}}{d_{2,n}}$$

$$\bar{\bar{x}} \pm Z_{\alpha/2} \frac{\bar{s}_{n-1}}{c_{4,n}}$$

donde  $c_2$ ,  $d_2$  y  $c_4$  son constantes que dependen del tamaño muestral,  $n$ , y que están tabuladas.

En el supuesto más clásico de usar criterios  $3\sigma$  suele utilizarse unas constantes  $E_1$ ,  $E_2$ ,  $E_3$ , resultando las fórmulas para gráficos individuales:

$$\bar{\bar{x}} \pm E_1\bar{s}_n$$

$$\bar{\bar{x}} \pm \bar{R}$$

$$\bar{\bar{x}} \pm E_3\bar{s}_{n-1}$$

Al igual que ocurría en los gráficos de variables, también en el caso de medidas individuales debe hacerse el gráfico para las desviaciones o recorridos, con objeto de controlar la variabilidad de los datos y no solo el valor nominal.

Otras alternativas para los gráficos de medidas individuales son los CUSUM y los EWMA. Los gráficos de sumas acumuladas o gráficos CUSUM y los gráficos con medias ponderadas exponencialmente o gráficos EWMA no se basan en las observaciones individuales, o promedios de una muestra de ellas, sino en la acumulación de la información. Por esta razón se les denomina también gráficos con memoria, los cuales veremos a continuación.

### Gráficos de control CUSUM

El gráfico de sumas acumuladas o gráficos CUSUM fue propuesto por Page (1954). Es la representación de las sumas acumuladas de las desviaciones entre cada dato, en este caso la media muestral ( $\bar{x}_i$ ), y un valor de referencia ( $\mu_0$ ).

Usará el estadístico  $s_r$ .

$$s_r = \sum_{i=1}^r (\bar{X}_i - \mu_0).$$

Este estadístico tiene en cuenta las  $r$  primeras submuestras para tomar la decisión sobre la muestra  $r$ -ésima.

Bajo supuestos de normalidad la distribución del estadístico es la siguiente:

$$s_r \in N \left( r(\mu - \mu_0), \frac{r\sigma_0^2}{n} \right).$$

Cuando sea cierta la hipótesis nula  $\mu = \mu_0$ , las sumas acumuladas  $s_r$  tendrán media 0, con lo que es de esperar que oscilen alrededor de la recta  $y = 0$ . En el caso en que la hipótesis no sea cierta los valores graficados de las sumas  $s_r$  serán crecientes o decrecientes.

El objetivo de este tipo de gráfico de control es controlar la evolución de una variable  $X \in N(\mu, \sigma)$  donde  $\sigma$  es conocida o se posee una estimación.

El CUSUM algorítmico (o tabular) calcula las desviaciones de cada valor respecto del valor objetivo o nominal  $\mu$ . Se distinguirá entre desviaciones positivas y negativas, puesto que en muchas situaciones ambas desviaciones no tienen la misma repercusión.

Es necesario definir un valor  $K$  a partir del cual se considera que la desviación acumulada es significativa. Este valor determinará la sensibilidad del gráfico de control. Si la suma acumulada hasta la observación  $i$ -ésima es menor que cierto umbral  $K$ , se considera que la desviación acumulada es cero.

Existen dos tipos de sumas significativas acumuladas:  $C_i^+$  para las desviaciones positivas y  $C_i^-$  para las negativas:

- $C_i^+$  acumulación de desviaciones positivas significativas:

$$C_i^+ = \max\{0, C_{i-1}^+ + (X_i - \mu) - K\}$$

De esta forma, si el proceso está bajo control, la variable  $X$  tomará valores cercanos al nominal y el término  $C_i^+$  tenderá a tomar valores nulos.

- $C_i^-$  = acumulación de desviaciones negativas significativas:

$$C_i^- = \min\{0, C_{i-1}^+ - (X_i - \mu) - K\}$$

El valor de  $K$  se suele elegir en función de la desviación que se quiera detectar. Supongamos que el valor nominal de la variable de interés es  $\mu$ .

Los valores de las sumas acumuladas se han de comparar con el valor máximo admisible. Este valor, denominado valor de decisión  $H$  suele ser  $H = h\sigma$ , donde se recomienda que  $h = 5$  y por tanto  $H = 5\sigma$  (algunos autores también recomiendan valores de  $h = 4$ ). Los valores  $\pm H$  son los límites de control de este gráfico.

### Gráficos de control EWMA

El gráfico de medias móviles con ponderación exponencial o gráfico EWMA (Exponentially Weighted Moving Average, por sus siglas en inglés), fue propuesto por primera vez por Roberts (1959). En este tipo de gráfico las muestras más cercanas tienen un mayor peso en el cálculo de la media para un punto dado.

El gráfico EWMA es muy utilizado en el caso de observaciones individuales. Las observaciones individuales pueden ser medias (cuando las observaciones individuales de las que provienen las medias no están disponibles), lecturas individuales, cocientes, proporciones o medidas similares.

A diferencia de los gráficos de control de Shewhart, el gráfico EWMA es un gráfico con memoria, ya que utiliza la información de las medias de las observaciones pasadas, dando más peso a las medias más cercanas. De esta forma, si lo que interesa detectar son cambios pequeños en la característica de calidad estudiada puede ser más recomendable utilizar este tipo de gráficos, ya que son más sensibles a los mismos al acumular información de muchos períodos, de modo que el cambio podría apreciarse.

El estadístico es la media ponderada de la observación actual y todas las observaciones precedentes, donde la media previa recibe más peso, es decir,

$$Z_t = \lambda x_t + (1 - \lambda)Z_{t-1}, \quad 0 < \lambda < 1.$$

Donde:

$Z_0 = \mu$  el valor inicial que se necesita cuando  $t = 1$  es la media del proceso.

$Z_t$  media móvil ponderada exponencialmente en el tiempo actual  $t$ .

$Z_{t-1}$  media móvil ponderada exponencialmente en el tiempo inmediatamente precedente  $t - 1$ .

$x_t$  = la observación presente.

$\lambda$  es el factor de ponderación para la observación actual.

Si se cumplen las hipótesis de independencia y estabilidad sobre los  $x_t$ , se tiene que:

$$E(Z_t) = \mu,$$

y

$$Var(Z_t) = \sigma^2 \frac{\lambda(1 - \lambda)^{2t}}{2 - \lambda}.$$

Cuando  $t$  aumenta, el último término entre paréntesis en la parte derecha converge rápidamente hacia uno, y la correspondiente expresión para la varianza se transforma en  $Var(Z_t) = \sigma^2[\lambda/(2 - \lambda)]$ .

Eligiendo  $\lambda = 2/(t + 1)$ , la expresión aproximada de la varianza se transforma en  $Var(Z_t) \approx \sigma^2/t$  (varianza de medias de muestras de  $t$  elementos).

Bajo estas condiciones, los límites de control se transforman en  $\hat{\mu} \pm 3\sqrt{\frac{\hat{\sigma}^2}{t}}$ . Para otros valores de  $\lambda$ , los límites de control son:

$$LSC = \hat{\mu} + 3\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$LIC = \hat{\mu} - 3\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

Para las primeras observaciones, se debe utilizar la primera ecuación para la varianza. Si no se tiene una buena estimación para  $\sigma$ , se debe utilizar un gráfico de recorridos y estimar por  $\frac{\bar{R}}{d_2\sqrt{n}}$ . En el caso de valores individuales, el recorrido móvil medio se puede utilizar en la forma en que se hizo para los gráficos de control de valores individuales.

El gráfico EWMA puede expresarse:

$$Z_{t,k} = \sum_{i=1}^k (W_i \bar{X}_{t-i+1} + a)$$

En donde  $t$  es el instante actual,  $k$  es el número de medias consideradas en cada punto y que se llama orden del gráfico,  $w_i$  son los  $k$  pesos considerados y  $a$  es una constante que puede ser nula.

- Si  $k = 1$ ,  $w_i = 1$  y  $a = 0$  obtendríamos el gráfico de Shewart para la media.
- Si  $w_i = 1$  y  $a = -k\mu_0$  se obtiene el gráfico CUSUM.

Al igual que el gráfico CUSUM, el gráfico EWMA es más efectivo que el gráfico de medias para detectar pequeños cambios en la media de la característica medida y, por tanto, en la calidad del proceso de producción; sin embargo, ambos gráficos funcionan peor que el gráfico de medias para grandes cambios.

### 2.1.2. Gráficos de control multivariante

La idea de desarrollar gráficas de control univariantes para controlar el estado de cada una de las características de calidad observadas de forma independiente, resulta engañosa. Normalmente las variables de un proceso complejo suelen estar muy relacionadas entre sí y con este procedimiento se perdería el estudio de las correlaciones latentes del proceso.

Actualmente es muy frecuente que se precise estudiar más de una variable de manera simultánea. Esta parte del control de la calidad, que tiene por objeto el estudio conjunto de datos multidimensionales, se conoce como control multivariante.

La mayor parte de las técnicas del control multivariante de procesos han sido desarrollados a lo largo del siglo XX. Inicia con los trabajos de Harol Hotelling, que estudió las miras de los aviones de combate que intervinieron en las batallas aéreas de la segunda guerra mundial (1947), hasta las aportaciones de matemáticos como el hindú Mahalanobis entre otros. La mayoría de las aportaciones tuvieron como hilo conductor crear extensiones de gráficos y métodos de control del caso univariante al caso multidimensional.

Supondremos un vector de  $p$  componentes,  $\vec{X} = (X_1, X_2, \dots, X_p)$  y sería un conjunto de  $p$  variables aleatorias unidimensionales definidas sobre el mismo espacio de probabilidad.

La esperanza del vector aleatorio  $\vec{X}$ , también llamada vector de medias, es el vector formado por las medias individuales:

$$E(\vec{X}) = \begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix} = \vec{\mu}.$$

La extensión del concepto de varianza de una variable aleatoria real al caso multidimensional se lleva a cabo mediante la llamada matriz de varianzas-covarianzas, definida por

$$Cov = (\vec{X}) = \sum = (\sigma_{ij}) = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1p} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \cdots & \sigma_{pp} \end{pmatrix} = E = \left( (\vec{X} - \vec{\mu})(\vec{X} - \vec{\mu})^t \right),$$

donde  $\sigma_{ij} = Cov(X_i, X_j)$ .

Con respecto a la estimación del vector de medias y de la matriz de varianzas-covarianzas, supongamos que se dispone de una muestra de datos  $x_{ijk}$  correspondientes a la observación  $i$ -ésima ( $i = 1, 2, \dots, p$ ) de la  $j$ -ésima característica de calidad ( $j = 1, 2, \dots, p$ ) en la  $k$ -ésima muestra ( $k = 1, 2, \dots, m$ ), que pueden estructurarse mediante la matriz.

El vector de medias poblacional,  $\vec{\mu}$ , puede estimarse mediante el llamado vector de medias muestral dado por:

$$\bar{x}_{jk} = \frac{1}{n} \sum_{j=1}^n x_{ijk},$$

y la estimación de la varianza para cada elemento del vector aleatorio viene dada por:

$$s_{jk}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ijk} - \bar{X}_{jk})^2.$$

La matriz de varianzas-covarianzas poblacional puede estimarse también mediante su análoga muestral al igual que el vector de medias. Así dadas las covarianzas entre la característica de calidad  $j$ -ésima y la característica de calidad  $h$ -ésima para la  $k$ -ésima muestra podrá estimarse mediante la expresión:

$$s_{jhk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ijk} - \bar{x}_{jk})(x_{ihk} - \bar{x}_{hk}), \quad k = 1, 2, \dots, m; j \neq h.$$

En base a los anteriores estimadores se definen los siguientes estimadores:

$$\bar{\bar{x}}_j = \frac{1}{m} \sum_{k=1}^m \bar{X}_{jk}, \quad j = 1, 2, \dots, p.$$

$$\bar{\bar{s}}_j^2 = \frac{1}{m} \sum_{k=1}^m s_{jk}^2, \quad j = 1, 2, \dots, p.$$

$$\bar{\bar{s}}_{jh} = \frac{1}{m} \sum_{k=1}^m s_{jhk}, \quad j \neq h.$$

Estos cálculos pueden reescribirse en forma matricial:

$$\bar{\bar{x}} = \begin{pmatrix} \bar{\bar{x}}_1 \\ \vdots \\ \bar{\bar{x}}_p \end{pmatrix},$$

$$S = \begin{pmatrix} \bar{s}_1^2 & \bar{s}_{12} & \bar{s}_{13} & \dots & \bar{s}_{1p} \\ & \bar{s}_2^2 & \bar{s}_{23} & \dots & \bar{s}_{2p} \\ & & \bar{s}_3^2 & \ddots & \vdots \\ & & & \ddots & \bar{s}_p^2 \end{pmatrix}_{p \times p}.$$

La distribución multivariante más utilizada (como ya ocurría también en el caso univariante) es la normal multidimensional. Se trata de una distribución continua, que se denota por  $N_p(\vec{\mu}, \Sigma)$ , cuya densidad viene dada por

$$f(\vec{x}) = (2\pi)^{-p/2} \det(\Sigma)^{-1/2} \exp \left[ -\frac{1}{2} (\vec{x} - \vec{\mu})^t \Sigma^{-1} (\vec{x} - \vec{\mu}) \right],$$

siendo  $\vec{\mu}$  un vector de  $p$  componentes y  $\Sigma$  una matriz de dimensión  $p \times p$ , simétrica y definida positiva. Donde  $\vec{\mu}$  resulta ser el vector medias de esta distribución y que  $\Sigma$  es la matriz de varianzas-covarianzas.

Los gráficos de control multivariantes parten de la hipótesis de distribución de normalidad multivariante para las variables CTQ estudiadas. Una forma de sobrellevar estas limitaciones es mediante el control no paramétrico de procesos multivariantes, que se basa en el concepto de profundidad de datos.

### La $T^2$ Hotelling

Hotelling en 1947 (Montgomery, 2009) desarrolló el estadístico denominado  $T^2$  y su extensión a los gráficos de control, esta distribución permite analizar variables multidimensionales en las que exista dependencia entre ellas.

Consideremos un proceso de control en el que se observan  $p$  variables que se distribuyen según una Normal  $p$ -variante con un vector de medias  $\mu_0$  y una matriz  $\Sigma_0$  de varianzas-covarianzas según una distribución normal multidimensional  $N_p(\mu_0, \sigma_0)$ .

Podemos construir el estadístico:

$$T^2 = n(\bar{X} - \mu_0)^t \Sigma_0^{-1} (\bar{X} - \mu_0)$$

que representa la distancia de Mahalanobis entre el vector que representa la media muestral y el vector que representa la media poblacional.

En el caso univariante se calculaban dos límites de control, pero en el caso multivariante se calcula un único límite de control superior dado por  $LCS = \chi_{\alpha, p}^2$

El valor del estadístico  $T^2$  se estimará mediante los correspondientes valores muestrales:

$$T^2 = n(\bar{X} - \bar{\bar{X}})^t S^{-1} (\bar{X} - \bar{\bar{X}})$$

La distribución que sigue en este caso es la de una  $F$  de Snedecor con  $p$  y  $(mn - m - p + 1)$  grados de libertad, para un nivel de significación del  $\alpha$ .

Al igual que antes, y dado que tanto la distribución  $\chi^2$  y la  $F$  son asimétricas con sesgo a la derecha, se tienen los límites:

$$LCS = \frac{p(m-1)(n-1)}{mn - m - p + 1} F_{\alpha, p, mn - m - p + 1}$$

$$LCI = 0$$

En el caso de observaciones individuales, es decir  $n = 1$ , suponiendo que se tienen un total de  $m$  muestras, para evaluar  $p$  variables de control, el correspondiente gráfico de control se construirá mediante:

$$LCS = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha, p, m-p}$$

$$LCI = 0$$

Además, para procesos en los que el  $m$  es suficientemente grande,  $m > 100$ , se puede aproximar estos límites por  $LCS = \chi_{\alpha, p}^2$ .

Al igual que sucede en el caso univariante, cuando se precisa detectar desplazamientos pequeños con respecto a la media del proceso, existen alternativas a los gráficos  $T^2$  de Hotelling, dentro de los denominados gráficos con memoria. Las opciones más utilizadas son las versiones multivariantes de los gráficos CUSUM y EWMA, que se introducen a continuación.

### Gráfico MCUSUM

Los gráficos CUSUM pueden extenderse también al caso multivariante, aunque no hay una única forma de hacerlo. Según Hamed et al. (2016), los primeros en realizar estudios en la materia fueron Woodall y Ncube en 1985, usando un esquema basado en múltiples ( $p$ ) CUSUM univariantes.

Healy (1987) desarrolló un modelo de gráfico MCUSUM, diseñado de forma que detectara los cambios lo más rápidamente posible, teniendo al mismo tiempo el mejor comportamiento posible en términos del recorrido medio de racha (ARL):

$$G_i = \text{máx}\{(G_{i-1} + a^t(x_i - \mu_0) - 0.5\lambda\mu_1), 0\}.$$

A continuación se muestra el significado de cada uno de sus elementos constitutivos:

$\mu_0$ : Vector de medias cuando el proceso está bajo control.

$\Sigma_0$ : Matriz de varianzas-covarianzas bajo control.

$\mu_1$ : Vector de medias cuando el proceso no está bajo control.

$\lambda\mu_1$ : Raíz cuadrada del parámetro de descentrado.

Finalmente,  $a^t$  es el cociente  $A/\lambda\mu_1$ , siendo  $A = (\mu_1 - \mu_0)^t \Sigma_0^{-1}$ .

Para interpretar el gráfico se considera que se produce una señal de alarma cuando  $G_i$  supera un cierto valor de referencia  $H$ , convenientemente elegido y que se encuentra tabulado.

Hawkins (1991) desarrolla un modelo para variables ajustadas mediante regresión multivariante, en base a la suposición de que cuando se produce un cambio en la media, se produce de una determinada manera. Concretamente se asume que el cambio se debe a una variación de magnitud  $\delta$  en la media de una de las variables. Otras propuestas de gráficos MCUSUM son los de Croisier (1988) o los de Pignatiello y Runger (1990).

### Gráfico MEWMA

El gráfico de medias móviles ponderadas exponencialmente para el caso multivariante (MEWMA) es una extensión a esta nueva situación del correspondiente gráfico univariante. La extensión multivariante, propuesta por Lowry et al. (1992), adopta la forma:

$$Z_i = \Lambda X_i + (I - \Lambda)Z_{i-1}.$$

Donde:

$X_i$ : Vector de medias muestrales.

$\Lambda$ : Matriz diagonal formada por los valores  $\lambda$  para las distintas variables.

$I$ : Matriz identidad.

Como valor inicial se toma  $Z_0 = 0$ , mientras que la información que proporcionan los  $Z_i$  se recoge en el estadístico:

$$T_i^2 = Z_i^t \Sigma_{Z_i}^{-1} Z_i,$$

siendo  $\Sigma_{Z_i}^{-1}$  la inversa matriz de varianzas-covarianzas de los  $Z_i$ .

En lo que respecta al límite de control (superior) empleado, en Alfaro-Navarro (2002) se indica que Runger y Prabhu (1996) plantean una aproximación que permite estudiar el desempeño del gráfico MEWMA en términos del ARL. Proporcionan además una guía para la selección del límite superior de control, dependiente del cambio en la media que queramos detectar, del valor lambda seleccionado y del número de variables que estemos analizando.

## 2.2. Análisis de capacidad de procesos

Además de controlar todas las mejoras conseguidas en un proceso mediante la utilización de gráficos de control, es necesario evaluar en qué grado todas las mejoras implementadas ayudan a cumplir con las especificaciones marcadas por gerencia, clientes o normativa. De esta tarea se ocupa el denominado análisis de capacidad de procesos.

### 2.2.1. Análisis de capacidad univariante

Se define el análisis de capacidad de un proceso como un estudio estadístico y de ingeniería orientado a estimar su capacidad, es decir, la cuantificación de la variabilidad del proceso y el análisis de esta variabilidad en comparación con las especificaciones del producto final.

El análisis de la capacidad de un proceso es vital dentro de los programas de mejora de la calidad y, por extensión, de la metodología Seis Sigma, específicamente en las etapas Analizar y Mejorar.

Si bien en el análisis de capacidad de procesos se utilizan herramientas como son los histogramas y los gráficos qqplot, para verificar el cumplimiento de las hipótesis de partida y para comparar la distribución de las observaciones con los límites de especificación, las principales herramientas de este tipo de análisis son los denominados índices de capacidad. Para calcular los índices de capacidad de un proceso se han de tener en cuenta dos premisas:

- El proceso ha de estar bajo control.
- Se supone que se conoce la distribución de probabilidad de la característica de calidad del proceso. La suposición más común es que los datos se distribuyen según una distribución normal.

Los índices más utilizados son los índices  $C_p$ ,  $C_{pu}$ ,  $C_{pl}$ ,  $C_{pk}$ ,  $C_{pm}$  y  $C_{pmk}$ , que se definirán a continuación. Estos índices dependerán del límite inferior de especificación ( $LSL$ ) y el límite superior de especificación ( $USL$ ).

El índice  $C_p$  es un índice de primera generación que se define como:

$$C_p = \frac{USL - LSL}{6\sigma} = \frac{USL - LSL}{X_{0.99865} - X_{0.00135}}.$$

Atendiendo al valor del índice de capacidad, se presentan tres casos diferenciados:

- $C_p < 1$ . El proceso no puede cumplir las especificaciones requeridas. Cuanto más pequeño sea el índice de capacidad mayor sería la proporción de piezas, productos o servicios no conformes.
- $C_p = 1$ . El proceso fabricará aproximadamente un 0.3% de defectuosos. Un tipo de proceso que cumpliera esta característica se consideraba justamente apto para la fabricación. No obstante, en la actualidad este porcentaje de defectuosos puede ser inaceptable en muchos productos (cuyas proporciones de defectuosos se cuentan en uno o dos por millón).
- $C_p > 1$ . En este caso el porcentaje de productos defectuosos es realmente muy pequeño, aunque esto siempre es relativo ya que depende del sector industrial concreto. Cuando  $C_p = 2$ , se considera que el proceso estudiado es capaz a un nivel Seis Sigma.

En la práctica  $\sigma$  es casi siempre desconocida, por lo que se sustituye por su estimador,  $\hat{\sigma}$ :

$$\hat{C}_p = \frac{USL - LSL}{6\hat{\sigma}}.$$

El índice  $C_p$  se puede definir como la medida de la capacidad potencial del proceso.  $C_p$  estima el potencial del proceso, pues no tiene en cuenta la localización del mismo (en términos de la distribución de probabilidad de la característica de calidad).

Además, también se define el índice,  $C_{pk}$ , índice de segunda generación que describe el desempeño real de la capacidad del proceso. Además de la variabilidad, también tiene en cuenta la localización del proceso (la media de su característica de calidad) con respecto a los límites de especificación (tolerancia). Este índice solo tiene en cuenta el desplazamiento con respecto a los límites de especificación como indica su expresión. Se construye a partir de los índices de capacidad inferior ( $C_{pl}$ ) y superior ( $C_{pu}$ ):

$$C_{pu} = \frac{USL - \mu}{3\sigma},$$

$$C_{pl} = \frac{\mu - LSL}{3\sigma},$$

$$C_{pk} = \min\{C_{pu}, C_{pl}\}.$$

Ya que en la mayor parte de los casos no se conocen  $\sigma$  ni  $\mu$ , se deben estimar, teniendo entonces:

$$\hat{C}_{pu} = \frac{USL - \hat{\mu}}{3\hat{\sigma}},$$

$$\hat{C}_{pl} = \frac{\hat{\mu} - LSL}{3\hat{\sigma}},$$

$$\hat{C}_{pk} = \min\{\hat{C}_{pu}, \hat{C}_{pl}\}.$$

La combinación de los índices  $C_p$  y  $C_{pk}$  se utiliza para comparar, tanto la capacidad potencial del proceso (hasta dónde puede llegar en lo que a variabilidad se refiere), como la estimación de la capacidad real. Mientras más parecidos sean, más centrado estará el proceso dentro de los límites de especificación.

Se suelen utilizar también los índices de segunda generación como el  $C_{pm}$ , que considera el desplazamiento de la media del proceso con respecto al *Target* (T) o valor nominal. Los índices de capacidad de tercera generación miden la variabilidad del proceso se mide según dos componentes, la propia varianza de la variable CTQ y el sesgo al cuadrado con respecto al *target* u objetivo.

$$C_{pm} = \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}},$$

$$\hat{C}_{pm} = \frac{USL - LSL}{6\sqrt{\hat{\sigma}^2 + (\hat{\mu} - T)^2}}.$$

En Pearn et al. (1992) se introducen índices de tercera generación como el  $C_{pmk}$ , nace como corrección del  $C_{pk}$ , con el motivo de aumentar la sensibilidad de los desplazamientos de la media del proceso con respecto al *Target*.

$$C_{pkm} = \frac{d - |\mu - m|}{3\sqrt{\sigma^2 + (\mu - T)^2}},$$

$$\hat{C}_{pkm} = \frac{d - |\mu - m|}{3\sqrt{\hat{\sigma}^2 + (\hat{\mu} - T)^2}}.$$

Siendo  $d$  la mitad de la amplitud del intervalo de especificación  $d = \frac{USL - LSL}{2}$  y  $m$  el punto medio de dicho intervalo  $m = \frac{USL + LSL}{2}$ .

Por otro lado, en Vännman (1995) se propone generalizar en una expresión dependiente de los parámetros  $u$  y  $v$  todos los índices anteriores:

$$C_p(u, v) = \frac{d - u \cdot |\mu - m|}{3\sqrt{\hat{\sigma}^2 + v \cdot (\mu - T)^2}}$$

$$C_p(0, 0) = C_p, \quad C_p(1, 0) = C_{pk}, \quad C_p(0, 1) = C_{pm}, \quad C_p(1, 1) = C_{pkm}.$$

Cuando  $C_p$  y  $C_{pk}$  son iguales, el proceso está centrado, la media de la distribución coincide con el *Target*. Los procesos se empiezan a considerar capaces con  $C_p > 1$ ; cuando  $C_p = 1$  la amplitud de tolerancia coincide con la distancia  $6\sigma$ . Aunque hay autores que ponen este límite en 1.33 (menos de 96 no conformes por millón, ppm) cuando se dan dos límites de tolerancia y 1.25 para límites unilaterales.

