



Trabajo Fin de Máster

Uso de técnicas SPC para el estudio y control de los indicadores del COVID-19

Claudia Torviso Rodríguez

Máster en Técnicas Estadísticas

Curso 2021 – 2022

Propuesta de Trabajo Fin de Máster

Título en galego: Uso de técnicas SPC para o estudo e control dos indicadores da COVID-19

Título en español: Uso de técnicas SPC para el estudio y control de los indicadores del COVID-19

English title: Use of SPC techniques for the study and monitoring of COVID-19 indicators

Modalidad: Modalidad A

Autor/a: Claudia Torviso Rodríguez, Universidad de A Coruña

Director/a: Javier Tarrío Saavedra, Universidad de A Coruña; Salvador Naya Fernández, Universidad de A Coruña

Breve resumen del trabajo:

En este trabajo se ha realizado un análisis del uso de herramientas del Control Estadístico de Procesos para el estudio de datos epidemiológicos, como es el caso de la COVID-19, a través de una revisión de la literatura. Además, se han aplicado dichas técnicas a un caso práctico que ha sido el estudio de la COVID-19 en el área metropolitana de A Coruña.

Recomendaciones:

Otras observaciones:

Don Salvador Naya Fernández, catedrático de universidad de la Universidad de A Coruña, don Javier Tarrío Saavedra, profesor contratado doctor de la Universidad de A Coruña, informan que el Trabajo Fin de Máster titulado

Uso de técnicas SPC para el estudio y control de los indicadores del COVID-19

fue realizado bajo su dirección por doña Claudia Torviso Rodríguez 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 31 de agosto de 2022.

El director:

Don Salvador Naya Fernández

El director:

Javier Tarrío Saavedra

La autora:

Claudia Torviso Rodríguez

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, [Disposición 2978 del BOE núm. 48 de 2022](#)), la autora declara que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas, ...)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración, sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Agradecimientos

En primer lugar, me gustaría agradecer a mis tutores, Javier y Salvador, por haberme guiado y ayudado tanto durante la elaboración de este trabajo de fin de máster, sobre todo hacia el final, que parecía que nunca iba a llegar. También a los responsables del proyecto COVIDBENS por proporcionarnos los datos reales de la carga vírica que se han usado a lo largo del presente trabajo, en especial a los investigadores Margarita Poza, Ricardo Cao y Susana Ladra, y a Carlos Lamora, de la empresa pública Edar Bens.

También me gustaría darles las gracias a mis padres, Geni y JM, y a mi hermana Sara, por enseñarme a ser la persona que soy ahora, responsable y muy trabajadora, como todos ellos, y por darme la oportunidad tan grande que ha sido formarme en la universidad. Estaré eternamente agradecida.

Me gustaría darles las gracias a todas mis amigas, aquellas que siguen ahí a pesar de estos últimos años que han sido tan difíciles para todas. Gracias por todos esos jueves que, aunque han sido menos de los que me hubiese gustado y sin estar todas por culpa de la distancia, me han servido para desconectar, como siempre, casi llorando de la risa. Estoy muy orgullosa de todas vosotras y de cada uno de los pasos que dais y que daréis.

En especial, me gustaría darle las gracias a Marta, mi hermana postiza, por apoyarme tanto y siempre, desde hace tantos años que ya no sé ni contarlos. Ojalá que me sigas acompañando y poder acompañarte yo a ti en todos tus logros, que sé que no serán pocos.

También a Tania, mi compañera de máster. Gracias por hacer que todos esos kilómetros no parecieran nada con unas cuantas videollamadas y tantas y tantas notas de voz de *Whatsapp*. No podría haber elegido una mejor compi para todos esos trabajos en grupo.

Por último, pero no menos importante, a Pablo, por creer tanto en mí, aun cuando yo no lo hago, por animarme a dar lo mejor de mí en cada cosa que hago y por darme siempre el impulso necesario para seguir.

Índice general

Resumen	XVII
Introducción	1
1.1. Breve reseña histórica acerca del Coronavirus	1
1.2. Problemática y soluciones.....	2
Control Estadístico de Procesos y Gráficos de Control	6
2.1. Control Estadístico de Procesos	6
2.2. Gráficos de Control.....	8
2.2.1. ¿Qué son los gráficos de control y cómo se construyen?	8
2.2.2. ¿Cómo se interpreta un gráfico de control?	11
2.2.3. Tipos de gráficos de control	14
2.3. Gráficos de control en el ámbito de la salud pública.....	30
2.3.1. Diferencias entre el SPC en la industria y en la medicina	30
2.3.2. Gráficos y métodos más utilizados en medicina	32
2.3.3. Nuevas formas y métodos surgidos con el COVID-19.....	37
Caso de estudio de la zona metropolitana de A Coruña	45
3.1. Proyecto COVIDBENS	45
3.2. Descripción de los datos	49
3.3. Gráficos de control para el estudio del COVID-19 con datos de COVIDBENS	50
3.4. Aplicación de gráficos de control específicos en epidemiología.....	78
3.4.1. Vigilancia de una serie temporal de datos de recuento mediante el método EARS C1, C2 o C3	78
Conclusiones.....	91
Código de R utilizado	94
Bibliografía.....	107

Índice de figuras

Figura 1: Casos diarios confirmados de COVID-19 en España a 25/04/2021	3
Figura 2: Construcción de la máscara V para un gráfico CUSUM.....	26
Figura 3: Gráfico de embudo para las muertes después de un tratamiento para la fractura de cadera.....	37
Figura 4: Fases de la evolución de la pandemia de COVID-19 en Galicia entre marzo y junio de 2020.....	38
Figura 5: Número diario de muertes reportadas por COVID-19 en el mundo entre enero de 2020 y febrero de 2021	42
Figura 6: Número de contagios diarios por COVID-19 en Los Ángeles, EEUU, entre marzo de 2020 y noviembre de 2020	44
Figura 7: Mapa del área metropolitana de A Coruña	46
Figura 8: Porcentaje de las variantes encontradas en las muestras de aguas residuales	47
Figura 9: Estimación del número real de casos activos de COVID-19 mediante dos modelos de regresión lineal.	49
Figura 10: Gráfico de medidas individuales para la variable “Número de casos activos”	51
Figura 11: Gráfico EWMA para la variable “Número de casos activos”	52
Figura 12: Gráfico CUSUM para la variable “Número de casos activos”	53
Figura 13: Gráfico de medidas individuales para los residuos de la variable “Número de casos”	55
Figura 14: Gráfico EWMA para los residuos de la variable “Número de casos”	56
Figura 15: Gráfico CUSUM de los residuos del “Número de casos activos”	57
Figura 16: Gráfico de dispersión de la variable “Número de casos”	58
Figura 17: Gráfico de medidas individuales con muestra de calibrado	59
Figura 18: Gráfico EWMA para los residuos de la variable “Número de casos” con muestra de calibrado	60
Figura 19: Gráfico CUSUM para los residuos del “Número de casos” con muestra de calibrado	61
Figura 20: Gráfico de dispersión del número de casos activos	62
Figura 21: Gráfico de medias para medidas individuales para la variable “Carga vírica medida”	63
Figura 22: Gráfico EWMA para la variable “Carga vírica medida”	64
Figura 23: Gráfico CUSUM para la variable “Carga vírica medida”	66
Figura 24: Gráfico de medias para medidas individuales para los residuos de la carga vírica ..	68

Figura 25: Gráfico EWMA para los residuos del modelo ajustado a la carga vírica.....	69
Figura 26: Gráfico CUSUM de los residuos del modelo ajustado a la carga vírica	70
Figura 27: Gráfico de dispersión de la carga vírica medida en las aguas residuales.....	71
Figura 28: Gráfico de medidas individuales de los residuos de la carga viral medida con una muestra inicial de calibrado.....	73
Figura 29: Gráfico EWMA de los residuos de la carga vírica con una muestra de calibrado.....	74
Figura 30: Gráfico CUSUM de los residuos de la carga vírica con una muestra de calibrado...	75
Figura 31: Gráfico de dispersión de la carga vírica medida en las aguas residuales del área de A Coruña.....	76
Figura 32: Gráfico de detección temprana de anomalías (EARS) utilizando el método C1 con la variable “Número de casos activos”	81
Figura 33: Gráfico de detección temprana de anomalías (EARS) utilizando el método C1 con la variable “Carga vírica”	83
Figura 34: Gráfico de detección temprana de anomalías (EARS) utilizando el método C2 con la variable “Número de casos activos”	85
Figura 35: Gráfico de detección temprana de anomalías (EARS) utilizando el método C2 con la variable “Carga vírica”	86
Figura 36: Gráfico de detección temprana de anomalías (EARS) utilizando el método C3 con la variable “Número de casos activos”	88
Figura 37: Gráfico de detección temprana de anomalías (EARS) utilizando el método C3 con la variable “Carga vírica”	89

Índice de tablas

Tabla 1: Casos de COVID-19 que han precisado hospitalización, ingreso en UCI y fallecidos por CCAA en España a 27/04/2020	4
---	---

Resumen

Resumen en español

En este trabajo se hace un análisis del uso de herramientas propias del Control Estadístico de Procesos para el estudio de la evolución de enfermedades, como es el caso de la COVID-19.

Se lleva a cabo una revisión de la literatura sobre el SPC y de los Gráficos de Control más utilizados. Se analizan las diferencias entre el SPC en la industria y en el ámbito de la salud pública, y se estudian los gráficos y métodos específicos más utilizados en este último. Además, se han analizado los nuevos métodos de estudio surgidos a raíz de la irrupción del SARS-CoV-2 en el mundo.

Como caso de estudio, se han aplicado algunas de las técnicas descritas a lo largo del trabajo a un caso concreto que es el estudio del COVID-19 en las aguas residuales del área metropolitana de A Coruña. Estos datos han sido cedidos por el proyecto de COVIDBENS, del cual se hace una breve introducción al trabajo que han llevado a cabo. Por último, se aplican gráficos de control específicos en epidemiología a estos datos.

Resumo en galego

Neste traballo faise unha análise do uso de ferramentas propias do Control Estatístico de Procesos para o estudo da evolución de enfermidades, como é o caso da COVID-19.

Lévase a cabo unha revisión da literatura sobre o SPC e dos Gráficos de Control máis utilizados. Analízanse as diferenzas entre o SPC na industria e no ámbito da saúde pública, e estúdanse os gráficos e métodos específicos máis utilizados neste último. Ademais, analizáronse os novos métodos de estudo surxidos a raíz da irrupción do SARS-CoV-2 no mundo.

Como caso de estudo, aplicáronse algunhas das técnicas descritas ao longo do traballo a un caso concreto que é o estudo da COVID-19 nas augas residuais na área metropolitana de A Coruña. Estes datos foron cedidos polo proxecto de COVIDBENS,

do cal faise unha breve introdución ao traballo levado a cabo. Por último, aplícanse os gráficos de control específicos en epidemioloxía a estes datos.

English abstract

This paper analyses the use of Statistical Process Control tools for the study of the evolution of diseases, such as COVID-19.

A review of the literature on SPC and the most commonly used Control Charts is carried out. The differences between SPC in industry and in public health are analysed, and the specific charts and methods most commonly used in the latter are studied. In addition, the new study methods that have arisen as a result of the outbreak of SARS-CoV-2 in the world have been analysed.

As a case study, some of the techniques described throughout the work have been applied to a specific case which is the study of COVID-19 in the wastewater of the metropolitan area of A Coruña. These data have been provided by the COVIDBENS project, of which a brief introduction to the work carried out is given. Finally, specific control charts in epidemiology are applied to these data.

Capítulo 1

Introducción

1.1. Breve reseña histórica acerca del Coronavirus

En diciembre de 2019 surge un brote de neumonía de origen desconocido en la provincia china de Hubei, más concretamente en la capital, Wuhan. Después de numerosas investigaciones se descubre la secuencia del patógeno de un nuevo coronavirus (Ministerio de Sanidad, 2020a), llamado SARS-CoV-2 y posteriormente conocido como COVID-19, diferente al resto de coronavirus ya conocidos, como el SARS-CoV-1 y el MERS-CoV (Vargas et al., 2020).

En enero de 2020, la OMS declara un estado de emergencia para abordar el brote de coronavirus que se extiende por el país y que comienza a extenderse por el extranjero (Güell, 2020). Es en este mes cuando se producen las primeras defunciones producidas por COVID-19 y cuando se declara la primera cuarentena en el mundo debido a este virus en la provincia de Hubei. También en enero se detecta el primer caso en España (Ministerio de Sanidad, 2020b).

En febrero de 2020 se declara la primera muerte fuera de China y comienzan a disminuir los casos activos tras la cuarentena domiciliaria en la provincia de Hubei, demostrando, así, la eficacia de la cuarentena para combatir el virus y contener los contagios que se extienden por todo el mundo (Vidal Liy, 2020).

En marzo de 2020, el virus ya está muy extendido por todo el mundo. La OMS considera que el brote de coronavirus es ya una pandemia (Organización Mundial de la Salud, 2020). En España comienzan a cancelar congresos, ferias y cursos, poniendo hincapié en aquellos que impliquen la participación de personal sanitario, el grupo más afectado y más necesario al comienzo de la pandemia (Useros, 2020). El 13 de marzo, el Gobierno anuncia un confinamiento en el municipio de Igualada (Cataluña) y otros tres municipios más (Larraz, 2020), además de que muchas comunidades autónomas

anuncian el cierre de todos los centros educativos (guarderías, colegios, institutos y universidades) después de la recomendación del Gobierno durante, en principio, 15 días (García, 2020). Pocos días después, el 14 de marzo, se declara el estado de alarma por el cual se cancelará toda aquella actividad no esencial y se confinará a la población en sus domicilios. Solo se podrá dejar el domicilio para la compra de alimentos, suministros médicos, realizar aquellos trabajos esenciales o para emergencias (Real Decreto 463/2020, 2020).

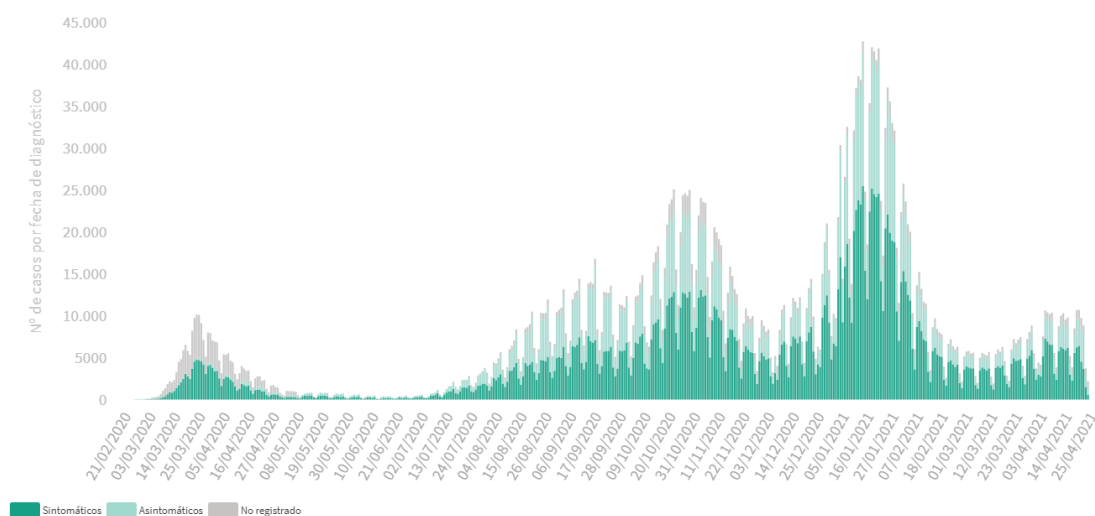
A finales de abril y durante el mes de mayo se comienza a hablar en España de la “desescalada”, que fueron una serie de medidas tomadas por el Gobierno Central junto con las comunidades autónomas durante varias semanas para dejar la cuarentena domiciliaria y volver, poco a poco, a la normalidad y siempre teniendo en cuenta la evolución de la enfermedad en cada territorio (La Moncloa, 2020a). Comenzaron a dejar salir a la calle a los niños, dejar pasear a la población durante unas horas al día, abrir comercios no esenciales, practicar deporte al aire libre, etc. Otras medidas que se tomaron más adelante también fueron el uso obligatorio de la mascarilla en lugares cerrados y abiertos si no se mantenía una distancia interpersonal de, al menos, 1.5m o los cierres perimetrales, es decir, la prohibición de moverse libremente entre comunidades y, en algunos casos, entre provincias. La “desescalada” terminó en el mes de junio de 2020, aunque hubo nuevos confinamientos y nuevas restricciones durante el resto de 2020 y 2021, pero esta vez en solo algunos territorios en los que la pandemia volvía a tener una incidencia muy alta (La Moncloa, 2020b). Todas estas medidas se basaron en datos sobre la incidencia del coronavirus en nuestro país, así como otros datos también importantes acerca de la ocupación hospitalaria, ocupación de camas UCI o muertes por coronavirus. En el resto del mundo se comienza a actuar de forma similar a España, realizando una desescalada gradual y siempre teniendo en cuenta la incidencia del virus en cada momento para tomar las decisiones oportunas (Aldama et al., 2020).

1.2. Problemática y soluciones

Durante estos meses del comienzo de la pandemia y, sobre todo, durante el confinamiento domiciliario y la posterior “desescalada”, una gran parte de la población ha estado pendiente de las novedades acerca del virus. Todos los días, las autoridades

sanitarias nacionales e internacionales mostraban datos de nuevos contagios, hospitalizaciones, ingresos en UCI, muertes, etc. Esta información, a pesar de ser muy necesaria para entender la evolución de la pandemia, muchas veces se difundía de forma errónea. Se publicaban gráficos en los que solamente podíamos ver si los contagios o las muertes disminuían o aumentaban, un ejemplo lo tenemos en los gráficos que publicaba el Ministerio de Sanidad del Gobierno de España en sus informes de situación diarios (Figura 1).

Figura 1: Casos diarios confirmados de COVID-19 en España a 25/04/2021



Fuente: Centro de Coordinación de Alertas y Emergencias Sanitarias. Ministerio de Sanidad. Gobierno de España.

También se llegaron a presentar tablas que simplemente nos informaban del número de nuevos contagios, hospitalizaciones o muertes por región, como la tabla siguiente en la que se muestran los casos de COVID-19 que han precisado hospitalización, ingreso en UCI y fallecidos (**Tabla 1**):

Tabla 1: Casos de COVID-19 que han precisado hospitalización, ingreso en UCI y fallecidos por CCAA en España a 27/04/2020

CCAA	Casos que han precisado hospitalización		Casos que han ingresado en UCI		Fallecidos		Curados	
	Total	Nuevos	Total	Nuevos	Total	Nuevos	Total	Nuevos
Andalucía	5.768	20	720	4	1.157	12	4.890	149
Aragón	2.403	8	257	0	726	14	2.010	50
Asturias	1.814	24	137	3	253	4	759	10
Baleares	1.062	0	166	0	179	4	1.136	13
Canarias	881	1	171	0	131	0	1.047	
Cantabria	995	4	78	0	186	3	1.241	114
Castilla La Mancha	8.444	27	579	14	2.365	35	5.282	86
Castilla y León	7.703	50	518	3	1.690	24	6.272	64
Cataluña	25.799	134	2.724	141	4.699	133	17.897	891
Ceuta	10	0	4	0	4	0	107	3
C. Valenciana	5.032	19	656	2	1.189	3	6.317	74
Extremadura	1.490	10	108	0	428	6	1.652	70
Galicia	2.746	11	83		405	11	1.816	13
Madrid	39.472		3.309		7.986	64	35.565	198
Melilla	44	0	3	0	2	0	87	6
Murcia	627	0	106	0	128	1	990	70
Navarra	1.942	0	130	1	432	1	1.918	83
País Vasco	6.457	31	538	5	1.241	11	9.840	238
La Rioja	1.392	12	85	1	320	5	2.049	12
ESPAÑA	114.081		10.289		23.521	331	100.875	2.144

Fuente: Centro de Coordinación de Alertas y Emergencias Sanitarias. Ministerio de Sanidad. Gobierno de España.

El exceso de información al que se vio sometida la población durante la cuarentena domiciliar, sumado a la falta de rigor por parte de los medios de comunicación a la hora de notificar los datos sobre la pandemia mundial ha supuesto un aumento de la preocupación que ya había por la desconocida enfermedad que se expandía por el mundo. Algunos estudios llegaron a la conclusión de que esta excesiva información implicó niveles altos de ansiedad, más preocupación por el coronavirus y problemas para conciliar el sueño y/o mantenerlo (Roy et al., 2020).

Si no existe un método que permita identificar señales de alarma ante un gran aumento de los contagios o de las muertes de forma rápida, todos estos gráficos y datos no aportan nada, ni a la población en general ni a las autoridades responsables de la toma de decisiones. La información transmitida debe ser clara, fácilmente interpretable por la población y desglosada para que las autoridades políticas puedan considerarla y puedan utilizarla para llevar a cabo la toma de decisiones acerca del control de la pandemia (Inkelas et al., 2021).

El método que proponemos en este trabajo es el uso de Gráficos de Control, en un principio industriales, y otras técnicas del Control Estadístico de la Calidad, para estudiar las variaciones en la incidencia del coronavirus, con el objetivo de detectar señales de alarma que signifiquen un cambio significativo en la pandemia.

Para la parte práctica de este trabajo en la que se estudiarán estos métodos derivados del Control Estadístico de Procesos, utilizaremos datos acerca de los contagios y de la carga vírica del SARS-CoV-2 en el área de A Coruña. Estos datos fueron cedidos por el proyecto COVIDBENS que estuvo activo desde abril de 2020 estudiando la incidencia de COVID-19 en las aguas residuales de todo el área de A Coruña (COVIDBENS, 2022a). Proyectos similares de estudio de la incidencia de la enfermedad en las aguas residuales se comenzaron a llevar a cabo a partir de los meses de marzo y abril de 2020 en otros lugares de España como la Región de Murcia o el área metropolitana de Valencia (Santos, 2020). El estudio de la incidencia de la enfermedad se llevó a cabo en otros países como Estados Unidos, con el proyecto de Biobot Analytics, una startup desarrollada por estudiantes de doctorado del Instituto de Tecnología de Massachusetts (*Biobot Analytics · About*, s. f.), que estudió muestras de plantas de tratamiento de aguas residuales en todo el país (Snyder & Cullinane, 2020). También se llevaron a cabo estudios en Países Bajos al detectarse la presencia del virus en febrero de 2020, antes de comenzar a confirmarse los contagios por coronavirus en ese país (Medema et al., 2020).

Estos proyectos ayudaron a entender mejor la evolución de la pandemia en las regiones estudiadas, algo que resultaba difícil dado que los únicos datos de los que se disponían, como ya hemos dicho, eran de contagios confirmados a través de las pruebas PCR, antígenos u otras. Los estudios en las aguas residuales consiguieron detectar material genético del virus días antes de que comiencen a aparecer los nuevos contagios al detectarse estos mismos antes de que comiencen a presentar síntomas e, incluso, detectan contagios de personas asintomáticas. Además, el estudio de las aguas residuales permite conocer el origen del brote y permite detectar nuevas variantes y su incidencia en las localidades. Estos datos, por lo tanto, sirvieron a las autoridades sanitarias para conocer la tendencia (creciente o decreciente) del coronavirus y en el seguimiento de la pandemia en las localidades estudiadas (Ministerio de Sanidad, 2021).

Capítulo 2

Control Estadístico de Procesos y Gráficos de Control

2.1. Control Estadístico de Procesos

El Control Estadístico de Procesos (CEP o SPC, del inglés, *Statistical Process Control*) como tal, nació en la década de los años 20, aunque las primeras civilizaciones ya sabían la importancia de controlar la calidad de sus elaboraciones y ya utilizaban técnicas, aunque muy rudimentarias en este momento, para que todo aquello que fabricaban tuviera la calidad deseada. En el año 1924, W. A. Shewhart comienza a utilizar los conocidos como gráficos de control en los laboratorios de la *Bell Telephone* y es en el año 1931 cuando publica su libro "*Economic Control of Quality of Manufactured Products*" en donde marca las pautas del Control Estadístico de Procesos y los Gráficos de Control, que posteriormente serán conocidos como Gráficos de Control de Shewhart. Con la Segunda Guerra Mundial y la implantación de departamentos de calidad en muchas industrias armamentísticas, comienzan a expandirse estos métodos por diferentes industrias y comienzan a crearse institutos sobre calidad. En el año 1946 se crea el instituto americano para la calidad, el *American Society for Quality Control*. No solo se expande por EEUU, comienza a expandirse por Japón y se crea la Unión de Científicos e Ingenieros Japoneses (JUSE). A partir de los años 50, comienzan a desarrollarse e introducirse nuevos gráficos de control y surgen varios textos sobre el Control de Calidad que se siguen utilizando y teniendo en cuenta hoy en día. En la década de los 90, comienzan a publicarse estándares y normas sobre el Control de Calidad. Desde los 2000 a la actualidad, el Control Estadístico de Procesos y los Gráficos de Control ha ido evolucionando y se ha empezado a utilizar en otros ámbitos diferentes al industrial, como veremos más adelante (Montgomery, 2009).

Definimos el Control Estadístico de Procesos como un conjunto de herramientas estadísticas que son útiles para conseguir que un proceso se mantenga estable y mejorar su capacidad reduciendo la variabilidad (Montgomery, 2009). Los principales

objetivos del SPC son minimizar los defectos en la producción o el proceso que se esté analizando, conseguir una mejora continua del proceso en cuestión y comparar la producción respecto a los requisitos, ya sea del cliente o de la propia empresa para poder tomar medidas en caso de necesitar alguna corrección en el proceso. Todos estos objetivos están siempre ligados a la estabilidad del proceso, es decir, a reducir la variabilidad del mismo. Podemos poner un ejemplo para entender mejor el concepto de variabilidad de un proceso. Supongamos el caso de una empresa que fabrica bolígrafos. Nunca habrá dos bolígrafos exactamente iguales ya que todos están sometidos a muchas fuentes distintas de variación. A pesar de que está programado que cada bolígrafo tenga una capacidad de escritura de 2000 metros, puede existir variabilidad en el proceso de fabricación de los mismos que provoque que algunos bolígrafos tengan una capacidad de 1998 metros y que otros bolígrafos de otro lote, tengan una capacidad de escritura de 2003 metros. También puede ocurrir que la largura de los bolígrafos varíe en algunos milímetros entre un lote u otro de bolígrafos o que el color de los mismos sea un poco más oscuro en unos que en otros, a pesar de que se venden como si fueran del mismo color.