Además, desde la metodología Six Sigma estos índices deberían ser como mínimo de 2 para considerarse un proceso como capaz. Se resumen los valores mínimos y los errores por millón que equivalen a cada valor en la siguiente tabla:

Procesos	Especificaciones_bilaterales	Especificaciones_unilaterales
Procesos existentes	1.33	1.25
Procesos nuevos	1.50	1.45
Procesos existentes (seguridad, resistencia, parámetros críticos)	1.50	1.45
Procesos nuevos (seguridad, resistencia, parámetros críticos)	1.67	1.60

Tabla 2.1 Valores mínimos recomendados del índice de capacidad para considerar un proceso capaz.

Valores de 1.33, 1.50 denotan procesos capaces según los estándares establecidos (Tabla 1.1). El valor de 2 es el buscado en un proceso Six Sigma (0.0018 ppm).

Con respecto a los intervalos de confianza, dado que el valor de los valores de los índices de capacidad dependen de la muestra, se pueden entender como estadísticos. Por ello es conveniente hacer las estimaciones de los índices poblacionales mediante intervalos de confianza para tener una idea acerca de la precisión de la estimación. Si se asume la normalidad de la distribución de la variable aleatoria medida, se obtienen las siguientes expresiones:

$$\hat{C}_p \sqrt{\frac{\chi_{1-\frac{\alpha}{2}, n-1}^2}{n-1}} \leq C_p \leq \hat{C}_p \sqrt{\frac{\chi_{\frac{\alpha}{2}, n-1}^2}{n-1}}$$

### 2.2.2. Análisis de capacidad multivariante

De forma breve, a continuación, se presenta la notación, los antecedentes teóricos y algunas de las soluciones multivariantes existentes. Para aplicarlas, se utilizará el paquete MPCÍ (Santos y Scagliarini, 2015).

Sea  $X$  el vector de  $v$  características de calidad de interés con vector de media  $\mu$  y matriz de covarianza  $\Sigma$ . Suponemos que la distribución de probabilidad conjunta de las características de calidad  $v$  es la distribución normal multivariante,  $X \sim N(\mu, \Sigma)$ .

Vamos a denotar mediante  $LSL^\top = [LSL_1, LSL_2, \dots, LSL_v]$ ,  $USL^\top = [USL_1, USL_2, \dots, USL_v]$  y  $T^\top = [T_1, T_2, \dots, T_v]$ , a los valores de los vectores  $v$  con los límites de especificación inferior, los límites de especificación superior y los valores objetivo, respectivamente.

#### Vector multivariante de capacidad

El vector de capacidad multivariante fue propuesto por Shahriari et al.(1995), basado a su vez en el trabajo en el trabajo original de Hubele, Shahriari y Cheng (1991). El vector contiene tres componentes y se denota como  $(C_{pM}, PV, LI)$ .

$C_{pM}$  es la primera componente del vector, representando una relación de volúmenes. El numerador es el volumen definido por la región de tolerancia de ingeniería, mientras que el denominador es el área o volumen de una “región de proceso modificada”, definida como la región más pequeña de forma similar a la región de tolerancia de ingeniería, circunscrita alrededor de un contorno de probabilidad:

$$C_{pM} = \left[ \frac{Vol(\text{Región de Tolerancia})}{Vol(\text{Región de Proceso Modificado})} \right]^{1/v}$$

El volumen de la región de tolerancia de ingeniería es

$$\prod_{i=1}^v (USL_i - LSL_i).$$

Para calcular el volumen de la región del proceso modificado, es importante recordar que bajo la hipótesis de normalidad multivariante, el estadístico

$$(X - \mu)^\top \Sigma (X - \mu) = g(x)$$

sigue una distribución  $\chi^2$  con  $v$  grado de libertad. Por lo tanto, los límites de la región de proceso superior,  $UPL_i$ , e inferior,  $LPL_i$ , con  $(i = 1, 2, \dots, v)$  se derivan de la proyección del elipsoide de probabilidad sobre los respectivos ejes. Específicamente, se determinan resolviendo los sistemas de ecuaciones de la primera derivada, con respecto a cada  $X_i$  de la forma cuadrática (Nickerson 1994):

$$(X - \mu)^\top \Sigma (X - \mu) = \chi^2(v, \alpha),$$

en la que  $\chi^2_{(v, \alpha)}$  es el percentil  $100(1 - \alpha)$ -ésimo de una distribución  $\chi^2$  con  $v$  grados de libertad asociados con el contorno de probabilidad. Por lo general, en analogía con el “ $6\sigma$ ” en el denominador de índices univariados,  $\alpha = 0.0027$ . Las soluciones (dos para cada dimensión  $i$ ) de la ecuación anterior son:

$$UPL_i = \mu_i + \sqrt{\frac{\chi^2_{(v, \alpha)} \det(\Sigma_i^{-1})}{\det(\Sigma^{-1})}}$$

y

$$LPL_i = \mu_i - \sqrt{\frac{\chi_{(v,\alpha)det(\Sigma_i^{-1})}^2}{det(\Sigma^{-1})}}$$

Donde  $i = 1, 2, \dots, v$  y  $det(\Sigma_i^{-1})$  es el determinante de una matriz obtenida a partir de  $\Sigma^{-1}$  al eliminar la  $i$ -ésima fila y columna. Por tanto, el volumen de la región de proceso modificada es:

$$\prod_{i=1}^v (UPL_i - LPL_i).$$

En la práctica  $\mu$  y  $\Sigma$  se desconocen. Dada una muestra aleatoria,  $X_1, X_2, \dots, X_n$ , de tamaño  $n$  del proceso, a estimar  $\mu$  y  $\Sigma$  la media muestral, siendo

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i,$$

mientras que la matriz de covarianza de la muestra es

$$S = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})(X_i - \bar{X})^\top$$

Los valores de  $C_{pM}$  superiores a 1 indican que la región del proceso modificado es más pequeña que la región de tolerancia de ingeniería, por lo que tenemos una alta probabilidad de que los artículos producidos se clasifiquen como conformes.

Con respecto al segundo vector  $PV$ , se define como el nivel de significación del estadístico  $T^2$  de Hotelling, calculado bajo el supuesto de que el centro de las especificaciones de ingeniería se considera la verdadera media subyacente del proceso:

$$PV = Pr \left( F_{(v, n-v)} > \frac{(n-v)}{v-(n-1)} T^2 \right),$$

siendo

$$T^2 = n(\bar{X} - \mu)^\top S^{-1}(\bar{X} - \mu)$$

y  $F_{(v, n-v)}$  la distribución  $F$  con  $v$  y  $n-v$  grados de libertad. Valores de  $PV$  cercanos a cero indica que el centro del proceso está lejos del valor objetivo de ingeniería.

La tercera componente del vector,  $LI$ , resume una comparación de la ubicación de la región de proceso modificada y la región de tolerancia. Indica si alguna parte de la región del proceso modificado queda fuera de las especificaciones de ingeniería. Tiene un valor de 1 si toda la región de proceso modificada está contenida dentro de la región de tolerancia y, en caso contrario, un valor de 0.

### El índice de capacidad multivariante MCpm

El índice  $MC_{pm}$  fue propuesto por Taam et al.(1993) y se define como una relación de dos volúmenes. El numerador es el volumen de la región de tolerancia modificada  $R_1$  y el denominador es el volumen de la escala del 99.73% de la región  $R_2$  del proceso. Bajo la hipótesis de normalidad multivariante,  $R_2$  es una región de proceso elipsoidal, mientras que la región de tolerancia modificada  $R_1$  es el elipsoide más grande que está centrado en el objetivo completamente dentro de la región de tolerancia original. En el caso general de las  $v$  características, la región  $R_1$  es un hiperelipsoide con un volumen dado por:

$$Vol(R_1) = \frac{2\pi^{v/2} \prod_{i=1}^v a_i}{v\Gamma(\frac{v}{2})}$$

donde  $a_i (i = 1, 2, \dots, v)$  son las longitudes de los semiejes.

Entonces, el índice de capacidad multivariante se escribe como

$$MC_{pm} = \frac{Vol(\text{Región de Tolerancia Modificada})}{Vol((X - \mu)^\top \Sigma_T^{-1} (X - \mu) \leq K(m))}$$

donde  $X$  es el vector  $(v \times 1)$  de mediciones de una distribución normal multivariante con vector medio  $\mu$  y matriz de covarianza  $\Sigma$ ,  $\Sigma_T = E[(X - T)(X - T)^\top]$  es la matriz de error cuadrático medio del proceso,  $T$  es un vector de valores objetivo, y  $K(m)$  es un percentil 99.73 de una  $\chi^2$  con  $v$  grados de libertad.

$MC_{pm}$  se puede reescribir como:

$$MC_{pm} = \frac{Vol(R_1)}{Vol(R_3)} \times \frac{1}{[1 + (\mu - T)^\top \Sigma^{-1} (\mu - T)]^{1/2}} = \frac{C_p}{D}$$

Es una función de dos componentes:  $C_p$  representa la variabilidad del proceso en relación con las regiones de tolerancia modificadas;  $D$ , detecta el proceso de desviación del objetivo. Dada una muestra aleatoria de  $n$  mediciones  $(X_1, X_2, \dots, X_n)$ , para cada una de las dimensiones  $v$ , el estimador para  $MC_{pm}$  es dado por:

$$\hat{M}C_{pm} = \frac{\hat{C}_p}{\hat{D}}$$

Cuando el vector medio del proceso es igual al vector objetivo, y el índice tiene el valor 1, entonces el 99.73 % de los valores del proceso se encuentran dentro de la región de tolerancia modificada.

En este capítulo se utilizaron principalmente las referencias Alfaro-Navarro (2002), Bass (2007), Kane (1986), Mitra (2016), Montgomery (2009), Prat et al. (1998), Santos y Scagliarini (2012), Vännman (1995).

## Capítulo 3

# Análisis exploratorio y estimación del efecto del aprendizaje

En este capítulo se realiza un estudio descriptivo de los datos y se estima el efecto del entrenamiento en el tiempo de tránsito de los buques a través del Canal, teniendo en cuenta también los posibles efectos del tipo de barco, la esclusa, la dirección de tránsito y variables relacionadas a la dimensión del buque, como la manga y la eslora.

En la primera parte del capítulo se estima la curva de aprendizaje del tiempo de tránsito del buque Baroque, navegado a través de la esclusa Agua clara. Este buque, de tipo Bulk carrier (granelero), fue utilizado en un programa de aprendizaje operativo como formación para los pilotos. A continuación, pasamos al análisis y estimación de las curvas de aprendizaje correspondientes al tiempo de operación real para cruzar las esclusas del Canal Ampliado de Panamá, Agua Clara (Atlántico) y Cocolí (Pacífico). Se realiza un análisis descriptivo y exploratorio de los datos, por último se utiliza un modelo de regresión multivariado para definir de qué forma y en qué medida influyen cada una de las variables anteriormente mencionadas (dirección de tránsito, tipo de buque, dimensiones del buque y esclusa) en el tiempo de tránsito.

### 3.1. Descripción de datos

#### 3.1.1. Base de datos correspondiente al buque Baroque

La apertura del Canal de Panamá Ampliado requirió emprender un proceso de entrenamiento o aprendizaje previo, impartido a un grupo limitado de pilotos experimentados. Como complemento al entrenamiento personal, se desarrolló un programa de aprendizaje operativo que se implementó en fechas inmediatamente posteriores a la inauguración.

Se destinó al granelero Baroque para la formación de la organización. Esta embarcación fue maniobrada por un grupo de pilotos que estaban totalmente certificados para el canal existente, a través de la esclusa Agua Clara. Se realizaron un total de 44 tránsitos, desde el 25 de junio al 5 de agosto de 2016.

La base de datos contiene 44 observaciones y 3 variables.

```
'data.frame': 40 obs. of 3 variables:
 $ Date          : chr  "07-25-16" "06-25-16" "06-29-16" "06-29-16" ...
 $ Days_of_Experience: num  1 1 4 4 4 6 6 7 7 7 ...
 $ Transit_Time   : num  270 240 237 239 205 235 207 252 225 174 ...
```

- **Date:** Corresponde a la fecha en que se realiza el tránsito.

- **Days\_of\_Experience:** Es el tiempo (medido en días) transcurrido desde el comienzo de las actividades de capacitación.
- **Transit\_Time:** Contiene información sobre el tiempo de tránsito (en minutos) que tarda el buque Baroque en cruzar la esclusa Agua Clara en el Canal de Panamá Ampliado.

### 3.1.2. Base de datos correspondiente a los datos reales

Los datos reales a estudiar contienen un total de 818 observaciones (cada una correspondiente a un tránsito de un buque a través de una de las esclusas del Canal Ampliado) y 8 variables (que caracterizan a cada uno de los tránsitos).

```
'data.frame': 818 obs. of  8 variables:
 $ Fecha      : chr  "28/6/16" "28/6/16" "30/6/16" "1/7/16" ...
 $ lock       : Factor w/  2 levels "Agua_Clara","Cocoli": 2 2 2 2 2 2 2 2 2 ...
 $ direction  : Factor w/  2 levels "North","South": 1 1 1 1 1 1 1 1 1 ...
 $ Type       : Factor w/  3 levels "CONTENEDORES",...: 3 3 3 1 1 3 1 1 1 ...
 $ Time       : num  1 1 3 4 6 6 7 10 11 12 ...
 $ transit_time: int  161 191 158 199 201 169 216 211 122 258 ...
 $ B          : num  120 120 120 158 131 ...
 $ LOA        : num  739 755 739 1106 998 ...
 - attr(*, "na.action")= 'omit' Named int [1:8] 154 167 251 359 512 720 733 817
 ..- attr(*, "names")= chr [1:8] "154" "168" "252" "360" ...
```

La variable **transit\_time**, que será la variable crítica para medir la calidad de cada uno de los tránsitos, contiene información sobre el tiempo de tránsito (en minutos) que tardan los buques en cruzar las esclusas Cocolí y Agua Clara, del Canal de Panamá Ampliado.

Con el objetivo de identificar y analizar los factores que influyen en el tiempo de tránsito de los buques por las esclusas, se estudia la variable **transit\_time** con respecto a las siguientes variables:

- **Fecha:** Fecha en que se realiza el tránsito, desde junio del 2016 hasta octubre 2019.
- **lock:** Esclusas del Canal de Panamá Ampliado.

-Cocolí

-Agua\_Clara

- **direction:** Dirección hacia donde se dirigen los barcos.

-Sur: De Atlántico a Pacífico

-Norte: De Pacífico a Atlántico

- **Type:** Tipo de barco.

-Contenedores

-LNG: Cargan Gas Natural Licuado (GNL)

-LPG: Cargan Gas Licuado de Petróleo (GLP)

Según informaciones obtenidas en la página web de la Autoridad del Canal de Panamá (2020), estos son los tres tipos principales de embarcaciones que cruzan el Canal de Panamá Ampliado: los “Portacontenedores” conforman el 50.3% del total de buques, mientras que los buques de GLP y GNL representan el 32.3% y 9.1% respectivamente. El porcentaje restante lo componen los graneleros, portavehículos, refrigerados, tanqueros y de pasajeros.

- **Time:** Tiempo (medido en días) transcurrido desde la inauguración del Canal de Panamá Ampliado.
- **B:** Manga del buque.
- **LOA:** Eslora del buque (length overall).

## 3.2. Estimación de las curvas de aprendizaje

### 3.2.1. Curva de aprendizaje para las instalaciones y pilotos maniobrando el buque Baroque

Inicialmente se desconoce el efecto de la experiencia (medida como el tiempo transcurrido desde que se comienza el proceso de entrenamiento) sobre el tiempo de tránsito de los buques por las nuevas esclusas. Este efecto de la experiencia también se conoce como curva de aprendizaje (Carral et al., 2021). Por tanto, con el fin de estimar este efecto aprendizaje sobre el tiempo en tránsito del buque, se propone el ajuste de un modelo de regresión GAM (Wood, 2017), a partir de una base de thin plate splines. La aplicación de este modelo permite estimar la magnitud y el tipo de efecto (lineal o no lineal) del efecto del tiempo de experiencia.

Suponiendo que la variable respuesta, el tiempo de tránsito, se distribuye normalmente, se estima un modelo GAM en función del tiempo de experiencia. Abajo se muestra un resumen con las estimaciones del modelo, el correspondiente análisis de significación del efecto de la experiencia y las medidas de bondad de ajuste. Mientras que en la Figura 3.1 se representa el efecto suavizado de la variable experiencia. Se han incluido las observaciones reales y el intervalo de estimación (para la media condicionada) a un nivel de confianza del 95%.

```

Family: gaussian
Link function: identity

Formula:
Transit_Time ~ s(Days_of_Experience)

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   194.80      3.18    61.2   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df   F p-value
s(Days_of_Experience) 2.9   3.61 10.9 9.1e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.497  Deviance explained = 53.4%
GCV = 448.7  Scale est. = 404.96    n = 40

```

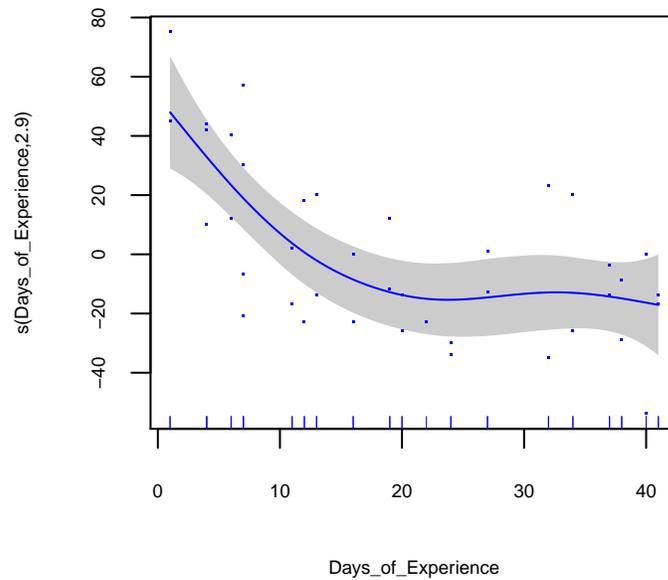


Figura 3.1: Dispersión del tiempo de tránsito de la embarcación Baroque frente al tiempo desde el inicio del entrenamiento.

Al estudiar la curva de aprendizaje de la formación de los pilotos maniobrando el buque Baroque se confirma una mejora en el pilotaje a lo largo del tiempo. Se puede observar una tendencia decreciente de tipo asintótica, no lineal. Esto quiere decir que a medida que pasa el tiempo desde que iniciaron las capacitaciones, la velocidad a la que disminuye el tiempo de tránsito se hace cada vez menor, hasta alcanzar una zona de saturación. Esta tendencia puede ser debida a que, al principio, hay un amplio margen de mejora en el pilotaje y las maniobras a efectuar. Pero según va pasando el tiempo, con las capacidades de pilotos e instalaciones, las posibilidades de mejora disminuyen hasta llegar a un límite de tiempo marcado por la asíntota horizontal.

En el resumen del modelo se muestra que el efecto de la experiencia sobre el tiempo de tránsito es significativo ( $P\text{-value} < 0.05$ ). El coeficiente de determinación es  $R^2 = 53.4\%$ , es decir, el tiempo transcurrido desde el comienzo de las actividades de capacitación explican el 53.4% de la variabilidad del tiempo de tránsito a través de la esclusa.

### 3.2.1.1. Estimación de la posición y variabilidad del tiempo de tránsito

Se aplican gráficos de control para estimar la posición y variabilidad del tiempo de tránsito del buque Baroque a través de la esclusa Agua Clara. Además, la aplicación de gráficos de control tipo Shewhart para medidas individuales, en la Fase I, nos permite estimar si el proceso de tránsito a través de la esclusa está bajo control.

Luego de comprobar que se cumple la hipótesis de normalidad e independencia entre observaciones, mediante la aplicación de los contrastes de Shapiro-Wilks, Lilliefors, gráfico qqplot, Box-Pierce y Ljung-Box, a un nivel de significación del 0.05, supuestos necesarios para que se apliquen de manera fiable los gráficos de control, aplicamos la técnica univariante Shewart, comprobando que el proceso está bajo control.

En la Figura 3.2 se contruye un gráfico de control para medidas individuales, a partir de los 25 primeros tránsitos a través de la esclusa (muestra de calibrado). Se observa que el proceso no está bajo control, presenta

una observación fuera de los límites naturales de control y tres rachas. Las observaciones fuera de control se eliminan y los límites de control natural se vuelven a calcular hasta que no se detectan más valores atípicos. Seguidamente se procede a aplicar la Fase II de los gráficos de control, es decir, se monitorizan el resto de tránsitos (muestra de monitorizado) y se comparan con los límites de control estimados en la etapa anterior (Figura 3.3). Se observa un patrón de cambio paulatino de nivel, apareciendo observaciones fuera de control. El proceso ha cambiado, siendo la causa asignable el aprendizaje de pilotos e instalaciones. Por tanto, el siguiente paso sería volver a calcular los límites de control naturales con la muestra de monitorizado.

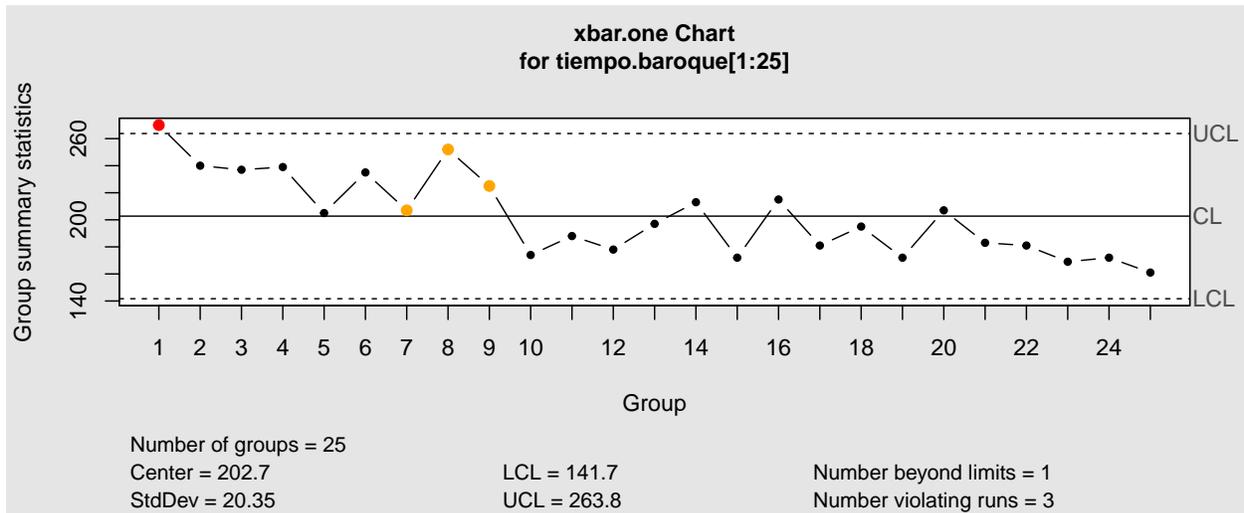


Figura 3.2: Gráfico de control para medidas individuales del tránsito del buque Baroque: Fase I, estimación de los límites de control a partir de la muestra de calibrado.

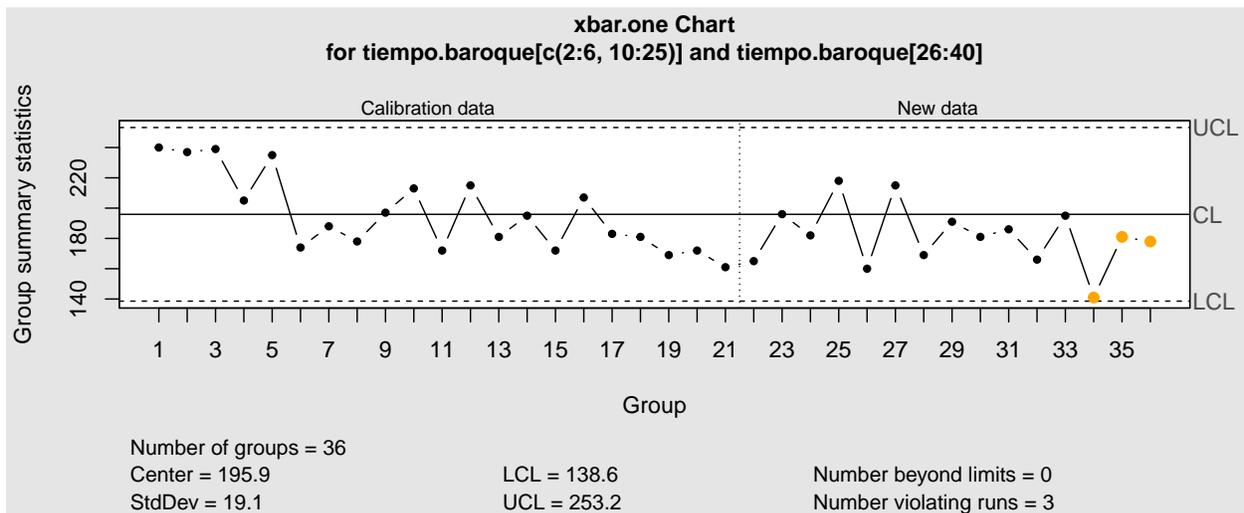


Figura 3.3: Gráfico de control para medidas individuales del tránsito del buque Baroque: Fase II, monitorización de tránsitos comparando su escala con los límites de control natural estimados en la Fase I.

En el capítulo 4 se explicarán de forma más detallada los pasos dados para realizar los gráficos de control implementados a la base de datos real.

### 3.2.2. Estimación de la curva de aprendizaje correspondiente al tiempo de operación real

Una vez indentificado un efecto significativo de la experiencia o aprendizaje en el tiempo de tránsito del buque Baroque, se extiende este análisis a los tránsitos reales a través del Canal Ampliado.

En la primera parte de esta sección hacemos un estudio descriptivo y exploratorio de todas las variables. Además, se estudia la dependencia entre el tiempo de tránsito y las diversas variables cualitativas, incluyendo un análisis de la varianza del tiempo de tránsito en función de estos factores.