Para controlar el proceso o alguna de sus características, se utilizan una serie de herramientas estadísticas. Las principales herramientas que se utilizan en el control estadístico de procesos son las siguientes:

- Histograma o, de forma similar, un gráfico de tallos y hojas
- Hoja de control
- Gráfico de Pareto
- Diagrama de causa y efecto
- Diagrama de concentración de defectos
- Diagrama de dispersión
- Gráfico de control

En este trabajo trataremos, sobre todo, de los gráficos de control.

2.2. Gráficos de Control

Para decidir si la variabilidad observada en un proceso es inherente a él o si, en cambio, es una variación especial y anormal que puede afectar en gran medida a la calidad del proceso, mediremos la característica de calidad del proceso tomada de la muestra y lo representaremos en un diagrama o gráfica ordenada por una variable temporal. Con este diagrama resultante, podremos elaborar lo que conocemos como gráfico de control gracias a Shewhart en 1924. Los primeros gráficos de control utilizados por Shewhart fueron para monitorizar el promedio ($\bar{x}$) y la variación (R o S), aplicados para variables cuantitativas. Posteriormente comenzaron a utilizarse otros gráficos para variables cualitativas como el gráfico c , para monitorizar el número total de defectos por lotes y el gráfico p , para monitorizar la proporción de defectos por lotes. Estos últimos no consideran la dependencia en el tiempo de las variables cualitativas.

Otros gráficos propuestos posteriormente que sí consideran la dependencia en el tiempo son el gráfico de sumas acumuladas (*CUSUM*) propuesto por E. S. Page (1954) y el gráfico de medias móviles exponencialmente ponderadas (*EWMA*) propuesto por S. W. Roberts (1959). Estos gráficos funcionan mejor que los gráficos de control de Shewhart cuando los cambios en el proceso son pequeños ya que consideran información anterior del proceso.

Actualmente comienzan a considerar que en la calidad final de un producto influyen una o más características del proceso. Para poder monitorizar varias características al mismo tiempo han surgido técnicas multivariantes de control como el gráfico de control T^2 de Hotelling y los gráficos multivariantes *MCUSUM* y *MEWMA* que son extensiones al caso multivariante de los gráficos *CUSUM* y *EWMA* univariantes.

2.2.1. ¿Qué son los gráficos de control y cómo se construyen?

Un Gráfico de Control es un gráfico que representa cómo se comporta un proceso o alguna variable del mismo registrando los datos ordenados en el tiempo. El objetivo principal de un gráfico de control será, entonces, detectar lo más rápido posible cualquier cambio o tendencia en un proceso que pueda afectar a la calidad del mismo. Eso se

consigue minimizando el tiempo que pasa entre que se produce ese cambio, por muy pequeño que sea, hasta que lo detectamos (Prat Bartés et al., 2015).

Los gráficos de control se construyen en dos fases:

- En la **fase I**, se recoge un conjunto de datos del proceso, al que llamaremos muestra preliminar, y se analizan estimando unos límites de control de prueba para determinar si el proceso ha estado bajo control durante el tiempo en que se recogieron los datos. Se estudia si se pueden establecer unos límites de control fiables, que nos sirvan para supervisar los procesos y la producción futura.
- En la **fase II**, una vez tengamos un conjunto de datos limpio, estable y representativo del proceso en estado de control, se extraen nuevos datos del proceso y se representan cada una de esas muestras posteriores en un gráfico con los límites de control que se han calculado previamente.

En el gráfico se representarán los valores muestrales de forma ordenada en el tiempo con el objetivo de monitorizar el proceso. Normalmente se unen esos valores mediante líneas para facilitar la visualización de los datos. En el eje de abscisas se representa el tiempo y en el eje de ordenadas la característica de calidad que se va a estudiar. Los gráficos de control estándar cuentan con una línea central o límite de control (del inglés, CL) que coincidirá con el valor medio de la característica a estudiar y dos líneas que representan el límite superior de control (UCL, del inglés, *Upper Control Limit*) y el inferior (LCL, del inglés, *Lower Control Limit*) que, normalmente, se sitúan a ± 3 desviaciones típicas de la línea central de control, cubriendo el 99,7% de la variación. Esto es así porque estamos teniendo en cuenta que las variaciones naturales o habituales del proceso; es decir, aquellas que son inherentes a él, siguen una distribución normal. En ocasiones, es frecuente utilizar otros límites de control que se sitúan a ± 2 desviaciones típicas, cubriendo el 95% de la variación. Las observaciones de la característica a analizar que se sitúen fuera de los límites de control pueden ser observaciones atípicas o un indicador de que el proceso está fuera de control. Una vez identificadas las observaciones fuera de los límites, debemos identificar las causas de esa variabilidad.

Podemos ver un gráfico de control como si fuese un contraste de hipótesis en donde la hipótesis nula planteada, H_0 , sería que el proceso está bajo control en todas las

muestras seleccionadas. De no aceptarse la hipótesis nula, estaríamos asumiendo que la hipótesis alternativa, H_1 , de que el proceso está fuera de control es cierta. Para verificar que el proceso está bajo control, el contraste de hipótesis anterior se verificará para cada submuestra.

Si rechazamos la hipótesis nula cuando es cierta, es decir, suponemos que el proceso está fuera de control cuando realmente está bajo control, estaríamos cometiendo un error de tipo I, también denominado en este contexto de control de calidad como riesgo del vendedor o probabilidad de falsa alarma. Si, por el contrario, no rechazamos la hipótesis nula cuando ésta es falsa, estaríamos cometiendo un error de tipo II, también conocido en el contexto del control de calidad como riesgo del comprador. La probabilidad de cometer los errores de tipo I y II se denominan riesgo α y riesgo β , respectivamente.

Cuando estamos monitorizando un proceso pueden surgir puntos fuera de los límites de control aun cuando el proceso está bajo control, lo que daría lugar a falsas alarmas (o errores de tipo I). Para asegurarnos de que el gráfico de control está funcionando como debería, debemos garantizar un bajo porcentaje de falsas alarmas y, al mismo tiempo, la rapidez suficiente para detectar cualquier tipo de cambio en el proceso. Para ello, introducimos el concepto de longitud media de racha (ARL, del inglés, *Average Run Length*). La longitud media de racha es el promedio de puntos que necesitamos representar en el gráfico hasta que el primer punto muestre una señal de estar fuera de control o que caiga fuera de los límites de control. O lo que es lo mismo, el número de muestras necesarias para detectar la primera fuera de control. Este valor se calcula como:

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

Sabiendo esto, si queremos tener un bajo porcentaje de falsas alarmas el ARL bajo control deberá ser alto y si queremos obtener una rapidez suficiente para la detección de cambios, el ARL fuera de control tendrá que ser bajo.

También se puede utilizar el tiempo promedio hasta la señal (ATS, del inglés, *Average Time to Signal*) que es el tiempo medio que tardará el gráfico hasta que indique una

señal de falta de control. Se calcula de manera sencilla utilizando el ya calculado ARL, como $ATS = ARL \times h$, siendo h el tiempo entre muestras.

Para medir la potencia de este contraste de hipótesis podemos utilizar la denominada curva Característica de Operación o curva OC (del inglés, *Operating Characteristics*) en la que se representa el error de tipo II en función de distintos valores para la hipótesis alternativa.

En el diseño de gráficos de control es necesario especificar el tamaño muestral y la frecuencia del muestreo. Para especificar el tamaño muestral, se suele hacer uso de las curvas OC. Las muestras grandes facilitan la detección de cambios pequeños, esto lo podemos ver si comparamos curvas OC con diferentes tamaños muestrales. Sin embargo, se suelen utilizar muestras pequeñas de tamaño 4 o 5 en intervalos cortos de tiempo. La frecuencia del muestreo depende de muchas cuestiones, generalmente económicas. Lo ideal sería tomar muestras grandes con mucha frecuencia, pero esto no es viable económicamente. En general, se tiende a tomar muestras pequeñas en intervalos cortos de tiempo o muestras más grandes en intervalos más largos. Debemos intentar encontrar un equilibrio entre el tamaño muestral y la frecuencia de la toma de muestras. Otra forma de decidir el tamaño y la frecuencia de las muestras es a través de los conceptos de ARL y ATS que hemos visto.

2.2.2. ¿Cómo se interpreta un gráfico de control?

Se dice que un proceso está bajo control si la única variabilidad presente en el mismo es la debida a causas comunes, también llamadas aleatorias o no asignables, es decir, aquellas que son impredecibles y son inherentes a la incertidumbre del proceso como, por ejemplo, una alteración de la temperatura, la humedad u otras cuestiones ambientales. El proceso bajo control no presentará causas asignables, que son aquellas que producen efectos predecibles y aumentan la variabilidad. Si volvemos al ejemplo de la fábrica de bolígrafos, algunas causas asignables en la variación del proceso pueden ser, por ejemplo, un fallo grande de la máquina de llenado de tinta y por eso los bolígrafos no tienen la misma capacidad; o un cambio sustancial en el plástico que se utiliza para la fabricación y por ello los bolígrafos son de diferente color. Estas causas,

normalmente, son muy influyentes en el proceso, por ello tiene que encontrarse la forma de eliminarlas en el caso de que sean detectadas en algún proceso.

Los gráficos de control no solo nos sirven para saber si el proceso se ha desviado del objetivo que deseamos, sino que también nos servirán para detectar algún tipo de tendencia en la característica del proceso que estamos controlando. Si, por ejemplo, detectamos que la variable tiende a empeorar, podremos establecer una alarma y estudiar el proceso antes de que esté fuera de control. Si, en cambio, detectamos una tendencia a mejorar el proceso, podremos estudiar y analizar qué es lo que hace que el proceso mejore para así, en un futuro, continuar mejorando la calidad del mismo.

Un cambio en el proceso que estemos estudiando se puede detectar antes de que se salga de los límites de control. Para ello, estaremos atentos a cualquier patrón como los siguientes, los cuales nos darán una pista sobre qué está pasando en el proceso:

- **Cambios bruscos en la media y/o en la variabilidad:** se verán reflejados a través de puntos muy extremos fuera de los límites de control.
- **Tendencias o rachas:** se producen cuando la media o la varianza se desplaza de forma continua en el tiempo. Una racha se da cuando varios puntos consecutivos se encuentran en el mismo lado de la línea central.
- **Periodicidades,** es decir, repeticiones cíclicas en la característica que estemos estudiando.
- **Inestabilidad:** se dan grandes fluctuaciones o variaciones que pueden provocar que algún valor se salga fuera de los límites de control.
- **Sobreestabilidad:** al contrario que en la inestabilidad, se da cuando la variabilidad es menor de la que cabría esperar.

Además de tener en cuenta los patrones que pueden surgir en un gráfico de control, para detectar que un proceso está fuera de control existen unas reglas propuestas por la *Western Electric Handbook* (Western Electric, 1956). Un proceso está fuera de control si, en un gráfico de control detectamos alguno de los siguientes casos:

- Un punto cae fuera de los límites 3 sigma.
- Dos de tres puntos caen fuera de los límites 2 sigma.
- Cuatro de cinco puntos consecutivos se encuentran a una distancia de un sigma o más de la línea central.

- Ocho puntos consecutivos se encuentran en un mismo lado de la línea central, es decir, una racha de al menos ocho puntos.

En ocasiones, que un proceso tenga observaciones fuera de control no significa que se esté perdiendo calidad, hay que tener en cuenta la variable que estamos estudiando. Por ejemplo, en la fábrica de bolígrafos, que incluyan más o menos tinta de la que cabe esperar sí que es una pérdida de calidad. Por un lado, están los bolígrafos con menos cantidad de tinta que no van a satisfacer las necesidades del cliente que busca una capacidad de escritura de 2000 metros. Por otro lado, están los bolígrafos con una mayor cantidad de tinta que, aunque están satisfaciendo a los clientes, es muy probable que la empresa incurra en gastos y acabe soportando pérdidas innecesarias. Sin embargo, si estudiamos otra característica de los bolígrafos, como puede ser el número de veces que se acciona el pulsador antes de que empiece a producir errores, si el proceso está fuera de control por tener muchas observaciones por encima de los límites superiores, podría ser debido a que el nuevo material es mejor, a que el procedimiento de fabricación está mejorando u otros. Si, por ejemplo, utilizamos gráficos de control para monitorizar el comportamiento de cierta enfermedad, como haremos en este trabajo; es deseable que el número de contagios se encuentre por debajo de los límites inferiores de control ya que podría significar que esa enfermedad en cuestión se está erradicando.

Aunque en un principio los gráficos de control estaban pensados para su uso en el control de los procesos de producción industrial y, en general, en todo el ámbito industrial, estos tienen aplicaciones en muchas otras áreas (Frisén, 2009). Se han utilizado, por ejemplo, en el ámbito de la economía para detectar cambios en momentos de expansión o recesión económica o en el ciclo de la economía. También comienzan a utilizarse para monitorizar y detectar cambios importantes en el medioambiente, controlando la biodiversidad y algunas especies. Otra área de expansión del uso de los gráficos de control y el control estadístico de procesos es el de la salud pública, que es el ámbito que nos concierne en el presente trabajo.

2.2.3. Tipos de gráficos de control

Gráficos de control para variables

Muchas características de un producto o proceso se pueden medir en términos numéricos, como por ejemplo el peso, el volumen, o cualquier otra característica continua. En este caso, estamos hablando variables cuantitativas, también llamadas simplemente variables (Montgomery, 2009).

Para el caso de tener que controlar una variable cuantitativa, suele ser necesario controlar su valor medio y su variabilidad. Existen dos tipos de gráficos: aquellos que nos sirven para controlar medidas de posición de la variable (media o mediana) y aquellos que nos sirven para el control de la dispersión o variabilidad (gráficos de desviaciones típicas, varianzas o rangos).

Para construir un gráfico de control para la media, debemos observar los gráficos para la dispersión, ya sea los de rangos o los de desviaciones típicas. Esto es así ya que los límites de control inferior y superior en el gráfico de la media se construyen a partir de la estimación de la desviación estándar de la variable que estemos estudiando:

$$\hat{\sigma} = \frac{\bar{R}}{d_2},$$

utilizando el rango medio, o:

$$\hat{\sigma} = \frac{\bar{s}}{c_4},$$

utilizando la desviación típica muestral. Los valores que se le pueden dar a d_2 se encuentran tabulados. Los valores de c_4 dependen del tamaño muestral y aprenderemos a calcularlos a continuación.

Por lo tanto, primero se establecerán los límites de control de los gráficos de rangos (R) o del de desviaciones típicas (s) y, después, se podrá graficar el de las medias.

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

Para elaborar el gráfico de medias utilizando las desviaciones típicas muestrales, calcularemos la media ($\bar{x}_i$) y la desviación típica de cada muestra (s_i) para cada i ($i = 1, 2, \dots, k$).

La media global es un estimador insesgado de la media teórica, μ . Se puede calcular mediante la media de todos los datos de cada muestra o promediando el total de los datos.

$$\bar{\bar{x}} = \frac{\sum_i \bar{x}_i}{k} = \frac{\sum_i \sum_j x_{ij}}{N},$$

Con $N = kn$.

La cuasidesviación típica muestral es un estimador sesgado de la desviación típica teórica ya que $E[s_i] = c_4$, en donde c_4 es una constante que depende del tamaño muestral, n . Tendremos, entonces, el siguiente estadístico como estimador insesgado de la desviación típica:

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

Podemos calcular el valor de la constante c_4 como sigue:

$$c_4 = \sqrt{\frac{2}{n-1}} \frac{\left[\left(\frac{n}{2} \right) \right]}{\left[\left(\frac{n-1}{2} \right) \right]}$$

Una vez calculada la estimación de la desviación típica muestral, podemos calcular los límites de control que se representarán en el gráfico de medias:

$$LCL = \bar{\bar{x}} - A_1 \bar{s}$$

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

$$UCL = \bar{\bar{x}} + A_1 \bar{s}$$

En donde $A_1 = \frac{3}{\sqrt{nc_4}}$.

En el caso de querer utilizar las cuasidesviaciones típicas en lugar de las desviaciones típicas muestrales (es bastante común y en muchos paquetes de R de control estadístico de la calidad se utilizan), los estimadores y los límites quedarían así:

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

$$LCL = \bar{\bar{x}} - A_3 \bar{\hat{s}}$$

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

$$UCL = \bar{\bar{x}} + A_3 \bar{s}$$

En donde $A_3 = \frac{3}{\sqrt{n}c_4}$.

El gráfico de desviaciones típicas, s , se puede construir con el intervalo $(B_3\bar{s}, B_4\bar{s})$, que contiene a la desviación típica muestral s_i con una probabilidad aproximada de 0.9974. En el intervalo, las constantes B_3 y B_4 dependen del tamaño muestral y están tabuladas. También las podemos calcular siguiendo las siguientes fórmulas:

$$B_3 = 1 - \frac{3}{c_4} \sqrt{1 - c_4^2}$$

$$B_4 = 1 + \frac{3}{c_4} \sqrt{1 - c_4^2}$$

Con ese intervalo, se pueden representar los límites del gráfico de control de desviaciones típicas, que serán:

$$LCL = B_3\bar{s}$$

$$CL = \bar{s}$$

$$UCL = B_4\bar{s}$$

▪ Gráficos de control $(\bar{x}, R)$

Otro de los gráficos de control que podemos utilizar es el basado en recorridos o rangos. El rango de una muestra es la diferencia entre la mayor y la menor de sus observaciones (Montgomery, 2009). Es decir, si tenemos una muestra de tamaño n , $x_1, x_2, \dots, x_n$, podemos calcular el rango de esa muestra:

$$R = x_{max} - x_{min}$$

Este estadístico nos proporciona una estimación de la desviación típica de la población de la que procede la muestra, por lo que puede sustituirla para el cálculo de los gráficos de control de la media. Esto es así porque $\frac{R_i}{d_2}$ es un estimador insesgado de σ y, por lo

tanto, su media a lo largo de todas las muestras también lo será. Tomando los rangos de k muestras, $R_1, R_2, \dots, R_k$, obtenemos el promedio de los rangos:

$$\bar{R} = \frac{R_1 + R_2 + \dots + R_k}{k} = \frac{\sum_i R_i}{k}$$

Podremos estimar, entonces, la desviación típica como $\frac{\bar{R}}{d_2}$ y construir los límites de control:

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

$$CL = \bar{x}$$

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

Siendo $A_2 = \frac{3\bar{R}}{\sqrt{n}d_2}$.

Como en el caso del gráfico de control $\bar{x}$, podemos construir el gráfico de control de los rangos muestrales tomando como línea central el valor esperado del rango muestral $d_2\sigma$ y como límites de control los dados por el intervalo $(D_1\sigma, D_2\sigma)$. Los valores D_1 y D_2 están tabulados y su intervalo contiene el rango muestral con una probabilidad de 0.9974 para una población normal de varianza 1.

▪ Gráficos de medidas individuales

Podemos hacer uso de gráficos de medidas individuales cuando tenemos una única observación en cada momento del tiempo. Por ejemplo, cuando solo podemos estudiar una observación en cada lote de productos o materiales, en procesos continuos cuando no tiene sentido hablar de "individuos", o cuando queremos realizar una comparación con las especificaciones de calidad.

El problema recae en la estimación de la varianza. Para ello se pueden utilizar dos métodos: rangos móviles o agrupación por bloques.

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

El método de agrupación de los datos en bloques consiste en formar grupos con esos datos.

Este gráfico se construye de forma similar al gráfico de medias usando rangos $(\bar{x}, R)$, pero teniendo en cuenta que $n = 1$ y realizando las correcciones oportunas. La estimación de σ 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 típicas $\overline{s_{n-1}}$. De esta forma, podemos construir los límites de control para el gráfico de medidas individuales:

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

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

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

Siendo c_2 , d_2 y c_4 constantes que dependen del tamaño de la muestra y que están tabuladas.

Si estamos utilizando los criterios clásicos de 3σ , se utilizan unas constantes (E_1, E_2, E_3) para estimar los límites del gráfico de control:

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

$$\bar{\bar{x}} \pm E_2 \bar{R}$$

$$\bar{\bar{x}} \pm E_3 \overline{s_{n-1}}$$

En el caso de tener medidas individuales, también debemos evaluar conjuntamente con el gráfico de las desviaciones típicas o el gráfico de rangos, para así evaluar también la variabilidad de los datos.

Gráficos de control para atributos

En ocasiones, algunas características de calidad no pueden expresarse ni medirse de forma numérica como con las variables. En ese caso, simplemente se observará si presenta o no alguna propiedad o si es conforme o no conforme a las especificaciones existentes sobre esa característica de calidad, por ejemplo: un producto es defectuoso

o no, una pieza encaja o no en otra, un mecanismo funciona o no funciona, etc. En control de calidad es habitual utilizar los términos conforme y no conforme en lugar de defectuoso o no defectuoso. En ese caso, estaremos hablando de atributos o variables cualitativas.

Dentro de esta clasificación, distinguimos dos tipos de gráficos de control para atributos: los que estudian la proporción de unidades defectuosas o no conformes (gráficos p y np) y los que estudian el número de defectos o no conformidades observadas (gráficos c y u).

▪ Gráfico p

El gráfico p se utiliza cuando la característica de calidad que estemos estudiando la podemos clasificar en defectuosos y no defectuosos (o conformes y disconformes) y deseamos controlar la proporción (que será p) de alguno de estos grupos.

Este gráfico se construye representando la proporción de unidades no conformes en el gráfico y calculando los límites de control de forma similar a como hemos explicado para otros gráficos de control (Prat Bartés et al., 2015).

Tomando $\hat{p} = \frac{\sum_{i=1}^k r_i}{kn}$, siendo k el número de muestras, n el número de elementos total y r el número de defectuosos en cada muestra; podemos obtener los límites de control:

$$LCL = \bar{p} - 3 \frac{\sqrt{\bar{p}(1 - \bar{p})}}{\sqrt{n}}$$

$$CL = \bar{p}$$

$$UCL = \bar{p} + 3 \frac{\sqrt{\bar{p}(1 - \bar{p})}}{\sqrt{n}}$$

En este gráfico, se calculan los límites de control utilizando la distribución binomial que, por el teorema central del límite, se aproxima a una normal. Tenemos que el número de disconformidades se distribuye como una $B(n, p)$ que se puede aproximar a una

$N(np, \sqrt{np(1 - p)})$. Entonces, la proporción media $\bar{p}$ se distribuirá como una $N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$.

▪ Gráfico np

El gráfico np lo podemos utilizar en problemas similares en los que se utiliza el gráfico p , pero esta vez, utilizando el número de unidades disconformes que se puede observar en la muestra en lugar de la proporción de las mismas. Este gráfico es más sencillo de construir que el gráfico p , al no ser necesaria la proporción de unidades defectuosas y pudiéndose interpretar en términos de proporciones simplemente escalando el gráfico dividiendo entre n la escala vertical del mismo (Prat Bartés et al., 2015).

Podemos calcular los límites de control para este gráfico de la forma:

$$LCL = n\bar{p} - 3\sqrt{n\bar{p}(1 - \bar{p})}$$

$$CL = n\bar{p}$$

$$UCL = n\bar{p} + 3\sqrt{n\bar{p}(1 - \bar{p})}$$

En donde $\bar{p}$ es la proporción de unidades defectuosas en la muestra y, por lo tanto, $n\bar{p}$ es el número de unidades defectuosas en la muestra.

En este gráfico, al igual que en el gráfico p , para construir los límites de control nos basamos en la distribución binomial. Por lo tanto, este gráfico se debe utilizar si se producen unidades defectuosas en más del 5% de las unidades supervisadas y no cuando la producción de unidades defectuosas es poco común.

▪ Gráfico c

En el gráfico c lo que tendremos en cuenta, en lugar del número de individuos o productos no conformes en la muestra, será el número de defectos o disconformidades que se pueden encontrar en cada individuo, producto o, de forma más genérica, en cada una de las observaciones que tomemos como muestra. Este tipo de gráficos de control puede resultarnos más útil en algunas ocasiones ya que el producto puede no ser defectuoso aun teniendo defectos y el carácter defectuoso de un producto puede ser de distinta magnitud dependiendo del número de disconformidades que presente. Estos gráficos también serán útiles en un flujo continuo cuando los defectos no van asociados a unidades (Prat Bartés et al., 2015). Por ejemplo, el número de defectos de tinte en una tela, número de burbujas en un cristal, número de camas ocupadas en un hospital por

semana, etc. Todos estos procesos se pueden considerar teniendo en cuenta diferentes unidades de medida: longitud, área, peso, hora, semana, etc. En cualquiera de estos dos casos podemos tomar como referencia una distribución de Poisson (λ), en donde λ representa el número medio de ocurrencias por unidad de medida: tiempo, superficie, espacio o cualquier otra unidad de medida (Benneyan, 1998). Cuando tenemos un λ lo suficientemente grande, podemos aproximar la distribución por una normal.

Siendo c la media y la varianza de una distribución de Poisson (del número de defectos por unidad de medida), que se puede aproximar a una distribución normal cuando $c > 5$, tendremos que:

$$LCL = \bar{c} - 3\sqrt{\bar{c}}$$

$$CL = \bar{c}$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}}$$

▪ Gráfico u

De forma similar a los gráficos np y p , que se podían utilizar en el mismo tipo de problemas, esto también ocurre con los gráficos c y u . El gráfico u lo podemos utilizar para el mismo tipo de problemas que el gráfico c , pero en aquellos casos en los que no tenemos unidades del mismo tamaño para realizar el control del número de defectos (Montgomery, 2009). Por ejemplo, en el caso que nombramos antes de los defectos del tinte en una tela, puede ser difícil tomar unidades de tela del mismo tamaño exacto por lo que, si se utiliza este gráfico, se pueden estudiar varias unidades de un tamaño similar.

Si tenemos x_i el número de defectos en la muestra i , n_i el número de unidades revisadas en la muestra i , podemos obtener el número medio de disconformidades en la muestra:

$$\bar{u} = \frac{\sum_i x_i}{\sum_i n_i}$$

$$i = 1, 2, \dots, k$$

Podemos, entonces, construir los límites del gráfico de control:

$$LCL = \bar{u} - 3 \sqrt{\frac{\bar{u}}{n}}$$

$$CL = \bar{u}$$

$$UCL = \bar{u} + 3 \sqrt{\frac{\bar{u}}{n}}$$

En donde $\bar{u}$ representa el número medio de disconformidades observadas por unidad de control en la muestra.

Otros gráficos de control: gráficos CUSUM y gráficos EWMA

Los gráficos de control anteriores son los llamados gráficos de control de Shewhart. Estos son muy útiles en la fase I del Control Estadístico de Procesos para detectar si el proceso está o no fuera de control y aquellas causas asignables que puedan causar cambios significativos en el mismo. También son útiles para detectar la naturaleza de dichas causas asignables, detectando patrones y tendencias.

Sin embargo, los gráficos de control de Shewhart solamente utilizan la información de la última observación de la muestra y no utilizan el resto de información de la muestra. Esto provoca que los gráficos no lleguen a detectar pequeños cambios en el proceso, lo que hace que sean menos útiles en la segunda fase del SPC, en la que el proceso suele estar ya bajo control y las causas asignables no provocan grandes cambios (Montgomery, 2009).

Existen gráficos alternativos si lo que se quiere es detectar cambios mucho más pequeños en el proceso y para la fase II del SPC: el gráfico de sumas acumuladas o acumulativas (CUSUM) y el gráfico de control de medias móviles ponderadas exponencialmente (EWMA). Hablaremos un poco más de cada uno de ellos a continuación.

Estos gráficos, al ir acumulando información sobre las observaciones en periodos de tiempo anteriores, se conocen como gráficos con memoria.

- **Gráficos de sumas acumuladas (CUSUM)**

Como alternativa a los gráficos de Shewhart, Page (1954) propone el gráfico CUSUM (*Cumulative Sum*) o de sumas acumulativas, el cual basa la decisión analizando toda o una gran parte de la muestra, a diferencia de los gráficos de Shewhart como hemos comentado antes.

Este gráfico utiliza el estadístico s_r :

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

Siendo $\bar{x}_i$ la media muestral de la muestra i -ésima y μ_0 el valor de referencia u objetivo para la media del proceso.

Este estadístico tiene en cuenta las r primeras submuestras para tomar la decisión sobre la muestra r -ésima del proceso. El gráfico CUSUM representa las sumas acumuladas de las desviaciones entre cada dato, en este caso la media muestral ($\bar{x}_i$) y un valor de referencia (μ_0).

Como vemos, para la construcción de este estadístico se utiliza información acerca de varias muestras, por lo que será más eficaz que un gráfico de control de Shewhart para detectar cambios pequeños. También son eficaces con muestras de tamaño $n = 1$, es decir, muestras individuales como las que podemos obtener, por ejemplo, en la industria química.

Bajo normalidad, la distribución del estadístico s_r será:

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

Bajo el supuesto de que la hipótesis nula $\mu = \mu_0$ sea cierta, las sumas acumuladas del estadístico s_r tendrán media 0 por lo que cabe esperar que en el gráfico CUSUM, las sumas oscilen alrededor de la recta $y = 0$. En el caso contrario, suponiendo que no aceptamos como cierta la hipótesis nula, los valores de s_r representados en el gráfico CUSUM tendrán una tendencia creciente o decreciente.

Existen dos formas de representar los gráficos CUSUM: el CUSUM tabular o algorítmico y la máscara o plantilla en forma de V.

Supongamos que queremos controlar la evolución de una variable $X \sim N(\mu, \sigma)$, en donde la desviación típica es conocida o podemos estimarla fácilmente. El **CUSUM tabular o algorítmico** calcula las desviaciones de cada valor respecto del valor objetivo μ . Vamos a distinguir entre desviaciones positivas y negativas ya que pueden no tener la misma importancia en el gráfico.

Primero, definiremos el valor K a partir del cual se considera que la desviación acumulada es significativa y el cual determinará la sensibilidad del gráfico de control. Si la suma acumulada hasta la observación i -ésima es menor que el K que escojamos, se considera que la desviación acumulada es 0 ya que no será significativa. Si el gráfico no es sensible a estas desviaciones que no son significativas será sencillo visualizar aquellas desviaciones que sí lo sean.

Definimos los dos tipos de sumas acumuladas significativas:

- C_i^+ = acumulación de desviaciones positivas significativas:

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

Si el proceso está bajo control, la variable X tomará valores cercanos al valor que definimos como objetivo y la acumulación de desviaciones positivas significativas, C_i^+ , tenderá a tomar valores nulos.

- C_i^- = acumulación de desviaciones negativas significativas:

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

Los valores obtenidos de las sumas acumuladas se compararán con el valor de decisión H . Se suele recomendar que se tomen unos valores para el valor de referencia K y el intervalo de decisión H que proporcionen un buen ARL. El valor de H se define como $H = h\sigma$ y se suele recomendar que $h = 5$ o $h = 4$, es decir, se tomará un H igual a 4 o 5 veces la desviación típica. En el caso del valor de referencia K , se define como $K = k\sigma$ y se recomienda que el valor de k sea igual a $\frac{1}{2}$. Si tomamos estos valores es muy probable que tengamos un gráfico CUSUM con buenas propiedades ARL.

La otra forma de representar el gráfico CUSUM es el de la **máscara o plantilla V**, que es el más popular a la hora de dibujar este tipo de gráficos. La máscara V se basa en la representación gráfica del estadístico CUSUM:

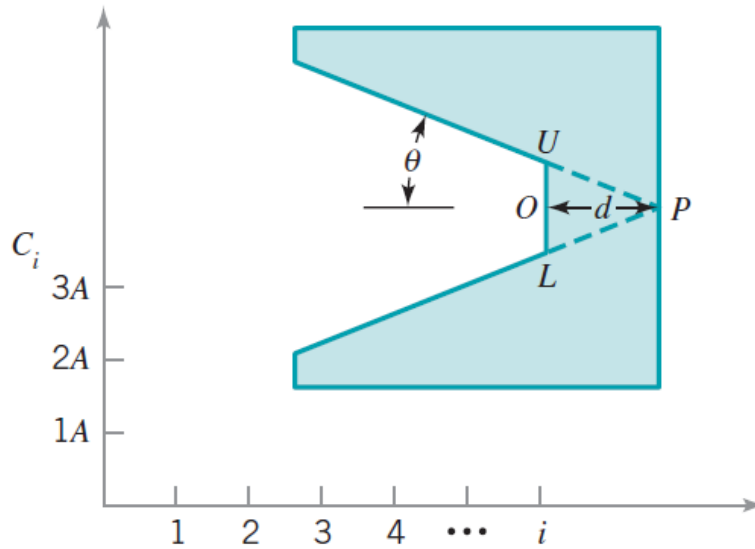
$$C_i = \sum_{j=1}^i (\bar{x}_j - \mu_0)$$

También se puede utilizar la versión estandarizada del estadístico:

$$C_i = \sum_{j=1}^i \left(\frac{\bar{x}_j - \mu_0}{\sigma} \right)$$

En cualquiera de los dos estadísticos tenemos que μ_0 es el valor de referencia y σ la desviación típica, ya sea conocida o la podamos estimar, de la variable que estemos controlando.

La máscara V consiste en un ángulo y una línea paralela al eje horizontal (OP) que se coloca sobre el gráfico de control CUSUM. En este gráfico, la distancia OP y el ángulo θ determinan su sensibilidad. Si las sumas acumulativas $C_1, C_2, \dots, C_i$ se encuentran entre las dos líneas que forman el ángulo de la máscara V, el proceso estará bajo control. Si, en cambio, alguna de las sumas acumuladas se encuentra por fuera de esas líneas el gráfico sugiere que se ha producido algún cambio y el proceso podría estar fuera de control. Los puntos por encima de la línea superior indicarían que la media del proceso está decreciendo, mientras que, los puntos por debajo de la línea inferior indicarían que la media del proceso se está incrementando. Esto se ve mejor gráficamente en la **Figura 2**.

Figura 2: Construcción de la máscara V para un gráfico CUSUM

Construcción de la máscara V para un gráfico de sumas acumuladas CUSUM. Extraído de Montgomery (2009).

Hay diversas formas de calcular tanto θ como la distancia d (distancia del segmento OP). En varios paquetes se utiliza la forma que propone Johnson (1961), [véase también (Johnson & Leone, 1962)]. Este autor propone fijar dichos parámetros de la siguiente forma:

$$d = \frac{h}{K} = \frac{2}{\delta^2} \frac{\sigma^2}{n} \ln\left(\frac{1-\beta}{\alpha/2}\right)$$

$$\theta = \arctan\left(\frac{\delta}{2A}\right)$$

Donde:

- A es la distancia horizontal entre puntos sucesivos del eje de abscisas medida en las unidades del eje de ordenadas. Dependerá, por tanto, de la escala del gráfico. Si el gráfico se realiza con algún programa estadístico ya viene predeterminado y no necesita calcularse.
- δ es la sensibilidad del análisis y el nivel de los cambios o desajustes que se quieren detectar, en número de desviaciones típicas.

- α es la probabilidad de una falsa alarma cuando el proceso se encuentra bajo control.
- β es la probabilidad de no detectar una desviación δ .
- $\bar{n}$ es el promedio de los tamaños muestrales. Si estamos ante observaciones individuales, $\bar{n} = 1$.

Algunos autores (Montgomery, 2009) no recomiendan el uso de este último procedimiento ya que algunas funciones muy útiles complementarias a los gráficos CUSUM no se pueden usar si utilizamos la máscara V. Además, a veces es difícil saber hasta dónde prolongar las líneas que forman el ángulo V. Otro de los problemas es que los valores de α y β pueden resultar algo ambiguos.

▪ **Gráficos de control de medias móviles ponderadas exponencialmente (EWMA)**

Los gráficos de control de medias móviles ponderadas exponencialmente o gráfico EWMA (del inglés, *Exponentially Weighted Moving Average*) fueron propuestos en su inicio por Roberts (1959) y fueron utilizados en procesos industriales o en finanzas por Wortham y Ringer (1971).

Como los gráficos CUSUM, los gráficos EWMA se utilizan para detectar cambios pequeños que los gráficos de control de Shewhart no detectan ya que utilizan mucha más información que estos últimos. Normalmente, en este tipo de gráficos se utilizan observaciones individuales.

El estadístico que se representa en el gráfico para cada periodo de tiempo es la media ponderada de la observación actual y todas las observaciones anteriores, en donde la media inmediatamente anterior recibe más peso. Es decir, se representa el siguiente estadístico:

$$Z_t = \lambda x_t + (1 - \lambda)Z_{t-1}$$

$$0 < \lambda < 1$$

En donde:

- $Z_0 = \mu$
- Z_t = media móvil ponderada exponencialmente en el periodo actual, t .

- Z_{t-1} = media móvil ponderada exponencialmente en el periodo inmediatamente anterior $t - 1$.
- x_t = la observación actual.
- λ es el factor de ponderación para la observación actual.

Por lo tanto, en cada momento del tiempo se está ponderando la observación actual x_t con el valor de la media móvil anterior Z_{t-1} .

Si se cumplen las hipótesis de independencia y estabilidad sobre las observaciones x_t , tendremos que:

$$E(Z - t) = \mu$$

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

Si nos fijamos en la fórmula de la varianza, el término $(1 - \lambda)^{2t}$ converge hacia 1, por lo que la varianza nos quedaría:

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

Si elegimos un $\lambda = \frac{2}{(t+1)}$, podemos aproximar la expresión de la varianza por $Var(Z_t) \approx \frac{\sigma^2}{t}$. Bajo estos supuestos, podemos construir los límites de control a través de la siguiente expresión:

$$\hat{\mu} \pm 3 \sqrt{\frac{\sigma^2}{t}}$$

Para otros valores diferentes de λ , los límites de control serán:

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

$$CL = \hat{\mu}$$

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

En ocasiones, con la fórmula de la varianza no se obtiene una buena estimación, por lo que se tiene que utilizar un gráfico de rangos y estimar la varianza con la fórmula ya conocida $\frac{\bar{R}}{d_2\sqrt{n}}$. Si tenemos valores individuales, el rango medio móvil se puede calcular como en los gráficos de control para medidas individuales.

La efectividad de los gráficos EWMA para detectar estos cambios pequeños depende del valor que le demos a λ . Si $\lambda \rightarrow 1$, los valores dibujados en el gráfico EWMA dependerán únicamente de las observaciones recientes y el gráfico será similar al gráfico de Shewhart. Al contrario, cuando $\lambda \rightarrow 0$, se está dando más importancia a valores anteriores de la característica estudiada en el proceso y nos acercamos al funcionamiento del gráfico CUSUM (Prat Bartés et al., 2015). Podemos decir, entonces, que si λ es próximo a 1, el gráfico tendrá poca memoria, mientras que, si es próximo a 0 tendrá mucha más memoria. La elección de este parámetro depende totalmente de los cambios que quiera detectar investigador, por lo que si se quiere detectar un cambio pequeño elegirá un λ pequeño para poder acumular y utilizar la información de muchos periodos anteriores al instante actual.

Una forma más genérica de expresar el gráfico EWMA es de la siguiente forma:

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

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

De forma similar a la elección de λ que hemos explicado anteriormente, a través de esta expresión, podemos:

- Obtener un gráfico de Shewhart para la media si $k = 1$, $w_i = 1$ y $a = 0$.
- Obtener un gráfico CUSUM si $w_i = 1$ y $a = -k\mu_0$.

Al igual que el gráfico CUSUM, el gráfico EWMA es más efectivo para detectar pequeños cambios que cualquiera de los gráficos de Shewhart estudiados. Sin embargo, estos últimos son mejores que los gráficos CUSUM y EWMA para detectar cambios más grandes en el proceso.

2.3. Gráficos de control en el ámbito de la salud pública

Los gráficos de control en el ámbito de la salud pública comienzan a utilizarse más tarde que en la industria. En la década de 1950 comienzan a utilizarse en laboratorios para la monitorización y el estudio de los análisis clínicos y en los bancos de sangre de los hospitales (Levey & Jennings, 1950). Levey y Jennings concluyen que los gráficos de control pueden aplicarse a los resultados de cualquier análisis químico de los que se realizan de forma rutinaria en los laboratorios y no solo a los resultados de los análisis de sangre, objeto principal de su trabajo. En los últimos años se ha producido un creciente interés en la reducción de costes y su control a niveles bajos, ya sea en el marco de la sanidad pública o de la sanidad privada. Como consecuencia, dada su estrecha relación con la mejora continua en los procesos, también ha aumentado la preocupación por el control y mejora de la calidad de los servicios sanitarios, con el objeto de obtener los mejores resultados a un menor coste (Shortell et al., 1998). Esto ha llevado a una aplicación cada vez más usual e intensa de técnicas estadísticas (como las pertenecientes a la rama del control estadístico de la calidad) que ayuden significativamente a la mejora de la calidad en el ámbito sanitario, además del desarrollo de nuevas herramientas específicas para la resolución de problemas en este marco.

2.3.1. Diferencias entre el SPC en la industria y en la medicina

Existen grandes diferencias en el control de calidad y el control estadístico de procesos dependiendo de que el ámbito estudiado sea la salud pública o la industria, sector en el que comenzaron a utilizarse este tipo de técnicas en un primer lugar. Debido a estas diferencias, el uso de SQC y SPC es relativamente reciente en el sector sanitario, sobre todo si se compara con otros ámbitos como la industria o la empresa en general.

En el ámbito de la salud, es muy habitual el uso de variables cualitativas o atributos, a diferencia de la industria (Woodall, 2006). El sexo, el grupo sanguíneo, la presencia o no de alguna enfermedad, presencia de algún factor de riesgo como puede ser el consumo de alcohol, tabaco, sedentarismo, etc., son variables cualitativas muy utilizadas en este ámbito.

En numerosas ocasiones, en el ámbito sanitario, es necesario tener en cuenta el riesgo inherente a los datos de los que se disponen antes de construir un gráfico de control (Woodall, 2006). Por ejemplo, no se pueden comparar pacientes de diferentes edades, condiciones físicas o hábitos (consumo de alcohol, tabaco, sedentarismo, etc.) ya que no tienen la misma probabilidad de supervivencia de alguna enfermedad, es decir, el riesgo de padecer esa enfermedad no será el mismo para cada paciente.

Otro hecho a tener en cuenta es que, en el caso de tener datos de alguna variable sanitaria, supongamos el número de contagios de COVID-19, al aplicar un proceso iterativo de eliminación de observaciones fuera de control, siguen apareciendo más observaciones fuera de control. En estos casos, un gráfico de control puede proporcionarnos varias alarmas después de detectar la primera (Woodall, 2006). Esto puede ser debido a que el proceso a que todo el proceso está de por sí fuera de control o, lo más usual, a que las observaciones de la variable no son independientes. Por ejemplo, y siguiendo con el caso del COVID-19, en los últimos dos años la enfermedad ha ido evolucionando en el tiempo, teniendo épocas de auge con muchos contagios diarios, seguidas de épocas de contracción de la enfermedad y viceversa, dando lugar a varias olas de la enfermedad. Si viéramos los datos en un gráfico de control es muy probable que éste detecte muchas alarmas, cada vez que una ola está por comenzar e, incluso, cuando simplemente aumenta más de lo habitual el número de contagios, hospitalizaciones u otras variables a estudiar. Esto que ocurre no es lo más deseable ya que, en ocasiones, las autoridades encargadas de proporcionar los datos no los proporcionan diariamente. Por ejemplo, durante los primeros meses de la pandemia, los datos de nuevos contagios eran proporcionados por los consejos sanitarios autonómicos y el Ministerio de Sanidad del Gobierno de España de forma diaria a excepción de los fines de semana, días festivos, etc. Además, no todas las Comunidades Autónomas notificaban los contagios a la misma velocidad, pudiendo producir desajustes en los datos a nivel estatal. Los lunes y días posteriores a alguna festividad, por lo tanto, solía haber un pico de contagios debido a eso. Es probable que un gráfico de control estándar proporcione una alarma cada vez que se producen estos picos de contagios, aunque al día siguiente los contagios vuelvan a estabilizarse.

2.3.2. Gráficos y métodos más utilizados en medicina

Los **gráficos de control estándar o gráficos de control de Shewhart** de los que hemos hablado anteriormente, han sido utilizados frecuentemente para monitorizar algunos indicadores de la salud pública tales como muertes, hospitalizaciones o ingresos en plantas UCI, entre otros muchos (Benneyan, 1998). Como ya hemos visto en el apartado anterior, existen diversos tipos de gráficos de control, por lo que se puede utilizar un tipo u otro dependiendo del tipo de variable que estemos tratando: gráficos para variables continuas como el tipo de ingreso en un hospital, gráficos para variables binomiales como puede ser la mortalidad, etc.

Sin embargo, estos gráficos no son los más adecuados para el seguimiento de observaciones individuales, sobre todo cuando no disponemos de demasiados datos. En este caso, uno de los gráficos más utilizados serán los **gráficos de sumas acumuladas o CUSUM** (Noyez, 2009). Los gráficos CUSUM son muy utilizados en este ámbito principalmente por su sencillez y su interpretación visual tan intuitiva, aun cuando no se tienen estudios específicos sobre estadística o control estadístico de la calidad. Otra de las ventajas de estos gráficos es su capacidad para detectar pequeños cambios y tendencias antes de que los procesos se salgan de control. Además, estos gráficos pueden ser utilizados en casos raros y con poca incidencia al conseguir monitorizar las enfermedades en tiempo real desde su comienzo, lo que los convierte en una buena herramienta en el ámbito de la medicina cuando se pretenden estudiar enfermedades raras o detectar brotes de enfermedades nuevas de las que aún no se tienen muchos datos o no hay demasiada incidencia en la población (Forteza-Sanchis y Escrig-Sos, 2019).

Ya hemos explicado las bases de cómo construir un gráfico CUSUM algorítmico o con máscara, pero existen algunas variaciones de los mismos que, como es el caso de este trabajo, se suelen utilizar en el ámbito de la medicina. Hablaremos **del gráfico CUSUM de fallos** y **del gráfico CUSUM ajustado al riesgo**:

- **Gráfico CUSUM de fallos:**

Este gráfico se construye representando la acumulación del número de eventos (llamados fallos) en el eje vertical y, el número total del procedimientos (o pacientes) en el eje horizontal. A diferencia del gráfico CUSUM clásico que ya hemos estudiado, este

gráfico va formando una línea ascendente con una pendiente de aproximadamente 45° a medida que van surgiendo fallos. En el caso de no tener ningún evento o fallo, la línea que se forma no tendrá pendiente, es decir, será horizontal (Noyez, 2009).

Como en los gráficos CUSUM clásicos, podemos representar unos límites de control que se construirán utilizando la tasa de fallos aceptada (p_0) y la no aceptada (p_1), de acuerdo con el error de tipo I y el error de tipo II, respectivamente.

Para construir los límites de control del gráfico CUSUM de fallos, primero debemos definir una serie de parámetros (Bolsin y Colson, 2000):

- P_0 = tasa aceptable de fallos
- P_1 = tasa inaceptable de fallos
- α = tasa de error de tipo I
- β = tasa de error de tipo II

Con estas definiciones, podemos obtener los siguientes valores, que utilizaremos para la construcción de los límites del gráfico CUSUM:

$$a = \ln\{(1 - \beta)/\alpha\},$$

$$b = \ln\{(1 - \alpha)/\beta\},$$

$$P = \ln(p_1/p_0),$$

$$Q = \ln\{(1 - p_0)/(1 - p_1)\},$$

$$s = Q/(P + Q).$$

Siendo s la disminución hacia abajo del gráfico CUSUM con cada éxito obtenido, mientras que el incremento hacia arriba con cada fracaso vendrá denotado por $(1 - s)$.

Con estos valores, podemos construir unos límites similares a los límites de control:

$$CF_{acceptable} = sN - b/(p + Q)$$

$$CF_{inacceptable} = sN + a/(P + Q)$$

En donde N es el número de intentos.

$CF_{acceptable}$ define el rendimiento aceptable en un gráfico CUSUM en fallos, mientras que $CF_{inacceptable}$ define el rendimiento inaceptable.

- **Gráfico CUSUM ajustado al riesgo:**

En ocasiones, y, sobre todo, en el ámbito médico, es necesario ajustar los gráficos al riesgo. Esto es así porque no podemos comparar a pacientes con distintas características físicas o médicas. Por ejemplo, la probabilidad de contraer alguna enfermedad no será la misma en un paciente de 20 años totalmente sano que en un paciente de 80 años. El ajuste al riesgo cobra un papel fundamental en este ámbito, mucho mayor que el que puede tomar en otros sectores como el industrial (Woodall, 2006).

Para este gráfico también es posible estimar unos límites de control, pero al depender del riesgo diferente de cada paciente, se convierte en una cuestión compleja.

Muchas veces nos encontramos con la necesidad de comparar datos como, por ejemplo, datos de las tasas de mortalidad de diferentes hospitales, diferentes médicos o enfermedades. Para llevar a cabo esta tarea, se utilizan unos tipos de tablas y gráficos diferentes a los usuales en el marco del control estadístico de la calidad. Estamos hablando de las **tablas de clasificación** (también llamados **cuadros o gráficos comparativos**) y los **gráficos de embudo**.

Las **tablas de clasificación o gráficos comparativos** consisten en representar los datos que vamos a comparar teniendo en cuenta la característica de estudio, cuyos valores están ordenados de mejor a peor. Es decir, si estamos comparando la tasa de mortalidad de los pacientes tratados por diferentes médicos, se representarán primero los médicos con unas tasas de mortalidad bajas y después, los médicos con unas tasas de mortalidad más altas. En el eje vertical se representará la tasa o característica que se esté estudiando y en el eje horizontal el médico, hospital o similar. A pesar de la fácil interpretación de estas tablas, estos gráficos han sido criticados por muchos autores (Adab et al., 2002), en parte, por la forma de construirlo: se ordenan las unidades de mejor a peor teniendo en cuenta características muchas veces subjetivas que no son medibles, en lugar de características medibles que sí tienen importancia estadística. Además, estos resultados también dependerán mucho de la calidad de los datos a los

que se pueden acceder. Por otro lado, es muy común que estos datos sean proporcionados directamente por los hospitales u otros organismos sanitarios, como pueden ser los ministerios o consejerías de sanidad, y no se pueda acceder a ellos de forma directa.

Los **gráficos de embudo** (en inglés, *funnel plots*) han sido propuestos por Spiegelhalter (2002) como una alternativa a los criticados gráficos comparativos. En los gráficos de embudo se representa en el eje horizontal el número de pacientes correspondientes a cada médico, hospital o, alternativamente, aquella característica que estemos comparando y, en el eje vertical, la tasa de interés (mortalidad, nacimientos, contagios, etc.) . De forma similar a otros gráficos de control que ya hemos estudiado, en los gráficos de embudo se representa una línea central horizontal θ_0 , que representa el objetivo del estudio y que es la media de todas las unidades que estemos comparando, ponderadas según el volumen de casos que cada una de ellas aporte al estudio. Es común que, si estamos comparando médicos u hospitales diferentes, cada uno de estos aporten un número diferente de pacientes o casos al análisis, por diversas razones. Si, por ejemplo, tenemos un médico que aporta 50 casos mientras que el resto aportan 100 o 150, cada uno, estos resultados serán ponderados en función del número de casos aportados por cada médico. En estos gráficos, además, se representarán también unas líneas de decisión que podemos analizar como si se tratara de los ya conocidos límites de control, basadas en un múltiplo del error estándar sobre la tasa media, calculadas con una confianza del 99.8%, tomando 3 desviaciones típicas. En ocasiones, se añaden otros límites utilizando 2 desviaciones típicas, y que se corresponden con una confianza del 95%, representando, así, 4 límites de control. Estos límites van haciéndose más estrechos a medida que avanzamos en el gráfico, dando lugar a la forma de embudo que le da nombre al gráfico.

Dada una serie de I observaciones y_i con las precisiones asociadas ρ_i , un gráfico de embudo consiste en un gráfico de y_i frente a i , con el objetivo θ_0 representado por una línea horizontal y unos límites de control dibujados antes que los propios datos y que no dependerán de los mismos. El parámetro de decisión ρ determina la exactitud con la que se mide el indicador de manera que la distribución de y dentro de la hipótesis nula de Y , dada una institución en control en la que se alcanza el objetivo, se toma como $p(y|\theta_0, \rho)$. En este contexto, ρ se toma como proporcional a la varianza inversa, $\rho =$

$g(\theta_0)/\text{Var}(Y|\theta_0)$, para una función g . A veces, la precisión se toma de forma aleatoria, pero lo mejor es seleccionar una medida que sea fácilmente interpretable.