En la segunda parte se ajusta un modelo un GAM para definir de qué forma y en qué medida influyen cada una de las variables en el tiempo de tránsito. La elección de este tipo de modelo se debe a la suposición de que el efecto de la experiencia puede ser de tipo no lineal, asintótico, por lo que precisamos un modelo flexible que, a parte de ser capaz de estimar el efecto de factores como la dirección o el tipo de barco, permita incluir efectos suaves de las variables explicativas cuantitativas.

#### 3.2.2.1. Análisis exploratorio de datos

##### Tipo de buque

El 61.61 % de los buques que pasaron a través del Canal en el período de estudio corresponde a los Contenedores, seguidos por los que cargan GLP con un 28.36 %, correspondiendo el resto de tránsitos a los GNL, cuyo análisis estadístico no será siempre posible debido al pequeño número de tránsitos realizados por este tipo de barcos (ver Figura 3.4).

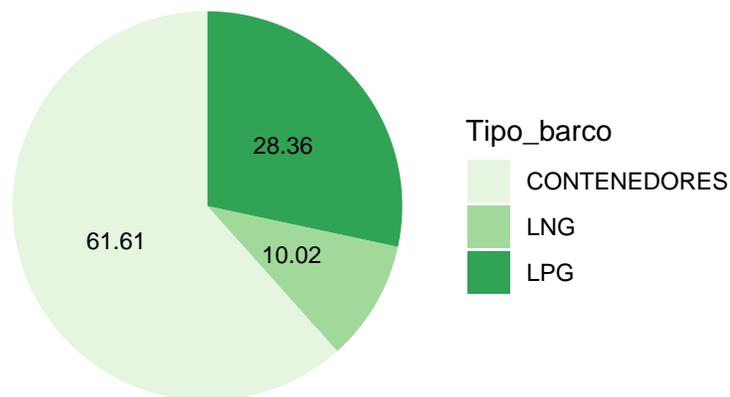


Figura 3.4: Tipo barco.

##### Tipo de dirección

Con respecto al factor dirección de tránsito, la Figura 3.5 muestra que los barcos recorrieron en mayor medida la dirección Pacífico a Atlántico, representando un 62.84 % de los tránsitos. El hecho de que la dirección norte

sea predominante puede tener su causa en la importancia del tráfico marítimo que arranca desde China y otros países asiáticos como Corea y Japón (Carral et al., 2018).

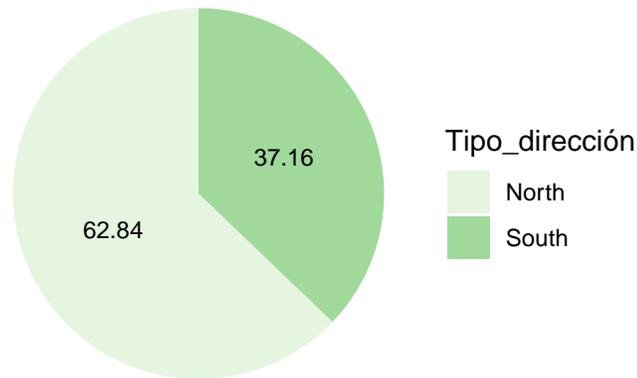


Figura 3.5: Tipo dirección.

### VARIABLES CUANTITATIVAS: Eslora, Manga, Tiempo de tránsito

Cuadro 3.1: Resumen estadístico de las variables Manga, Eslora y Tiempo de Tránsito

	media	median	sd
B	139.3	142.5	16.06
LOA	951.1	984.0	146.74
Transit_Time	154.4	151.0	35.17

La Tabla 3.1 muestra que el tiempo de tránsito a través de una esclusa, definido a partir de su media y su desviación típica, es 154.4+-35.17 min, mientras que la eslora es 951.1+-146.74 ft y la manga 139.3+-16.06 ft. Hay una mucho menor dispersión en la manga que en la eslora, dado que ésta es la principal restricción de paso a través del Canal. De hecho, los buques se diseñan con el objeto de poder atravesar esta instalación.

### Tiempo de tránsito en función de las variables cualitativas: tipo de barco, tipo de esclusa y dirección

Se presenta un pequeño resumen descriptivo que consta del cálculo de medidas de dispersión y posición del tiempo de tránsito en función del valor de las variables cualitativas dirección, esclusa y tipo de barco. El objetivo es observar si existen diferencias en el tiempo de tránsito dependiendo de los niveles de los mencionados factores.

En los resultados anteriores se verifica que entre los tiempos de tránsito según el tipo de esclusa y el tipo de dirección, las diferencias son pequeñas, si acaso el tiempo medio de tránsito a través de la esclusa Cocolí es mayor que el tiempo medio necesario para cruzar la esclusa de Agua Clara. No ocurre lo mismo con el tipo de embarcación. El tiempo de tránsito medio de los Contenedores es claramente el mayor, seguido por

Cuadro 3.2: Resumen estadístico del tiempo de tránsito en función de la esclusa.

lock	transit_time.mean	transit_time.median	transit_time.sd
Agua Clara	153.4	148	35.20
Cocoli	155.5	153	35.15

Cuadro 3.3: Resumen estadístico del tiempo de tránsito en función de la dirección.

direction	transit_time.mean	transit_time.median	transit_time.sd
North	154.4	150.5	35.21
South	154.5	151.0	35.16

Cuadro 3.4: Resumen estadístico del tiempo de tránsito con respecto al tipo de barco

Type	transit_time.mean	transit_time.median	transit_time.sd
CONTENEDORES	162.7	162.0	33.03
LNG	128.9	130.5	26.17
LPG	145.4	141.5	36.03

el correspondiente a los buques GLP. Los buques GNL son a los que les corresponde un tiempo medio más bajo. De acuerdo a Carral et al. (2017) estos resultados se deben a las condiciones especiales de tránsito de los buques GNL y GLP, además, de sus diferentes dimensiones y carga. Asimismo, en Carral et al. (2019) señalan que los buques de GNL tienen preferencia durante su tránsito por el Canal de Panamá, por lo que el tiempo de tránsito esperado debería ser más corto.

En la Figura 3.6 los diagramas de caja, que incluyen un intervalo de confianza para la mediana, además de las observaciones reales (puntos en negro), indican que existen diferencias en la mediana y la dispersión del tiempo de tránsito según los valores de tipo de barco, esclusa y dirección del barco. En todas las combinaciones y a los niveles de factores el tiempo de tránsito de los buques Contenedores es mayor, seguido por el correspondiente a los GLP, y a los GNL, en ese orden.

Los efectos en el tiempo de tránsito de las variables dirección y esclusa no son tan claros como en el caso del tipo de embarcación. La razón es que se superponen con el efecto de su iteración. Se puede verificar el efecto del cruce de estas variables por medio del diagrama de caja en la Figura 3.7.

La Figura 3.7 muestra que los tiempos de tránsito más altos corresponden a la combinación de, por un lado, la esclusa Agua Clara y la dirección Sur y, por otro lado, la esclusa Cocolí y la dirección Norte. En otras palabras, parece que las embarcaciones tienden a tardar más en pasar por las esclusas cuando provienen directamente del mar. De acuerdo a Carral et al. (2017) una causa asignable de esta tendencia son todas las maniobras necesarias para pasar del océano al Canal de Panamá Ampliado.

Esta información se completa con los resultados del resumen estadístico de las medidas de posición y dispersión del tiempo de tránsito para cada combinación de los factores esclusa, dirección y tipo de embarcación, resultados mostrados en la Tabla 3.5.

### **Análisis de la varianza**

Para probar si estas diferencias en el tiempo son estadísticamente significativas, se realiza un análisis ANOVA de tres factores (dirección de tránsito, la esclusa y tipo de embarcación) y una interacción (entre dirección y esclusa). Esta interacción se ha incluido teniendo en cuenta la información que se muestra en la Figura 3.7,

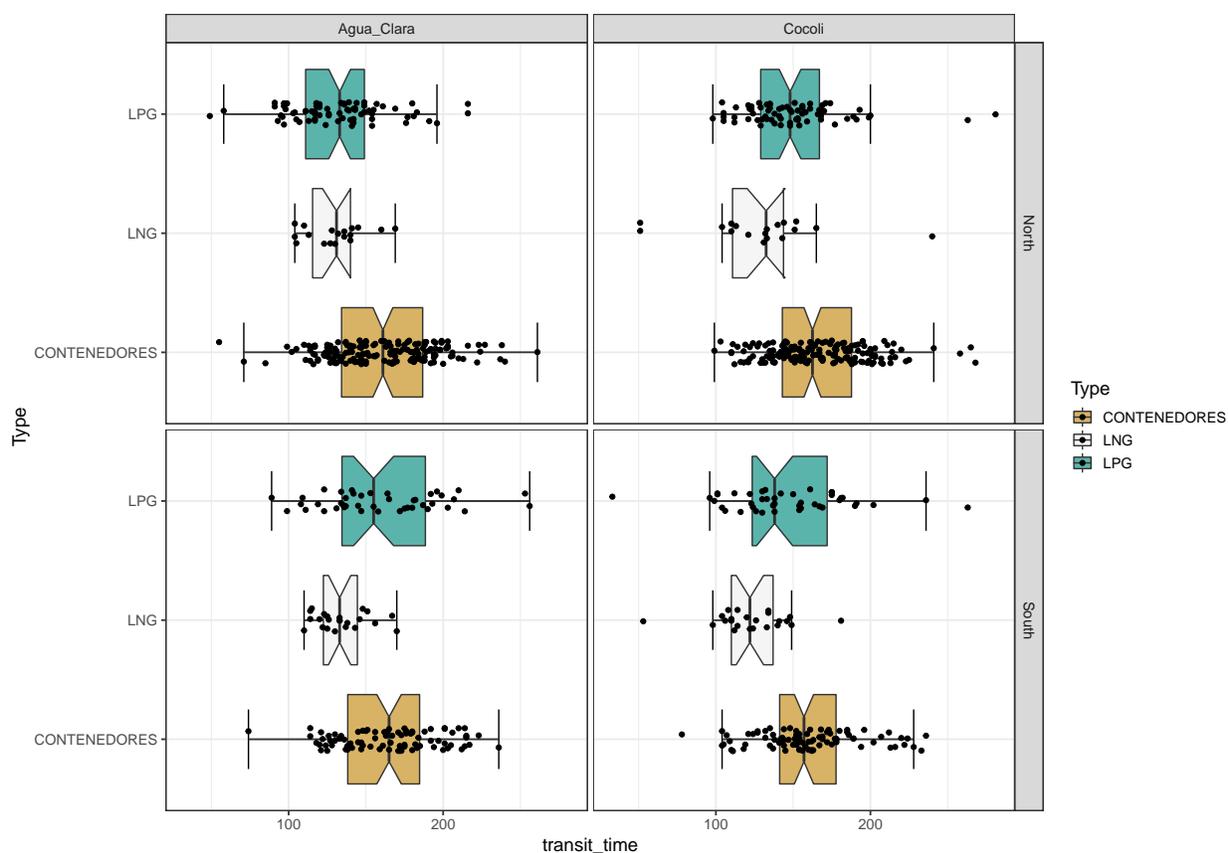


Figura 3.6: Diagramas de caja para el tiempo de tránsito en función del tipo de embarcación, esclusa y dirección de tránsito.

Cuadro 3.5: Resumen estadístico del tiempo de tránsito para cada combinación de esclusa, dirección y tipo de embarcación.

lock	direction	Type	transit_time.mean	transit_time.sd	transit_time.median
Agua_Clara	North	CONTENEDORES	160.5	34.64	161.0
Agua_Clara	North	LNG	130.2	18.42	131.0
Agua_Clara	North	LPG	132.0	32.03	133.0
Agua_Clara	South	CONTENEDORES	164.2	31.95	165.0
Agua_Clara	South	LNG	133.9	16.58	133.0
Agua_Clara	South	LPG	160.7	38.60	155.0
Cocoli	North	CONTENEDORES	165.8	32.05	162.5
Cocoli	North	LNG	129.1	41.34	132.5
Cocoli	North	LPG	149.4	31.59	148.0
Cocoli	South	CONTENEDORES	159.7	32.75	157.0
Cocoli	South	LNG	122.9	24.49	122.0
Cocoli	South	LPG	146.3	40.02	138.0

donde se visualiza que el efecto de la dirección del tránsito en el tiempo varía según la esclusa; por lo tanto, se espera un efecto de interacción significativo.

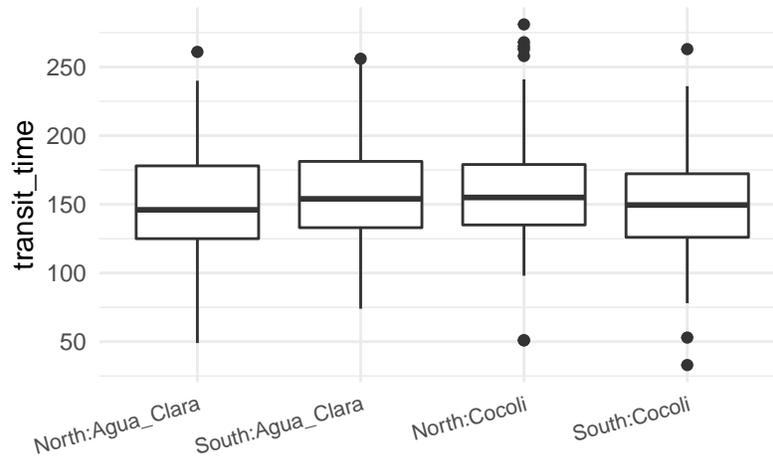


Figura 3.7: Tiempo de tránsito en función interacción de la esclusa y la dirección.

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Type	2	106521	53261	48.67	< 2e-16 ***
direction	1	1507	1507	1.38	0.24090
lock	1	932	932	0.85	0.35642
direction:lock	1	13092	13092	11.96	0.00057 ***
Residuals	812	888527	1094		

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Se implementa la prueba de hipótesis basada en la distribución  $F$  y, como resultado, los efectos del tipo de buque y la interacción de la dirección con la esclusa en la respuesta del tiempo de tránsito son estadísticamente significativos a un nivel de confianza del 95 % (p-valores < 0.05).

Por otra parte, no resultan estadísticamente significativos los efectos principales esclusa y dirección.

### Correlación entre las variables correspondientes a las dimensiones del buque y el tiempo de tránsito por las diferentes esclusas

Se utiliza la función `ggpairs()` del paquete `GGally` (Schloerke et al., 2021). Se puede sacar desde el propio R mediante `library(GGally)` que permite obtener salidas gráficas que, simultáneamente, incluyen una matriz de diagramas de dispersión, la estimación tipo núcleo de las funciones de densidad de las variables cuantitativas estudiadas y, además, los coeficientes de correlación de Pearson correspondientes a cada par de variables.

La Figura 3.8. muestra que existe una correlación lineal entre débil y moderada entre los tiempos que tarda un buque en pasar por la esclusa de Cocolí y la de Agua Clara. Entre las variables de las dimensiones, manga y eslora existe una asociación alta de 0.77, lo que tiene sentido.

En lo que se refiere al grado de dependencia lineal entre los tiempos de tránsito por ambas esclusas y las dimensiones de los buques, es relativamente bajo, siendo más alto cuando se relaciona la eslora con los tiempos de tránsito a través de las esclusas.

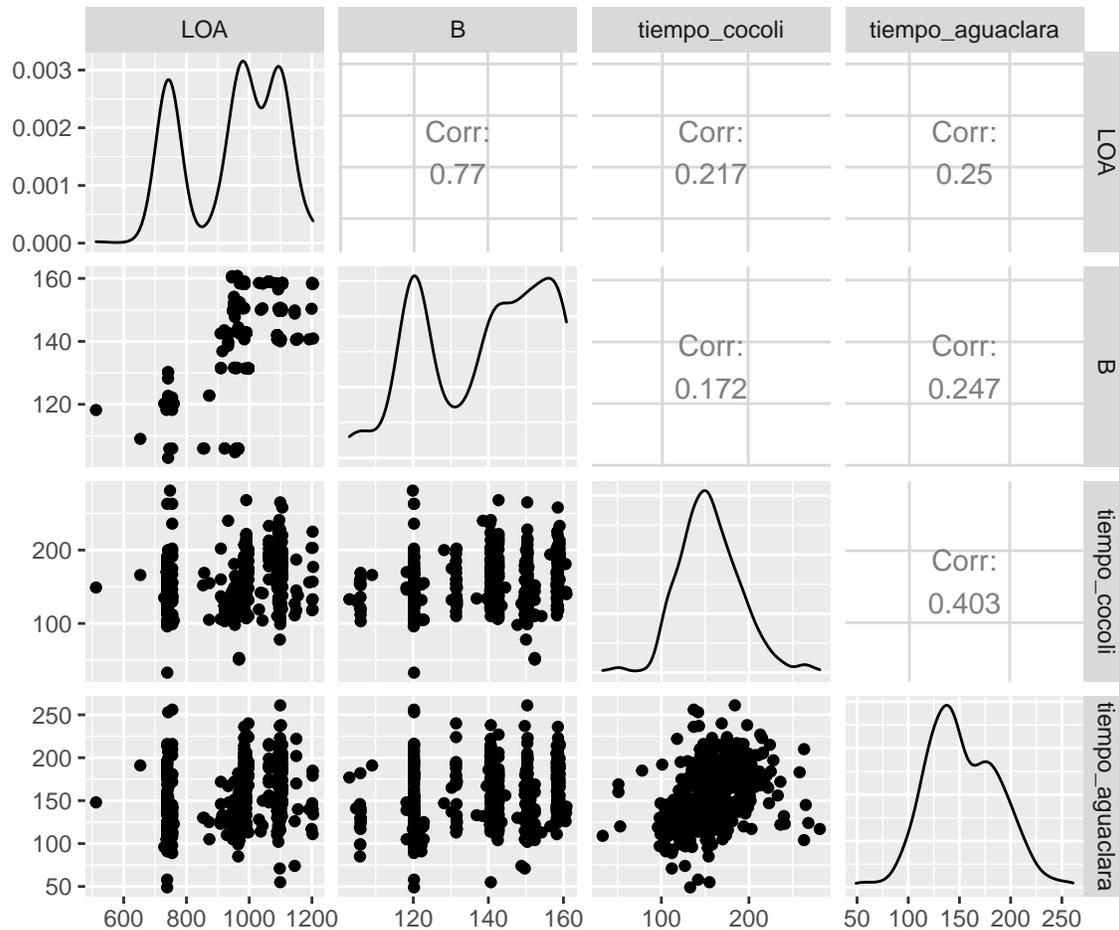


Figura 3.8: Estudio de correlaciones entre las variables LOA, B y tiempo de tránsito por Cocolí y Agua Clara. El panel superior muestra la correlación entre las variables, el inferior los diagramas de dispersión, mientras que en la diagonal se muestran la estimación no paramétrica de la función de densidad de cada una de las variables.

### 3.2.3. Ajuste de modelos aditivos generalizados

En esta sección estimamos los efectos del entrenamiento o experiencia, dirección de tránsito, tipo de barco, esclusa, eslora y manga sobre el tiempo de tránsito a través de las esclusas. Primero estimamos el tiempo de tránsito en función del tiempo transcurrido desde la inauguración del Canal Ampliado. Después ajustaremos el modelo añadiendo el resto de variables. Para el modelo suponemos una distribución  $t$  de student (Wood, 2021).

#### 3.2.3.1. Tiempo de tránsito en función de tiempo desde la inauguración del Canal Ampliado

En el resumen del ajuste del modelo GAM, utilizando una base de thin plate splines con 4 elementos, se muestra que el efecto de la experiencia o entrenamiento sobre el tiempo de tránsito de los buques es significativo ( $p$ -valor  $< 0.05$ ). El coeficiente de determinación es  $R^2=25.3$ , es decir, el tiempo transcurrido desde el comienzo de las actividades de capacitación explica el 25.3 % de la variabilidad del tiempo de tránsito a través de las esclusas. Por tanto, se observa que, aunque el efecto del entrenamiento es importante, hay más fuentes de variación que afectan al tránsito de los buques a través de las esclusas. En la siguiente sección

se estudiarán los efectos de diversos factores como son la dirección, la esclusa y el tipo de barco, por un lado, y las variables cuantitativas relativas a las dimensiones de los buques (eslora y manga).

Family: Scaled t(6.87,25.39)

Link function: identity

Formula:

transit\_time ~ s(Time, k = 4)

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	153.721	0.994	155	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:

	edf	Ref.df	Chi.sq	p-value
s(Time)	2.8	2.97	340	<2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.264 Deviance explained = 25.3%

-REML = 3929.4 Scale est. = 1 n = 818

### 3.2.3.2. Tiempo de tránsito en función de los efectos del entrenamiento, dirección de tránsito, tipo de barco, esclusa, eslora y manga

Una vez que se ha realizado un análisis descriptivo del conjunto de datos y considerando los resultados obtenidos en la sección 3.2.3.1, se estima el tipo y la magnitud de los efectos de la variable experiencia, del tipo de barco, de la esclusa, de la dirección en el tiempo necesario para cruzar una esclusa en el Canal de Panamá Ampliado.

Suponiendo que la respuesta, el tiempo de tránsito, se distribuye según una  $t$  de Student, se estima un modelo GAM, en el que se han asumido efectos lineales para todas las variables salvo para la experiencia o entrenamiento, que como ya se había visto en anteriores secciones, puede ser de tipo no lineal asintótico. Las estimaciones del modelo, así como el estudio de significación de los efectos y las medidas de bondad de ajuste, se muestran a continuación.

Family: Scaled t(4.57,20.913)

Link function: identity

Formula:

transit\_time ~ s(Time, k = 4) + (B) + (LOA) + Type + direction \*  
lock

Parametric coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	38.6267	17.1616	2.25	0.02440 *
B	0.4667	0.0888	5.26	1.5e-07 ***
LOA	0.0476	0.0163	2.92	0.00351 **
TypeLNG	-20.9274	3.4570	-6.05	1.4e-09 ***

```

TypeLPG                8.2788      5.0263      1.65  0.09954 .
directionSouth         8.9473      2.5277      3.54  0.00040 ***
lockCocoli             7.7034      2.1507      3.58  0.00034 ***
directionSouth:lockCocoli -17.4460    3.5280     -4.95  7.6e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
      edf Ref.df Chi.sq p-value
s(Time) 2.84  2.98   434 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.383  Deviance explained = 36.7%
-REML = 3828.1  Scale est. = 1          n = 818

```

Todas las variables tienen efectos significativos sobre la respuesta ( $p$ -valores  $< 0.05$ ). Únicamente el efecto del nivel GLP, de la variable tipo de buque, resulta no significativo, es decir, que no se puede distinguir del tiempo de tránsito empleado por los buques contenedores (usado como referencia al evaluar los efectos). El modelo GAM estima que los buques de GNL tardan 20.92 minutos (en valores medios) menos en pasar por las esclusas que los Contenedores.

En cuanto a los resultados obtenidos con respecto a la variable dirección y esclusa, a la dirección Sur le corresponden tiempos de tránsito significativamente mayores a lo largo de las esclusas si se comparan con los tiempos de los trayectos con dirección Norte. Por otro lado, en lo que respecta a la esclusa atravesada, el tiempo de tránsito estimado para cruzar la esclusa Cocolí tiende a ser mayor (7.7 min) que el correspondiente a atravesar la esclusa Agua Clara. A este respecto, el efecto de la marea en el Pacífico puede ser una de las causas (en el Caribe, la variación en el nivel del mar es de menor magnitud).

Otro resultado importante del modelo es que el efecto de la interacción entre la dirección y la variable esclusa en el tiempo de tránsito es significativo ( $p$ -valor  $< 0.05$ ).

Con respecto al efecto del tiempo de experiencia en el tiempo de tránsito por las esclusas, es estadísticamente significativo. Se observa que hay un efecto de aprendizaje.

En las Figuras 3.9 y 3.10 se muestran los efectos principales de los predictores, además del efecto de la interacción de la dirección y la esclusa. En los paneles superiores, de izquierda a derecha, se muestra el efecto del aprendizaje, seguido por el efecto lineal de la manga y el efecto lineal de la eslora de la embarcación sobre el tiempo de tránsito a través de las esclusas. En los paneles inferiores se muestran los efectos de los niveles de tipo de buque, tomando como referencia los buques portacontenedores, a continuación, los efectos de los niveles de dirección del tránsito, con el Norte como referencia y por último el efecto del factor esclusa, con Agua Clara como referencia.

Se verifica que el efecto aprendizaje (curva de aprendizaje) sobre el tiempo de tránsito es de tipo no lineal, específicamente asintótico. El efecto asintótico del tiempo de experiencia sobre el tiempo de tránsito por esclusas hace que, en un inicio, existiera un amplio margen de mejora. Sin embargo, a medida que pasaba el tiempo, la velocidad a la que disminuía el tiempo de tránsito se hacía cada vez menor, hasta llegar a una zona de saturación. Es decir, la organización tiende al controlar el tiempo de tránsito a través del canal, eliminando fuentes de variación asignables. La organización, dados unos recursos, mejora hasta un punto, para reducir el tiempo de tránsito por debajo de los límites alcanzados implicaría cambios en los recursos y el procedimiento empleado.