En ocasiones, se podrá asumir una distribución normal de los datos tal que

$$Y|\theta_0, \rho \sim N[\theta_0, g(\theta_0)/\rho].$$

Los límites de control se dibujarán, entonces, siguiendo la expresión

$$y_p(\theta_0, \rho) = \theta_0 + z_p \sqrt{g(\theta_0)/\rho},$$

donde z_p es aquel valor tal que $\Phi(z_p) = P(Z \leq z_p) = p$, para una variable normal estándar Z . Por lo tanto, los puntos representados en el gráfico de embudo estarán trazados por encima de los límites de control si $y_i \geq y_p$.

Los gráficos de embudo son preferibles a los gráficos comparativos porque, aunque ocurre lo mismo con la significación estadística de la clasificación, en el caso de los gráficos de embudo las unidades no se ordenan de mejor a peor o de forma cualitativa, como sí se suele hacer en las tablas de clasificación. La **Figura 3** muestra un ejemplo que propone Spiegelhalter (2002). En este gráfico, podemos ver los límites de control trazados con una confianza del 95% y del 99.8%, formando ese embudo dentro del cual se sitúa el porcentaje de muertes de cada hospital. Cada uno de los puntos representados en el gráfico corresponde al dato que se proporciona de cada hospital. El objetivo del porcentaje de muertes se sitúa en el 9.3%.

Figura 3: Gráfico de embudo para las muertes después de un tratamiento para la fractura de cadera.

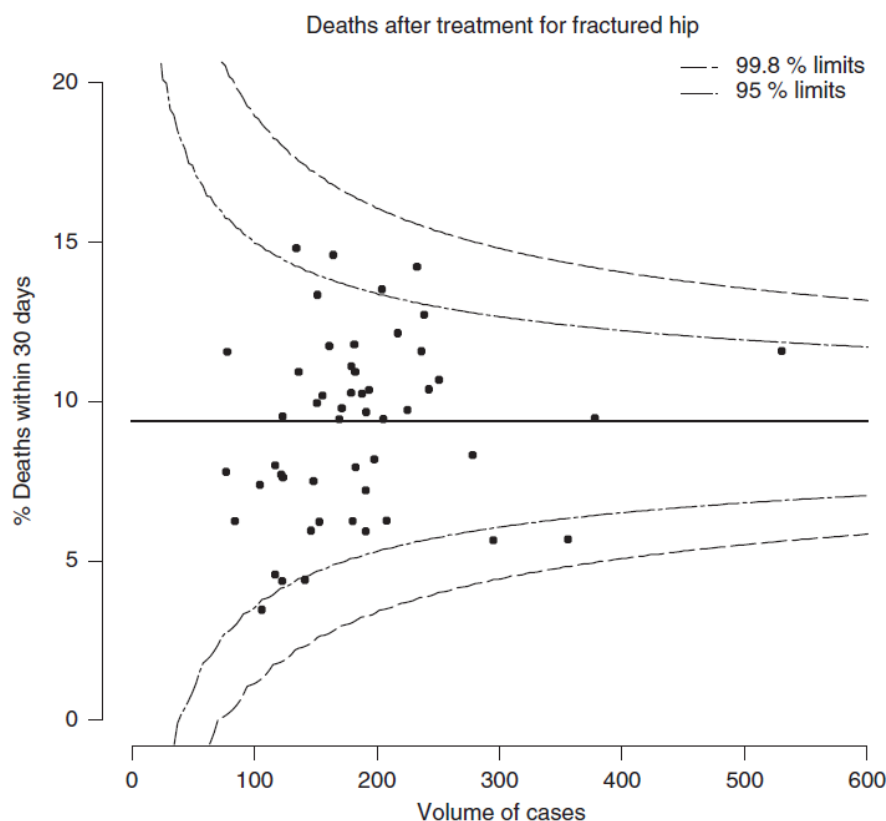


Gráfico de embudo de la tasa de mortalidad estandarizadas por edad y sexo a 30 días tras el tratamiento por rotura de cadera de mayores de 65 años en 51 hospitales y centros sanitarios de Inglaterra entre los años 2000 y 2001. Extraído de Spiegelhalter (2002).

2.3.3. Nuevas formas y métodos surgidos con el COVID-19

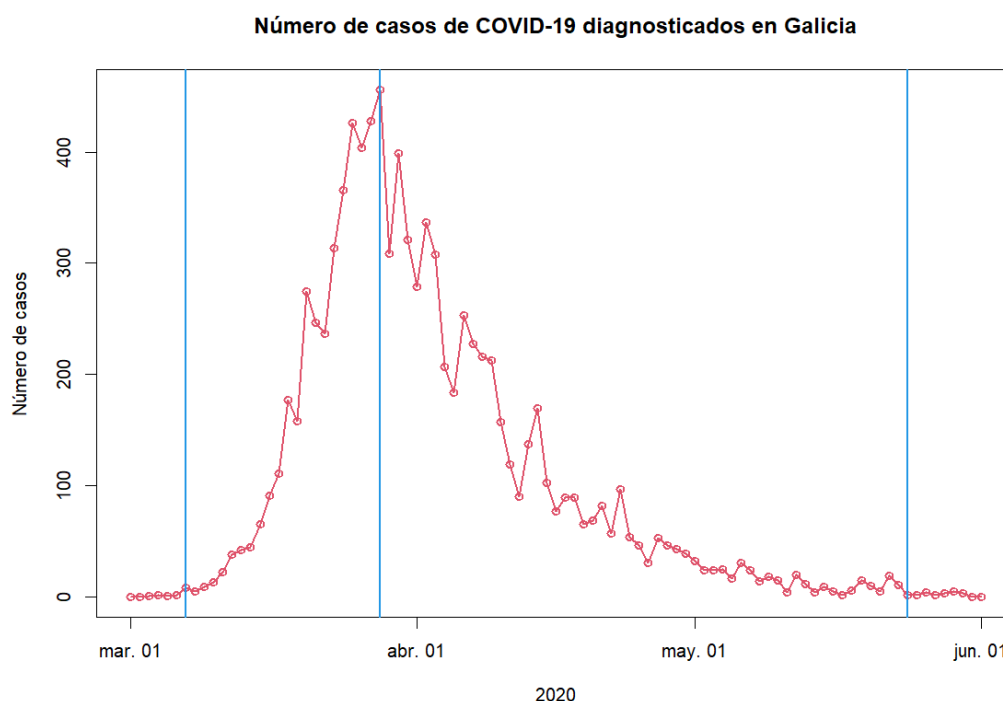
Con el auge del coronavirus, ha surgido la necesidad de crear nuevas formas y métodos para estudiar la evolución de este virus. Como hemos dicho en la introducción, la forma de comunicar los datos del coronavirus ha sido bastante ineficiente en el transcurso de la pandemia, es por ello por lo que algunos autores han dedicado esfuerzos a estudiar este tipo de datos y a crear nuevos métodos gráficos para su mejor comprensión, como veremos más adelante.

Algunos autores proponen el uso de algunos gráficos de control de Shewhart para controlar y monitorizar diferentes características del coronavirus. En el artículo de

Mandal et al. (2021) se estudia la variación del genoma del COVID-19 y se compara con otros virus de la familia de los coronavirus ya conocidos, como el SARS-Cov-1 o el MERS. Para ello utilizan gráficos de Shewhart y distinguen entre lo que llaman “hotspots” y “coldspots”, que son las zonas por encima del límite superior de control y las zonas más cercanas a la línea central de control, respectivamente. Fijándose en estas zonas en los gráficos con datos de los tres virus mencionados consiguen detectar alteraciones en el genoma del COVID-19, identificando, a su vez, posibles alteraciones que den lugar a nuevas variantes del virus. Estos gráficos, por tanto, son útiles para estudiar las alteraciones del virus y comprender mejor su comportamiento en el futuro.

Perla et al. (2021) proponen un gráfico de Shewhart híbrido para el estudio de las muertes diarias notificadas durante la pandemia del COVID-19. Este gráfico se basa en una curva epidemiológica típica, el número de muertes reportadas se representan en el eje vertical y alguna variable temporal (días, semanas, meses, etc.), en el eje horizontal. Este gráfico modela tres fases de la evolución de una epidemia: pre-crecimiento, crecimiento y post-crecimiento. Las podemos ver representadas en la **Figura 4**:

Figura 4: Fases de la evolución de la pandemia de COVID-19 en Galicia entre marzo y junio de 2020



Fuente: elaboración propia a través de los datos de la Red Nacional de Vigilancia Epidemiológica

En la parte izquierda del gráfico se representa la fase 1 que se corresponde con el pre-crecimiento, en la parte central tenemos la fase 2 de crecimiento de los contagios por COVID-19 y en la parte derecha tenemos la fase 3 de post-crecimiento o decrecimiento. Como explican los autores, estas fases de una epidemia serán modelizadas en el gráfico híbrido de Shewhart de forma diferente.

Existe también una cuarta fase del gráfico que se da cuando la etapa de decrecimiento comienza a estabilizarse en un mismo número de contagios o muertes, volviendo al nivel anterior al comienzo de la epidemia. Esta etapa se ve al final del gráfico de ejemplo (**Figura 4**), en el que se observa que el número de contagios por COVID-19 en Galicia vuelve a situarse en torno al 0.

Fase 1: crecimiento pre-exponencial

Esta fase comienza el primer día en el que ocurre el evento que se esté estudiando. En el caso de Perla et al. (2021), comienza el día en el que ocurre la primera muerte por COVID-19 mientras que, en el presente ejemplo, comienza el día en el que se da el primer caso de COVID-19 en Galicia (**Figura 4**). En esta fase, se enfocará el estudio en detectar cuándo el patrón de muertes o contagios sugiere que la epidemia está comenzando a crecer de forma exponencial.

Para ello, se representará en el gráfico de control el número de contagios o muertes y se calcularán los límites de control como en un gráfico C de Shewhart. La línea central del gráfico representa la media de eventos diarios y los límites de control se calculan de la forma típica en un gráfico C de control:

$$LCL = \bar{c} - 3\sqrt{\bar{c}}$$

$$CL = \bar{c}$$

$$UCL = \bar{c} + 3\sqrt{\bar{c}}$$

Estos límites servirán para detectar alguna señal de causa especial que puede dar comienzo a la fase 2 de la epidemia en la que la evolución empieza a ser exponencial.

La segunda fase del gráfico empezará cuando:

- Se da un total de al menos 8 observaciones sucesivas por encima o debajo de la línea central del gráfico de control.
- Se da una señal de causa especial, es decir, una observación se sale de los límites de control, lo cual nos indica un aumento grande de las observaciones.
- Tenemos una curva de regresión significativa y no negativa después de al menos 5 días de observaciones incluidas en el gráfico.

Fase 2: crecimiento exponencial

Esta fase comienza cuando el patrón del número de muertes o contagios deja de ser más o menos estable y se empieza a ver un crecimiento exponencial de las mismas. En esta fase se usa un gráfico I (gráfico X para medidas individuales o gráfico X-mr de rangos móviles) para monitorizar la fase de crecimiento de la epidemia y detectar alguna señal que nos indique que esta fase está terminando y los contagios o muertes comienzan a descender.

La fase 2 empieza el día en el que el gráfico C de la fase 1 indica una señal de causa especial de variación. Durante los 5 días (o unidades de tiempo que estemos utilizando) siguientes al comienzo de la fase se seguirán representando los datos como en el gráfico C e, inmediatamente después, se ajustará una regresión logarítmica del número de eventos diarios reportados (Y) sobre el número de días o unidades de tiempo en los que ocurre (X), empezando por los datos del día en el que finaliza la etapa de pre-crecimiento. Cuando el límite inferior de un intervalo con una confianza del 95% para la pendiente de regresión es no negativo, se empieza a modelizar la fase de crecimiento dibujando los eventos observados en este gráfico I, utilizando los coeficientes de regresión para el cálculo de la línea central y los límites de control. La línea central del gráfico de control I será la regresión lineal ajustada del logaritmo de los datos y los límites de control se construirán calculando el rango móvil medio de los residuos de la regresión ($M\bar{R}$) o, en su defecto, la mediana: $CL \pm 2.66 M\bar{R}$. Una vez calculados los límites y la línea central sobre los datos transformados logarítmicamente, se hará la transformación inversa para representarlos en el gráfico sobre los datos originales.

Cuando se observa que el crecimiento exponencial se mantiene por, al menos, 20 días (en general, después de 20 observaciones que crecen de forma exponencial) se “congelan” la línea central y los límites de control, es decir, se dejarán de incorporar

nuevos datos para el ajuste de la línea de regresión y los límites de control. Estos límites “congelados” se extienden en el gráfico como una guía para interpretar las observaciones posteriores que se representen en el gráfico. Las observaciones por encima del límite superior pueden indicar que la tasa de crecimiento exponencial está aumentando, mientras que las observaciones por debajo del límite inferior pueden indicar que la tasa de crecimiento está disminuyendo o, incluso, puede indicarnos el fin de la etapa de crecimiento exponencial para dar paso a la siguiente fase. El fin de esta fase se da, a diferencia de la fase 1, cuando tenemos 2 observaciones, como mínimo, fuera de los límites de control.

Fase 3: post-crecimiento y decrecimiento

Esta fase comienza cuando hay indicios de que ha terminado el crecimiento exponencial de los eventos y se espera que comiencen a descender y, posteriormente, a estabilizarse en un tiempo próximo.

En esta fase se representarán las observaciones en el gráfico y se actualizarán los límites de control. En esta fase es común encontrarse una zona en la que los eventos diarios se estabilicen antes de comenzar a descender. En este momento, la línea central será una línea horizontal y se calculará como la media de las observaciones. En cuanto los eventos diarios comiencen a descender, la línea central y los límites de control se calcularán basándose en la pendiente de la regresión, de manera similar a la fase de crecimiento exponencial, como en el gráfico I.

Esta etapa finalizará, al igual que la fase 2, cuando tenemos 2 observaciones fuera de los límites de control. La dirección que tomen esas observaciones nos indicará si volvemos a la fase 2 por un gran aumento de los contagios o muertes o si, en cambio, continúan descendiendo y estabilizándose, dando paso a una cuarta fase del gráfico.

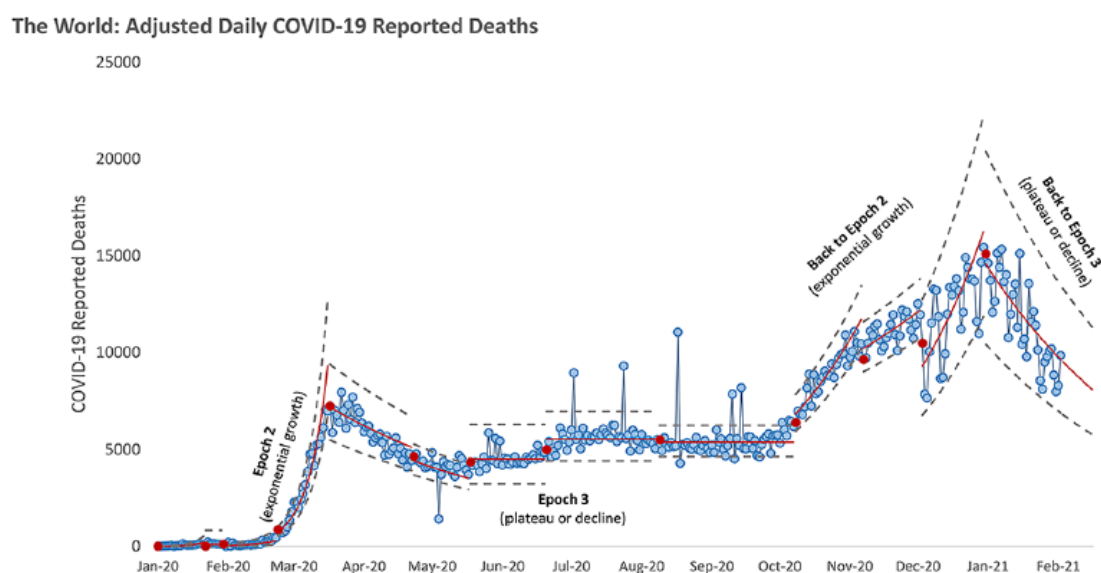
Fase 4: estabilidad después del descenso

Esta cuarta fase la explican estos mismos autores en un artículo posterior (Parry et al., 2021). Esta fase es la de estabilidad después del decrecimiento y se asemeja a la fase 1 en la que el número de eventos diarios es muy bajo y estable. Esta fase puede terminar si se detectan signos de causas especiales de variación en la que, de repente, los eventos diarios comienzan a aumentar y puede llegar a empezar, de nuevo, la fase 2 de crecimiento exponencial.

En la **Figura 5** se muestra un ejemplo del uso del gráfico híbrido de Shewhart para detectar las diferentes etapas de la epidemia de COVID-19 en el mundo. Se puede observar claramente el comienzo y el final de cada una de las etapas. La etapa de pre-crecimiento comienza en enero de 2020 al tener lugar la primera muerte por COVID-19. Después de varios días con muertes sucesivas por la enfermedad (con observaciones por encima de la línea central), el número de muertes comienza a crecer de forma exponencial dando lugar a la fase 2 durante marzo de 2020. En abril, el número de muertes comienza a descender, vemos en el gráfico varias observaciones por debajo de la línea central lo que nos indica un cambio de etapa. El número de muertes llega a estabilizarse durante el verano de ese mismo año. En esos meses se da la etapa 3 de decrecimiento. En este gráfico no se llega a ver la cuarta etapa de estabilidad porque el número de muertes no llega a números similares a la etapa 1. De hecho, la etapa 3 finaliza cuando el número de muertes comienza a ser, de nuevo, muy elevado por lo que se vuelve a la fase 2 de crecimiento exponencial. Después de esa etapa 2, comienzan a descender las muertes de nuevo, dando lugar a la fase de decrecimiento.

En la **Figura 5**, se pueden ver los diferentes límites de control correspondientes a cada una de las etapas de la epidemia, además de observarse cómo se van modificando según la etapa en la que nos encontremos.

Figura 5: Número diario de muertes reportadas por COVID-19 en el mundo entre enero de 2020 y febrero de 2021



Ejemplo del gráfico híbrido de Shewhart para el estudio de las diferentes fases de la epidemia de COVID-19 en el mundo. Extraído de Parry et al. (2021).

Este método sirve tanto para comunicar la evolución de la epidemia al público en general como para la toma de decisiones de las medidas de contención de las epidemias. Además, aunque en el ejemplo del artículo se utiliza para monitorizar las muertes del COVID-19, se podría utilizar para el estudio de otras variables epidemiológicas como los contagios, las hospitalizaciones o los ingresos en plantas UCI y UVI.

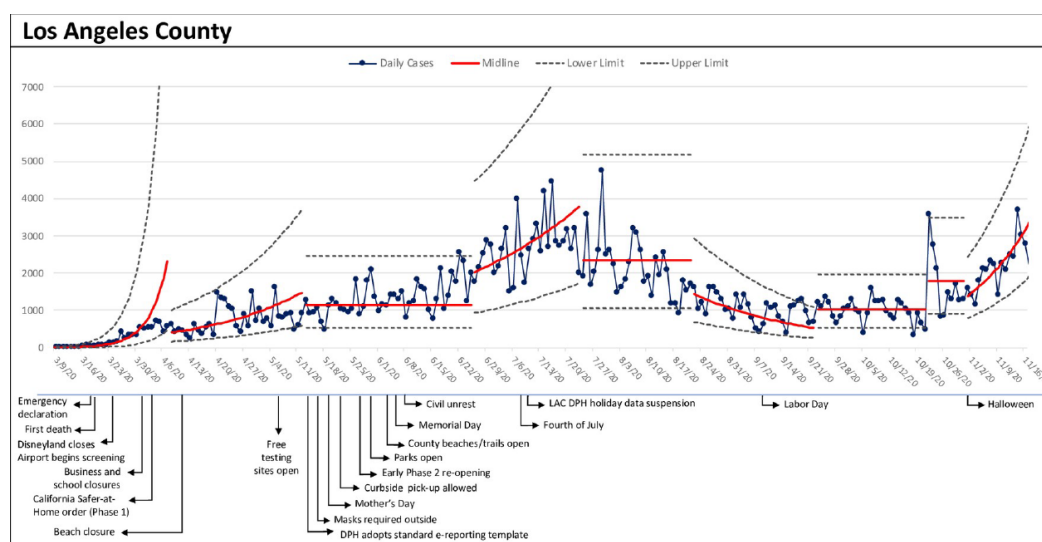
Este modelo de estudio de la COVID-19 permite detectar cualquier tipo de cambio significativo, ya sea bueno o malo, casi diariamente a medida que se van notificando los contagios. También permite detectar la fecha específica en la que se producen los cambios de fase de la epidemia, así como las épocas de transición entre cada una de ellas, lo que facilita la toma de decisiones que, en muchas ocasiones, tiene que ser muy rápida. No solo detecta los cambios propios de la evolución de una epidemia, sino que es capaz de detectar si las medidas tomadas para contener las enfermedades están teniendo algún efecto significativo en la evolución y cuándo se produce este efecto.

No obstante, este método depende de la disponibilidad de datos coherentes y fiables, por lo que a veces la variación puede ser diferente según el día de la semana y, por lo tanto, dar problemas. Tenemos como ejemplo la notificación del número de contagios en España durante el transcurso de los primeros meses de la pandemia. Durante esos meses se publicaron datos diarios sobre el número de contagios, muertes, hospitalizaciones y otras variables epidemiológicas, a excepción de los fines de semana y los días festivos, cuando no se notificaban datos. Esto implicó que, al volver a recogerse y notificarse datos después de varios días sin notificar nada, se produjeran valores excesivamente altos o excesivamente bajos. Es habitual, en estos casos, eliminar las causas de variación especiales de los fines de semana o similares para suavizar la variación. Además, también es necesario que este método sea aplicado por expertos en la materia y que se conozca la forma de notificar los datos.

Otros autores (Inkelas et al., 2021) han utilizado el gráfico híbrido de Shewhart creado por Perla et al. (2021) para estudiar la evolución de la infección por COVID-19 en algunas regiones de California. En el siguiente ejemplo extraído de dicho artículo (**Figura 6**) podemos ver las diferentes etapas de la pandemia del COVID-19 en el caso del condado de Los Ángeles. Se ve claramente cómo, en ocasiones, al finalizar una

etapa, puede ocurrir que se vuelva a la etapa anterior. Es decir, se puede pasar de una etapa estable en cuanto a los contagios y volver, de nuevo, a la etapa anterior de crecimiento exponencial. Esto ocurrió en todo el mundo durante la pandemia, surgieron diferentes olas de la enfermedad causadas por diferentes eventos: vacaciones, el invierno que se junta con el punto álgido de otras enfermedades como la gripe, etc. Debajo del eje horizontal del gráfico vemos destacadas algunos hechos importantes en el transcurso de la pandemia, como por ejemplo una cuarentena domiciliar o los cierres de escuelas, además de destacarse también algunas festividades. Esto nos ayuda a ver de qué manera funcionaron las medidas de contención de la pandemia y cómo algunas festividades en las que la población tiende a juntarse más a menudo pueden provocar el comienzo de una nueva etapa de la enfermedad.

Figura 6: Número de contagios diarios por COVID-19 en Los Ángeles, EEUU, entre marzo de 2020 y noviembre de 2020



Ejemplo del gráfico híbrido de Shewhart para estudiar la evolución y las diferentes etapas de la epidemia en el condado de Los Ángeles, EEUU. Extraído de Inkelas et al. (2021).

Estos autores concluyen que este gráfico de control híbrido puede ayudar a los responsables políticos y las autoridades sanitarias a actuar con rapidez. Además, con el uso de este método se puede reducir la sobre-reacción de la población y de los medios de comunicación a la variación aleatoria esperada de la evolución de la epidemia y fomentar la rápida toma de decisiones cuando esta variación indica un cambio de fase en la evolución de la misma.

Capítulo 3

Caso de estudio de la zona metropolitana de A Coruña

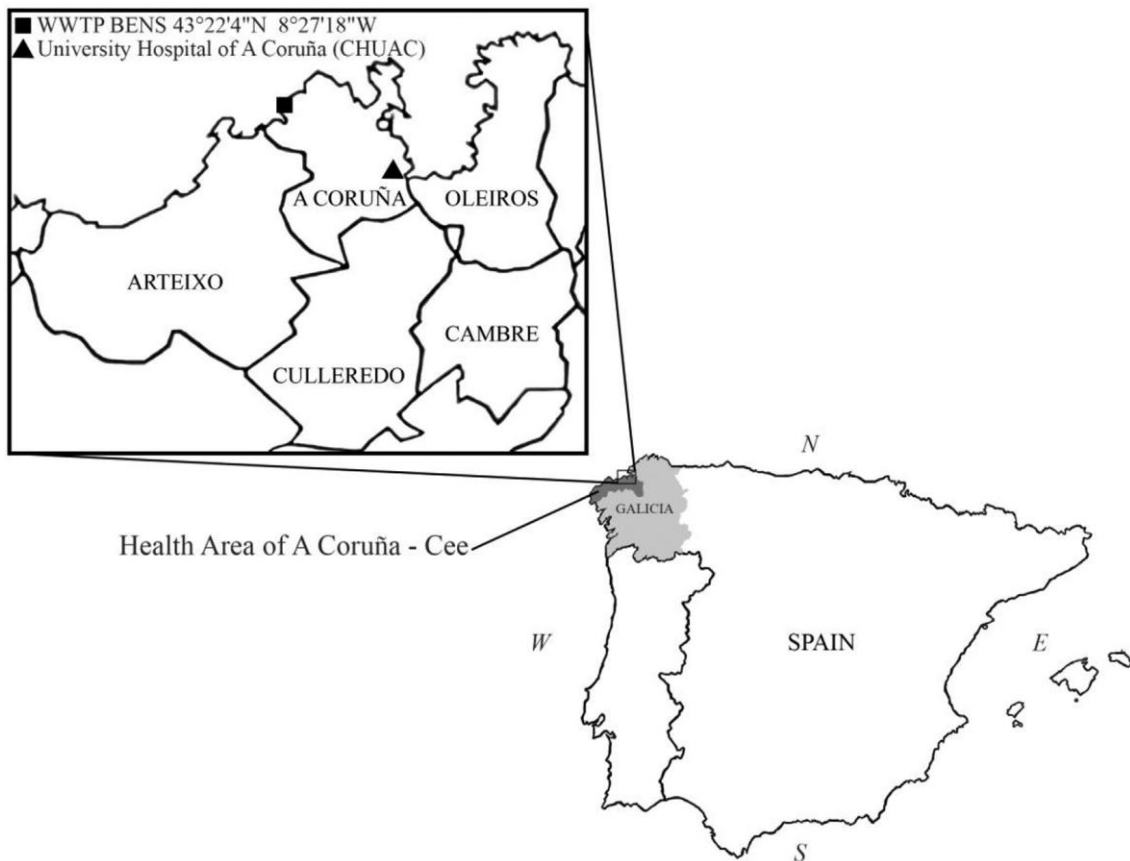
3.1. Proyecto COVIDBENS

En los últimos años, se han realizado muchos estudios acerca de la transmisión del COVID-19 y dónde se podían encontrar copias del virus. Se estudió la deposición del virus en alimentos y todo tipo de superficies (teléfonos, teclados y otros objetos de uso diario en oficinas, asientos del transporte público, aviones, etc.) con el fin de entender la forma de propagación del virus y la incidencia real de la enfermedad en el planeta (Goldman, 2020; Lema, 2020). También se estudiaron los residuos y las aguas residuales al descubrirse que muchos pacientes contagiados (asintomáticos y sintomáticos) daban positivo en pruebas fecales de ARN, a pesar de dar negativo en los test nasofaríngeos (Vallejo et al., 2022). Además, estos pacientes comenzaban a dar positivo en los primeros momentos de la enfermedad y hasta un mes después de haberse contagiado. Esto da una gran importancia a este tipo de datos, al poder adelantarse a la evolución de la enfermedad y a cualquier brote o nueva ola (medida en términos de número de casos de SARS-CoV-2 activos, por ejemplo).

En esto consistió el trabajo que se llevó a cabo en el proyecto **COVIDBENS** en la Estación Depuradora de Aguas Residuales (EDAR) de Bens a partir de abril del año 2020 (COVIDBENS, 2022a). El equipo de COVIDBENS está formado por investigadores de diferentes disciplinas (microbiología, química, estadística y diferentes ingenierías como la informática o la industrial) de la Universidad de A Coruña (UDC) y otros centros de investigación asociados a la universidad como son el Centro de Investigaciones Científicas Avanzadas (CICA), el Centro de Investigación TIC (CITIC), el Centro de Innovación Tecnológica en Edificación e Ingeniería Civil (CITEEC) y el Instituto de Investigación Biomédica de A Coruña (INIBIC – CHUAC), así como profesionales que pertenecen directamente a Edar Bens S.A. En este proyecto se analizaron las aguas residuales de todo el área metropolitana de A Coruña, que incluye los municipios de A

Coruña, Oleiros, Culleredo, Cambre y Arteixo, con el objetivo de analizar la cantidad de copias del COVID-19 en dicha área. Esta área comprende un total de 277.8 km^2 y 400,000 habitantes, aproximadamente.

Figura 7: Mapa del área metropolitana de A Coruña



Área metropolitana de A Coruña, incluyendo los municipios de A Coruña, Arteixo, Cambre, Culleredo y Oleiros, el área sanitaria de A Coruña – Cee, así como la localización específica de Edar Bens y el Complejo Hospitalario Universitario de A Coruña (CHUAC). Extraído de Vallejo et al. (2022).

El éxito del proyecto recae en la alerta temprana al conseguir la detección de nuevos brotes de coronavirus antes de que los datos oficiales de nuevos contagios proporcionados por el sistema sanitario notificasen esas alarmas. El proyecto demostró ser capaz de adelantarse hasta 18 días con respecto a los datos oficiales reportados por las autoridades sanitarias. Los modelos estadísticos epidemiológicos diseñados en exclusiva en el proyecto de **COVIDBENS** consiguieron estimar el número de portadores

de coronavirus, ya sean sintomáticos o asintomáticos, únicamente mediante el análisis de la cantidad de material genético del virus presente en las aguas residuales, siendo un proyecto pionero en el mundo.

Se estimó el número de contagios ya fueran sintomáticos o asintomáticos, a diferencia de los estudios del sistema sanitario que detectaban la gran mayoría de los contagios por aquellos pacientes que acuden a centros sanitarios al tener síntomas compatibles y realizarse un test, por cribados aleatorios para detectar asintomáticos, con el inconveniente de que al ser voluntario no todos los llamados al cribado acudían. En la actualidad se están detectando una gran parte de los contagios por las pruebas de autodiagnóstico que la población puede realizar en la comodidad de sus domicilios o en farmacias con la ayuda del personal de farmacia. Sin embargo, estas pruebas no son del todo fiables ya que al tomarse la muestra por el propio usuario puede estar tomándose de forma incorrecta, dando lugar a falsos negativos, interpretaciones incorrectas, falsos positivos e, incluso, podría tomarse la prueba antes de que la carga viral sea lo suficientemente alta como para detectarse, dando lugar a negativos en personas que sí son portadoras del virus (Casado, 2022; Llaneras et al. , 2022).

Además, el equipo diseñó un modelo estadístico que permitió determinar la frecuencia de las variantes y mutaciones del coronavirus (por ejemplo, DELTA u ÓMICRON) que se encontraban en la población (COVIDBENS, 2022b).

Figura 8: Porcentaje de las variantes encontradas en las muestras de aguas residuales

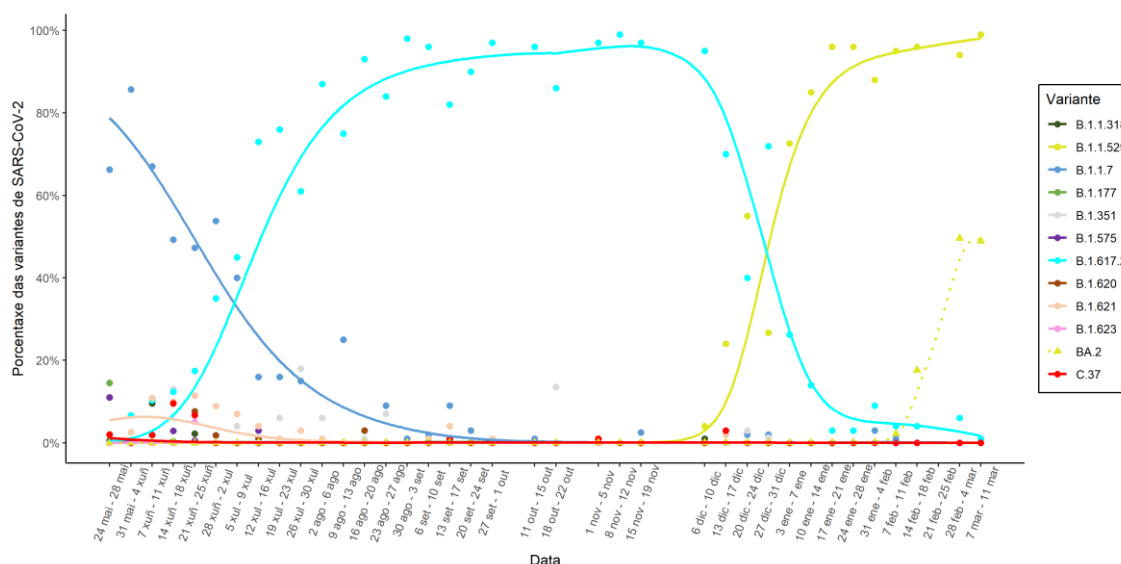


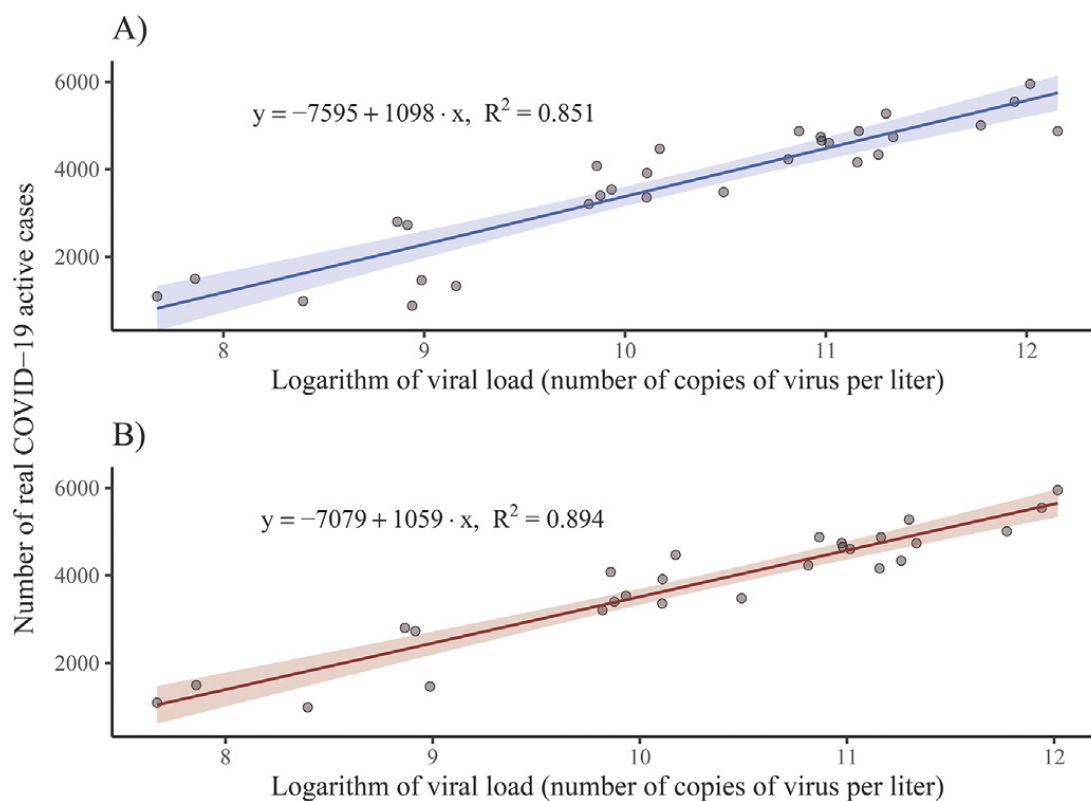
Gráfico del porcentaje de variantes de coronavirus encontradas en las aguas residuales de la comarca de A Coruña entre mayo de 2021 hasta el 11 de marzo de 2022. Extraído de la página web del proyecto COVIDBENS: <https://edarbens.es/covid19/variantes-encontradas/>

Para recoger los datos, se instalaron unos sistemas automáticos que obtienen muestras de las aguas residuales en la propia planta de EDAR Bens y en el Hospital Universitario de A Coruña, en las siete plantas específicas que se utilizaron para acoger a los pacientes de COVID-19 (Vallejo et al., 2022). Una vez obtenidas las muestras, se extrajo el ARN de las muestras y se determinó aquel que provenía del SARS-CoV-2. Se cuantificó la cantidad de ARN. Para complementar estos datos y crear el modelo estadístico, se obtuvieron otros datos ofrecidos por el Servicio Gallego de Salud (SERGAS), el Complejo Hospitalario Universitario de A Coruña (CHUAC) y la Dirección General de la Función Pública de Salud de la Xunta de Galicia, así como otros datos meteorológicos obtenidos de MeteoGalicia. A partir del tratamiento de estos datos, se creó un modelo estadístico en función de la carga viral medida en las aguas residuales, el caudal diario en la red de alcantarillado y otras variables ambientales tales como la lluvia, la temperatura o la humedad. Se obtuvo un modelo lineal múltiple que utilizaba la transformación logarítmica de la carga viral, el flujo diario de agua, las precipitaciones, la temperatura y la humedad. Este modelo se pudo simplificar en un modelo lineal simple en función del logaritmo de la carga viral ya que el R^2 no mejora de forma significativa al añadir las variables flujo, humedad, precipitaciones y temperatura. Solamente el logaritmo natural de la carga viral ajusta un buen modelo lineal ($R^2 = 0.851$). Después de eliminar tres valores atípicos, el ajuste mejoró un poco, obteniendo un $R^2 = 0.894$. El modelo lineal ajustado utilizado para predecir el número de casos activos por COVID-19 pasó a ser:

$$Y = -7079 + 1059 \ln X$$

En donde Y denota el número real de casos activos por COVID-19 y X, la carga viral (número de copias de ARN por litro). Podemos ver representados los ajustes de estos modelos en la **Figura 9**.

Figura 9: Estimación del número real de casos activos de COVID-19 mediante dos modelos de regresión lineal.



Gráficos de dispersión que representan el logaritmo de la carga viral medida en Edar Bens y el número de casos reales de COVID-19 estimados mediante el modelo de regresión lineal. En la figura (A) se representa el primer modelo estimado y en la figura (B) se representa el modelo mejorado después de eliminar los tres valores atípicos detectados. En los gráficos se incluye el ajuste lineal y el intervalo de confianza (zona sombreada alrededor de los ajustes). Extraído de Vallejo et al. (2022).

Este modelo ha permitido hacer predicciones del desarrollo de la pandemia con una antelación de hasta 18 días (Vallejo et al., 2022).

3.2. Descripción de los datos

Para el presente trabajo haremos uso de un conjunto de datos cedido por COVIDBENS que contiene 184 observaciones sobre 8 variables: la fecha en la que se realiza cada medición (**Fecha**), la carga vírica medida en las aguas residuales que se corresponde con el número de copias de ARN del virus por litro de agua residual estudiada (**Carga_medida**), carga vírica estimada (**Carga_estimada**), número de casos diarios

que son el número de casos activos por día según los datos reportados por el Sergas (**Casos**), número de vacunados (**Vacunados**) y porcentaje de vacunados (**Vacunados_pct**) entre el 19 de julio de 2020 y el 9 de marzo de 2022.

```
'data.frame': 184 obs. of  8 variables:
 $ Fecha          : Date, format: "2020-07-19" "2020-07-21" "2020-07-23" ...
 $ Carga_medida   : num  6787 12920 7339 10520 22803 ...
 $ Carga_estimada: num  1120 2120 3970 9790 17100 28200 54400 78400 105000 146000
 ...
 $ Casos          : int  358 1034 1699 2656 3246 3777 4473 4860 5170 5516 ...
 $ Lim_inferior   : int  NA 21 718 1707 2308 2843 3534 3915 4216 4552 ...
 $ Lim_superior   : int  1413 2048 2680 3604 4184 4711 5411 5806 6123 6481 ...
 $ Vacunados      : int   1 1 1 1 1 1 1 1 1 1 ...
 $ Vacunados_pct  : chr   "0%" "0%" "0%" "0%" ...
```

En este trabajo nos centraremos, sobre todo, en la información aportada por las variables de la carga vírica medida y el número de casos activos. Debemos fijarnos en que los datos no son diarios, sino que las muestras han sido tomadas durante 3 días a la semana en el transcurso del proyecto. Además, los datos de la carga vírica se corresponden con el valor medio obtenido a partir de seis muestras tomadas en esos 3 días de agua residual.

3.3. Gráficos de control para el estudio del COVID-19 con datos de COVIDBENS

Para comenzar con el estudio, utilizaremos los datos de las variables “Casos” y “Carga_medida” para construir algunos de los gráficos de control habituales de los que se han introducido en el apartado anterior del presente trabajo de fin de máster.

Estamos ante dos variables que, dada su naturaleza, encajan en el estudio de forma individual, a diferencia de otros estudios con variables que definen procesos industriales en donde se pueden obtener varias mediciones en un día dependiendo del material de medición o de las condiciones meteorológicas, por ejemplo. En este caso y al tener variables de tipo epidemiológico, las variables se construyen realizando una medición en cada uno de los días estudiados, ya sea del número de casos activos o la carga vírica

medida en las aguas residuales. Es por ello por lo que, en lugar de realizar un gráfico de medias común se ha decidido estudiar un gráfico de medias para datos de medidas individuales.

Figura 10: Gráfico de medidas individuales para la variable “Número de casos activos”

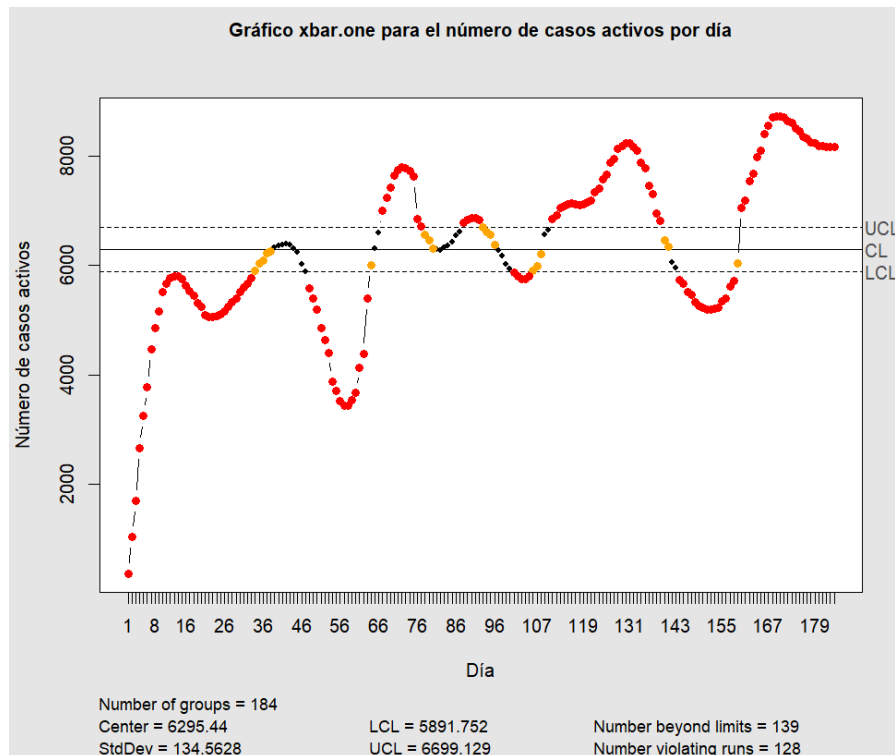


Gráfico de medidas individuales (xbar.one) obtenido mediante la aplicación de la función “qcc” del paquete “qcc” del software estadístico R, a partir de los datos cedidos por COVIDBENS, en concreto el número de casos activos.

Vemos cómo el gráfico de medidas individuales (**Figura 10**) detecta que la mayoría de las observaciones se encuentran fuera de los límites de control. Si nos fijamos en los puntos y en su evolución, este gráfico no nos resulta útil para el presente caso de estudio dado que se observa una fuerte dependencia entre las diferentes observaciones (autocorrelación), por lo que no se cumple una de las hipótesis de partida para el correcto uso de este tipo de gráficos de control. De hecho, la variabilidad está infraestimada, al no tenerse en cuenta la dependencia entre observaciones, por lo que los límites de control calculados son más estrechos de que deberían ser, dando lugar a un gran número de falsas alarmas. Desde otra perspectiva, se puede argumentar que en el periodo estudiado se alternan, sucesivamente, procesos de crecimiento,

decrecimiento y estabilización del número de casos activos, es decir, existen muy diferentes poblaciones de las que provienen los datos observados.

Acto seguido, se muestran los resultados obtenidos mediante la aplicación de los gráficos EWMA (**Figura 11**) y CUSUM (**Figura 12**).

Figura 11: Gráfico EWMA para la variable “Número de casos activos”

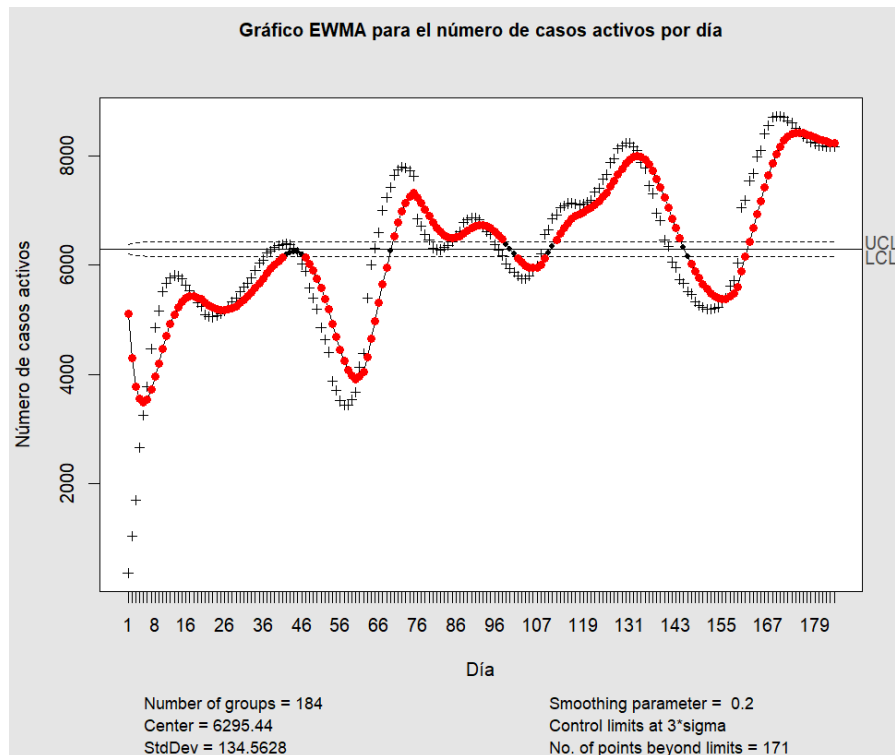


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R, a partir de los valores del “Número de casos activos”, pertenecientes al conjunto de datos cedidos por COVIDBENS.

Lo mismo ocurre en el caso de los gráficos EWMA (**Figura 11**) y CUSUM (**Figura 12**), en los cuales prácticamente todas las observaciones se encuentran fuera de los límites de control representados en el gráfico.

Viendo estos últimos gráficos en conjunto, como ya se ha comentado, es probable que las observaciones de la variable número de casos activos presenten una alta autocorrelación, razón por la cual no estaríamos obteniendo unos resultados fiables a la hora de construir los gráficos de control. Para contrastar si la variable número de casos

presenta autocorrelación, hemos realizado un contraste de **Ljung-Box** con un retardo, el cual nos muestra el siguiente resultado:

```
> Box.test(Casos, lag = 1, type = "Ljung-Box")
```

```
Box-Ljung test
```

```
data: Casos
```

```
X-squared = 164.16, df = 1, p-value < 2.2e-16
```

Se rechaza la independencia, por lo tanto, estaremos ante una variable con autocorrelación, de al menos, un retardo. Para corregir esto, primero construiremos un modelo ARIMA, tal y como se propone en Barbeito et al. (2017), con el cual se obtenga un buen ajuste de los datos de los que disponemos para, acto seguido, construir, de nuevo, los gráficos de control, pero esta vez para los residuos obtenidos del modelo ARIMA ajustado, en donde podremos estudiar si se detectan o no olas y/o brotes de la enfermedad.

Figura 12: Gráfico CUSUM para la variable “Número de casos activos”

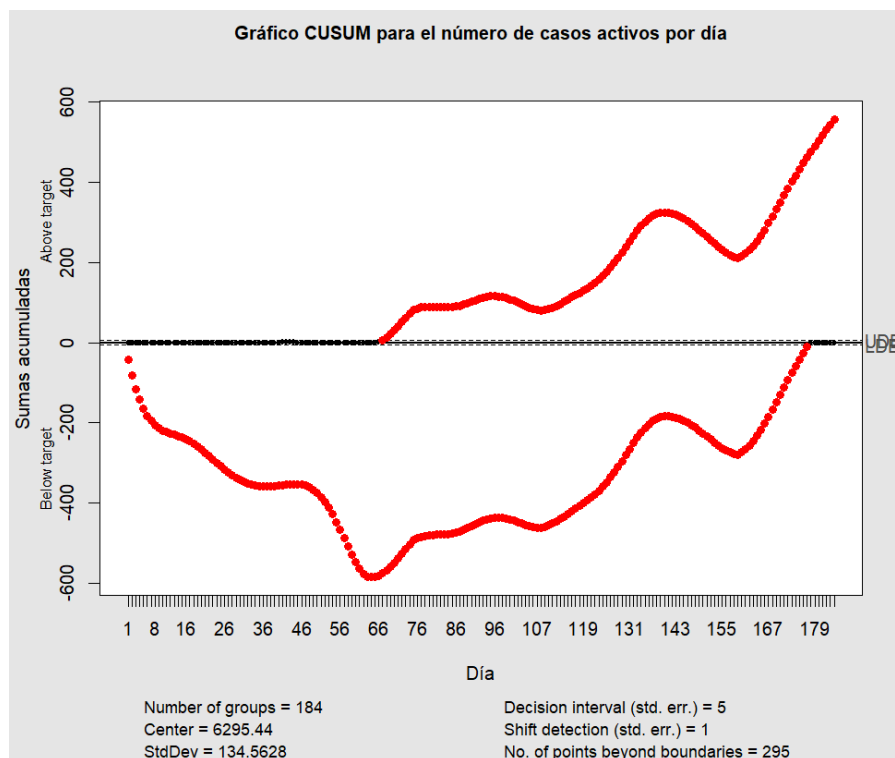


Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de la variable “Número de casos activos” del conjunto de datos cedidos por COVIDBENS.

Para construir el modelo ARIMA más adecuado, utilizaremos la función “auto.arima” del paquete “forecast” de R, la cual seleccionará el ajuste más adecuado para la serie de tiempo que estemos utilizando. En el caso de la variable “Casos”, la función nos selecciona un modelo **ARIMA(2,1,2)** sin constante y con una transformación Box Cox previa de los datos.

```
Best model: ARIMA(2,1,2)

Series: casos_ts

ARIMA(2,1,2)

Box Cox transformation: lambda= 1

Coefficients:

          ar1      ar2      ma1      ma2
      1.8575  -0.9102  -1.4986   0.6833
s.e.  0.0490   0.0486   0.0762   0.0688

sigma^2 estimated as 19114:  log likelihood=-1160.78

AIC=2331.57  AICc=2331.91  BIC=2347.61
```

Después de realizar la diagnosis del modelo y concluir que este modelo proporciona un buen ajuste de nuestros datos, construimos los gráficos de control haciendo uso de los residuos obtenidos con este modelo.