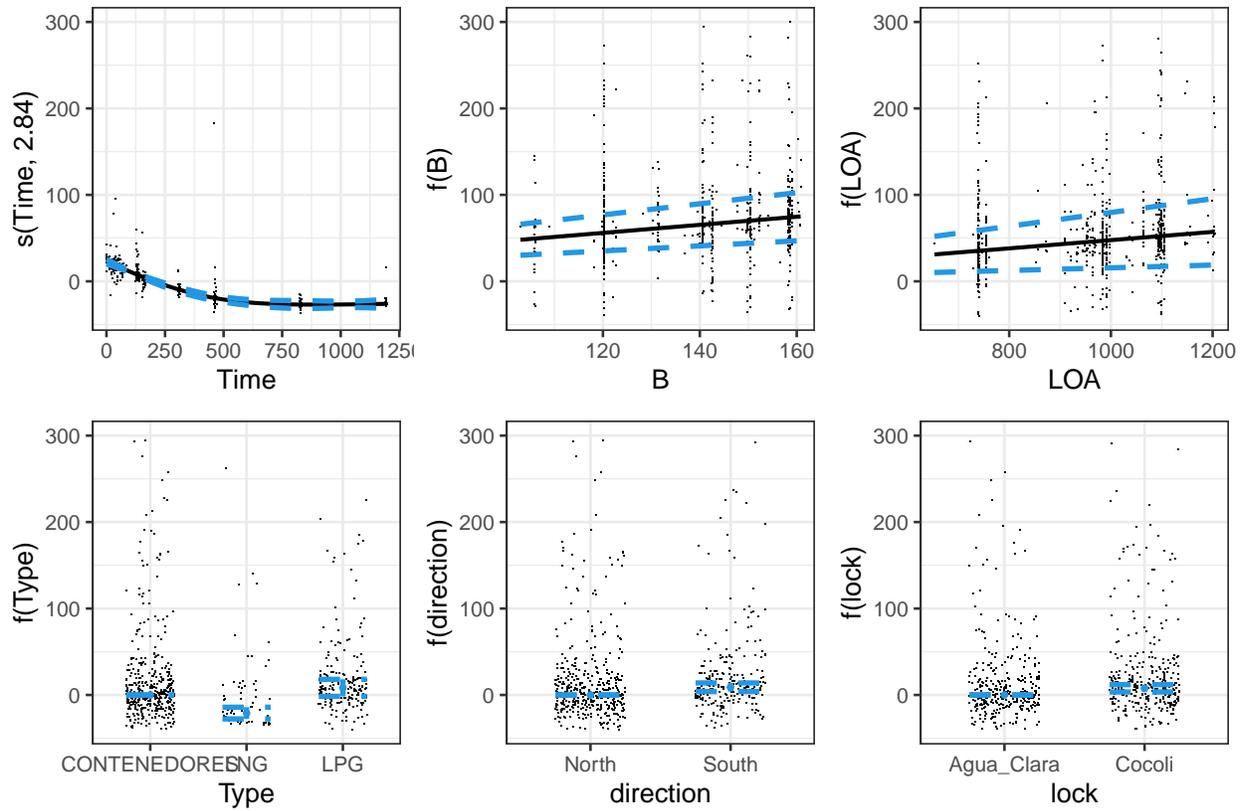


Figura 3.9: Efectos de la experiencia (panel superior izquierdo), manga (panel superior central), eslora (panel superior derecho), tipo de buque (panel inferior izquierdo), dirección (panel inferior central) y esclusa (panel inferior derecho) sobre el tiempo de tránsito, estimados mediante el ajuste de un modelo de regresión GAM.

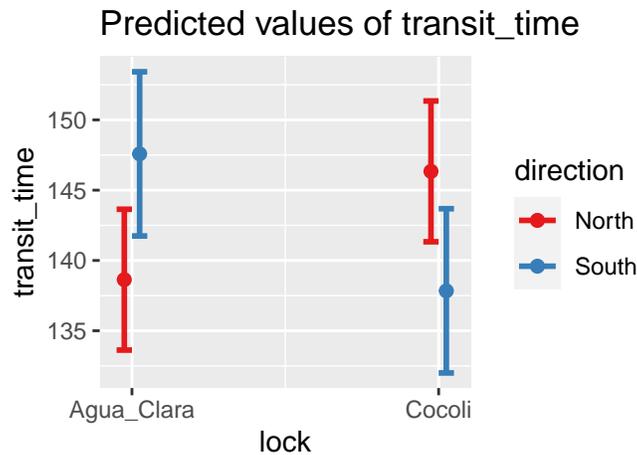


Figura 3.10: Efecto de la interacción entre dirección y la esclusa sobre el tiempo de tránsito, estimado mediante el ajuste del modelo GAM.

Con respecto a las estimaciones de la manga y la eslora de la embarcación sobre el tiempo de tránsito, su efecto es lineal y creciente. Es decir, cuando la manga del barco aumenta en 1 pie, el tiempo de tránsito a través de las esclusas tiende a aumentar en 0.46 min. En el caso de que aumente la eslora en 1 pie, el tiempo

de tránsito a través de las esclusas tiende a aumentar en 0.047 min. Es decir, cuanto mayor es el tamaño del buque, mayor es el tiempo que tarda en cruzar las esclusas.

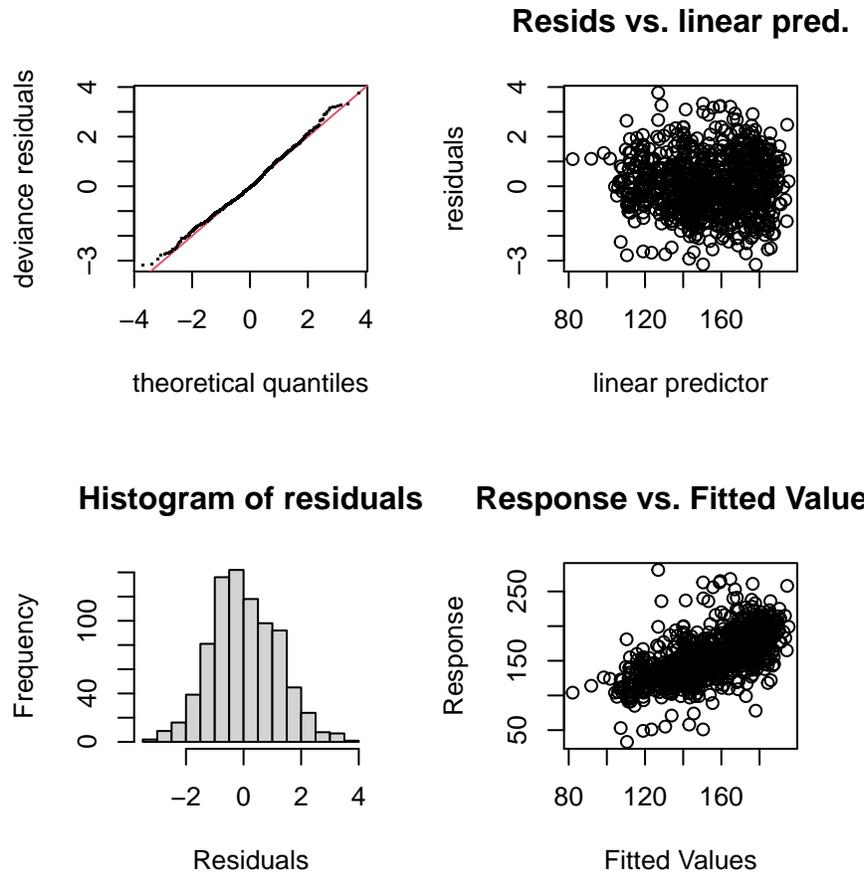


Figura 3.11: Diagn3sis del modelo GAM.

En cuanto a los efectos mostrados en los paneles inferiores de la Figura 3.9, la direcci3n Sur produce un tiempo de tr3nsito significativamente m3s largo a trav3s de las esclusas que en la direcci3n Norte. Por otro lado, el tiempo en tr3nsito tiende a ser mayor en la esclusa de Cocol3 que a trav3s de la esclusa de Agua Clara. Adem3s, el tiempo esperado en tr3nsito de los buques portacontenedores tiende a ser mayor que el de los buques de GNL.

Finalmente, agregamos el efecto, en el tiempo de tr3nsito, de la interacci3n entre direcci3n y la esclusa. El tiempo de tr3nsito es significativamente mayor cuando se cruza la esclusa entrando desde el oc3ano (Cocol3 y direcci3n norte, Agua Clara y direcci3n sur) (ver Figura 3.10).

Con estos resultados confirmamos las conclusiones a las que llegamos en el resumen del modelo.

Se realiza el diagn3stico de los supuestos de distribuci3n de probabilidad de partida y homogeneidad utilizando m3todos gr3ficos. Todos se cumplen de forma aproximada. El modelo no presenta un patr3n espec3fico de variabilidad cambiante al representar los residuos frente a los valores ajustados. En el gr3fico de cuantiles la parte central de la curva se ajusta razonablemente a la diagonal. En lo que se refiere a la bondad de ajuste, el modelo GAM explica un 36.7% de la variabilidad del tiempo de tr3nsito. Es decir, siendo significativos los efectos de las variables explicativas estudiadas, existen m3s fuentes de variaci3n que afectan al tiempo

de tránsito, como son, por ejemplo, las variables meteorológicas, las mareas, la existencia de averías en las instalaciones o el tráfico, entre otras muchas.

En este capítulo se utilizaron principalmente las referencias Carral et al. (2021), Carral et al. (2018), Carral et al. (2017) y Wood (2017).

## Capítulo 4

# Control estadístico de procesos y análisis de capacidad

Después de identificar aquellos factores que influyen en el tiempo de tránsito se han utilizado herramientas de control estadístico de procesos, como gráficos de control y análisis de capacidad como propuesta de aplicación de la metodología Seis Sigma en el Canal de Panamá. El objetivo es aportar una metodología para controlar en todo momento el proceso de tránsito a través de las esclusas, estimando unos límites de control naturales (variabilidad y posición) para las variables CTQ, representativas del proceso (tiempo de tránsito a través de las esclusas), permitiendo la identificación de estados fuera de control, anomalías, alarmas, patrones que puedan indicar tanto averías como procesos de aprendizaje, además de medir su magnitud.

En definitiva, se pretende evaluar la utilidad del uso de herramientas como los gráficos de control para promover la mejora continua en el proceso de tránsito a través del Canal Ampliado. Por otro lado, una vez estimada la variabilidad natural de las características que definen el tránsito, es importante observar si estas cumplen las especificaciones marcadas por la Autoridad del Canal, la empresa o los clientes. Por tanto, también se propone, en este capítulo, la aplicación de técnicas de análisis de capacidad de procesos.

Es importante destacar que la Autoridad del Canal, de acuerdo con los sindicatos, tiene una regulación por la que asigna complementos salariales a aquellos pilotos cuyos tránsitos sean menores que cierto valor (también compensaciones a los armadores si el trayecto es demasiado lento), que depende también de las dimensiones del buque a pilotar. Pero estos límites no están todavía completamente definidos en el Canal Ampliado, no existiendo una normativa que los indique para las dos nuevas esclusas. Por tanto, este estudio también pretende proponer unos límites tentativos y evaluar a a partir de ellos la capacidad del proceso de tránsito a través de las nuevas esclusas.

### 4.1. Técnicas de gráficos de control, Fase I

#### 4.1.1. Gráficos de control univariantes

Dada la estructura del conjunto de datos estudiados, se implementan gráficos de control Shewart para medidas individuales con el objetivo de identificar patrones de cambio en el tiempo de paso por las esclusas, y evaluar, en general, la posibilidad de aplicación de este tipo de herramientas en el ámbito del Canal de Panamá

Considerando que este tiempo es diferente dependiendo del tipo de barco, la esclusa y dirección, se aplican técnicas de gráficos de control tomando en cuenta cada uno de los niveles de los factores y de esta forma verificar qué diferencias existen. En este análisis se deja fuera los buques de gas natural licuado (GNL) por no contar, en el momento del estudio, con observaciones suficientes para la realización de un análisis fiable.

En un futuro, teniendo un acceso continuado y fluido a los datos de tiempos de tránsito a través del Canal, podremos extender el estudio no sólo a los buques GNL, sino también a otros cuyo tránsito es menos frecuente como buques petroleros, por ejemplo.

Al final de la sección, teniendo en cuenta que la diferencia en los tiempos de tránsito obedece en gran medida al tamaño del buque, se dividirán los tránsitos según grupos formados a partir de la magnitud de la variable eslora. Dividiremos los valores de eslora en varios tres grupos: bajo, medio, alto, y aplicaremos gráficos de control a las bases de datos resultantes.

Otros gráficos de control, como los EWMA y CUSUM, son utilizados para identificar la existencia de pequeños cambios, menores que dos desviaciones estándar con respecto a la media del proceso.

#### 4.1.1.1. Clasificación según esclusa, tipo de barco y dirección de tránsito

##### Gráficos de medidas individuales Shewart

Después de separar los datos del tiempo de tránsito según tipo de buque, esclusa y dirección, se inicia la Fase I en la que los límites naturales de control se estiman utilizando una muestra preliminar, en este caso las observaciones de los tiempos de tránsito correspondientes a los 25 primeros buques, que se corresponden más o menos con el primer mes de operación del Canal Ampliado.

Antes de la aplicación de los gráficos se ha comprobado el cumplimiento de los supuestos de normalidad y ausencia de autocorrelación para cada grupo, mediante la aplicación de los contrastes de Shapiro-Wilks, Lilliefors, gráfico qqplot, Box–Pierce y Ljung–Box, supuestos necesarios para que se apliquen de manera confiable los gráficos de control, especialmente los de Shewhart.

Para la generación de los gráficos se utiliza la librería `qcc`. Primero se crea un objeto `qcc.groups`, utilizando la función con el mismo nombre, con los siguientes argumentos: las observaciones en formato de vector y la identificación de las mismas como pertenecientes a un grupo. Seguido creamos el objeto `qcc`.

A modo de ejemplo, mostramos cómo creamos el `qcc.groups` y el objeto `qcc` con los datos de los buques que transportan GLP, cruzando por la esclusa Agua Clara en dirección Sur.

```
muestra1<- 1:length(tiempo.aguaclara.s.lpg)
aguaclara.s.lpg<- qcc.groups(tiempo.aguaclara.s.lpg, muestra1)
aguaclara.s.lpg.qcc<-qcc(data=aguaclara.s.lpg[1:25], type="xbar.one", nsigmas=3, plot=TRUE)
```

Se muestran los principales argumentos de la función utilizados para crear el objeto `qcc`:

- `data`: Con el objeto `qcc.groups` como primer argumento. En nuestro caso, para el calibrado, seleccionamos las 25 primeras muestras.
- `type`: En el segundo parámetro especificamos un *string value*, que especifica el tipo de control que deseamos aplicar, en nuestro caso “xbar.one” porque estamos tratando muestras individuales.
- `nsigmas`: El tercer argumento es un valor numérico que especifica el número de sigmas a utilizar para calcular los límites de control.
- `plot`: Al especificar “TRUE” se obtiene el gráfico de control. Además, esta función proporciona un resumen que incluye la estadística descriptiva referida a los grupos, tamaño de la submuestra, número de grupos, media del estadístico del gráfico de control, desviación típica, resultado de dividir el rango medio entre  $d_2$  (tabulado).

Una vez creado el objeto `qcc` podemos obtener tanto un gráfico de control como un resumen estadístico. En la Figura 4.1 se muestran los resultados al realizar lo antes descrito para cada población estudiada. En las dos primeras filas de paneles, de arriba a abajo y de izquierda a derecha, se muestra el tiempo de paso de los buques de GLP cruzando las esclusas de Agua Clara (izq) y Cocolí (dcha) en dirección sur (primera fila contando desde arriba) y norte (segunda fila contando desde arriba). En la tercera y cuarta fila, empezando desde arriba, se observa el tiempo de paso de los buques Contenedores cruzando ambas esclusas en dirección Sur (tercera fila) y Norte (cuarta fila), situándose a la izquierda los paneles que hacen referencia a la esclusa Agua Clara y a la derecha los que se refieren a la esclusa Cocolí.

Tenemos en el eje  $Y$  los tiempos de tránsito y en el eje  $X$  cada uno de los buques que transitan, ordenados de forma cronológica. La línea media de los gráficos representa el tiempo medio y las líneas laterales punteadas a cada lado de la línea media son los límites de control del tiempo, se encuentran a una distancia de tres veces la desviación estándar de tiempo desde la línea media. La distancia entre los límites de control superior e inferior es el intervalo de confianza a un nivel de confianza del 99.73 % bajo el supuesto de distribución normal.

Los puntos representan las observaciones del tiempo de tránsito de cada buque a través de la esclusa en cuestión, en orden cronológico. Los puntos de color rojo representan las observaciones del tiempo de tránsito fuera de los límites de control y pueden sugerir que el proceso está fuera de control debido a causas asignables más que al azar. Estas observaciones pueden estar relacionadas con alarmas o anomalías o con efectos positivos como puede ser una reducción del tiempo de tránsito debido al aprendizaje o a la aplicación de una mejora en las instalaciones.

Los puntos amarillos se denominan rachas, indican tendencias y patrones que no son debidos a la propia aleatoriedad del proceso. En este caso, las rachas se identifican cuando se observan más de seis tránsitos a un lado de la línea central del gráfico de control, siguiendo las recomendaciones de Strucca (2017).

En la Fase I de la aplicación de gráficos de control, se estiman los límites de control naturales a partir de una muestra de calibrado. Poda dicha estimación, esta fase, se eliminan o descartan para su cálculo todas aquellas observaciones atípicas, diferentes a las que definen el normal funcionamiento del proceso. De esta forma, se estimarán de forma fiable la posición y variabilidad del proceso y, consecuentemente, los límites de control (Montgomery, 2009). El objetivo es quedarse con un conjunto de observaciones que pertenezcan a una misma población objetivo.

Respecto al comportamiento mostrado de los buques Contenedores y GLP, en ninguno de los casos parece existir alguna tendencia, más bien se observa un patrón aleatorio, natural.

En algunos casos, se detectan tiempos fuera de control. En la mayoría de los casos se trata de situaciones puntuales que no indican cambios en el proceso (especiales dificultades puntuales por condiciones extremas de tráfico, lluvias, marea, etc.). En el caso del tiempo de tránsito de las embarcaciones GLP por la esclusa Cocolí en dirección Sur, presenta una racha en la observación 25, podríamos decir que esta relacionada con un patrón de cambio de nivel (menor tiempo). Esto indicaría el comienzo de un cambio en el proceso tránsito de los barcos por las esclusas, por lo tanto se elimina la muestra del cálculo de los límites de control.

Las observaciones fuera de control, habiendo identificado las causas asignables u observando que estos cambios son puntuales, se eliminan consecutivamente y los límites de control natural se vuelven a calcular hasta que no se detectan más valores atípicos (fuera de control, en este contexto). De esta forma, se puede suponer que la población de los tiempos de tránsito, de buques Contenedores o LPG, a través de las esclusas Cocolí o Agua Clara, y en dirección Sur o Norte, está bajo control.

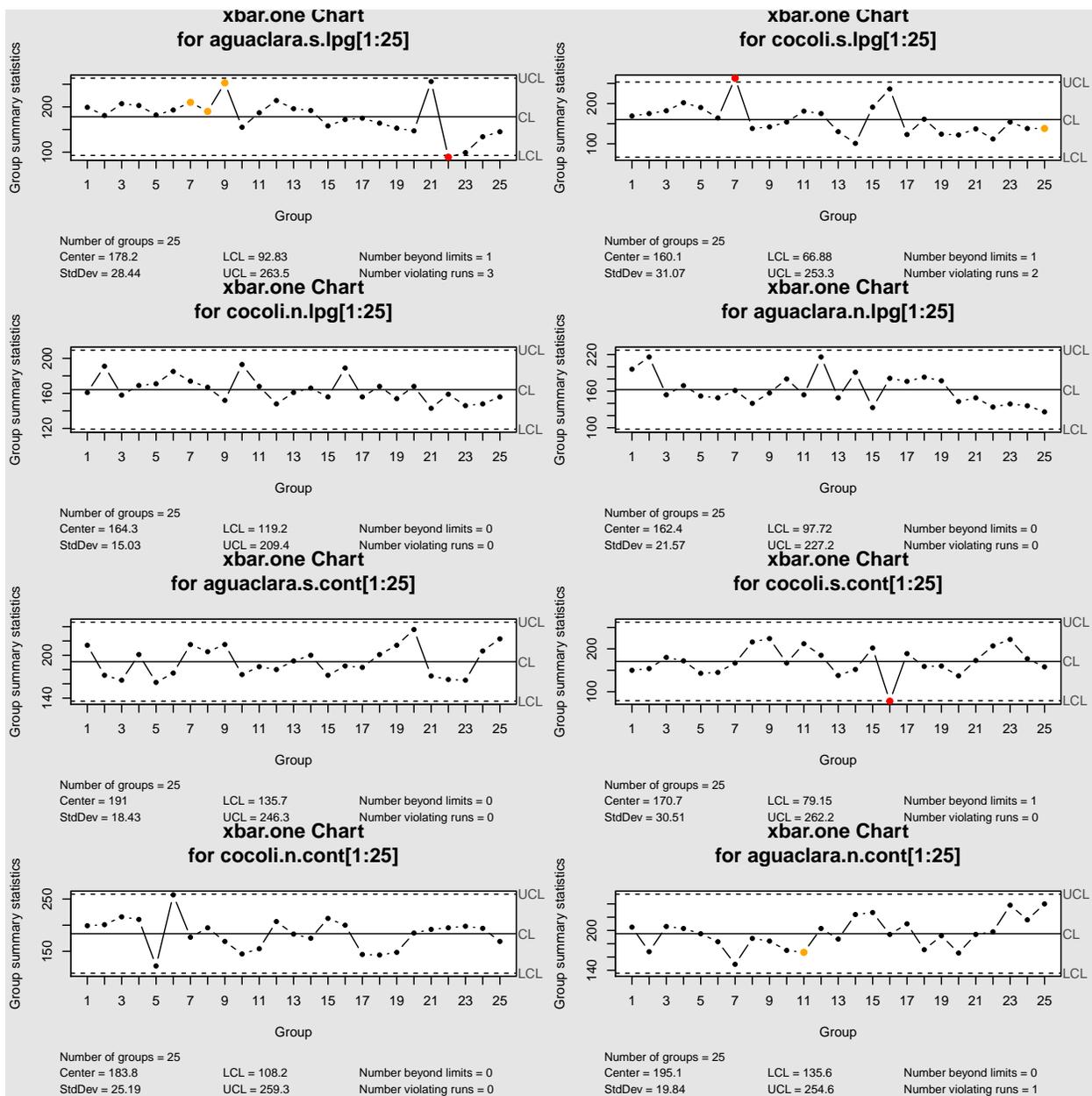


Figura 4.1: Gráficos de control para observaciones individuales del tránsito de buques GLP (en las dos filas superiores) y Contenedores (en las dos filas inferiores) cruzando las esclusas Cocolí y Agua Clara en dirección Sur (primera y tercera fila desde arriba) y Norte (segunda y cuarta fila desde arriba). Estimación de los límites de control en la Fase I (calibrado).

A modo de ejemplo, en la Figura 4.2 se muestra el gráfico de control para el tránsito de buques GLP por la esclusa Agua Clara en dirección Sur, una vez que se han descartado los valores atípicos y se han recalculado los límites de control natural. En este caso se descartan 7 observaciones de tiempo fuera de control mediante un proceso iterativo.

Se muestra la media del proceso ( $CL=181.1$  min) y sus límites de control natural ( $UCL=227.3$  min,  $LCL=134.8$  min). Suponiendo una distribución normal, el 99.7% de la distribución del tiempo de tránsito de estos barcos se encuentra entre UCL y LCL.

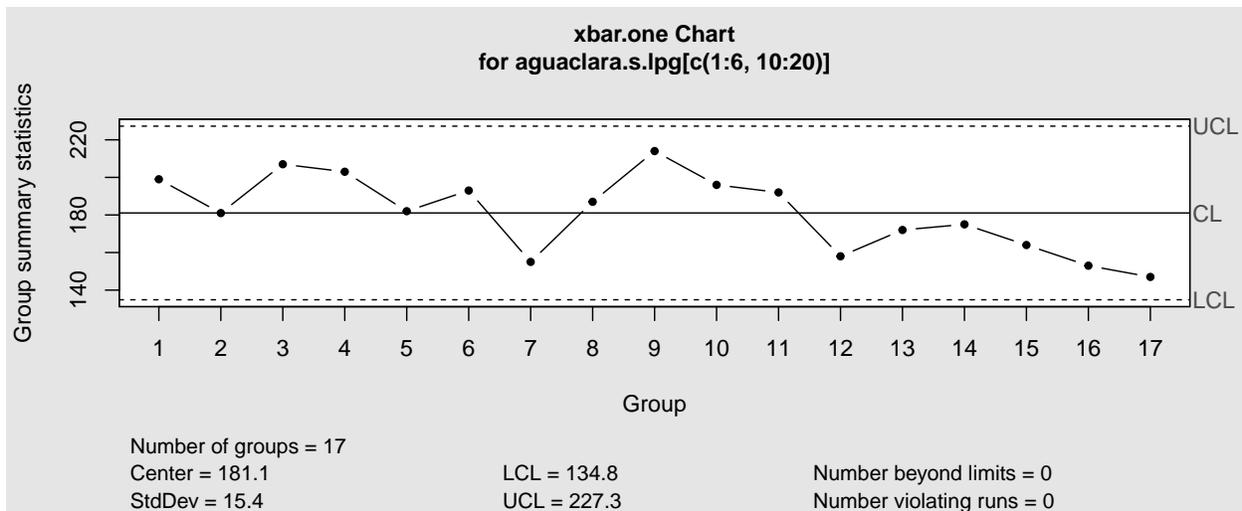


Figura 4.2: Ajuste de los límites de control del tiempo de tránsito de buque GLP por la esclusa Agua Clara dirección Sur.

Finalmente, en cada uno de los casos, no quedan observaciones fuera de control ni ningún punto que viole las reglas mostradas anteriormente. Por lo tanto, en cuanto a la dispersión del proceso medido, este está bajo control. En lo sucesivo, manteniendo los límites de control calculados, en la Fase II se contrastará si las nuevas observaciones pertenecen a la misma población que la muestra de calibrado, mediante la monitorización sucesiva de las mismas sobre el gráfico resultante de la Fase I.

Si estamos interesados en detectar pequeños cambios a lo largo de las sucesivas submuestras, podemos construir dos tipos de gráficos más sensibles que los  $\bar{x}$  tradicionales. Estos son los gráficos CUSUM y los gráficos EWMA. Sin embargo, ambos gráficos funcionan peor que el gráfico de medias cuando se trata de detectar grandes cambios (del orden de más de dos desviaciones típicas sobre la media del proceso).

## Gráficos CUSUM

Se estiman los límites naturales de control univariante con CUSUM aplicado a los tiempos de tránsito de los buques por las nuevas esclusas del Canal de Panamá.

Como hicimos con los gráficos Shewart, se empieza con la Fase I donde los límites naturales de control se estiman utilizando las 25 primeras observaciones.

Para crear los gráficos CUSUM se utiliza la librería `qcc`, creamos el objeto `qcc.groups` como lo hicimos anteriormente. La función utilizada en este caso es `cusum()`.

Se muestra un ejemplo con los datos de los buques GLP, cruzando por la esclusa Agua Clara en dirección Sur, donde presentamos los argumentos principales de dicha función.

```
aguaclara.s.lpg.cusum<-cusum(data=aguaclara.s.lpg[1:25],
                             decision.interval= 4, se.shift=1, plot=TRUE)
```

- `data`: El objeto `qcc.groups`.
- `se.shif`: Es el valor  $\delta = 2 \cdot k(\mu_1 = \mu + \delta\sigma)$  que da la sensibilidad del gráfico. Se detectarán cambios, desplazamientos con respecto a la media  $\delta$  veces mayores que la desviación típica estimada.
- `decision.interval`: Es el parámetro  $h$ , a partir de cual se construyen los límites de control:  $H = h \cdot \sigma$ . Se suele asignar 4 o 5. Por el momento se decide trabajar con un `decision.interval=4`, en general avisa antes de un cambio en el proceso.

- `plot`: Al igual que en el caso Shewart al especificar “TRUE” se obtiene el gráfico de control. Además, esta función proporciona un resumen que incluye la estadística descriptiva referida a los grupos.

Mostramos en la Figura 4.3 los resultados de la aplicación de la técnica CUSUM. Al calcular los límites de control con el método CUSUM se puede verificar la línea gruesa, que es la media=0, las diferencias positivas por encima de esa línea media y las diferencias negativas por debajo. Además, como en el caso de los gráficos Shewart, se puede observar un resumen de los principales estadísticos debajo de cada figura.

Se observa un comportamiento bastante similar a los gráficos Shewart. En el caso de los buques Contenedores y GLP, las observaciones presentan un patrón aleatorio, salvo en el caso de los buques GLP que atraviesan Cocolí en dirección Sur, para los que se comienza a identificar una disminución en los tiempos de tránsito.

Excepto en el caso del tiempo de tránsito del buque GLP por la esclusa Agua Clara cuando toma la dirección Sur, no hay que eliminar ninguna observación para estimar los límites de control porque el proceso estudiado esta bajo control. Son fiables los límites de control y se puede pasar a la Fase II.

En el caso del tiempo de tránsito debido al aprendizaje del buque GLP por la esclusa Agua Clara podríamos suponer que esta ocurriendo un cambio en el proceso debido al aprendizaje como causa asignable, por lo que se procede a eliminar estas muestras de los cálculos y a reajustar los límites. Se verá más adelante, cuando estudiemos la Fase II, en la que se mostrarán los gráficos de control así obtenidos, incluyendo las nuevos tránsitos a monitorizar correspondientes en su mayoría al periodo 2017-20, que existe un acusado cambio paulatino de nivel debido al aprendizaje.

### Gráficos EWMA

Al igual que en los casos anteriores construimos los gráficos de control EWMA con la librería `qcc`, creando el objeto `qcc.groups`. La función utilizada en este caso es `ewma()`,

A modo de ejemplo, mostramos como creamos el `qcc.groups` a partir del objeto `qcc` creado anteriormente con los datos de los buques que transportan GLP, cruzando por la esclusa Agua Clara en dirección Sur.

Los parámetros principales de dicha función son los siguientes:

- `data`: El objeto `qcc.groups`.
- `lambda`: El parámetro de suavizado,  $0 \leq \lambda \leq 1$ . Se suele tomar como buena práctica el asignar el valor de 0.2 al parámetro `lambda`.
- `nsigmas`: Especifica el número de sigmas que se utilizarán para calcular los límites de control.
- `plot`: Al especificar “TRUE” se obtiene el gráfico de control. Además, esta función proporciona un resumen que incluye la estadística descriptiva referida a los grupos.

En los gráficos EWMA los puntos son las medias ponderadas exponencialmente de cada una de las observaciones y las cruces son las observaciones reales de los tiempos de tránsito.

Al igual que en los gráficos anteriores, la salida gráfica de los gráficos EWMA proporciona el número de observaciones, media y desviación, además del parámetro de suavizado. Muestra que estamos haciendo los límites de control a 3 veces sigma y el número de observaciones fuera de los límites.

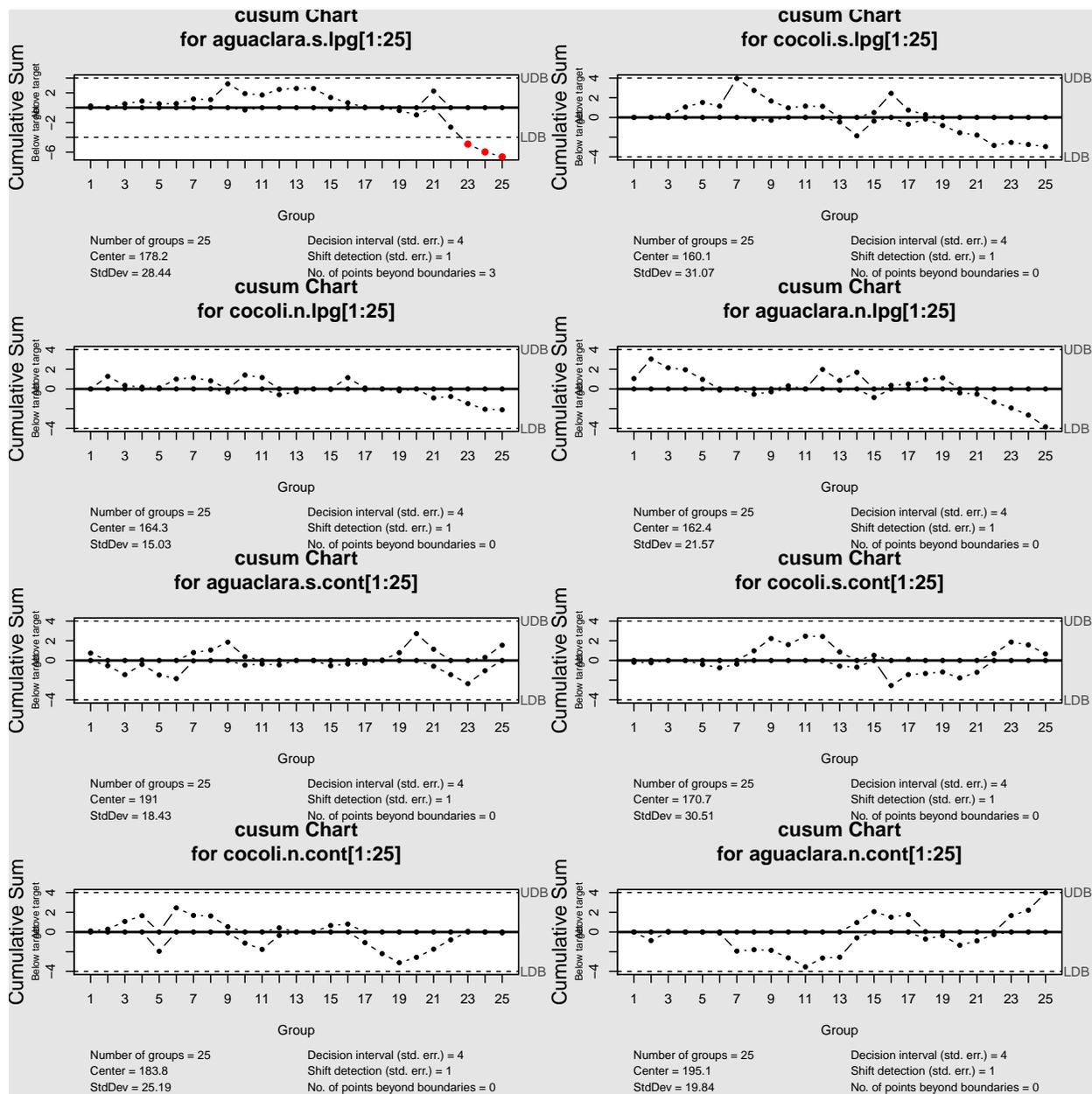


Figura 4.3: Gráficos CUSUM del tránsito de buques GLP (dos primeras filas empezando desde arriba) y Contenedores (tercera y cuarta filas empezando desde arriba) cruzando las esclusas Cocolí y Agua Clara en dirección Sur (primera y tercera fila empezando desde arriba) y Norte (segunda y cuarta filas empezando desde arriba). Estimación de los límites de control en la Fase I (calibrado).

Los paneles de la Figura 4.4 muestran que sólo se observan estados fuera de control, valores atípicos, en los casos donde los barcos GLP transitan por la esclusa Agua Clara o Cocolí en dirección Sur, observaciones que procedemos a eliminar (habiendo identificado sus causas asignables: patrón de aprendizaje en el caso del panel superior izquierdo y patrón tipo “freak” puntual, en el panel superior derecho), hasta resultar que el proceso está bajo control y así poder continuar con el estudio de monitorizado, que veremos más adelante en la sección correspondiente a Fase II.

```
aguaclara.s.lpg.ewma<-ewma(data=aguaclara.s.lpg[1:25], lambda=0.2, nsigas=3, plot=TRUE)
```

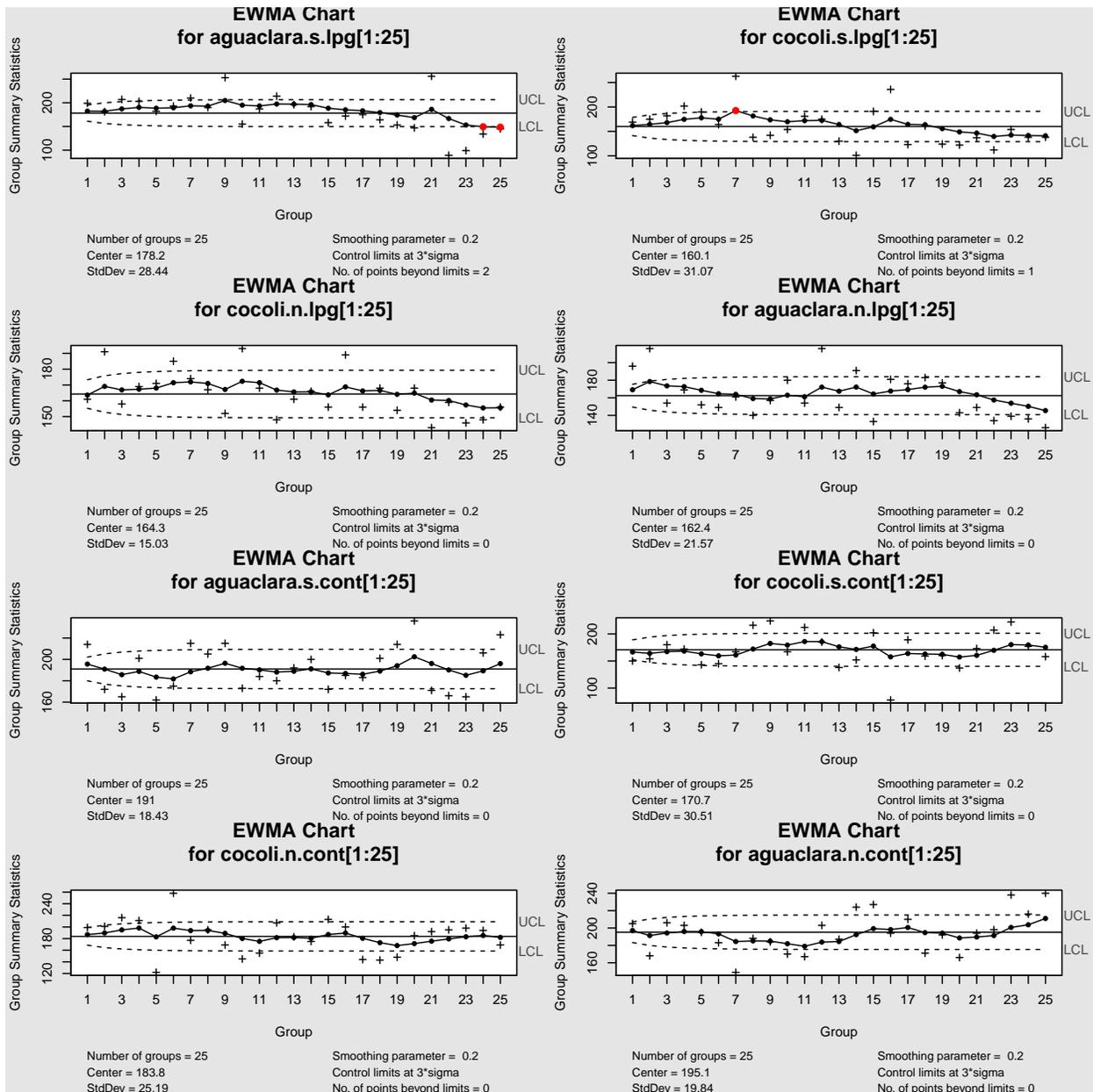


Figura 4.4: Gráficos EWMA del tránsito de buques GLP (dos primeras filas empezando desde arriba) y Contenedores (tercera y cuarta filas empezando desde arriba) cruzando las esclusas Cocolí y Agua Clara en dirección Sur (primera y tercera fila empezando desde arriba) y Norte (segunda y cuarta filas empezando desde arriba). Estimación de los límites de control en la Fase I (calibrado).

Los gráficos de control Shewhart, los CUSUM y los EWMA aportan resultados similares. El proceso de tránsito a través de las esclusas del Canal ampliado tiende a estar bajo control en las primeras semanas de apertura del Canal, en el 2016. Dado que el número de tránsitos GLP es sensiblemente menor que el de los buques Contenedores, con el fin de estimar convenientemente los límites de control a partir de una muestra lo suficientemente representativa (20 o más observaciones, siguiendo las indicaciones de Montgomery (2005)),

se han incluido tránsitos realizados a finales de 2016, fecha en la que se empieza a identificar una disminución paulatina del tiempo de tránsito en la esclusa Agua Clara, con dirección Sur.

Los gráficos de control permiten también estimar la posición y variabilidad de los tiempos de tránsito en las primeras semanas de funcionamiento del Canal Ampliado. Referencia a partir la cual identificaremos posibles cambios en el proceso.

#### 4.1.1.2. Clasificación según esclusa, medida de la eslora y dirección de tránsito.

##### Gráficos de medidas individuales Shewart

Probaremos a aplicar gráficos de control para estimar la variabilidad del tiempo de tránsito, sin embargo, en lugar de clasificar según tipo de la decisión de barco lo haremos según la eslora. Dividiremos los valores de eslora en varios grupos (alto, medio, bajo). La decisión de agrupar por eslora, se debe a que las dimensiones de los buques afectan al tiempo de tránsito, además de intuir que una parte importante de la variabilidad de los tiempos de tránsito debida al tipo de buque es debida a las diferentes dimensiones que estos tipos de buques suelen tener.

En este caso, se decide aplicar, a modo de ejemplo, solo los gráficos Shewart, dado que en el estudio de la monitorización del tiempo de tránsito clasificando por tipo de buque, esclusa y dirección (el cual veremos más adelante) los gráficos de control Shewart, EWMA y CUSUM, aportan resultados muy similares, incluso, en algunos casos, Shewart avisa primero del cambio en el proceso, por ser un cambio relativamente grande (superior en muchos casos a 2 desviaciones típicas con respecto a la media del proceso). Además, a continuación veremos que al clasificar la población por eslora, en vez de por tipo de barco, tendremos un comportamiento similar, por la relación existente entre estos dos factores.

En la Figura 4.5 se visualiza, tomando en cuenta las 25 primeras observaciones del proceso, que cuanto menor es la dimensión de la eslora del buque, el patrón de aprendizaje es más acentuado. En el caso de la dimensión de tamaño bajo se observa una ligera tendencia a la baja en el tiempo de tránsito, cuando se trata de una dimensión media o, especialmente alta se muestra un comportamiento representativo del patrón aleatorio. Este resultado está en completa relación con los obtenidos anteriormente, clasificando por tipo de buque. De hecho, los buques GLP tienden a tener una eslora significativamente menor que los Contenedores, por lo que se clasifican en su totalidad dentro del nivel bajo de eslora.

Por tanto, se observa un comportamiento del tiempo de tránsito del buque GLP bastante parecido al comportamiento de las tendencias del tiempo de tránsito de los buques con eslora de tamaño bajo. Igual ocurre al visualizar el comportamiento de los contenedores con la tendencia del tiempo de tránsito de los buques con una de tamaño medio y alto, especialmente en este último caso.

Decidimos estudiar la relación de ambas variables. En la Figura 4.6 se observa que el 100 % de los barcos GLP tienen tamaño de eslora bajo. El 95.12 % de los buques GNL tienen una eslora media y el resto pertenecen al grupo de eslora de tamaño bajo. Los buques Contenedores están repartidos entre esloras de tamaño medio y alto, conformando estos últimos el mayor porcentaje con un 57.14 %.

Se aplica el *test*  $\chi^2$  de independencia, también conocido como  $\chi^2$  de *Pearson* que se emplea para estudiar si existe asociación entre dos variables categóricas, es decir, si las proporciones de una variable son diferentes dependiendo del valor que adquiera la otra variable, cuando los datos son independientes.

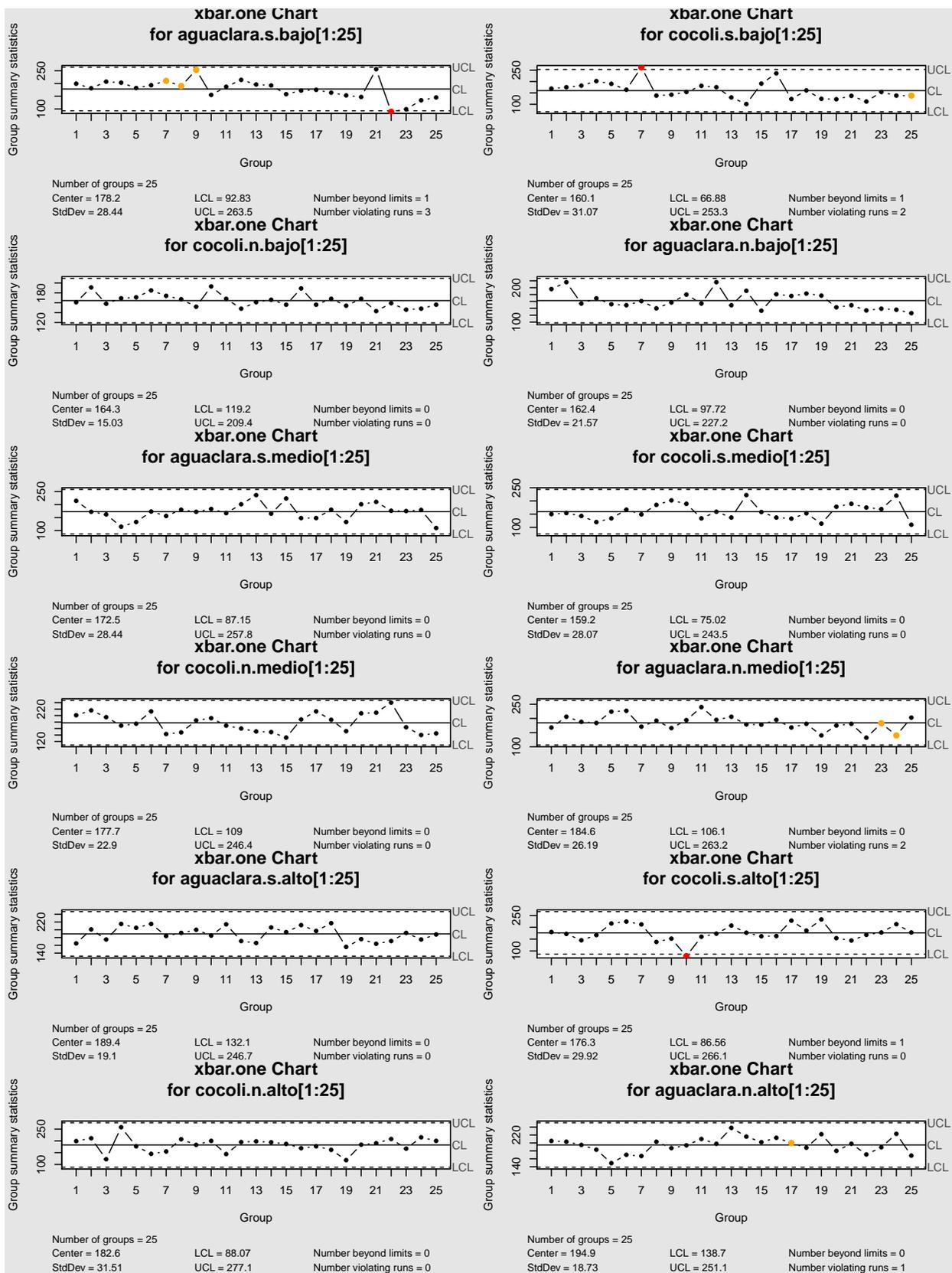


Figura 4.5: Gráfico de observaciones individuales del tiempo de tránsito según esclusa, tamaño de eslora (desde la parte superior, fila 1 y 2: nivel bajo de eslora, fila 3 y 4: nivel medio, fila 5 y 6: nivel alto) y dirección de tránsito (filas 1, 3 y 5: dirección Sur, filas 2, 4 y 6: dirección Norte). Estimación de los límites de control en la Fase I (calibrado).

La formulación de las hipótesis son las siguientes:

$H_0$ : Las variables son independientes por lo que una variable no varía entre los distintos niveles de la otra variable.

$H_a$ : Las variables son dependientes, una variable varía entre los distintos niveles de la otra variable.

Su fórmula estadística esta dada por:

$$\chi^2 = \sum_{i,j} \frac{(\text{observado}_{i,j} - \text{esperado}_{i,j})^2}{\text{esperado}_{i,j}}$$

El valor esperado de cada grupo se obtiene multiplicando las frecuencias marginales de la fila y columna en la que se encuentra la celda y dividiendo por el total de observaciones. Se suman las diferencias de todos los niveles. Elevar al cuadrado las diferencias permite hacerlas todas positivas y permite además magnificar aquellas más grandes.

Con la función `chisq.test()` aplicamos el estadístico en R. Se obtiene un *p-value* por debajo del nivel de significación de 0.05, rechazando la hipótesis nula de independencia entre las variables.

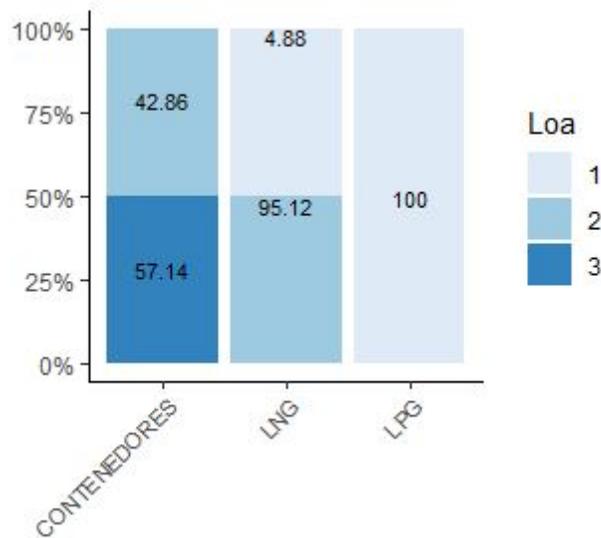


Figura 4.6: Tamaño de la eslora por tipo de buque.

	CONTENEDORES	LNG	LPG
1	0	4	232
2	216	78	0
3	288	0	0

Pearson's Chi-squared test

```
data: t_type_loa
X-squared = 918, df = 4, p-value <2e-16
```

Queremos estudiar el nivel de asociación de estas variables. Dado que el *test* contrasta si las variables están relacionadas, al tamaño del efecto se le conoce como fuerza de asociación. Existen múltiples medidas

de asociación, entre las que destacan: el **coeficiente phi**, que es el valor de la **chi cuadrado** entre el número de observaciones, un valor próximo a 0 indica independencia entre los factores, valores próximos o superiores a 1 implican relación entre los factores. El **coeficiente de contingencia** también es una medida de la intensidad de la relación basado en la **chi cuadrado** y toma valores entre 0 (independencia) y 1 (dependencia). La **V de Cramer** es muy habitual para medir la relación entre factores, es menos susceptible a valores muestrales. También 0 implica independencia y 1 una relación perfecta entre los factores.

En R se pueden calcular estas medidas mediante la función `assocstats()` del paquete `vcd` (Meyer et al., 2015).

```

                X^2 df P(> X^2)
Likelihood Ratio 1069.35 4      0
Pearson          918.24 4      0

Phi-Coefficient   : NA
Contingency Coeff.: 0.727
Cramer's V       : 0.749

```

Tenemos valores altos en todos los estadísticos calculados, concluyendo que ambos factores tienen un nivel de asociación alto. Con estos resultados confirmamos el comportamiento de estas tendencias al comparalas con los gráficos del comportamiento del tiempo por tipo de buque. Por tanto, en un futuro, se podrían definir límites de especificación superior, que marcaran un criterio de incentivos a pilotos o indemnizaciones a armadores, según el tipo de buque que transita por la esclusa o según su tamaño (eslora).

#### 4.1.2. Gráficos de control multivariantes

Teniendo en cuenta que la calidad del tránsito de cada buque por el Canal Ampliado está definida por dos variables, tiempo para cruzar Agua Clara y tiempo para cruzar Cocolí, se pueden aplicar, alternativamente al enfoque escalar, técnicas de control multivariante, como son los gráficos de control  $T^2$  de Hotelling y las versiones multivariantes de los gráficos CUSUM (MSUCUM) y EWMA (MEWMA). En esta sección se mostrará el resultado de su aplicación dentro de la Fase I o calibrado.

##### Gráficos $T^2$ de Hotelling

Con el objeto de ilustrar la aplicación del procedimiento, se tomarán los tránsitos de buques Contenedores en dirección Sur.

Primero se ha de comprobar si las variables aleatorias a controlar siguen una distribución multinormal. Para ello se puede utilizar el test de Mardia, implementado en la librería `MVN`, en particular su función `mardiaTest()` (Korkmaz et al., 2019).