Figura 13: Gráfico de medidas individuales para los residuos de la variable “Número de casos”

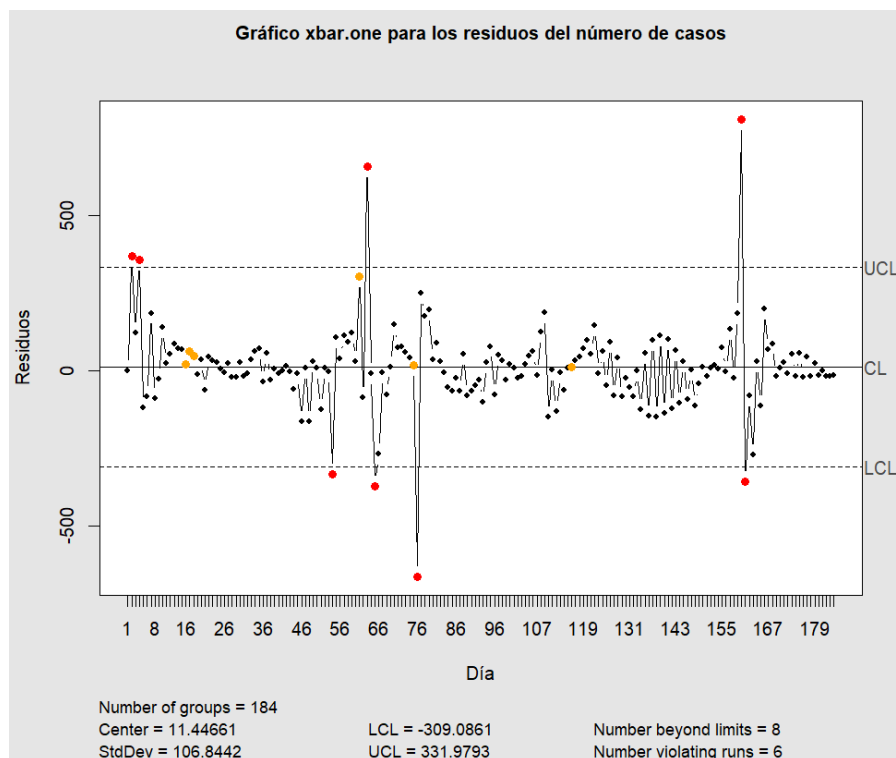


Gráfico de medidas individuales (xbar.one) elaborado con la función “qcc” del paquete “qcc” de R con los datos de los residuos obtenidos con el ajuste del modelo ARIMA(2,1,2) para la variable “Número de casos” de los datos cedidos por COVIDBENS.

Como se observa en la **Figura 13**, mediante el ajuste del modelo ARIMA, se ha conseguido eliminar en gran medida, la autocorrelación existente entre las observaciones. En este caso, únicamente 8 de los residuos correspondientes a observaciones del número de casos se encuentran fuera de los límites de control. Podemos comprobar cuáles son los residuos que caen fuera de los límites (**beyond.limits**), así como aquellos en los que se produce algún cambio de racha (**violating.runs**):

```
$beyond.limits
```

```
[1] 2 4 63 160 54 65 76 161
```

```
$violating.runs
```

```
[1] 16 17 18 61 75 116
```

Lo mismo ocurre con el gráfico EWMA (**Figura 14**), solamente 5 observaciones caen fuera de los límites de control.

Figura 14: Gráfico EWMA para los residuos de la variable “Número de casos”

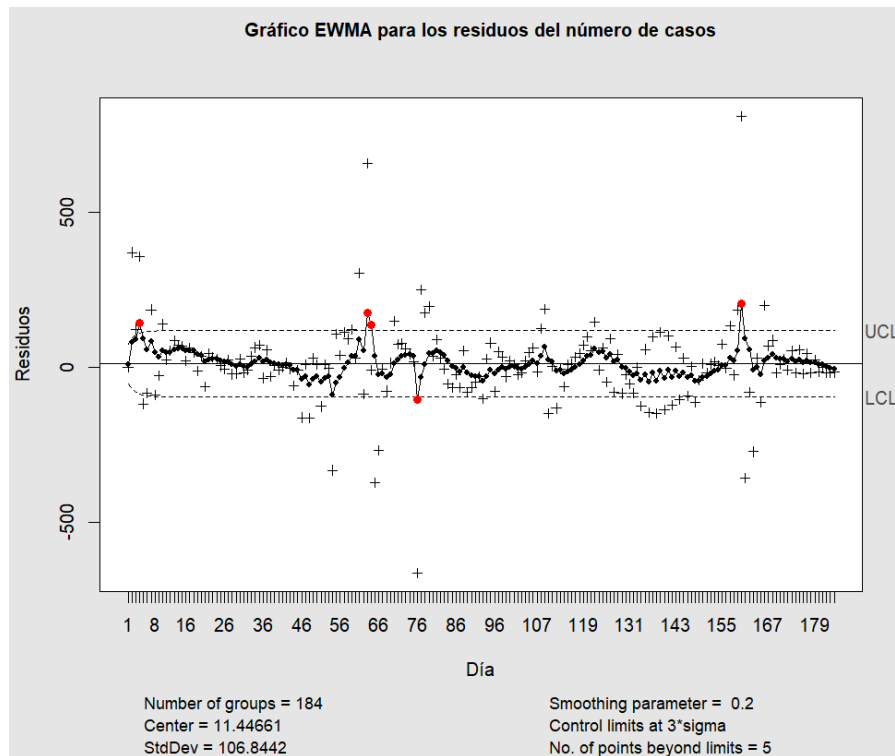


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R. Elaborado con datos sobre los residuos del “Número de casos activos” del conjunto de datos cedidos por COVIDBENS.

Estos puntos que caen fuera de los límites corresponden a los residuos de las observaciones:

```
4 63 64 76 160
```

En el caso del gráfico CUSUM (**Figura 15**), tenemos 9 observaciones que se encuentran fuera de los límites de control del gráfico, tanto del límite superior como del inferior. Estas observaciones son las siguientes:

```
$lower
```

```
[1] 66 68 76 163 165
```

\$upper

[1] 4 63 64 160

Figura 15: Gráfico CUSUM de los residuos del “Número de casos activos”

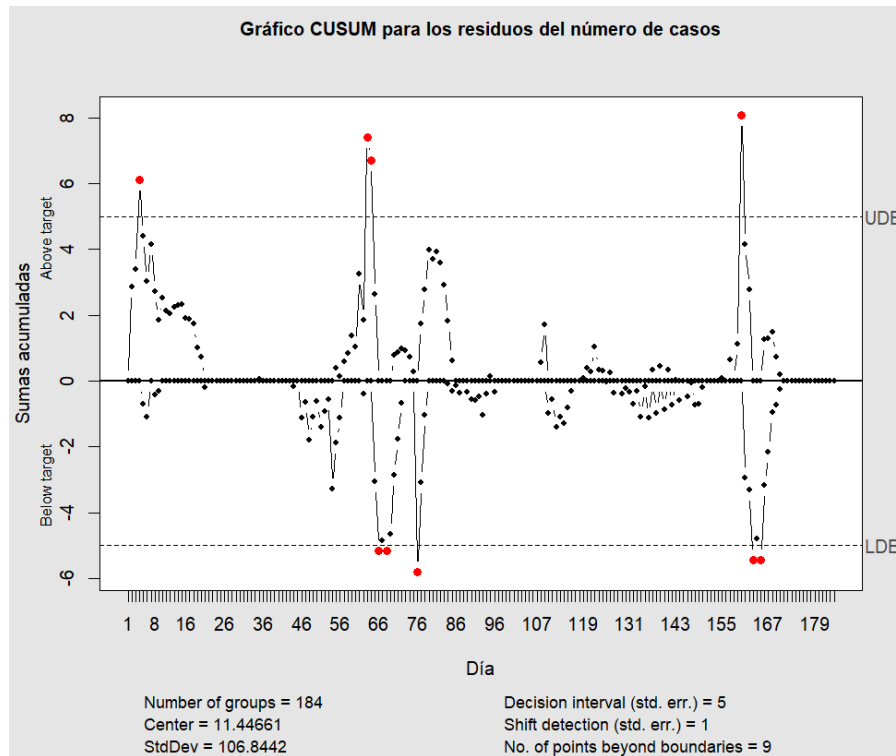


Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de los residuos del “Número de casos activos” del conjunto de datos cedidos por COVIDBENS.

Las diferencias que hemos podido encontrar entre la aplicación de gráficos Shewhart y EWMA o CUSUM se deben, en parte, a que los primeros están diseñados para detectar cambios grandes, mientras que los otros son especialmente eficientes para detectar cambios pequeños.

Para ver mejor si estas observaciones corresponden al inicio de algún brote o comienzo de una nueva ola de coronavirus, podemos plasmar dichas observaciones en el gráfico de dispersión de la variable “Número de casos”. Eso lo vemos en la **Figura 16**.

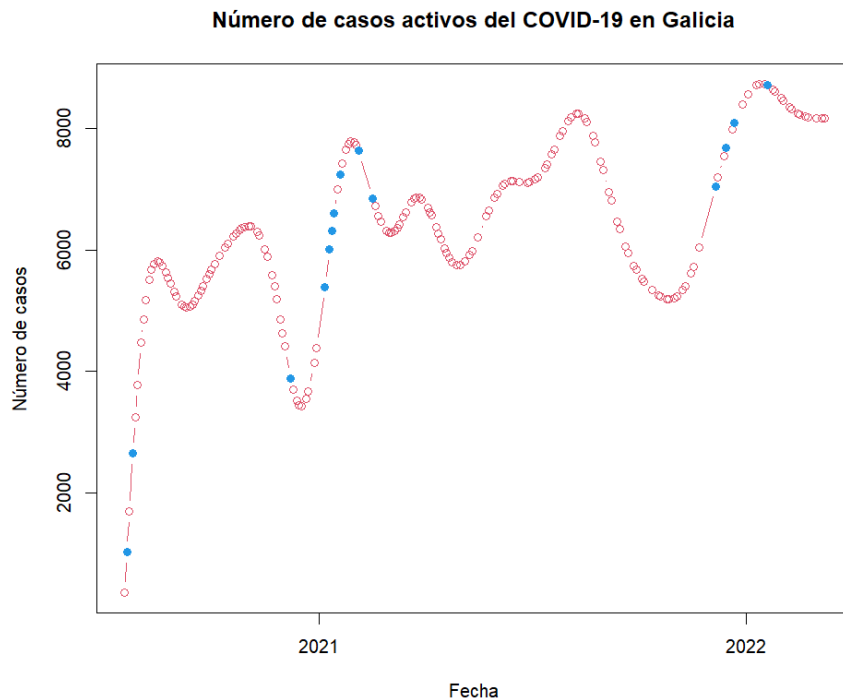
Figura 16: Gráfico de dispersión de la variable “Número de casos”

Gráfico de dispersión de la variable “Número de casos activos”. En azul, podemos ver los puntos que corresponden a aquellos que caen fuera de los límites en los gráficos de control en los gráficos de medias para medidas individuales, EWMA y CUSUM de los residuos del modelo ajustado a esta variable. Elaborado con el conjunto de datos cedidos por COVIDBENS.

Si nos fijamos en los puntos azules representados en el gráfico (**Figura 16**), estos parecen detectar algún cambio de tendencia. Sin embargo, esto no está claro. Si nos fijamos en los puntos rojos, el número de casos va fluctuando a lo largo del tiempo y los puntos azules no están detectando el comienzo ni el final de los brotes de coronavirus que vemos representados en el gráfico.

Anteriormente, se han considerado todos los datos como una muestra retrospectiva o de calibrado, es decir, se han aplicado gráficos de control para la Fase I del control estadístico de procesos. En una segunda fase del control estadístico de procesos, cogeremos una muestra inicial de calibrado para calcular los límites de control de cada uno de los gráficos de control que se han utilizado. Una vez calculados esos límites de control, se representarán en el gráfico junto con la muestra de calibrado y las observaciones posteriores para detectar qué puntos caen fuera de dichos límites

iniciales. Para el caso del número de contagios por COVID-19, utilizaremos, como hemos hecho en estos últimos gráficos, los residuos del modelo ARIMA que ajusta nuestros datos. Como muestra de calibrado se utilizarán las primeras 60 observaciones (los dos primeros meses) por tratarse de una muestra bastante estable.

Figura 17: Gráfico de medidas individuales con muestra de calibrado

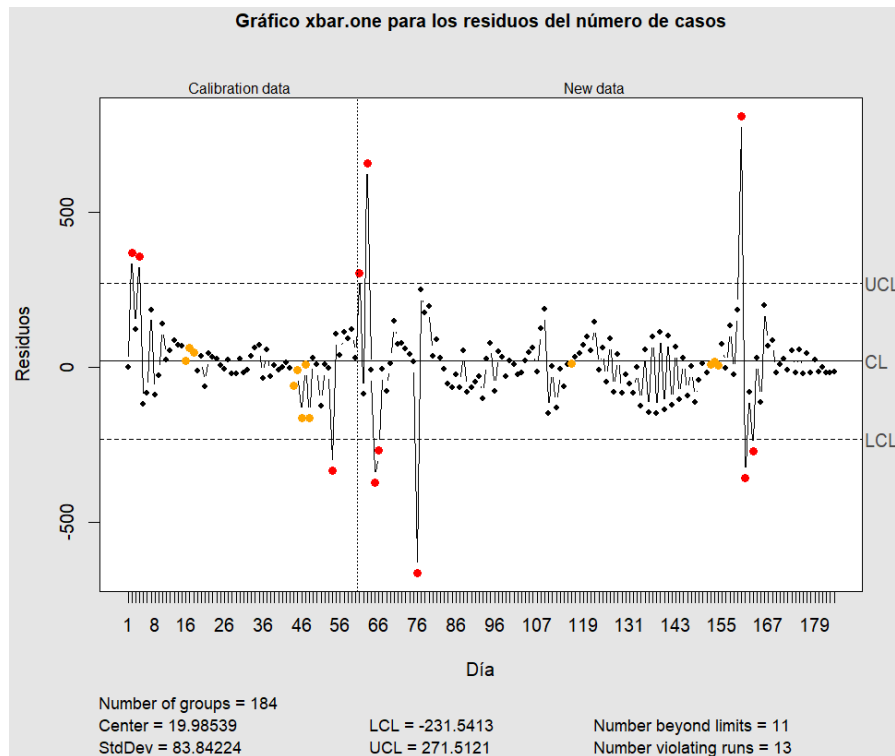


Gráfico de medidas individuales (xbar.one) elaborado con la función “qcc” del paquete “qcc” de R con los datos de los residuos obtenidos con el ajuste del modelo ARIMA(2,1,2) para la variable “Número de casos” de los datos cedidos por COVIDBENS. La muestra de calibrado comprende las 60 primeras observaciones.

Utilizando los límites calculados con las 60 primeras observaciones aparecen nuevos puntos fuera de los límites de control como podemos observar en la **Figura 17**. Estos puntos corresponden a las observaciones:

```
$beyond.limits
```

```
[1] 2 4 61 63 160 54 65 66 76 161 163
```

```
$violating.runs
```

```
[1] 16 17 18 61 44 45 46 47 48 116 152 153 154
```

En el gráfico EWMA (**Figura 18**), también obtenemos nuevos puntos que caen fuera de los límites de control. Estos puntos se corresponden con las observaciones:

2 3 4 54 64 65 77 161

Figura 18: Gráfico EWMA para los residuos de la variable “Número de casos” con muestra de calibrado

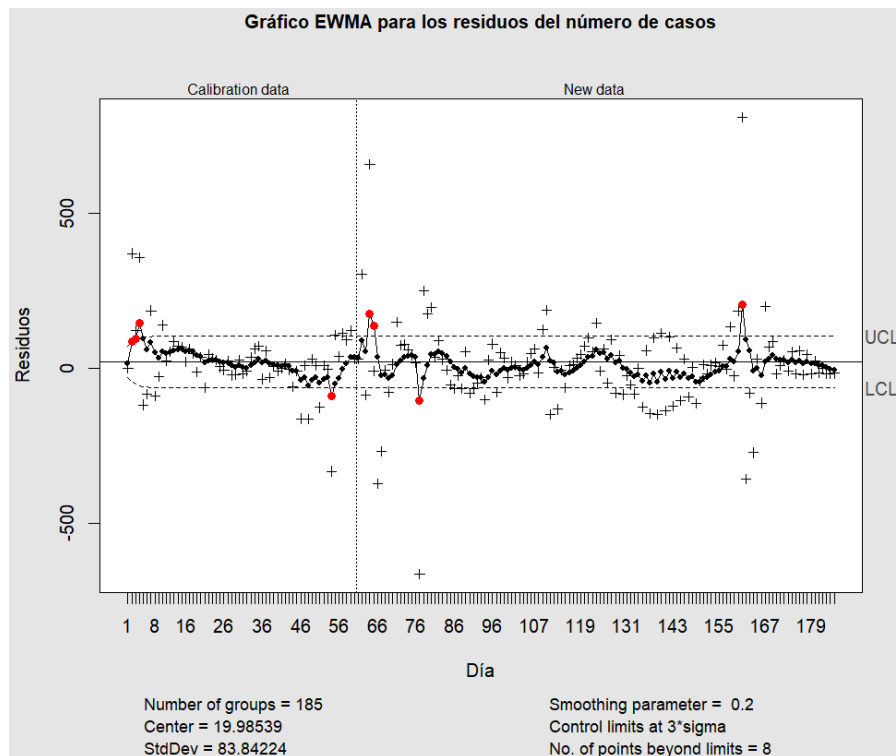


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R. Elaborado con datos sobre los residuos del “Número de casos activos” del conjunto de datos cedidos por COVIDBENS. La muestra de calibrado comprende las 60 primeras observaciones.

Al igual que en el caso de los gráficos de medias y EWMA con una muestra de calibrado, el gráfico CUSUM (**Figura 19**) nos muestra más puntos fuera de los límites de control. Estos puntos son:

\$lower

[1] 54 55 67 68 69 70 71 77 164 165 166 167

\$upper

[1] 4 5 7 64 65 80 82 161 162

Figura 19: Gráfico CUSUM para los residuos del “Número de casos” con muestra de calibrado

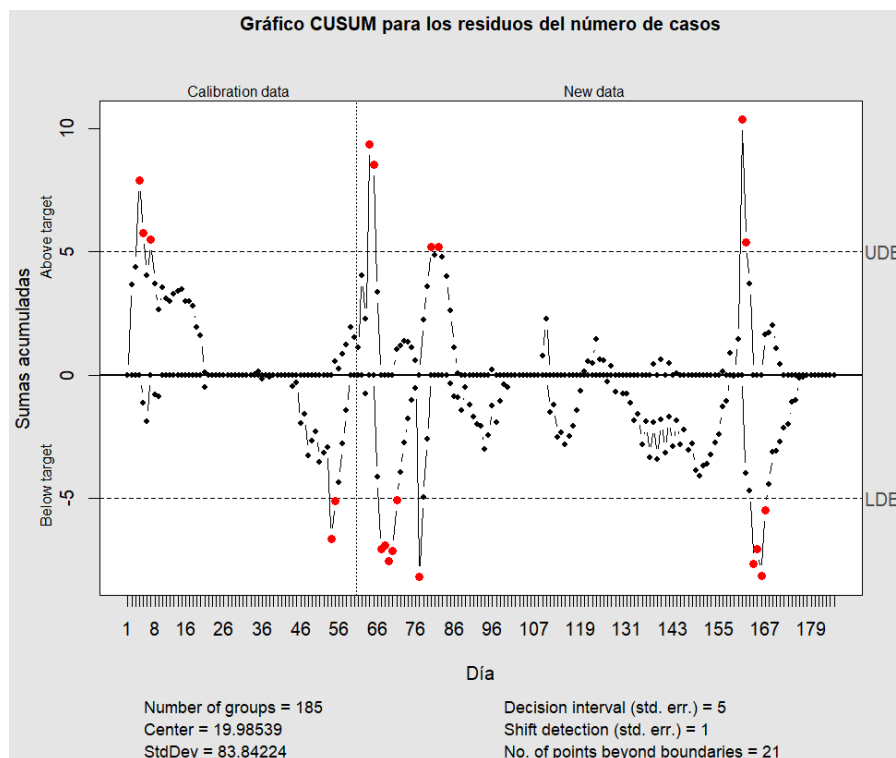


Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de los residuos del “Número de casos activos” del conjunto de datos cedidos por COVIDBENS. La muestra de calibrado comprende las 60 primeras observaciones.

Al igual que en el caso anterior, vamos a construir el gráfico de dispersión con estos puntos que salen fuera de los límites de control para ver si existe alguna diferencia con el gráfico anterior. Este gráfico lo podemos ver en la **Figura 20**.

En este caso, obtenemos unos resultados muy similares a los del caso anterior en los que no se diferenció entre muestra de calibrado y monitorizado para la construcción de los gráficos de medidas individuales, EWMA y CUSUM. Los puntos azules parecen situarse indicando cambios de tendencia en el número de contagios. Al tomar como referencia la primera ola en la muestra de calibrado, nuestro objetivo es detectar como fuera de control las sucesivas nuevas olas que presentan mayores diferencias con respecto a esta primera ola. Sin embargo, aún quedan varios brotes que se ven representados en el gráfico que no se están detectando a través de los gráficos de control.

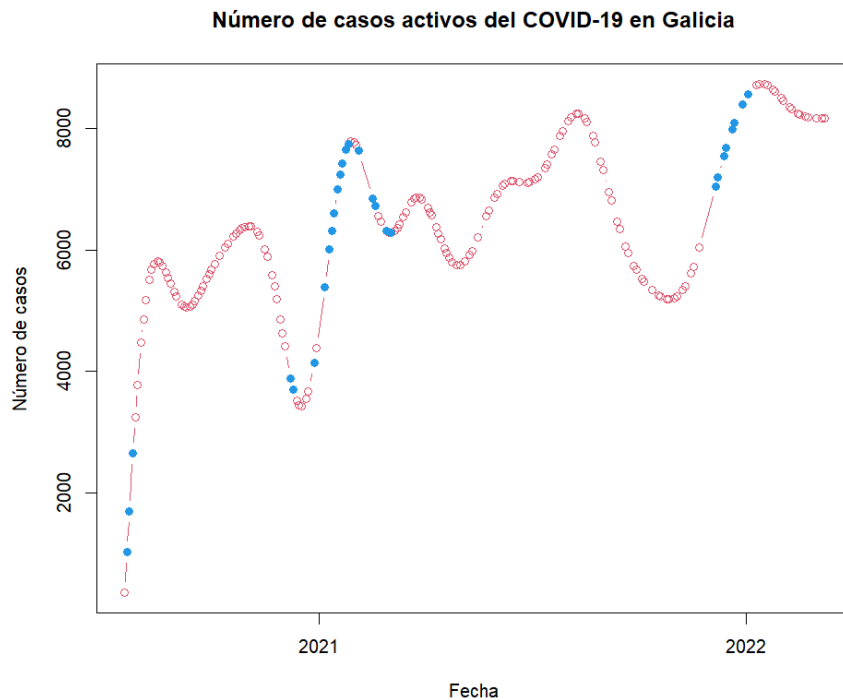
Figura 20: Gráfico de dispersión del número de casos activos

Gráfico de dispersión de la variable “Número de casos activos”. En azul, podemos ver los puntos que corresponden a aquellos que caen fuera de los límites en los gráficos de control en los gráficos de medias para medidas individuales, EWMA y CUSUM de los residuos del modelo ajustado a esta variable. Para la construcción de los gráficos y sus límites de control se utilizó una muestra inicial de calibrado de las primeras 60 observaciones. Elaborado con el conjunto de datos cedidos por COVIDBENS.

Procederemos a estudiar, de nuevo, estos gráficos de control, pero esta vez los construiremos utilizando los datos de la variable “Carga vírica medida” para ver si en este caso los gráficos son capaces de adelantarnos los cambios de tendencia en la carga vírica y de detectar nuevos brotes de la enfermedad. Comenzamos con el gráfico de medias para datos de medidas individuales (**Figura 21**).

Figura 21: Gráfico de medias para medidas individuales para la variable “Carga vírica medida”

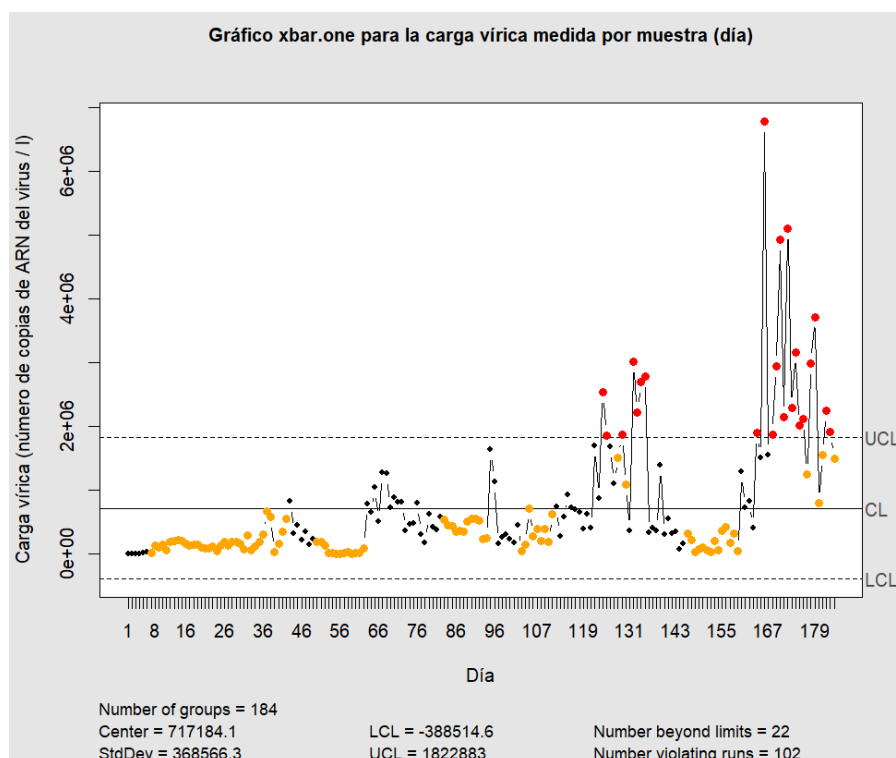


Gráfico de medidas individuales (xbar.one) elaborado con la función “qcc” del paquete “qcc” de R con los datos cedidos por COVIDBENS.