Aplicamos la función a la muestra de calibrado, compuesta por las 30 primeras observaciones del tránsito en dirección Sur, que a su vez se relacionan con el primer mes y medio de funcionamiento del Canal.

```

dist_mutinormal<-mvn(Panama_sur_cont[1:30,7:8], subset = NULL, mvnTest = "mardia",
                    multivariatePlot ="qq")

dist_mutinormal[[1]]

```

	Test	Statistic	p value	Result
1	Mardia Skewness	1.11249990720386	0.892283919017011	YES
2	Mardia Kurtosis	-0.308748440939623	0.757512892241709	YES
3	MVN	<NA>	<NA>	YES

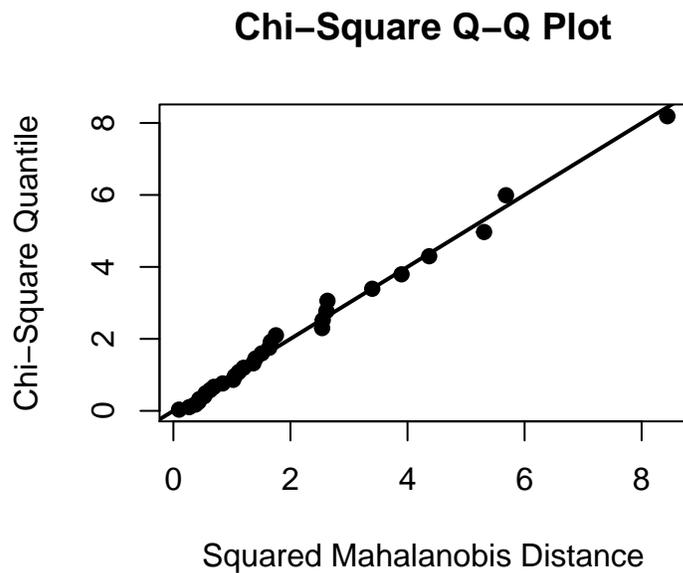


Figura 4.7: Test de distribución multinormal.

Si se tiene en cuenta los resultados de contrastes multivariantes y la Figura 4.7, no podemos rechazar que la distribución sea normal multivariante, ya que los  $p$  valores de las estadísticas de asimetría y curtosis son superiores a 0.05. En el gráfico se verifica que la distribución multinormal se cumple de forma aproximada.

Asumiendo observaciones independientes, para obtener el gráfico  $T^2$  de Hotelling para medidas individuales se empleará la función `mqqc`, de la librería `qcc`, agregando el parámetro `type= "T2.single"`, para observaciones individuales.

```
rownames(Panama_sur_cont)=1:dim(Panama_sur_cont)[1]
qq = mqqc(Panama_sur_cont[1:30,7:8], type = "T2.single", pred.limits = TRUE)
```

```
summary(qq)
```

Call:

```
mqqc(data = Panama_sur_cont[1:30, 7:8], type = "T2.single", pred.limits = TRUE)
```

T2.single chart for Panama\_sur\_cont[1:30, 7:8]

Summary of group statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.094	0.620	1.338	1.933	2.512	8.158

Number of variables: 2

Number of groups: 30

Group sample size: 1

Center:

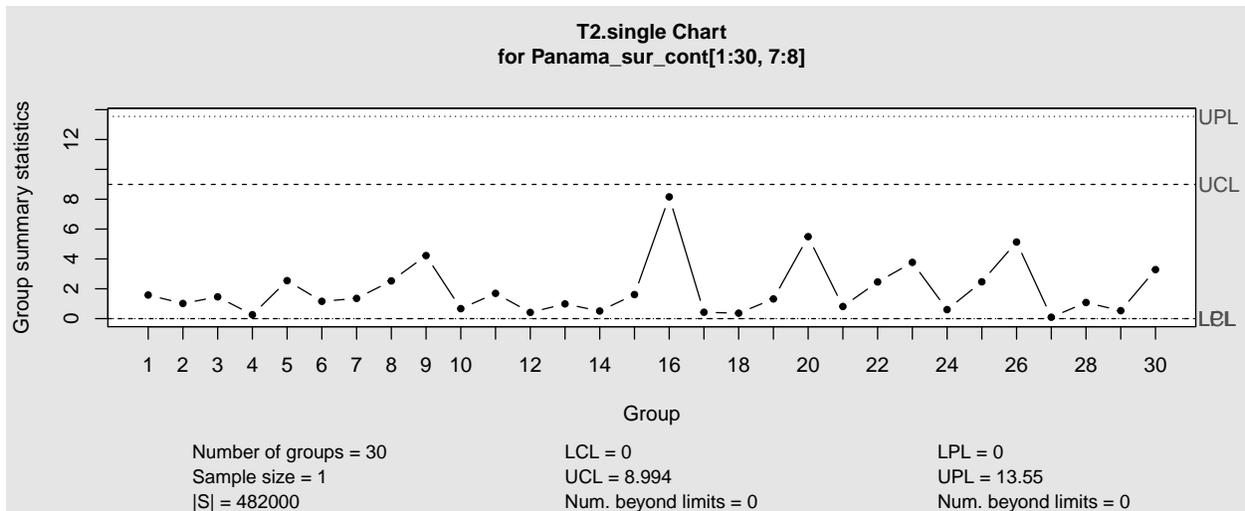


Figura 4.8: Gráfico control  $T^2$  de Hotelling para la etapa de calibrado..

```
TT_First_Lock TT_Second_Lock
      190.2      170.3
```

Covariance matrix:

```
          TT_First_Lock TT_Second_Lock
TT_First_Lock      454.79      -27.66
TT_Second_Lock     -27.66     1061.61
|S|: 482040
```

Control limits:

```
LCL  UCL
    0  8.994
```

Prediction limits:

```
LPL  UPL
    0 13.55
```

```
ellipse=ellipseChart(qq, show.id = FALSE, chart.all = TRUE)
```

La función proporciona medidas descriptivas del estadístico  $T^2$ , número de variables, número grupos, número de observaciones por grupo, vector de medias, matriz de covarianzas muestral y límites de control.

La Figura 4.8 muestra el gráfico de control multivariante para el tránsito de buques contenedores en dirección Sur. Se observa que el proceso de tránsito a través de las dos esclusas está bajo control, con lo que se estiman los límites de control natural, resultando LCL=0 y UCL=8.99. También se han estimado los límites de control de predicción, que serán utilizados para contrastar, en la Fase II, si una nueva observación pertenece a la misma población que la muestra de calibrado.

Igualmente, dado que se estudian dos variables, de forma alternativa al gráfico  $T^2$ , se puede contruir un gráfico de elipse (Figura 4.9), en el que la elipse representa los límites de control naturales para las dos variables de forma conjunta. Además, se han incluido los límites de control univariantes para las dos variables estudiadas.

También se pueden construir alternativas de gráficos con memoria de control para detectar cambios pequeños, como son los gráficos CUSUM y MEWMA, obteniendo resultados similares.

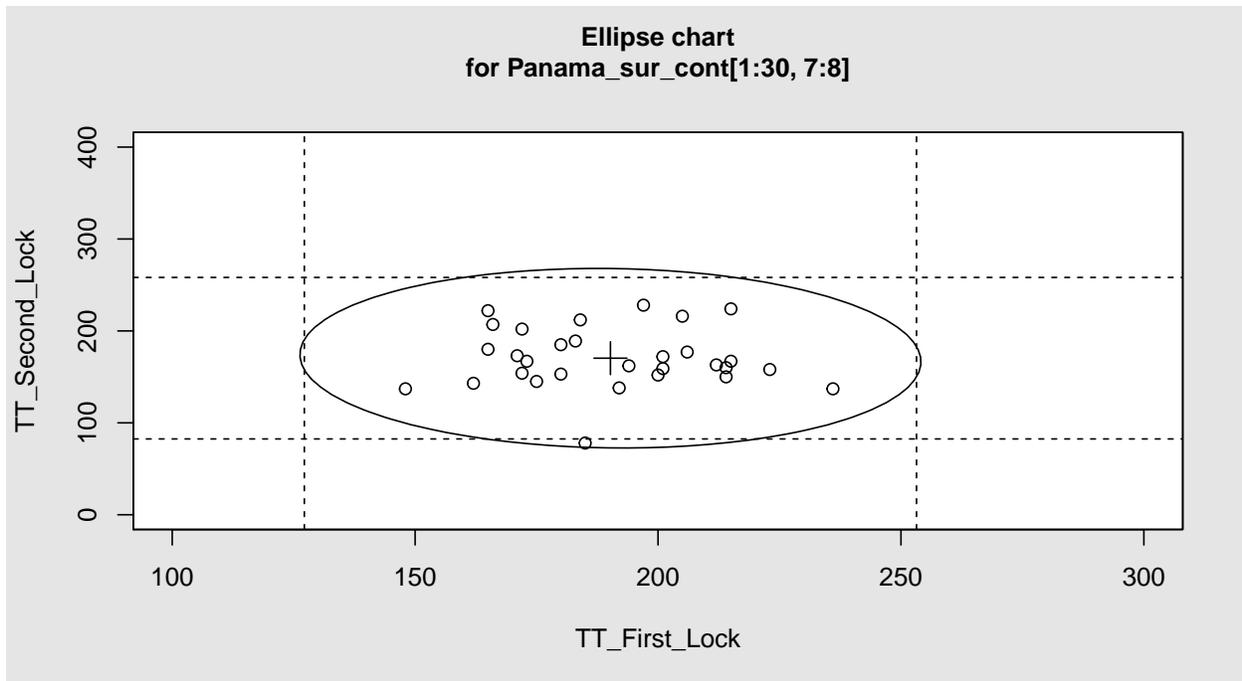


Figura 4.9: Gráfico de elipse representa los límites de control naturales para las dos variables de forma conjunta.

### MCUSUM

El gráfico MCUSUM es el gráfico de las diferencias acumuladas con respecto al vector de medias. La función `mqcd()` se utiliza para crear el objeto multivariante y `mqcs.mcusum()` para obtener el gráfico de control multivariante MCUSUM, de la siguiente forma:

```
data.mqcd <- mqcd(Panama_sur_cont[1:30,7:8])
res.mqcs <- mqcs.mcusum(data.mqcd)
```

A partir del mismo objeto se puede obtener la expresión gráfica del gráfico de control MCUSUM aplicando la función `plot()` (ver Figura 4.10).

Se puede obtener un resumen de las principales características del gráfico de control, incluyendo todos sus estadísticos (vector de medias, matriz de varianzas covarianzas, límites de control).

Summary of group statistics:

```
V1
Min.   :0.000
1st Qu.:0.484
Median :0.828
Mean   :0.956
3rd Qu.:1.356
Max.   :2.716
```

```
Number of quality characteristics: 2
Number of samples or observations: 30
Number of observations or sample size: 1
```

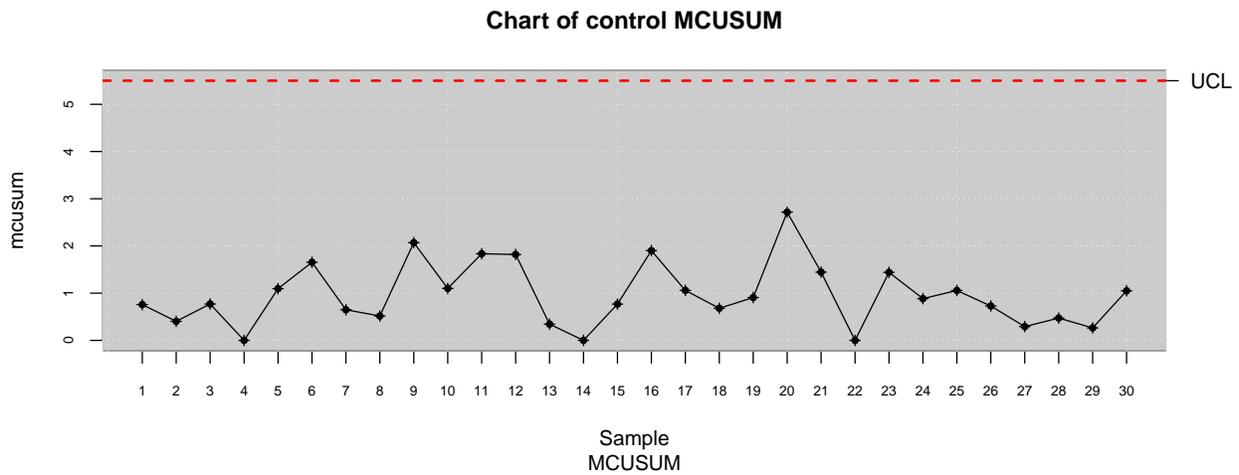


Figura 4.10: MCUSUM transits Contenedores Sur.

Mean Vector:

190.2 170.3

Covariance Matrix:

	TT_First_Lock	TT_Second_Lock
[1,]	454.79	-27.66
[2,]	-27.66	1061.61

Control limits:

lcl ucl  
0.0 5.5

Number beyond limits: 0

## MEWMA

El gráfico de control tipo MEWMA es la versión multivariante del gráfico de medias móviles ponderadas exponencialmente. Este tipo de gráfico se obtiene mediante la función `mqcd()`, que crea el objeto multivariante y `mqcs.mewma()` que proporciona la salida gráfica multivariante.

```
data.mqcd <- mqcd(Panama_sur_cont[1:30,7:8])
res.mqcs <- mqcs.mewma(data.mqcd)
```

A continuación, se muestran los resultados de la estimación del gráfico de control MEWMA a partir de la muestra de calibrado:

Summary of group statistics:

	V1
Min.	:0.0216
1st Qu.	:0.2024
Median	:0.4080
Mean	:0.6512
3rd Qu.	:0.9702

Max. :2.4740

Number of quality characteristics: 2  
 Number of samples or observations: 30  
 Number of observations or sample size: 1

Mean Vector:

190.2 170.3

Covariance Matrix:

	TT_First_Lock	TT_Second_Lock
[1,]	454.79	-27.66
[2,]	-27.66	1061.61

Control limits:

lcl	ucl
0.000	8.634

Number beyond limits: 0

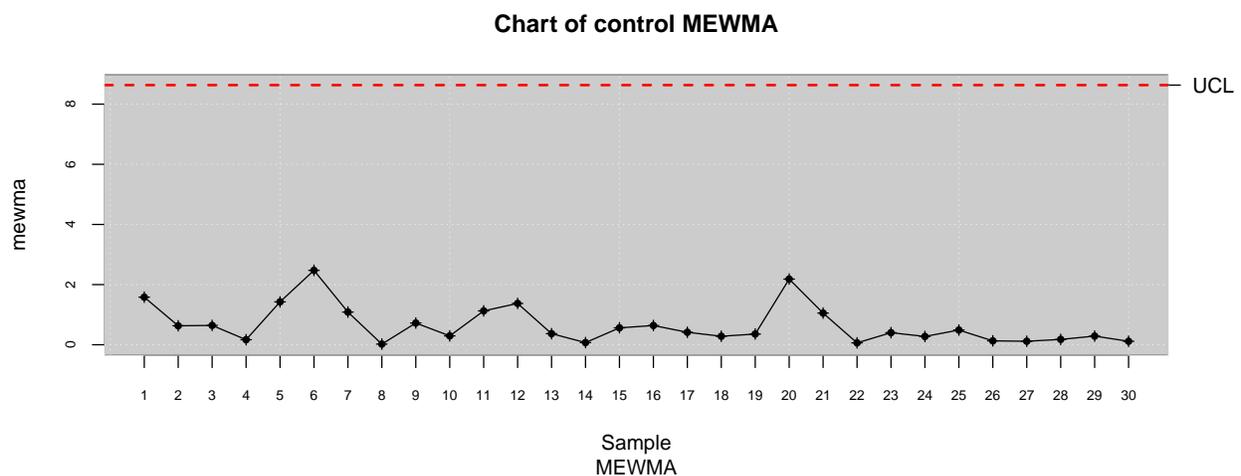


Figura 4.11: MEWMA tránsitos Contenedores Sur.

## 4.2. Técnicas de gráficos de control, Fase II

### 4.2.1. Gráficos de control univariantes

#### Gráficos de medidas individuales Shewart

En esta nueva sección se muestra el proceso estimado de monitoreo o control continuo del proceso. Una vez que se han descartado los valores atípicos y se han estimado los límites de control natural, en la Fase II o de monitorizado se contrasta si una nueva observación pertenece a la misma población que la muestra de calibrado. Si esto no sucede en una o varias observaciones, es indicativo de un cambio puntual o un cambio en el proceso. Si aconteciese un cambio en el proceso, deberían recalcularse los límites de control una vez el proceso se hubiese estabilizado. Se añaden las observaciones correspondientes a la muestra de monitorizado (en términos generales, obtenidas en el intervalo comprendido entre finales de 2016 y el año 2020) mediante el parámetro `newdata`, de la función `qcc`.

Mostramos la forma en como agregamos los nuevos datos para el proceso de monitoreo para el caso particular del tiempo de tránsito de los buques GLP cruzando la esclusa Agua Clara en dirección Sur.

```
aguaclara.s.lpg.mon<-qcc(aguaclara.s.lpg[c(1:6, 10:20)],type="xbar.one",
                        newdata=aguaclara.s.lpg[26:43],nsigmas=3,plot=FALSE)
```

Procedemos a realizar lo mismo para cada una de las poblaciones y obtenemos los resultados en la Figura 4.12.

En todos los casos se puede observar un patrón de cambio gradual en el nivel, presentando una tendencia a la baja, donde en muchos casos el proceso se va estabilizando. Las muestras de monitorizado presentan una media menor y en casi todos los casos muchos puntos fuera de control. La causa asignable a este patrón es el aprendizaje continuo, adquisición de habilidades por parte de los trabajadores y la mejora paulatina en las maniobras de paso a través de las esclusas.

En definitiva, se podría concluir que el tiempo de tránsito medio va disminuyendo. Habría que recalcular los límites de control cuando el proceso se estabilice.

Es importante destacar que en los casos de tránsito por la esclusa Cocolí, el tiempo de tránsito disminuye de forma más débil, en comparación con lo que ocurre cuando las embarcaciones cruzan la esclusa Agua Clara. En el caso de Cocolí, se observa que la disminución del tiempo de tránsito tiende a producirse antes y de forma más acusada cuando la dirección es Norte (entrando desde el océano a la esclusa), mientras que en el caso de Agua Clara, la disminución del tiempo de tránsito es significativamente mayor, independientemente de la dirección de tránsito y del tipo de buque estudiado.

En definitiva, se comprueba que la aplicación de gráficos de control tipo Shewhart es factible y muy útil para detectar anomalías y otros cambios en el proceso, como es el caso del patrón o curva de aprendizaje. Mediante el uso de estos gráficos se puede evaluar el grado de idoneidad de los procedimientos y personal asignado por la Autoridad del Canal para efectuar los tránsitos por el Canal Ampliado. En este caso específico, se observa que la gestión realizada por la Autoridad del Canal está permitiendo la mejora continua en los tránsitos, si bien está llegando al límite de la mejora potencial con los medios disponibles para la realización del proceso de tránsito.

### Gráficos de control CUSUM

En esta sección se aplicarán los gráficos CUSUM a las muestras de monitorizado (tomadas desde finales de 2016 al 2020) seleccionadas según la dirección, esclusa y tipo de buque. La intención última es evaluar su utilidad para la detección de cambios en el proceso, en este caso debidos al aprendizaje de pilotos e instalaciones. Al igual que los EWMA, suelen tener una mayor eficiencia para la detección de cambios pequeños en el proceso. A tal fin se utilizará la función `cusum()` del paquete `qcc`.

Al igual que en el caso de los gráficos Shewart se ha identificado, en general, un patrón de aprendizaje, efectivamente se ha producido una disminución paulatina en el tiempo de tránsito que induce a que el proceso de tránsito salga del control estadístico en todos los casos estudiados (si se comparan las observaciones nuevas con las correspondientes a la muestra de calibrado).

Como se ha mencionado anteriormente, cambios inferiores a  $2 \cdot \sigma$  los detectan mejor los gráficos CUSUM y EWMA, mientras que cambios superiores a  $2 \cdot \sigma$  los detectan mejor los gráficos de tipo Shewhart.

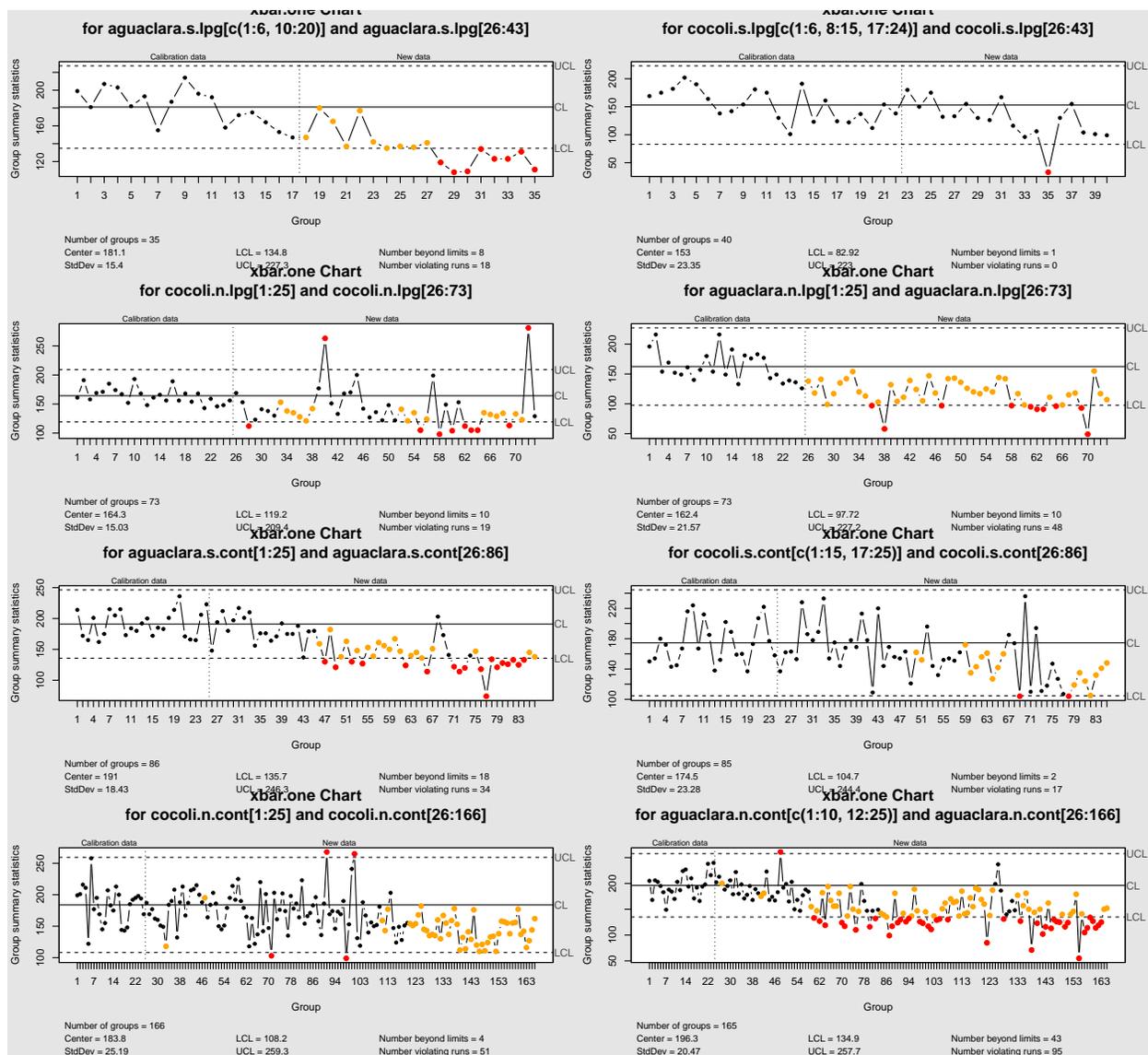


Figura 4.12: Gráficos Shewart del tránsito de buques GLP (primera y segunda filas superiores) y Contenedores (tercera y cuarta filas) cruzando las esclusas Cocolí y Agua Clara en dirección Sur (filas 1 y 3) y Norte (filas 2 y 3). Aplicación de los gráficos de control en la Fase II, etapa de monitorizado.

Al comparar gráficos Shewart con CUSUM se observa que en ambos resultados el comportamiento es bastante similar, el único caso donde podríamos decir que existe una diferencia importante es en el tiempo de tránsito de los buques Contenedores por la esclusa Cocolí en dirección Sur. En este caso los gráficos Shewart detectan mucho antes el cambio en el proceso, por lo que estamos ante un cambio grande. Por tanto, en el caso particular del Canal de Panamá Ampliado, no sería necesaria la aplicación de gráficos con memoria como los gráficos de control CUSUM.

### Gráficos de control EWMA

Al igual que se ha evaluado la pertinencia de aplicar gráficos CUSUM, se hace lo mismo con los gráficos EWMA, obteniendo los resultados de la Figura 4.14.

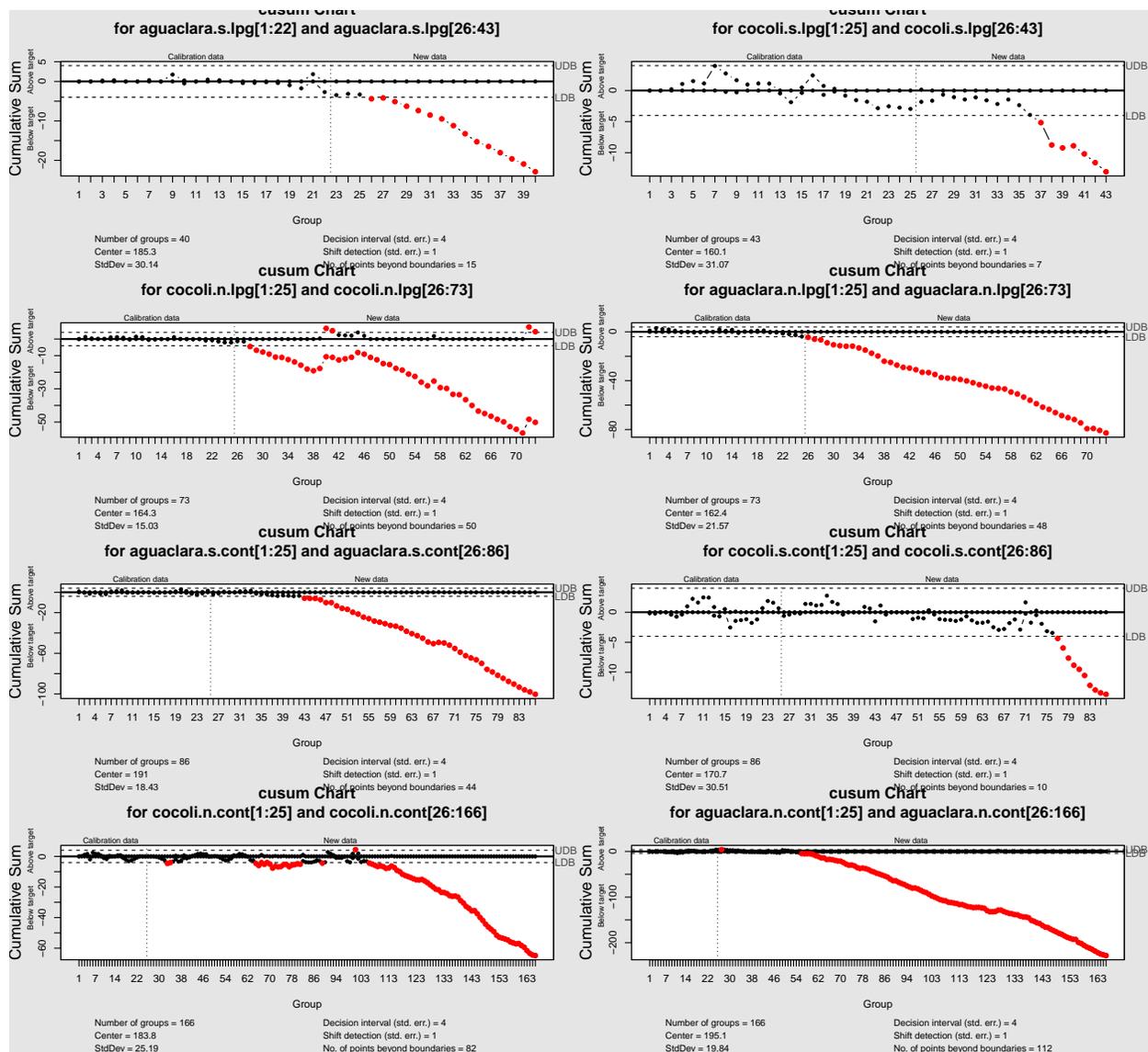


Figura 4.13: Gráficos CUSUM del tránsito de buques GLP (primera y segunda filas superiores) y Contenedores (tercera y cuarta filas) cruzando las esclusas Cocolí y Agua Clara en dirección Sur (filas 1 y 3) y Norte (filas 2 y 4). Aplicación de los gráficos de control en la Fase II, etapa de monitorizado.

Al comparar cuál detecta más rápido los puntos fuera de control entre los gráficos Shewart y EWMA, se verifica que el comportamiento es bastante similar, sin embargo, en los casos del tiempo de tránsito de Contenedores pasando por la esclusa Cocolí en dirección Sur y Contenedores pasando por la esclusa Agua Clara en dirección Norte, los gráficos Shewart detectan mucho antes el cambio en el proceso.

Al igual que en los casos anteriores (Shewart y CUSUM) habría que reconsiderar calcular los límites nueva vez porque se ha identificado un cambio en el proceso en la forma de un patrón de cambio paulatino que tiende a saturarse o estabilizarse.

Se observa también que, tal y como sucedía con los gráficos CUSUM, los gráficos EWMA no anticipan la identificación del estado fuera de control con respecto a los gráficos Shewart, su correspondientes ARL no tienden a ser menor que los relativos a los gráficos Shewart. Por ello, se recomienda la aplicación de estos últimos, con la ventaja adicional de ser más intuitivos y fáciles de interpretar, más si cabe en el ámbito de la empresa.

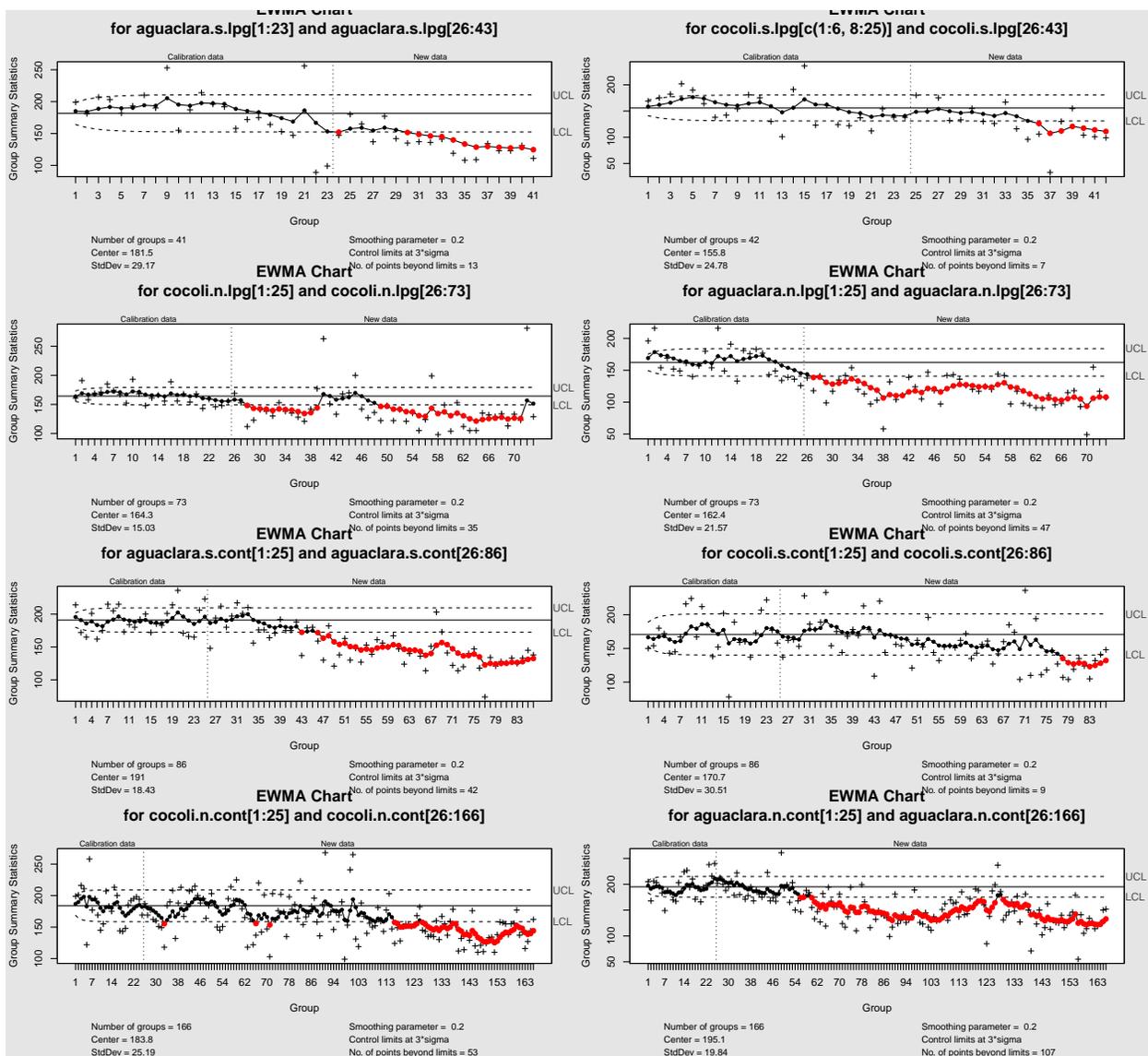


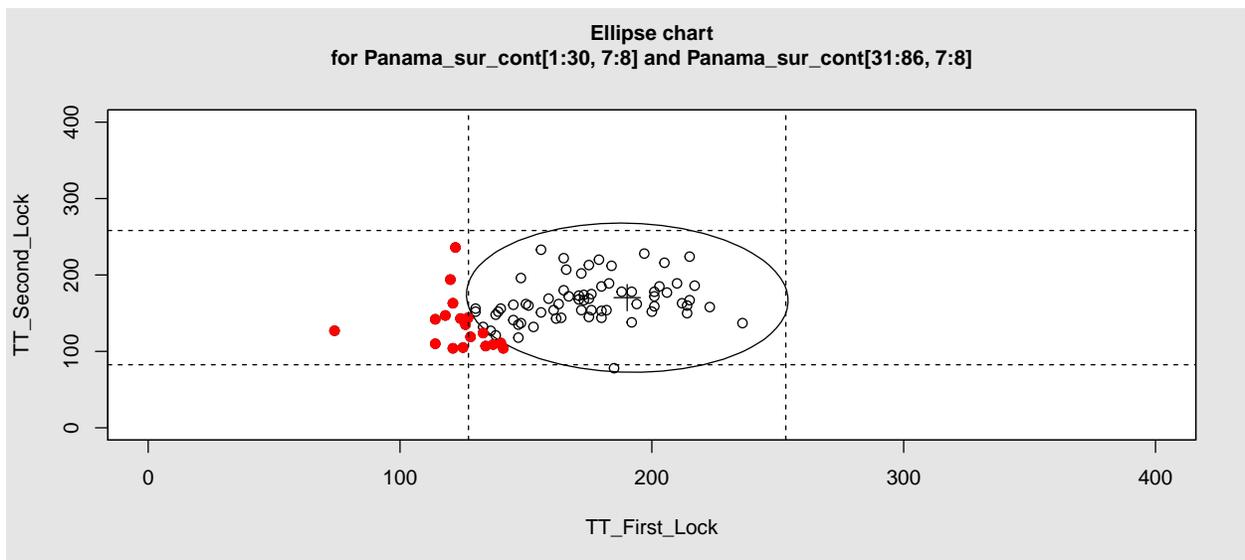
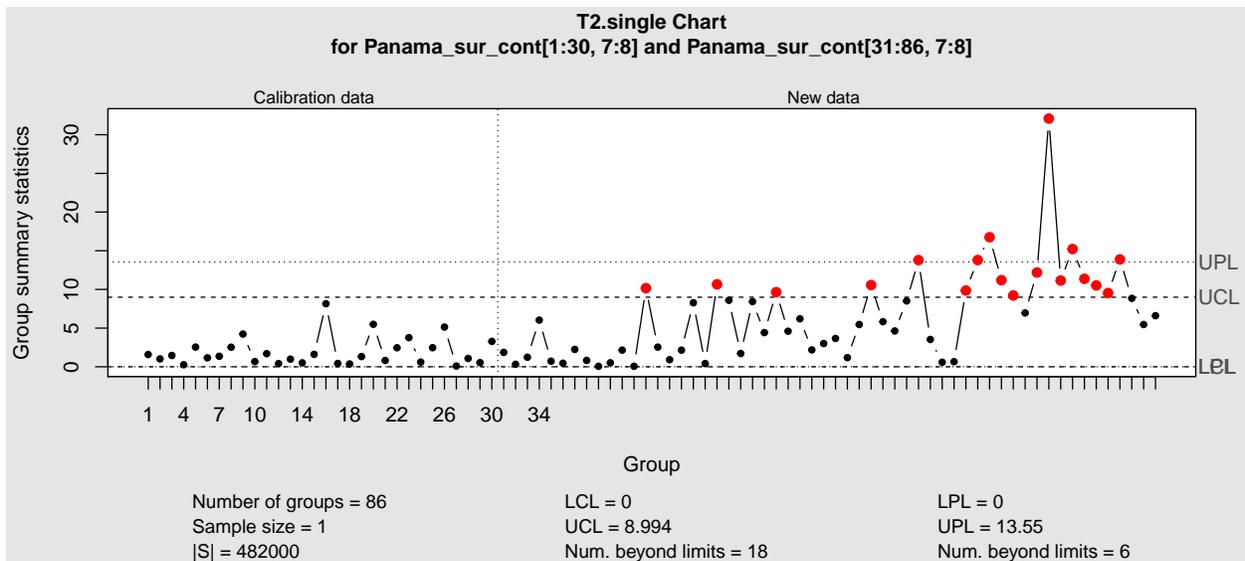
Figura 4.14: Gráficos EWMA de observaciones individuales del tránsito de buques GLP (primera y segunda filas superiores) y Contenedores (tercera y cuarta filas) cruzando las esclusas Cocoli y Agua Clara en dirección Sur (filas 1 y 3) y Norte (filas 2 y 4). Aplicación de los gráficos de control en la Fase II, etapa de monitorizado..

#### 4.2.2. Gráficos de control multivariante

Partiendo de los límites de control estimados en la Fase I, se monitorizan los tránsitos de los buques Contenedores con dirección Sur producidos desde el 9 de agosto de 2016.

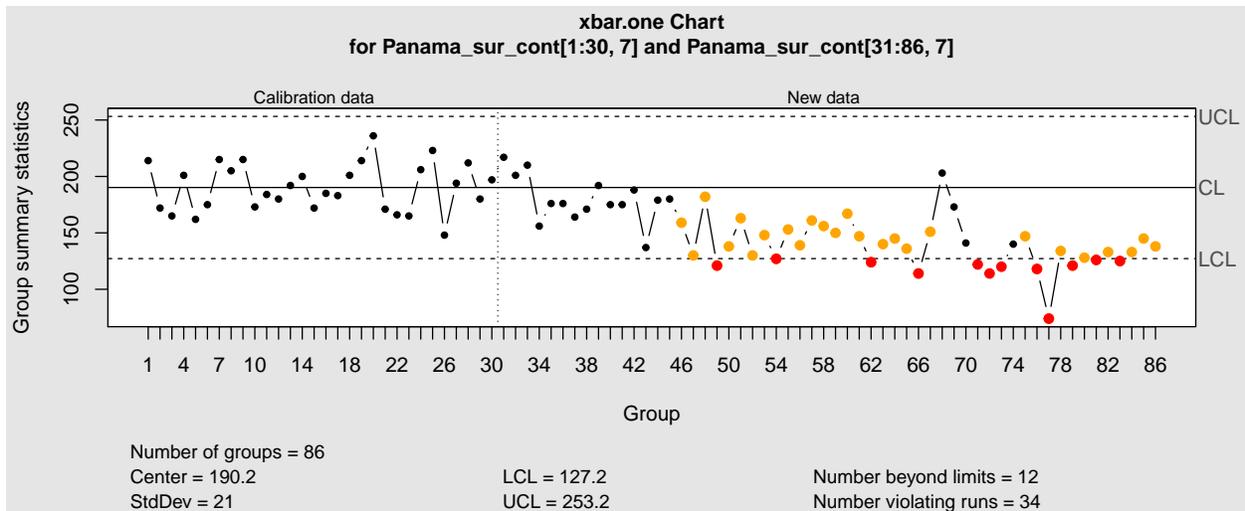
##### Gráficos T2 de Hotelling

```
qq = mqcc(Panama_sur_cont[1:30,7:8], type = "T2.single",
          newdata = Panama_sur_cont[31:86,7:8], pred.limits = TRUE)
```



Atendiendo a los límites de predicción, se observa que el proceso está fuera de control (indicador de la existencia de un patrón de aprendizaje) en 02/10/2017, sensiblemente más tarde que utilizando los gráficos de control univariantes (02/11/2016). Si se hace caso a los límites de calibrado, se detecta que el sistema está fuera de control el 31/10/2016, resultado muy similar al aportado por el enfoque univariante. En el gráfico de elipse se observa que todos los estados fuera de control se deben a una disminución del tiempo de tránsito por la primera esclusa. Es decir, que el aprendizaje es mucho más significativo en la primera esclusa, que es Agua Clara. Esta conclusión está de acuerdo con lo mostrado en secciones anteriores.

Un siguiente paso, para identificar qué variable ha originado los estados fuera de control, es construir de nuevo los gráficos de control univariantes pero, en este caso, aplicando una corrección tipo Bonferroni, para evitar el problema del contraste múltiple de hipótesis (estamos haciendo dos contrastes, uno por variable).



Call:

```
qcc(data = Panama_sur_cont[1:30, 7], type = "xbar.one", newdata = Panama_sur_cont[31:86,
```

```
7], conf
```

```
xbar.one chart for Panama_sur_cont[1:30, 7]
```

Summary of group statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
148.0	172.2	188.5	190.2	205.8	236.0

Group sample size: 1

Number of groups: 30

Center of group statistics: 190.2

Standard deviation: 21

Summary of group statistics in Panama\_sur\_cont[31:86, 7]:

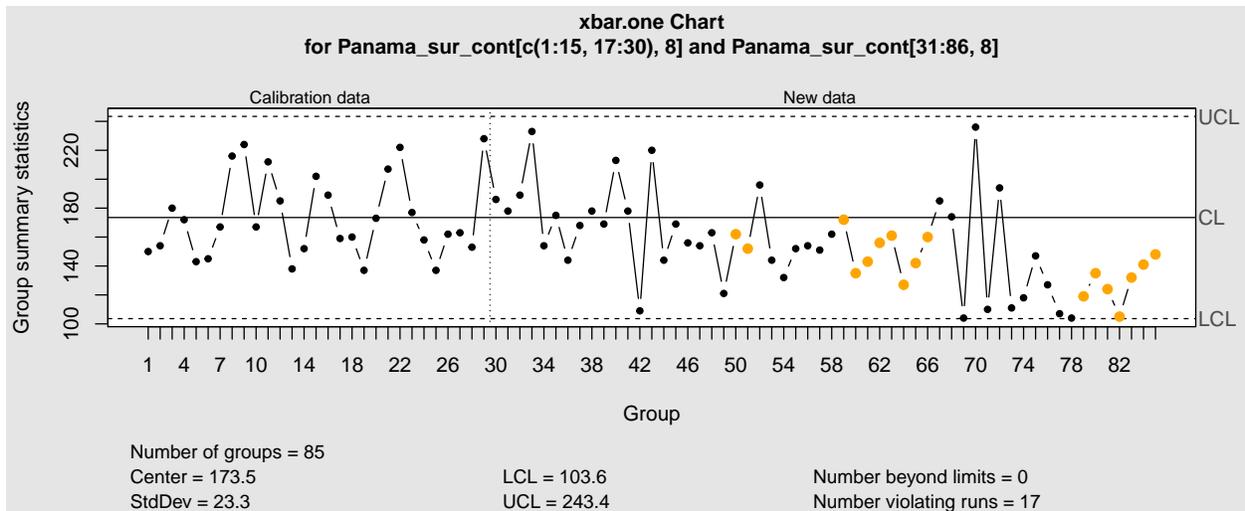
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
74.0	130.0	146.0	150.2	171.5	217.0

Group sample size: 1

Number of groups: 56

Control limits:

LCL	UCL
127.2	253.2



Call:

```
qcc(data = Panama_sur_cont[c(1:15, 17:30), 8], type = "xbar.one", newdata = Panama_sur_cont[31:86,
```

```
xbar.one chart for Panama_sur_cont[c(1:15, 17:30), 8]
```

Summary of group statistics:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
137.0	153.0	167.0	173.5	189.0	228.0

Group sample size: 1

Number of groups: 29

Center of group statistics: 173.5

Standard deviation: 23.3

Summary of group statistics in Panama\_sur\_cont[31:86, 8]:

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
104.0	132.0	153.0	154.0	172.5	236.0

Group sample size: 1

Number of groups: 56

Control limits:

LCL	UCL
103.6	243.4

El patrón de aprendizaje en Agua Clara, con dirección sur, es mucho más acentuado.

Seguidamente se construyen los gráficos de control MCUSUM y MEWMA para la Fase II de monitorizado.

### MCUSUM

Se creal el objeto multivariante con la muestra de monitorizado con la función `mqcd()`.

```
data.mqcd.new <- mqcd(Panama_sur_cont[31:86,7:8])
```

En este caso se selecciona el vector de medias y la matriz de covarianzas de la muestra de calibrado.

```
Xmv=matrix(res.mqcs$mean,byrow=TRUE)
S = data.frame(matrix(res.mqcs$S,byrow=FALSE,nc=2))
```

Se calcula el nuevo gráfico con los límites obtenidos en la muestra de calibrado:

```
res.mqcs.new <- mqcs.mcusum(data.mqcd.new,Xmv=Xmv,S=S)
```

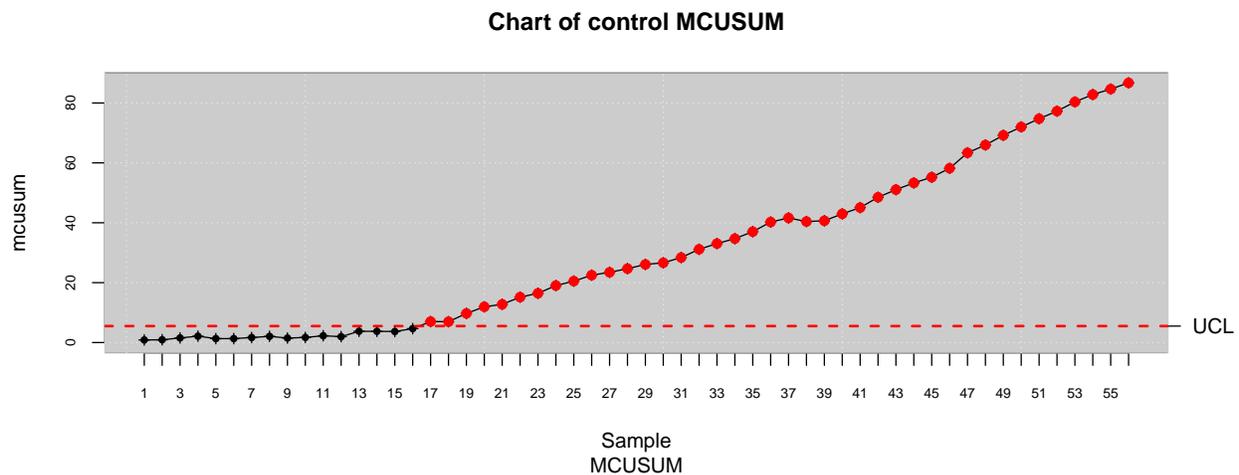


Figura 4.15: MCUSUM tránsitos Contenedores Sur.

### MEWMA

Con la función `mqcd()` se crea el objeto multivariante para la monitorización del resto de muestras.

```
data.mqcd.new <- mqcd(Panama_sur_cont[31:86,7:8])
```

Se selecciona el vector de medias y la matriz de covarianzas de la muestra de calibrado para representar los límites de control estimados, junto con la muestra que se quiere monitorizar:

```
Xmv=matrix(res.mqcs$mean,byrow=TRUE)
S = data.frame(matrix(res.mqcs$S,byrow=FALSE,nc=2))

res.mqcs.new <- mqcs.mewma(data.mqcd.new,Xmv=Xmv,S=S)
```

Mediante la aplicación de los gráficos de control MEWMA o CUSUM se consigue detectar, casi en la misma fecha que los gráficos de control univariantes (3/11/2016) que el proceso de tránsito está fuera de control debido al aprendizaje de pilotos e instalaciones. El resultado también es similar al  $T^2$  de Hotelling si se utilizan los límites de control de estimación.

## 4.3. Análisis de capacidad de las instalaciones del Canal

El análisis de capacidad proporciona información acerca del grado en que se cumplen las especificaciones marcadas por empresa, clientes o normativa. En este caso específico, para las esclusas del Canal antiguo, la Autoridad del Canal de Panamá tiene una política de incentivos para aquellos pilotos que hagan tránsitos

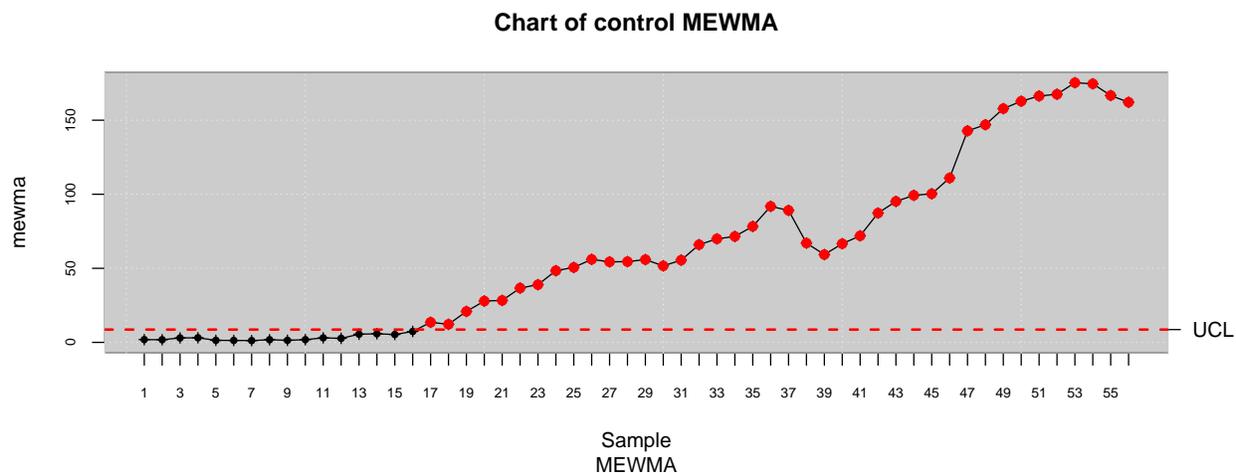


Figura 4.16: MCUSUM tránsitos Contenedores Sur.

a tiempos menores que un determinado límite. Estos límites varían según la esclusa y la eslora del barco. Además, también se fijan límites de tiempo para el paso de buques a partir de los cuales se ha de indemnizar a las compañías navieras. Por tanto, el análisis de capacidad de proceso, teniendo en cuenta un sólo límite de especificación superior, USL, proporciona una medida acerca del grado en el que la Autoridad del Canal no tendrá que pagar indemnizaciones o, alternativamente, el grado en el que los pilotos realizan trayectos lo suficientemente rápidos (recibiendo un suplemento de sueldo) para incrementar el número de tránsitos y el beneficio de la empresa.