Teniendo en cuenta la información mostrada en el gráfico de medidas individuales (**Figura 21**), la dependencia entre observaciones en el caso de la carga vírica parece no ser tan fuerte como la observada en el número de casos activos de COVID-19. De hecho, la tendencia de la carga vírica no es tan suave como la del número de casos, lo que hace que los gráficos de control tradicionales (asumiendo independencia) proporcionen resultados más interpretables y fiables. El gráfico detecta varios puntos en los que la carga viral es muy alta y supera los límites de control que se están representando en el gráfico. Dado que el gráfico de control de la **Figura 21** se ha construido con los datos originales de la carga vírica en aguas residuales, se pueden identificar las sucesivas olas, siendo las últimas (aquellas relacionadas con una mayor carga vírica y un mayor número de casos) las que se corresponden con observaciones fuera de los límites de control. Dado que el gráfico de control de la **Figura 21** se ha construido con los datos originales de la carga vírica en aguas residuales, se pueden

identificar las sucesivas olas, siendo las últimas (aquellas relacionadas con una mayor carga vírica y un mayor número de casos) las que se corresponden con observaciones fuera de los límites de control. Sin embargo, no parece detectar algún brote que podemos observar en torno a las observaciones 66 y 96 en las que se produce un aumento de la carga vírica que podría resultar significativo en el avance de la pandemia. De todos modos, se observa que los gráficos tradicionales tipo Shewhart podrían proporcionar resultados útiles cuando son aplicados a la carga vírica (mucho más que cuando se aplican al número de casos).

Estudiaremos, a continuación, los gráficos EWMA (**Figura 22**) y CUSUM (**Figura 23**) para ver si ocurre lo mismo que con el gráfico de medidas individuales.

Figura 22: Gráfico EWMA para la variable “Carga vírica medida”

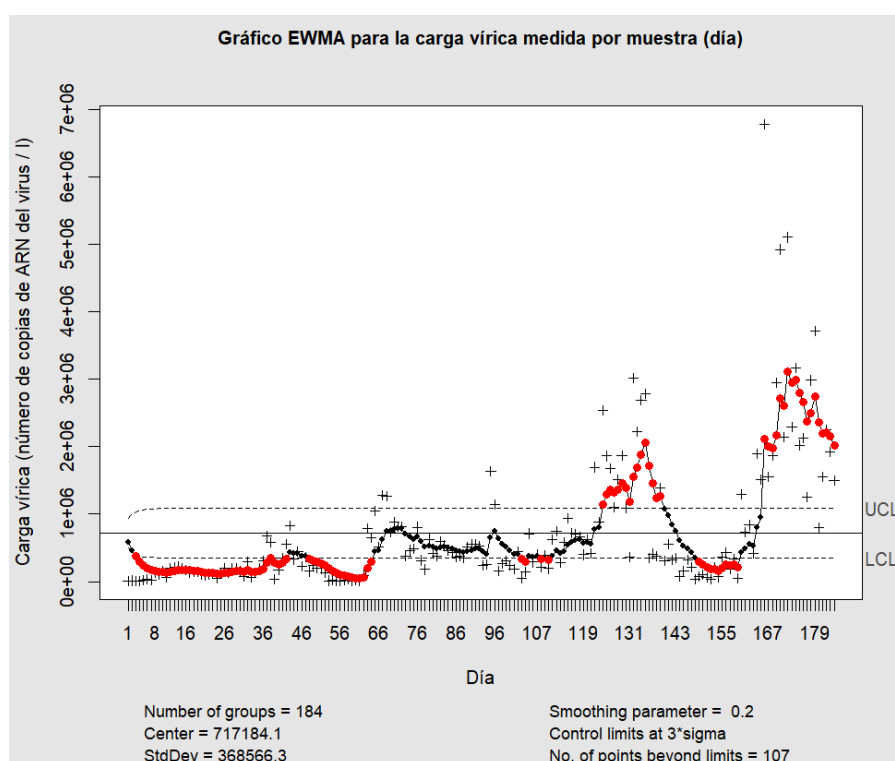


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R. Elaborado con datos cedidos por COVIDBENS.

En el gráfico EWMA (**Figura 22**) se vuelven a detectar dos grandes olas hacia el final del gráfico, en torno a las observaciones 131 y 179. En este gráfico, se detectan más observaciones fuera de los límites de control. Si observamos estos puntos vemos que

se encuentran por debajo del límite inferior. Teniendo en cuenta la naturaleza de los datos que estamos estudiando, lo deseable es que la carga vírica presente en las aguas residuales tenga valores bajos. Este fenómeno lo nombrábamos al comienzo del presente trabajo cuando estudiábamos por qué no parecían adecuados los gráficos de control que conocemos para el estudio de datos epidemiológicos. Por lo tanto, a pesar de que estos puntos se encuentren fuera de los límites parece razonable no tenerlos en cuenta, dada la necesidad pública de que la carga vírica esté cercana a 0. En este punto, es importante destacar que los gráficos de control de las **Figuras 21 y 22** (también los de la **Figura 23**) nos informan que el proceso-situación no está controlada, a través del alto número de puntos fuera de los límites de control o identificados como rachas. De hecho, en estos gráficos se observan todas las etapas por las que ha transcurrido la actual pandemia, muy diferentes unas de otras. En este punto es interesante destacar que, teniendo en cuenta toda la serie temporal de carga vírica, destaca el bajo nivel de carga vírica al comienzo de la pandemia, así como el altísimo nivel alcanzado en fechas recientes (**Figura 21**).

Construiremos, también, un gráfico de sumas acumuladas con los datos de la carga vírica para ver si este gráfico también llega a detectar las olas de coronavirus que hemos detectado con los otros dos gráficos. El gráfico CUSUM lo podemos ver en la **Figura 23**, que nos muestra que el proceso o situación sanitaria estaría fuera de control desde el principio de la pandemia.

Figura 23: Gráfico CUSUM para la variable “Carga vírica medida”

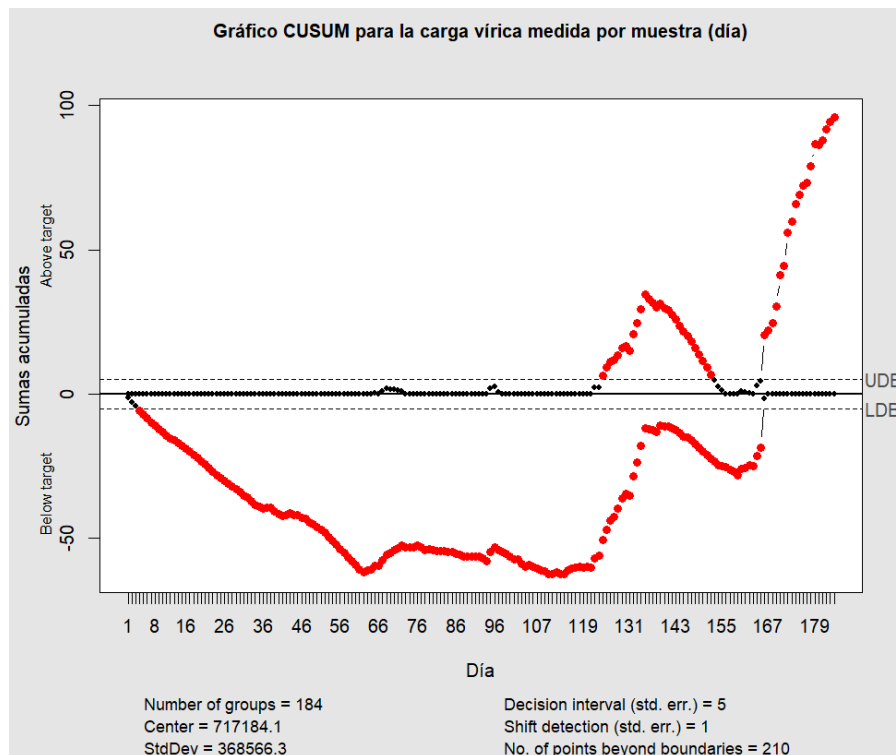


Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de la carga vírica medida por día (número de copias de ARN del virus/l) del conjunto de datos cedidos por COVIDBENS.

Los resultados observados (todas las observaciones identificadas como fuera de control) en la **Figura 23** pueden estar relacionados con la existencia de autocorrelación entre las observaciones. Esto lo comprobamos realizando el contraste de independencia de **Ljung-Box**:

```
> Box.test(Carga_medida, lag = 1, type = "Ljung-Box")
```

Box-Ljung test

data: Carga_medida

X-squared = 68.35, df = 1, p-value < 2.2e-16

Al rechazarse la independencia se ve la necesidad de ajustar los datos mediante un modelo ARIMA y de utilizar sus residuos para la construcción de los gráficos. Procederemos, por tanto, como en el caso del número de casos activos: realizaremos un ajuste a través de un modelo ARIMA y construiremos los gráficos de control con los residuos del modelo ajustado.

Para construir el modelo ARIMA más adecuado, utilizaremos la función “auto.arima” del paquete “forecast” de R, la cual seleccionará el ajuste más adecuado para la serie de tiempo que estemos utilizando. En el caso de la variable “Carga vírica medida”, la función nos selecciona un modelo **ARIMA(1,1,4)** sin constante y con una transformación Box Cox previa de los datos.

Best model: ARIMA(1,1,4)

Series: carga_ts

ARIMA(1,1,4)

Box Cox transformation: lambda= 1

Coefficients:

	ar1	ma1	ma2	ma3	ma4
	-0.8465	0.1899	-0.6266	-0.0656	0.3001
s.e.	0.0629	0.0940	0.0853	0.0694	0.0846

sigma^2 estimated as 4.466e+11: log likelihood=-2712.21

AIC=5436.42 AICc=5436.9 BIC=5455.68

Una vez realizada la diagnosis de los residuos obtenidos a través de este modelo podemos concluir que se está obteniendo un buen ajuste a los datos, por lo que podemos utilizar los residuos para construir los gráficos de control.

Construiremos, de nuevo, los gráficos de control de medidas individuales, EWMA y CUSUM para ver si, una vez eliminada la autocorrelación, los gráficos nos muestran resultados similares a los ya obtenidos con los datos originales.

En la **Figura 24** podemos observar el gráfico de medidas individuales, pero construido a través de los residuos del modelo ARIMA que se ajusta a los datos de la carga vírica

medida en las aguas residuales. Son 9 los puntos que caen fuera de los límites de control, los cuales se corresponden con las observaciones:

```
$beyond.limits
```

```
[1] 124 132 166 179 136 138 167 176 180
```

```
$violating.runs
```

```
[1] 22 23 24
```

Estas observaciones parecen corresponderse con las olas de coronavirus detectadas en los gráficos anteriores, por lo que en este caso sí se estarían detectando brotes de la enfermedad.

Figura 24: Gráfico de medias para medidas individuales para los residuos de la carga vírica

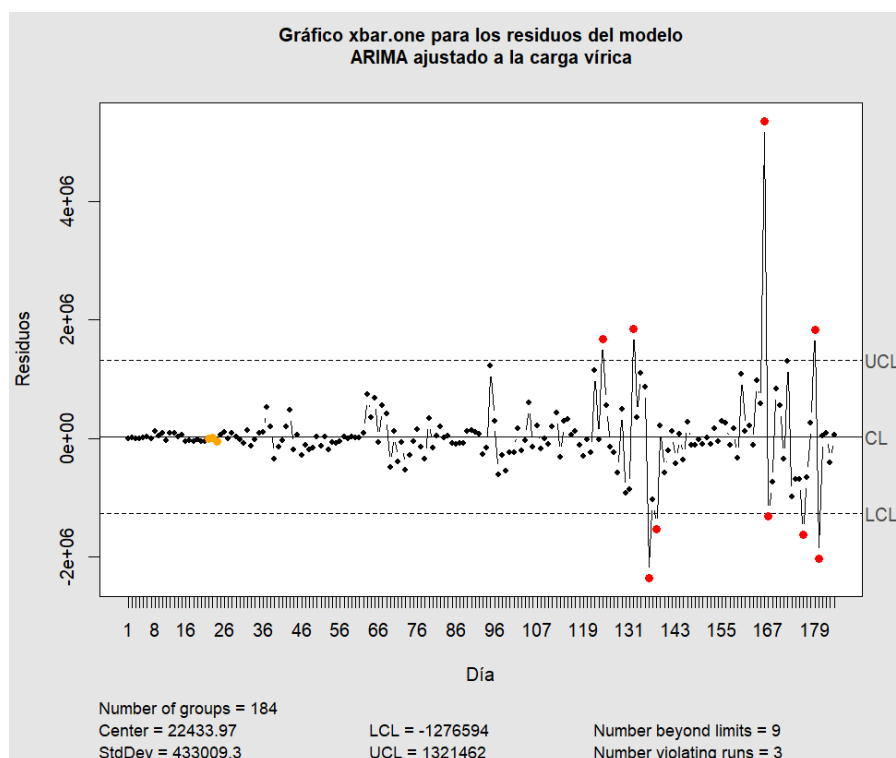


Gráfico de medias para medidas individuales (xbar.one) elaborado con la función “qcc” del paquete “qcc” de R con los datos de los residuos obtenidos con el ajuste del modelo ARIMA(1,1,4) para la variable “Carga vírica medida” de los datos cedidos por COVIDBENS.

En el gráfico EWMA (**Figura 25**), encontramos 11 puntos fuera de los límites de control. Estos puntos son los correspondientes a las observaciones:

125 135 138 140 166 167 168 169 170 172 177

Se obtienen resultados similares a los del gráfico de observaciones individuales que acabamos de estudiar. Las observaciones que están fuera de control se corresponden con las olas de coronavirus detectadas en los gráficos anteriores.

Figura 25: Gráfico EWMA para los residuos del modelo ajustado a la carga vírica

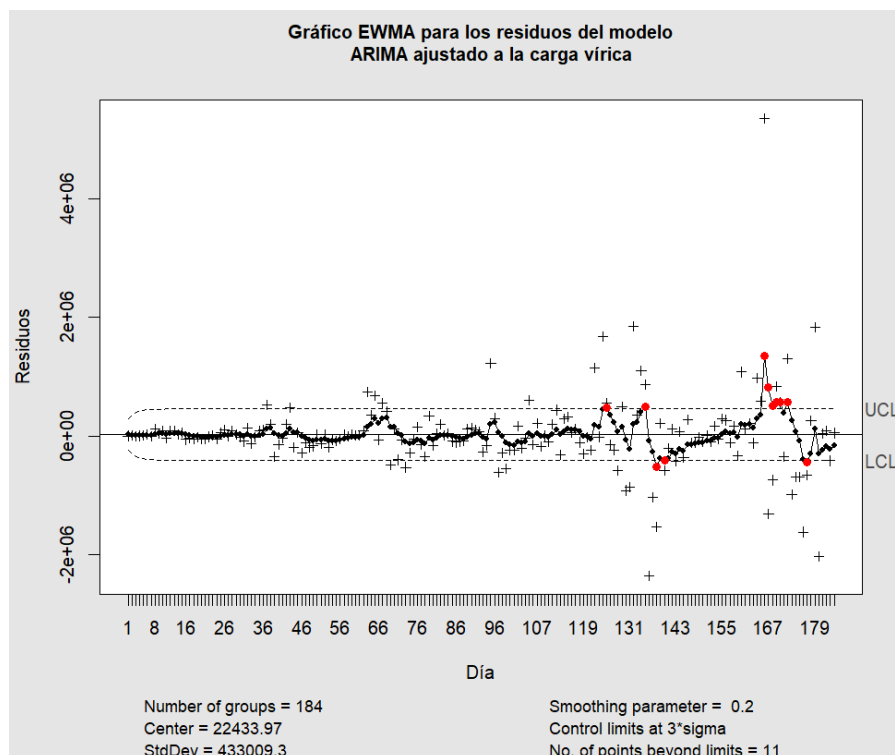


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R. Elaborado con datos sobre los residuos de la carga viral medida en las aguas residuales del conjunto de datos cedidos por COVIDBENS.

A continuación, se construye el gráfico CUSUM (**Figura 26**) de los residuos obtenidos con el modelo ARIMA ajustado a los datos de la carga vírica. En este caso se obtienen resultados mejores a los obtenidos a través de los datos originales. El gráfico ya no nos muestra la mayoría de los puntos fuera de control. Esta vez nos muestra más puntos fuera de los límites que en los gráficos de medias y EWMA, pero a simple vista parecen ser similares a los puntos que se encontraban fuera de los límites en dichos gráficos. Estos puntos son los correspondientes a las observaciones:

\$lower

[1] 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154
176

[21] 177 178 180 181 182 183 18

\$upper

[1] 125 134 135 166 167 168 169 170 171 172 173 174 175

Figura 26: Gráfico CUSUM de los residuos del modelo ajustado a la carga vírica

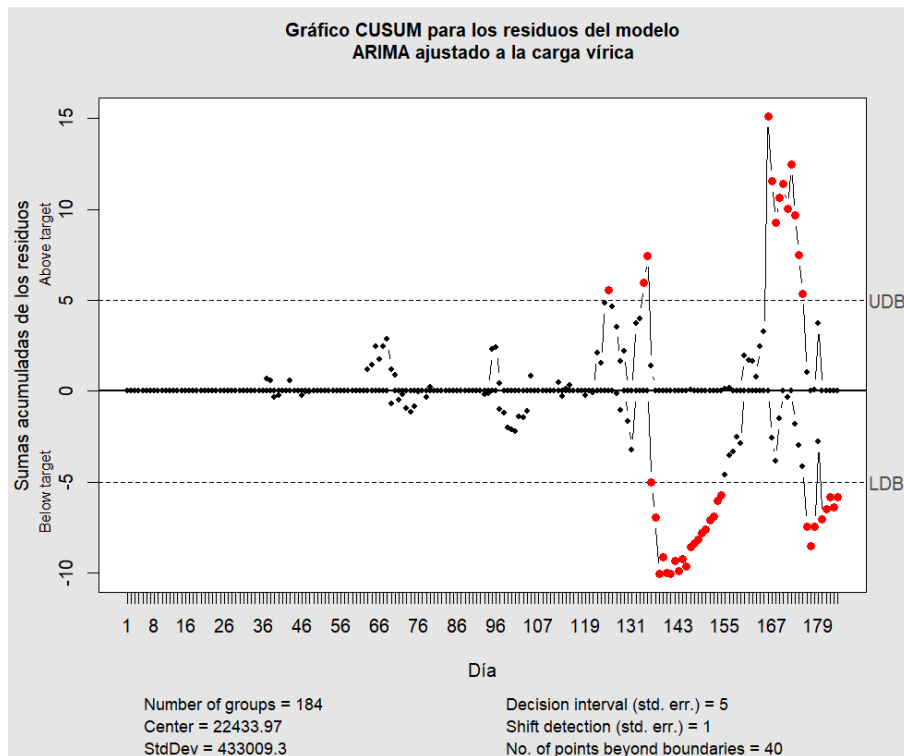


Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de los residuos de la carga vírica del conjunto de datos cedidos por COVIDBENS

Los gráficos de control de medidas individuales, EWMA y CUSUM, detectan que la situación o proceso está fuera de control prácticamente en las mismas fechas, es decir, detectan el comienzo de dos olas genéricas (una en la observación 125 y otra en la 166). Además, el gráfico CUSUM parece identificar como fuera de control la bajada en la carga vírica después de la primera de estas dos olas.

Para ver mejor los resultados, vamos a representar estos puntos en el gráfico de dispersión de la carga vírica medida en las aguas residuales (**Figura 27**). En este gráfico

se pueden ver coloreadas las dos olas de coronavirus que venimos detectando desde gráficos anteriores en las observaciones 131 y 170 aproximadamente, que se corresponden con las fechas 11 de agosto de 2021 y 17 de enero de 2022. Sin embargo, en el gráfico también vemos coloreadas muchas observaciones entre estos dos puntos. Si nos fijamos en el gráfico CUSUM anterior (**Figura 26**), estos puntos se corresponden con los que se encuentran fuera de control entre las observaciones 140 y 155, aproximadamente. Estas observaciones tienen unos niveles mucho más bajos de lo habitual, lo que supone que los puntos caigan fuera de los límites de control. Sin embargo, como hemos dicho anteriormente estos niveles tan bajos de la carga vírica son deseables desde el punto de vista epidemiológico. Si no tenemos en cuenta estos puntos, los gráficos detectan bien las dos olas de coronavirus. Siempre debemos tener en cuenta el tipo de datos que estamos estudiando porque, aunque el gráfico nos proporcione una alerta de que la carga vírica ha disminuido rápidamente, esto sería lo mejor desde el punto de vista de la sanidad pública.

Figura 27: Gráfico de dispersión de la carga vírica medida en las aguas residuales

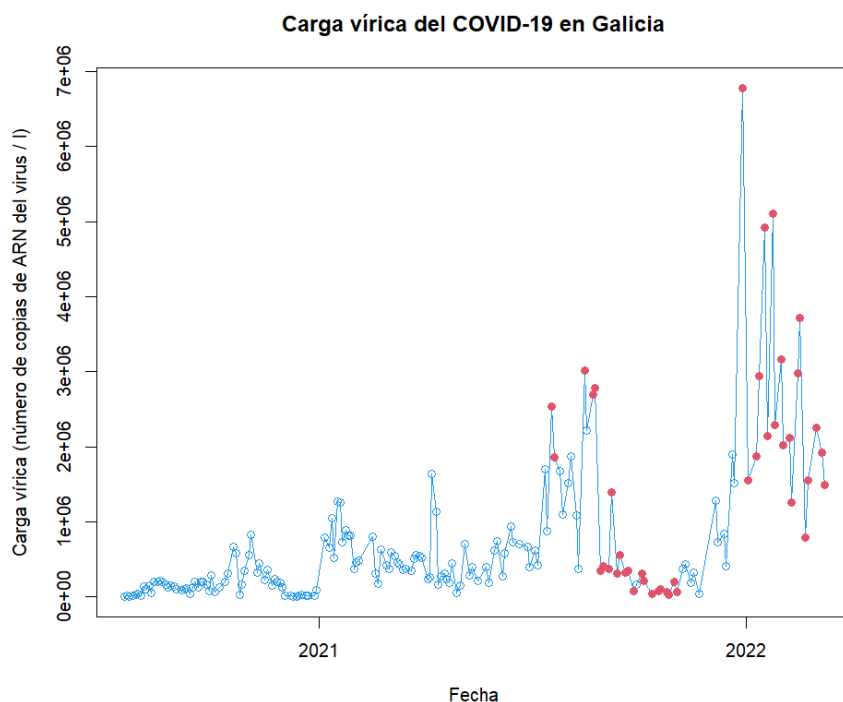


Gráfico de dispersión de la carga vírica medida en las aguas residuales del área de A Coruña (número de copias de ARN del virus por litro). En rojo se encuentran coloreados los puntos que

se han detectado que están fuera de control en los gráficos de medias, EWMA y CUSUM. Elaborado con el conjunto de datos cedido por COVIDBENS.

Teniendo en cuenta que se pretende la identificación de alarmas en tiempo real, no a posteriori, se aplicarán gráficos de control no sólo de Fase I, sino también de Fase II. Es decir, se estimarán los límites de control a partir de una muestra de calibrado, correspondiente a los primeros meses de la pandemia, para luego monitorizar la observaciones de carga vírica subsiguientes. En el momento en que una medida de la muestra de monitorizado caiga fuera de los límites de control, ésta debería estudiarse como una posible alarma de brote o comienzo de una nueva ola. La motivación de este estudio es ilustrar como se podrían aplicar los gráficos de control, independientemente de la muestra de calibrado o entrenamiento utilizada. De hecho, alternativamente, se podría haber elegido como muestra de calibrado o entrenamiento aquellas observaciones estabilizadas en torno a un valor bajo, o mismo cero, de carga vírica. En definitiva, en nuestro caso, detectaremos como estados fuera de control aquellos que sean significativamente diferentes a las observaciones correspondientes a los primeros meses de la pandemia.

Una vez contruidos los gráficos de control con los residuos, podemos coger una muestra más o menos estable de observaciones que será nuestra muestra de calibrado y construir los límites de control con esas observaciones escogidas. Una vez obtenidos y representados en nuestro gráfico los límites de control, pasaremos a representar el resto de las observaciones en el gráfico para poder compararlas con dichos límites.

Procederemos de forma similar a los gráficos de control del número de casos. Esta vez, hemos escogido una muestra de calibrado algo mayor (las primeras 100 unidades, que comprenden desde el 19 de julio de 2020 al 20 de abril de 2021) al presentarse más o menos estables estos residuos en el tiempo.

Figura 28: Gráfico de medidas individuales de los residuos de la carga viral medida con una muestra inicial de calibrado

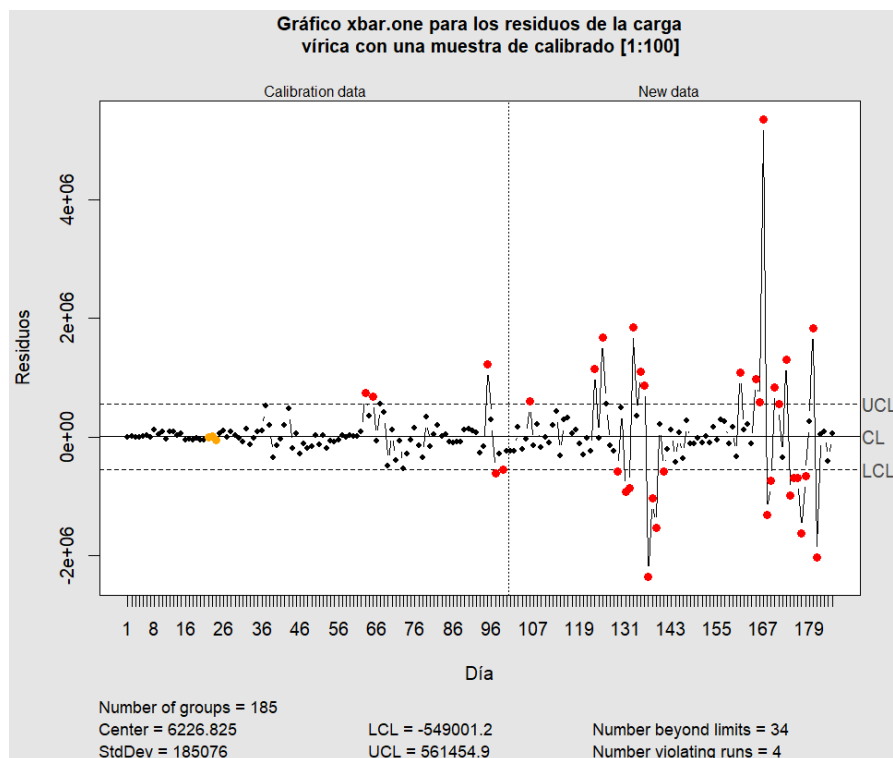


Gráfico de medidas individuales (xbar.one) elaborado con la función “qcc” del paquete “qcc” de R con los datos de los residuos obtenidos con el ajuste del modelo ARIMA(1,1,4) para la variable “Carga vírica medida” de los datos cedidos por COVIDBENS. La muestra de calibrado comprende las 100 primeras observaciones.