En el caso del Canal Ampliado, aunque todavía no hay una normativa fija de incentivos por rapidez en el tránsito, se está comenzando a utilizar  $USL=180$  min, es decir, 3 horas. El análisis de capacidad, en principio, lo hacemos a partir de un solo USL, no tendría tanto sentido añadir un LSL.

A partir de estas especificaciones se analizará si el proceso de tránsito es capaz de cumplir con las especificaciones marcadas, teniendo en cuenta las mediciones históricas de las esclusas. Se calculan los índices de capacidad mediante la comparación de las especificaciones con el rango de variación real de la variable CTQ.

Antes de realizar el análisis de capacidad, se comprueba que el proceso esté bajo control y que conocemos la distribución de la probabilidad de la variable de calidad del proceso.

Mediante la aplicación de gráficos de control de medidas individuales se detecta, especialmente al analizar la muestra de monitorización, que el proceso está fuera de control, como resultado de un patrón de cambio paulatino hacia niveles más bajos en el tiempo de tránsito, que se puede asociar a un proceso de aprendizaje de pilotos y organización. Al comparar ambas fases en los gráficos de arriba se observa que los tiempos están ya situados en otro nivel más bajo, sobre todo en el caso de los buques que cruzan Agua Clara. Por lo tanto, a la hora de hacer el análisis de capacidad de la instalación, se hará a partir de estas últimas observaciones en las que el proceso se ha estabilizado.

Entonces, dado que el proceso ha cambiado, se estiman de nuevo los límites de control a partir de estas últimas observaciones en las que el proceso se ha estabilizado. Hemos eliminado de los cálculos una serie de buques, por tener un tiempo de tránsito muy alto o muy bajo, debido a causas aparte de la propia aleatoriedad de los datos, como, por ejemplo, falta de agua puntual en las esclusas, mal tiempo, problemas puntuales en la maniobra o en las instalaciones.

La librería `qcc` incluye funciones útiles para llevar a cabo un estudio de capacidad de procesos. Esta librería permite hacer un estudio de capacidad de procesos a partir de los estadísticos obtenidos mediante la construcción de un gráfico de control de medidas individuales en nuestro caso.

Mostramos un ejemplo del paso del buque GLP por la esclusa Agua Clara.

```

aguaclara.s.lpg.pc<-qcc(aguaclara.s.lpg[c(26,29,31:43)],type="xbar.one",nsigmas=3,
                        restore.par=FALSE, plot=TRUE)

process.capability(aguaclara.s.lpg.pc, spec.limits=c(NA,180), restore.par = FALSE)

```

Se aplica este procedimiento a cada una de las poblaciones estudiadas y se obtienen los resultados de la Figura 4.17. En cada uno de sus paneles se muestra el histograma del tiempo de tránsito, la función de densidad asumiendo distribución normal (en línea de puntos, sustituyendo los parámetros por los correspondientes valores muestrales), el número de observaciones, la desviación típica y media muestral, los índices de capacidad (en el caso unilateral por la derecha sólo se calcula el  $C_{pu}$ ), la proporción de observaciones observadas que son mayores que el USL y la proporción de observaciones mayores que el USL esperadas asumiendo una distribución normal para el tiempo de tránsito.

Antes de analizar la información mostrada en la Figura 4.17, es importante destacar que el límite USL=180 min es el criterio que se está utilizando para primar el buen trabajo realizado a los pilotos. Por tanto, en este caso, si el resultado del análisis es que los pilotos son capaces de cumplir esta especificación, podría ser suficiente indicio para que la Autoridad del Canal comenzase gestiones para bajar el límite de especificación, impulsando a los pilotos a bajar los tiempos (pues habría margen de mejora) y, como consecuencia, el número de tránsitos, la satisfacción del cliente y el beneficio de la empresa.

Asumiendo que el proceso de tránsito es un proceso existente desde hace tres años, además de que es un proceso con un sólo límite de especificación, unilateral, este será capaz de cumplir especificaciones cuando su correspondiente índice de capacidad sea superior a 1.25 (Montgomery, 2005). Teniendo en cuenta este criterio, la Figura 4.17 muestra que el Canal y los pilotos son capaces de cumplir especificaciones en dos escenarios: en los trayectos de buques GLP realizados a través de la esclusa de Agua Clara en dirección Sur y en dirección Norte. Específicamente, para el caso del paso de Agua Clara en dirección sur, el proceso de tránsito es capaz incluso a un nivel Seis Sigma, con  $C_{pu}=2.1$ , y con una proporción esperada de trayectos no conformes ínfima, menor que 0.0009 por millón. En el caso de los trayectos por Agua Clara en dirección norte, el proceso sigue siendo capaz, pero no a un nivel Seis Sigma, siendo  $C_{pu}=1.27$  (menos de 48 trayectos no conformes por millón).

En el resto de casos el proceso no es capaz. Habría que reducir la variabilidad y/o reducir el tiempo de tránsito a través de las esclusas para que el proceso fuese capaz de cumplir especificaciones. Es interesante observar el caso particular de los buques contenedores a través de la esclusa Cocolí, dado que se estima que alrededor de un 18-19% de los tránsitos no cumplan especificaciones. Estos son los escenarios caracterizados por tiempos de tránsito más altos, siendo  $C_{pu} = 0.3$ .

Teniendo en cuenta los resultados de la Figura 4.17, podría recomendarse bajar el USL en el caso de los trayectos de buques GLP por la esclusa de Agua Clara, sobre todo en el caso de que el trayecto se realice en dirección Sur. En las condiciones actuales la práctica totalidad de estos tránsitos reciben incentivos, lo cuál debería corregirse con el fin de fomentar la reducción del tiempo de tránsito, parte del el plan de mejora continua del Canal. De hecho, si comparamos las diferencias entre los tiempos de tránsito según tipo de buque, dirección y esclusa, no es coherente que sólo haya un USL para todos los tránsitos. Por ello también podrían definirse diferentes USL para los buques Contenedores que cruzan la esclusa Cocolí (ya sea en dirección Norte o Sur), y otro USL diferente para el resto de escenarios definidos por  $C_{pu}$  de entre 0.53 y 0.92.

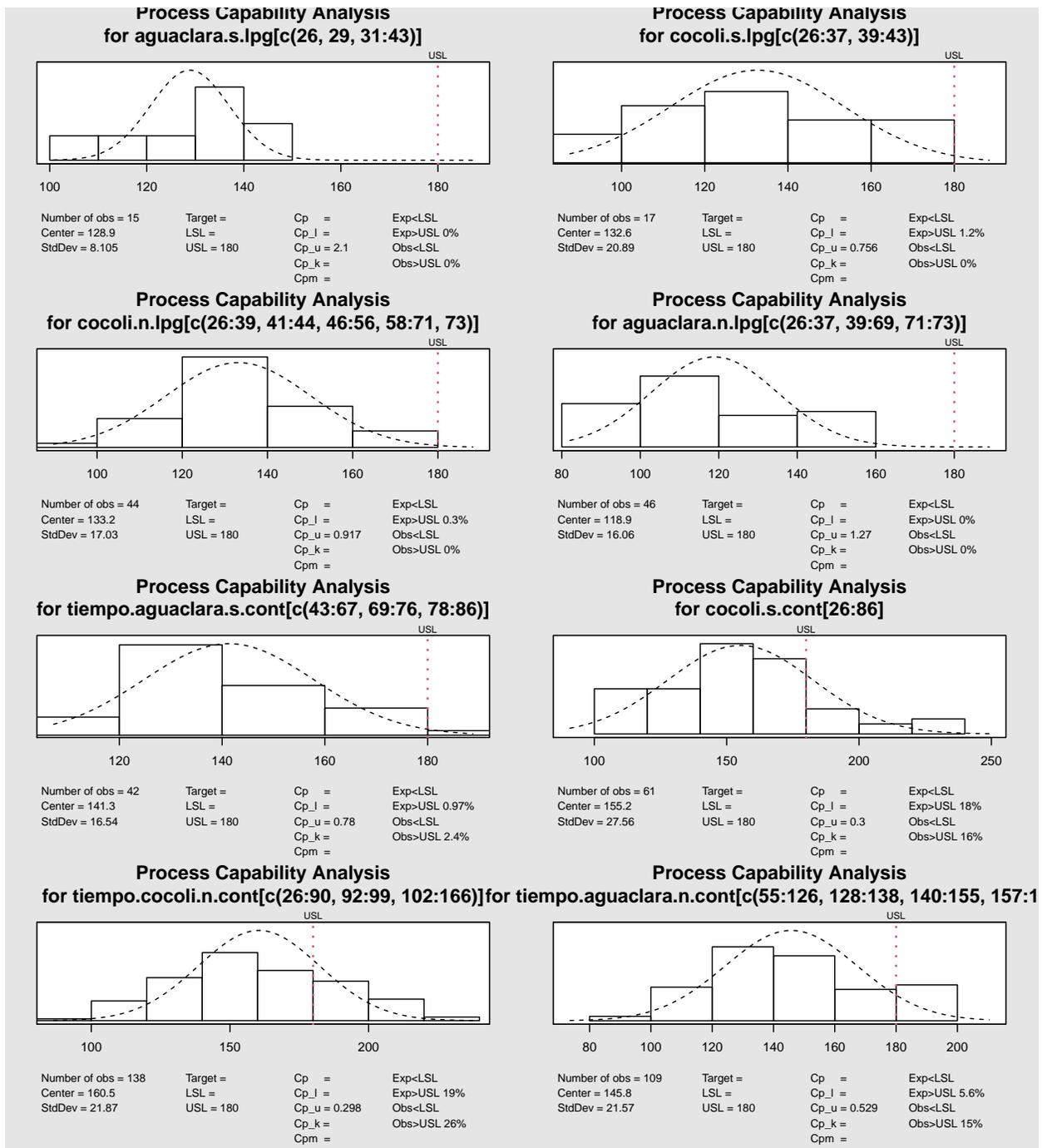


Figura 4.17: Análisis de capacidad univariante.

### Análisis de capacidad multivariante

En este caso utilizamos el paquete MPC1, que permite calcular índices de capacidad de procesos a partir de datos multivariantes. La sintaxis de la función `mpci` es la siguiente:

Los principales argumentos utilizados para calcular los indicadores incluidos en este paquete son los siguientes:

- Index: Especifica el índice que se calculará, en el caso del vector de capacidad multivariante utilizamos el índice “shah”, para el índice de capacidad multivariante MCpm utilizamos “taam”.
- x: La matriz de los datos (las características de calidad son las columnas).
- LSL: El vector de los límites inferiores de especificación.
- USL: El vector de los límites superiores de especificación.
- Target: El vector del objetivo del proceso. Si falta el argumento Target, el vector Target, el programa lo calcula como el punto medio del intervalo de especificación (valor predeterminado). alpha: Especifica la forma cuadrática, valor predeterminado es 0.0027.

Los argumentos npc y Method se refieren a los índices basados en PCA, pero que no serán discutidos en este trabajo. El gráfico de argumento lógico permite, en el caso bivariado, representaciones gráficas adicionales concernientes al proceso especificaciones.

Según la Autoridad del Canal de Panamá, tampoco sería recomendable hacer los tránsitos a través de las esclusas demasiado rápido, pues conllevaría aumentar la posibilidad de accidente, provocando el bloqueo o el daño de instalaciones y buque. Por ello, en este caso, para ilustrar el análisis de capacidad multivariante, incluyendo los valores de todos los índices implementados en el paquete MCPI, proponemos un límite de especificación LSL=50 min, basándonos en que tiempos menores a 50 min serían resultado de una temeridad (según la Autoridad del Canal) y podrían poner en riesgo el buque y la instalación.

El análisis de capacidad se hará a partir de las observaciones de los tránsitos de los buques contenedores a través de las esclusas, en dirección Sur, una vez el proceso se ha estabilizado a valores menores del tiempo de tránsito (intervalo 2017-2020).

```
library(MPCI)
LSL <- c(50, 50)
USL <- c(250.6, 250.6)
Target=c(135,135)
x=as.matrix(Panama_sur_cont[57:86,7:8])
a <- mpci(index="taam",x,LSL,USL,Target)
mpci(index="wang", x,LSL,USL,Target)
```

```
[[1]]
[1] "Wang and Chen (1998) Multivariate Process Capability Indices(PCI) based on PCA"

$`number of principal components`
[1] 2

$MCp
[1] 1.076

$MCpk
[1] 0.935

$MCpm
[1] 1.058

$MCpmk
[1] 0.9198
```

```
mpci(index="pan",x,LSL,USL,Target)
```

```
[[1]]
[1] "Pan and Lee (2010) Multivariate Capability Index (NMCpm)"

$NMCpm
      [,1]
[1,] 1.201
```

```
mpci(index="xeke",x,LSL,USL,
Target,Method = 1)
```

```
[[1]]
[1] "Xekalaki and Perakis (2002) Multivariate Process Capability Indices(PCI) based on PCA"

$`number of principal components`
[1] 2

$MCp
[1] 1.226

$MCpk
[1] 1.096

$MCpm
[1] 1.195

$MCpmk
[1] 1.068
```

Prácticamente todos los índices están en torno a 1, por lo que, en caso de seguir los criterios univariantes para los que el límite es  $C_p = 1.33$  (para procesos bilaterales), el proceso está justo por debajo del límite de ser capaz de cumplir especificaciones. El hecho de que los indicadores con subíndices  $p$  (capacidad potencia),  $pk$  y  $pm$  (capacidad real), sean muy similares, da a entender que el proceso está centrado y que sólo se podría aumentar su capacidad reduciendo la variabilidad de los tránsitos.

Las principales fuentes utilizadas en este capítulo son Carral et al. (2021), Carral et al. (2020), Flores (2020) Carral et al. (2019), Scrucca et al. (2017), Santos y Scagliarini (2015), Bersimis et al. (2005)

## Capítulo 5

# Conclusiones

En este trabajo se han introducido las curvas de aprendizaje y se ha considerado factible su estimación en el marco del pilotaje en el Canal de Panamá. Esto se puede hacer teniendo en cuenta que este proceso de pilotaje es un caso especial del campo del transporte donde el servicio/producto final es el resultado de maniobras repetitivas.

En este estudio, la variable que define la calidad son los tiempos de tránsito a través de las esclusas Cocolí y Agua Clara. Así, se han aplicado técnicas de control estadístico de calidad al tiempo en tránsito de los buques en función de cada esclusa, dirección, tipo de buque y variables sobre las dimensiones como eslora y manga, con el fin de identificar posibles patrones de aprendizaje, y, además, para estimar los modelos de regresión estadística que mejor los expliquen.

Previo al estudio de los datos reales se estudia la curva de aprendizaje del tiempo de tránsito del buque Baroque, navegado a través de la esclusa Agua clara y utilizado en un programa de aprendizaje operativo como formación para los pilotos. Se verifica que a medida que pasa el tiempo, desde que iniciaron las capacitaciones, la velocidad a la que disminuye el tiempo de tránsito se hace cada vez menor, hay una mejora en el pilotaje a lo largo del tiempo, hasta alcanzar una zona de saturación.

Se ha realizado un análisis estadístico exhaustivo a la base de datos reales para identificar los factores de influencia. De esta forma se obtiene un modelo para estimar y caracterizar el tiempo de tránsito por las nuevas esclusas de Cocolí y Agua Clara. Además, podemos tener una idea de cuáles variables deben tenerse en cuenta antes de aplicar técnicas de detección de patrones.

Al aplicar un modelo GAM hemos identificado que todos los efectos del modelo resultan significativos ( $p$ -valor  $< 0.05$ ), excepto la categoría del tipo de barco GLP. El modelo estima que los buques de GNL tardan 20.92 minutos (en valores medios) menos en pasar por las esclusas que los Contenedores. En cuanto a los resultados obtenidos con respecto a la variable dirección y esclusa, la dirección Sur produce un tiempo de tránsito significativamente mayor a lo largo de las esclusas frente a la dirección Norte, mientras que el tiempo esperado en tránsito tiende a ser mayor en la esclusa Cocolí si se compara con Agua Clara.

En el caso de la interacción entre dirección y esclusa, el tiempo de tránsito es mayor para la combinación de la esclusa Cocolí y dirección Norte, y la esclusa Agua Clara con dirección Sur.

Con respecto al tiempo de experiencia, cuando aumenta el número de maniobras de pilotaje el tiempo de tránsito disminuye. Se verifica que la curva de aprendizaje, es de tipo no lineal, específicamente asintótico. En un inicio, existiera un amplio margen de mejora. Sin embargo, a medida que pasaba el tiempo, la velocidad a la que disminuía el tiempo de tránsito se hacía cada vez menor, hasta llegar a una zona de saturación.

Con respecto a las estimaciones de la manga y la eslora de la embarcación sobre el tiempo de tránsito, su efecto es lineal y creciente.

También es importante destacar que el modelo multivariado explica el 36.7% del tiempo total en la variabilidad del tránsito. La variabilidad no explicada podría estar relacionada con causas asignables estacionales y meteorológicas, además de las diferentes habilidades de los pilotos, entre otras causas.

Los patrones de aprendizaje se han identificado mediante el uso de medidas individuales y gráficos de control CUSUM y EWMA. Se puede visualizar que en casi todos los casos se ha identificado un patrón de aprendizaje para el tiempo de tránsito, al observar la muestra de los nuevos datos correspondiente a las medidas obtenidas a partir de la 25 primeras observaciones. Habría que recalculer los límites de control porque el proceso ha cambiado. Observándose en algunos casos el proceso llega a un estado de saturación. Específicamente, el patrón de aprendizaje en Agua Clara, con dirección sur, es más acetuado.

En general, se pudieron observar comportamientos bastantes similares al utilizar una técnica u otra, aun así en algunos caso parecería que estos patrones de aprendizaje se identifican de forma más clara mediante el uso de gráficos de control de medias individuales Shewart.

Con respecto al análisis de capacidad univariante, solo los casos del buque GLP pasando por la esclusa Agua Clara, tanto en dirección norte como en dirección Sur, son capaces de cumplir con las especificaciones, en el resto de casos el proceso no es capaz. Habría que reducir la variabilidad y/o reducir el tiempo de tránsito a través de las esclusas para que el proceso fuese capaz de cumplir especificaciones.

En el caso multivariante, prácticamente todos los índices están en torno a 1, por lo que, en caso de seguir los criterios univariantes para los que el límite es  $C_p = 133$  (para procesos bilaterales), el proceso está justo por debajo del límite de ser capaz de cumplir especificaciones. El proceso está centrado y sólo se podría aumentar su capacidad reduciendo la variabilidad de los tránsitos.

# Bibliografía

- Agresti A (2007) An introduction to categorical data analysis. Wiley, New York.
- Alfaro-Navarro JL (2002) Métodos multivariantes en control estadístico de la calidad. Facultad de Ciencias Económicas y Empresariales de Albacete. Universidad de Castilla-La Mancha
- Página web de la Autoridad del Canal de Panamá. <https://micanaldepanama.com/>. Accedido 1 de octubre de 2020.
- Bass I (2007) Sig Sigma Statistics with Excel and Minitab. Mc Graw Hill, United States.
- Bates DM y Watts DG (2007). Nonlinear Regression Analysis and its Applications. New York: Wiley.
- Bersimis S, Psarakis S, Panaretos J (2005). Multivariate Statistical Process Control Charts and the Problem of Interpretation: A Short Overview and Some Applications in Industry. Research Gate.
- Canal de Panamá (2006) Plan Maestro del Canal de Panamá. 7 de junio de 2006. Apéndice A-pp. 1-54.
- Canal de Panamá (2019) Informe Anual 2019. <https://micanaldepanama.com/wp-content/uploads/2020/01/InformeAnual-2019-2.pdf>. Accedido 1 de octubre de 2020.
- Carral L, Tarrío-Saavedra J, Vega-Sáez A, Bogle J, Alemán G, Naya S (2021) Modelling operative and routine learning curves in manoeuvres in locks and in transit in the expanded Panama Canal. Journal of Navigation, 1-23.
- Carral L, Tarrío-Saavedra J, Rodríguez-Díaz A, Bogle J, Naya S (2020) Modeling of the learning curve corresponding to the transit through the new expanded Panama Canal. En: Proceeding of the VI International Ship Design y Naval Engineering Congress (CIDIN) and XXVI Pan-American Congress of Naval Engineering, Maritime Transportation and Port Engineering (COPINAVAL) pp.124-133.
- Carral L, Tarrío-Saavedra J, Alvarez-Feal JC, Naya S, Sabonge R (2019) Modelling and forecasting of Neopanamax vessel transit time for traffic management in the Panama Canal. Journal of Marine Science and Technology, 25:379–396.
- Carral L, Tarrío-Saavedra J, Castro-Santos L, Lamas-Galdo I, Sabonge R (2018) Effect of Inaugurating the Third Set of Locks in the Panama Canal on Vessel Size, Manoeuvring and Lockage Time. Promet – Traffic and Transportation, 30(2): 241-251.
- Carral L, Tarrío-Saavedra J, Naya S, Bogle J, Sabonge R (2017) Effect of Inaugurating the Third Set of Locks in the Panama Canal on Vessel Size, Manoeuvring and Lockage Time . Journal of Navigation, 70(6): 1205-1223.
- Flores M (2020) qcr: Quality Control Review. R package version 1.2. <https://CRAN.R-project.org/package=qcr>
- Gryna FM, Chua Rch, Defeo JA (2007) Método Juran. Análisis y planeación de la calidad. The McGraw-Hill Companies, México.

- Hamed M, Mansour M, Abd-Elrazik E (2016) Mcusum control chart procedure: Monitoring the process mean with application. *Journal of Statistics: Advances in Theory and Applications*, 16(1): 105-132.
- Haskin FJ (1913) *The Panama Canal*. Doubleday, Page and Company, Garden City, New York.
- Hubele NF, Vännman k (2004). The Effect of Pooled and Un-pooled Variance Estimators on Cpm When Using Subsamples. *Journal Quality Technology*, 36, 207-222.
- Huebner GG (1915) Economic aspects of the Panama Canal. *The American Economic Review* 5(4):816–829.
- Kane VE (1986) Process capability indices, *Journal Of Quality Technology*, 18 (1), pp.41-52.
- Korkmaz S, Goksuluk D, Zararsiz G (2019) MVN: Multivariate Normality Tests. R package version 3.3.3. <https://CRAN.R-project.org/package=MVN>. Accedido 1 de octubre de 2020.
- Lefcovich M (2003) Kaizen–La mejora continua y el cuadro de mando integral. Disponible en <http://studylib.es/doc/493795/t>
- Lowry CA, Woodall WH, Champ CW, Rigdon SE (1992) A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34:46-53.
- Martin J, Martin S, Pettit SJ (2015) Container ship size and the implications on port call Workload. *International Journal of Shipping and Transport Logistics*, 7(5), 553.
- Mitra A (2016) *Fundamentals of Quality Control and Improvement*. John Wiley y Sons, Inc., Hoboken, New Jersey.
- Meyer D, Zeileis A, Hornik K, Gerber F, Friendly M (2015) vcd: Visualizing Categorical Data. R package version 1.4-8. <https://CRAN.R-project.org/package=vcd> .Accedido 1 de abril de 2020.
- Montgomery DG (2009) *Introduction to Statistical Quality Control*, 6ed., Wiley, New York.
- Naya S, Devia A, Tarrío-Saavedra J, Flores M (2016) New robust capability ratios approaches for quality control. *DYNA* 83(198):94-101.
- Page ES (1954). Continuous inspection schemes. *Biometrika*, 41: 100-15.
- Pearn W, Kotz S, Johnson NL (1992) Distributional and inferential properties of process capability indices. *Journal of Quality Technology* 24:216-231.
- Peña D (2002) *Análisis de datos multivariantes*, McGraw-Hill, Madrid.
- Prat-Bartés A, Tort-Martorell X, Grima-Cintas P, Pozueta-Fernández L (1998) *Control y Mejora de la Calidad*. ed. UPC.
- Roberts S (1959) Control Chart Tests Based on Geometric Moving Averages. *Technometrics* 1:239-250.
- Sabonge R (2014) *The Panama Canal expansion: A driver of change for global trade flows*. United Nations, Economic Commission for Latin America and the Caribbean (CEPAL), Santiago de Chile.
- Saboange R, Sánchez RJ (2009) *El Canal de Panamá en la economía de América Latina y el Caribe*. Naciones Unidas, Comisión Económica para América Latina y el Caribe (CEPAL), Santiago de Chile.
- Santos-Fernández E, Scagliarini M (2015) MPCl: Multivariate Process Capability Indices (MPCl). R package version 1.0.7. <https://CRAN.R-project.org/package=MPCl>. Accedido 1 de octubre de 2020.
- Santos-Fernández E, Scagliarini M (2012) MPCl: An R Package for Computing Multivariate Process Capability Indices. *Journal of Statistical Software published by the American Statistical Association* 47(7).
- Schloerke B, Cook Di, Larmarange J, Briatte F, Marbach M, Thoen E, Elberg A, Toomet O, Crowley J, Hofmann H, Wickham H (2021) GGally: Extension to 'ggplot2'. R package version 2.1.1. <https://CRAN.R-project.org/package=GGally>. Accedido 1 de mayo de 2021.

- Scrucca L, Snow G, Bloomfield P (2017) qcc: Quality Control Charts. R package version 2.7. <https://CRAN.R-project.org/package=qcc>. Accedido 1 de octubre de 2020.
- Vännman K (1995) A Unified Approach to Capability Indices. *Statistica Sinica* 5:805-820.
- Wickham H, Chang W, Henry L, Lin-Pedersen T, Takahashi K, Wilke C, Woo K, Yutani H, Dunnington D (2016) ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics. R package version 3.3.3. <https://CRAN.R-project.org/package=ggplot2>. Accedido 1 de octubre de 2020.
- Wood S (2021). mgcv: Mixed GAM Computation Vehicle with Automatic Smoothness Estimation. R package version 1.8-35. <https://CRAN.R-project.org/package=mgcv>. Accedido 2 de Mayo de 2021.
- Wood SN (2017). *Generalized Additive Models: An Introduction With R*. CRC Press.
- Wu CW, Pearn WL, Kotz S (2009) An overview of theory and practice on process capability indices for quality assurance. *International Journal of Production Economics* 117:338-359.