En el gráfico de medias para datos de medidas individuales (**Figura 28**) obtenemos muchos más residuos que caen fuera de los límites de control. Estos residuos se corresponden con las observaciones:

```
$beyond.limits
```

```
[1] 63 65 95 106 123 125 133 135 136 161 165 166 167 170 171 173 180 97 99 129
```

```
[21] 131 132 137 138 139 141 168 169 174 175 176 177 178 181
```

```
$violating.runs
```

```
[1] 65 22 23 24
```

En el gráfico EWMA (**Figura 29**) nos ocurre lo mismo, al calcular los límites de control con únicamente las 100 primeras observaciones en lugar de toda la muestra, estos límites son más restrictivos por lo que un mayor número de residuos se saldrán fuera de los límites. En este caso, esos residuos se corresponden con las observaciones:

64 65 66 67 68 95 96 102 125 126 127 128 132 134 135 136 138 139 140 141
142 143 144 145 146 161 163 165 166 167 168 169 170 171 172 173 174 177 178 179
181 182 184

Figura 29: Gráfico EWMA de los residuos de la carga vírica con una muestra de calibrado

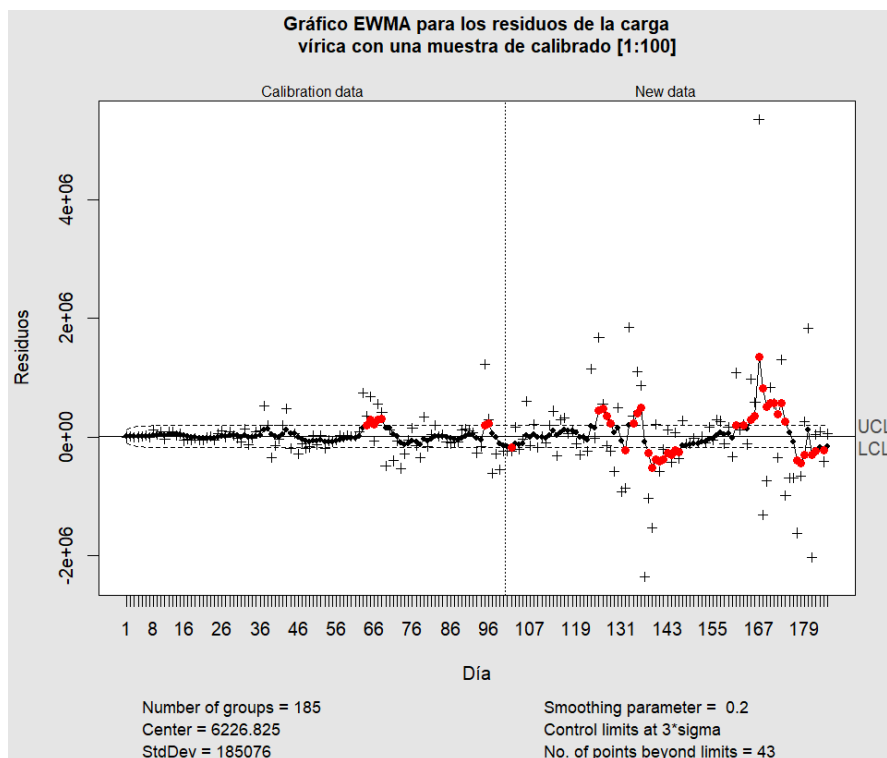


Gráfico de medias móviles con ponderación exponencial (EWMA) realizado con la función “ewma” del paquete “qcc” de R. Elaborado con datos sobre los residuos de “Carga vírica medida” del conjunto de datos cedidos por COVIDBENS. La muestra de calibrado comprende las 100 primeras observaciones.

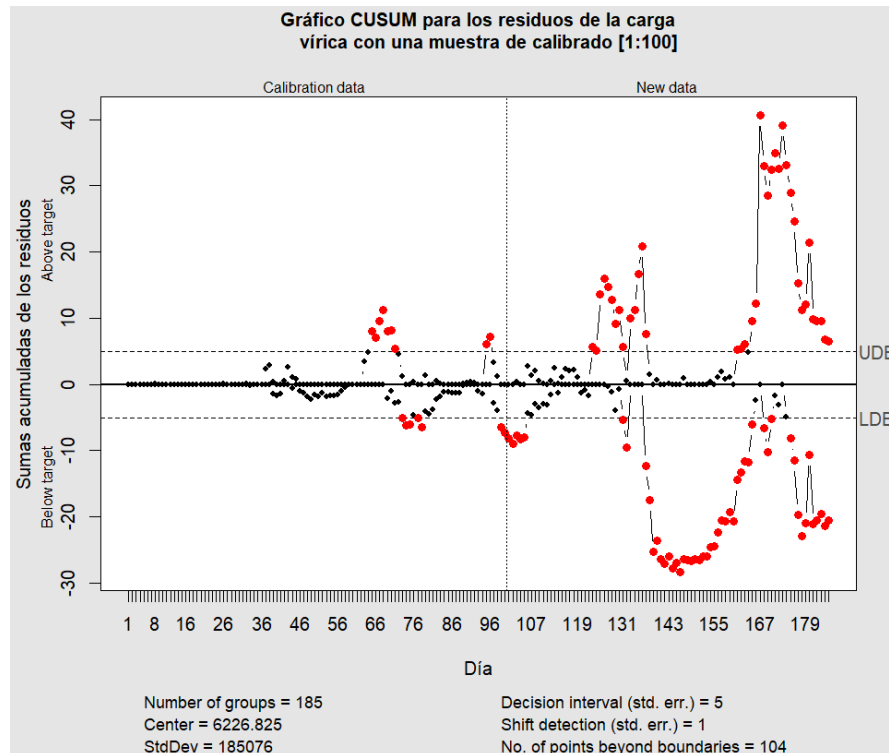
Figura 30: Gráfico CUSUM de los residuos de la carga vírica con una muestra de calibrado

Gráfico de sumas acumuladas (CUSUM) realizado con la función “cusum” del paquete “qcc” de R. Elaborado con datos de los residuos del “Número de casos activos” del conjunto de datos cedidos por COVIDBENS. La muestra de calibrado comprende las 60 primeras observaciones.

Como venía ocurriendo en los anteriores gráficos, en el gráfico CUSUM (**Figura 30**) también obtenemos muchos puntos fuera de los límites de control. Estos puntos se corresponden con las observaciones:

\$lower

[1] 73 74 75 77 78 99 100 101 102 103 104 105 131 132 137 138 139 140 141 142

[21] 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162

[41] 163 164 165 168 169 170 175 176 177 178 179 180 181 182 183 184 185

\$upper

[1] 65 66 67 68 69 70 71 95 96 123 124 125 126 127 128 129 130 131 133 134

[21] 135 136 137 161 162 163 165 166 167 168 169 170 171 172 173 174 175 176 177
178

[41] 179 180 181 182 183 184 185

El gráfico CUSUM (**Figura 30**), de la misma forma que con los datos de los residuos sin utilizar una muestra de calibrado, nos vuelve a mostrar demasiados puntos fuera de los límites de control. Eso puede ser debido a que se obtienen sumas acumuladas por debajo de los límites, algo que desde el punto de vista epidemiológico es una situación deseable, como ya hemos dicho en gráficos anteriores. Podemos comprobarlo llevando estos puntos al gráfico de dispersión de los datos originales acerca de la carga vírica medida en las aguas residuales (**Figura 31**).

Figura 31: Gráfico de dispersión de la carga vírica medida en las aguas residuales del área de A Coruña

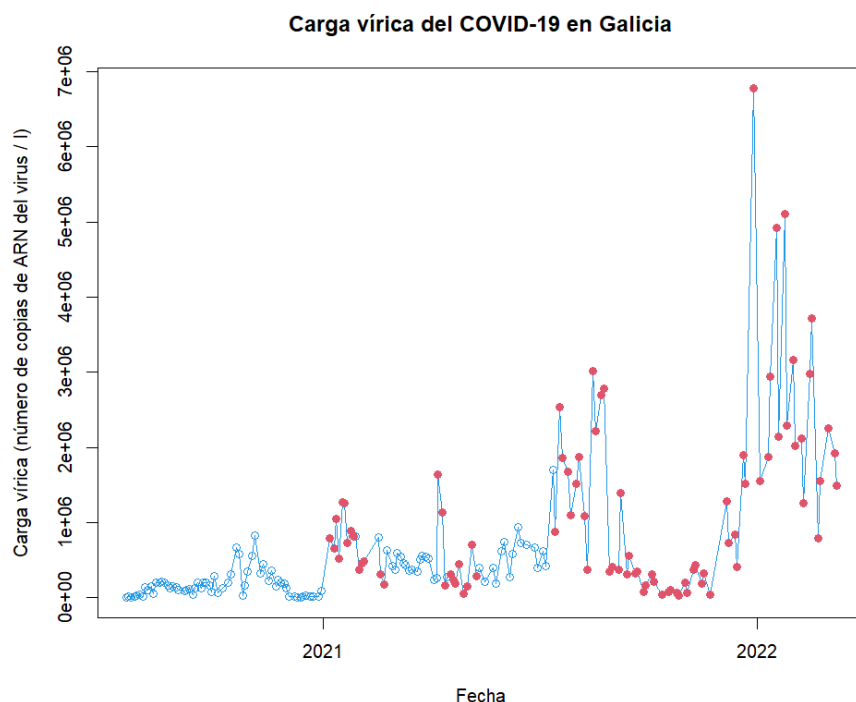


Gráfico de dispersión de la carga vírica (número de copias de ARN del virus por litro) medida en las aguas residuales del área de A Coruña. En rojo se representan los puntos que cayeron fuera de los límites de control en los gráficos de medias, EWMA y CUSUM contruidos con una muestra inicial de calibrado de las 100 primeras observaciones. Elaborado con el conjunto de datos cedidos por COVIDBENS.

En este gráfico (**Figura 31**) vemos una mejoría con respecto al gráfico de dispersión construido anteriormente (**Figura 27**). Vemos que ha sido capaz de detectar dos brotes más pequeños a principios de 2021. Además, a diferencia de los gráficos anteriores, en este gráfico están coloreados algunos puntos que se encuentran inmediatamente antes de los brotes de la enfermedad, por lo que parece que los gráficos se están adelantando a los acontecimientos. Sin embargo, el gráfico vuelve a detectar como fuera de control puntos que corresponden con una carga vírica muy baja, lo cual es deseable desde el punto de vista epidemiológico. Esto tiene relación con la muestra de calibrado que se ha definido: se detectan como fuera de control valores muy bajos de carga vírica (o que bajadas muy rápidas) porque son significativamente diferentes a los patrones de la muestra de calibrado. Los tres gráficos detectan al mismo tiempo las cuatro olas principales correspondientes a los puntos 23, 95, 123 y 161, que se corresponden con las fechas: 10 de septiembre de 2020, 7 de abril de 2021, 15 de julio de 2021 y 8 de diciembre de 2021, respectivamente.

En el siguiente apartado estudiaremos otros algoritmos y funciones de un paquete de R creado específicamente para el estudio de datos epidemiológicos.

3.4. Aplicación de gráficos de control específicos en epidemiología

La epidemiología es una disciplina científica perteneciente al área de la medicina que se ocupa del estudio cuantitativo de todas las características que definen y caracterizan a las enfermedades humanas. Por tanto, también se ocupa de su control y monitorización, tarea para la cual que se utilizan herramientas del Control Estadístico de Procesos, muchas veces específicamente diseñadas para este tipo especial de datos. Es por ello importante, en el marco de este TFM, mostrar algunas de estas alternativas de análisis, además de mostrar los resultados de su aplicación al caso de estudio de la evolución del número de casos activos de COVID-19 y cantidad de carga vírica en aguas residuales en la zona metropolitana de A Coruña. De entre todas las herramientas de control y detección de anomalías, destacan diversos gráficos de control (muchos de ellos modificaciones de gráficos Shewhart o con memoria) disponibles en el paquete de R “surveillance”. Este paquete está creado para el tratamiento y la modelización de datos de tipo epidemiológico y de enfermedades infecciosas (Salmon et al., 2016). Estos datos pueden ser series temporales de recuentos, proporciones y datos categóricos, así como procesos puntuales espaciotemporales (Höhle et al., 2022). El paquete contiene muchas funciones que podrían ser aplicadas en datos como los que estamos trabajando. Nos centraremos en aquellas funciones basadas en el Control Estadístico de Procesos y en algunos gráficos que son modificaciones, adaptadas a nuestro tipo de datos, de algunos de los que hemos hablado en puntos anteriores.

3.4.1. Vigilancia de una serie temporal de datos de recuento mediante el método EARS C1, C2 o C3

El método EARS o Sistema de Detección Temprana de Anomalías o Aberraciones (del inglés, Early Aberration Detection System) consiste en tomar los valores del rango de una serie temporal de datos epidemiológicos y calcular para cada uno de los puntos un umbral basado en los valores del pasado más reciente, es decir, las observaciones pasadas que están más cercanas al momento presente o de referencia, t . Ese umbral se comparará con el número de recuentos observado, es decir, el número de veces que se produce alguna variable epidemiológica como el número de contagios, el número de muertes, etc. Si la observación está por encima de un cuantil específico del intervalo de

predicción se producirá una señal de alarma. Este método es muy útil para datos en los que no existen muchos valores históricos o en los que no podemos acceder a ellos ya que solo necesitan los datos del pasado reciente. Por ello parece razonable estudiar esta función en el caso de nuestros datos. Al tratar con una enfermedad relativamente nueva como es el coronavirus, no disponemos de tantos datos como podríamos tener si estuviésemos estudiando una enfermedad que lleva entre la población cientos de años como la gripe.

El método EARS tiene tres variantes que calculan el valor esperado de manera diferente. Estas variantes son C1, C2 y C3.

En C1 y C2, el valor esperado es la media de los recuentos sobre la línea base. El intervalo de predicción depende de la desviación típica de los recuentos observados en esta ventana. Estos métodos y sus correspondientes gráficos pueden considerarse como gráficos de control de Shewhart.

En el método C3, el valor esperado se calcula como la suma en 3 momentos del tiempo (el instante que se está evaluando y los dos anteriores) de la diferencia entre las observaciones y las predicciones, calculándose las predicciones con el método C2. Este método podría considerarse como un método basado en el gráfico CUSUM, ya que se están sumando las diferencias entre predicciones y las observaciones en varios momentos temporales. Sin embargo, no es un CUSUM como tal, ya que se suman únicamente tres momentos temporales y no se está realizando la acumulación de las sumas de todo un rango, pero presentan muchas similitudes. Es por ello por lo que hemos considerado utilizar este método para el presente trabajo.

- **Método C1**

Para el método C1 la línea base o de partida (baseline) serán todos aquellos datos (la función escoge, por defecto, los 7 puntos anteriores al momento del tiempo de referencia) que se encuentren antes del momento de tiempo presente t , es decir, serán los datos del pasado reciente que se utilizarán para el cálculo del umbral o límite de control. El valor esperado será la media de la línea de partida. Se calcula un intervalo de predicción (bilateral) aproximado $(1 - \alpha)100\%$ basado en la suposición de que la diferencia entre el valor esperado y el valor real observado dividido entre la desviación

estándar de los recuentos a lo largo de la ventana, denominada $C_1(t)$, sigue una distribución normal estándar en ausencia de brotes de la enfermedad:

$$C_1(t) = \frac{Y(t) - \bar{Y}_1(t)}{S_1(t)}$$

Donde

$$\bar{Y}_1(t) = \frac{1}{baseline} \sum_{i=t-1}^{t-baseline} Y(i)$$

Y

$$S_1^2(t) = \frac{1}{baseline - 1} \sum_{i=t-1}^{t-baseline} [Y(i) - \bar{Y}_1(i)]^2$$

Entonces, bajo la hipótesis nula de que no hay brotes de la enfermedad,

$$C_1(t) \sim N(0,1)$$

Se producirá una alarma si

$$C_1(t) \geq z_{(1-\alpha)}$$

Siendo $z_{(1-\alpha)}$ el cuantil $(1 - \alpha)$ de la distribución normal estándar.

El límite superior $U_1(t)$ se define como:

$$U_1(t) = \bar{Y}_1(t) + z_{(1-\alpha)} S_1(t)$$

Para ver el gráfico que se crea con este método hemos utilizado los datos de COVIDBENS que ya hemos utilizado anteriormente. Primero para los datos del número de casos activos y después, para los datos de la carga vírica medida en las aguas residuales. Dada la construcción de las funciones del paquete, primero debemos transformar nuestros datos en una serie de tiempo de tipo “surveillance” con la función `sts` del paquete. Para transformar los datos, debemos indicar la fecha en la que empiezan los datos, la frecuencia de los mismos (en este caso 365 ya que los datos se los tenemos por días) e indicando que los datos incluyen la fecha en la que se realizaron las mediciones:

```
casos_sts <- sts(Casos, start = c(2020,7), frequency = 365,
```

```
epochAsDate = TRUE, epoch = as.numeric.Fecha))

carga_sts <- sts(Carga_medida, start = c(2020,7), frequency = 365,

epochAsDate = TRUE, epoch = as.numeric.Fecha))
```

Una vez obtenida nuestras series de tiempo de tipo surveillance podemos construir el gráfico con el sistema de detección de anomalías EARS mediante el método C1 a través de la función `earsC` e indicando que se utilizará el método C1:

```
monitoriz_casos1 <- earsC(casos_sts, control = list(method = "C1"))
```

Utilizando la función `plot` sobre este conjunto de datos, obtenemos el siguiente gráfico:

Figura 32: Gráfico de detección temprana de anomalías (EARS) utilizando el método C1 con la variable “Número de casos activos”

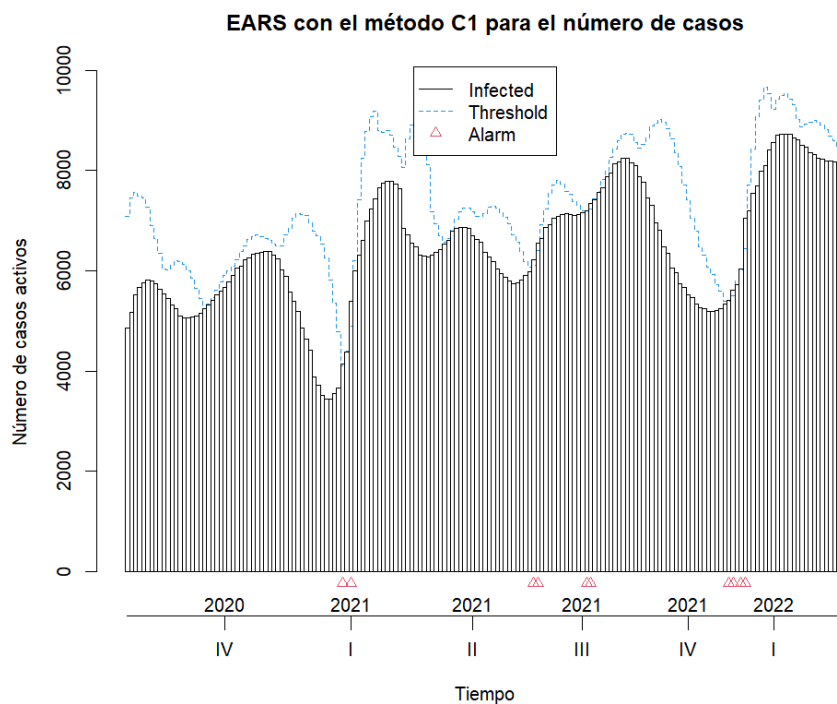


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C1. En negro tenemos representado el número de casos activos, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “`earsC`” del paquete “`surveillance`” de R y con el conjunto de datos cedido por COVIDBENS.

Este gráfico puede considerarse como un gráfico de control de Shewhart por lo que el umbral representado por la línea discontinua azul se corresponde con un límite de control superior. Estamos ante datos epidemiológicos que, por su naturaleza, nos interesa que se estabilicen en niveles bajos (es deseable que el número de casos activos de coronavirus en el área metropolitana de A Coruña se encuentren en niveles muy bajos porque significaría que se está llegando a la fase final de la pandemia) por lo que se tomarán como señales de alarma únicamente aquellos puntos que sobresalgan de este límite superior de control. Si nos fijamos en las señales de alarma del gráfico de detección de anomalías (**Figura 32**), se detectan aquellos grandes incrementos en el número de contagios, que podemos definir como olas de la enfermedad, justo antes de que sucedan (en concreto, se detectan cuatro de los grandes brotes u olas, justo al comienzo del crecimiento en el número de casos). Sin embargo, no se llegan a detectar las dos olas que vemos en el comienzo del gráfico, correspondientes al verano de 2020 y a octubre de ese mismo año.

Podemos hacer el mismo gráfico con los datos de la carga vírica medida (número de copias de ARN por litro de agua residual) para ver si también se detectan brotes u olas antes de que sucedan. Esto lo vemos en la **Figura 33**.

En el caso del gráfico construido con los datos de la carga vírica (**Figura 33**), se están detectando más señales de alarma que en el gráfico del número de casos (**Figura 32**), es decir, el gráfico de la carga vírica es más sensible a los cambios que el gráfico del número de casos activos. Esto parece razonable si tenemos en cuenta que estamos partiendo de unos niveles muy bajos al comienzo del gráfico (una muestra de calibrado con observaciones más estables y a un nivel más bajo que las tomadas anteriormente), por lo que cualquier incremento podría significar un brote de la enfermedad. En el gráfico de detección de anomalías de la carga vírica (**Figura 33**), nos ocurre algo similar al gráfico del número de casos (**Figura 32**), las señales de alarma se detectan instantes antes de que comience la nueva ola de COVID-19. Podemos verlo, por ejemplo, en la ola correspondiente a octubre de 2020 la cual se detecta como una señal de alarma en el gráfico de la carga vírica antes de que comiencen a subir el número de casos activos en la **Figura 32**, en la cual no se detecta como señal de alarma. También ocurre lo mismo a finales de 2021, se detectan señales de alarma en el gráfico de la carga vírica antes de que suceda la gran ola de 2022. A veces, en este tipo de gráficos, no se llegan

a detectar pequeños brotes de la enfermedad que pueden ser de menor importancia ya que son menos sensibles a cambios pequeños. Esto lo vemos en la **Figura 33** en el tercer trimestre de 2021 o a comienzos de 2022 en donde se produce una gran ola de la enfermedad, pero no se detectan otros brotes más pequeños.

Figura 33: Gráfico de detección temprana de anomalías (EARS) utilizando el método C1 con la variable “Carga vírica”

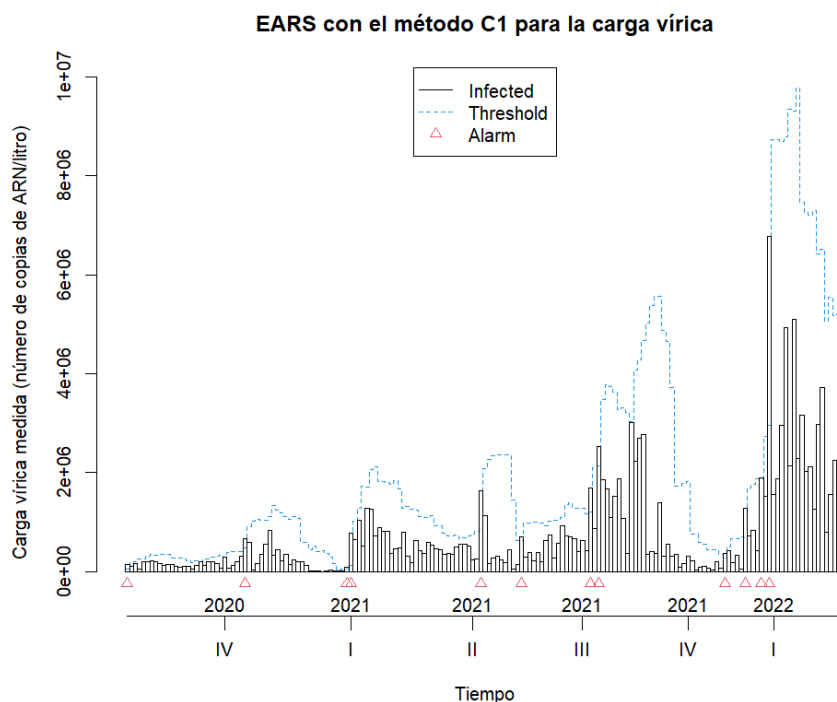


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C1. En negro tenemos representado la carga vírica medida en aguas residuales, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “earsc” del paquete “surveillance” de R y con el conjunto de datos cedido por COVIDBENS.

▪ Método C2

El método C2 es muy similar al método C1. Como ya hemos dicho, podríamos interpretarlos como gráficos de control de Shewhart, salvo por el desfase de 2 días en la construcción de la línea base (baseline). La línea base para C2 es la línea de base (por defecto se construye con 7 puntos, como en el método C1) con un desfase de 2

días antes del momento presente o de referencia t . Es decir, $(t - \text{baseline} - 2)$ en $t - 3$. Se puede calcular un intervalo de predicción (bilateral) aproximado $(1 - \alpha)100\%$ basado en el supuesto de que la diferencia entre el valor esperado y el valor real observado dividido por la desviación típica de los recuentos a lo largo de la ventana, denominada $C_2(t)$, sigue una distribución normal estándar en ausencia de brotes:

$$C_2(t) = \frac{Y(t) - \bar{Y}_2(t)}{S_2(t)}$$

Donde

$$\bar{Y}_2(t) = \frac{1}{\text{baseline}} \sum_{i=t-3}^{t-\text{baseline}-2} Y(i)$$

Y

$$S_2^2(t) = \frac{1}{\text{baseline} - 1} \sum_{i=t-3}^{t-\text{baseline}-2} [Y(i) - \bar{Y}_2(i)]^2$$

Bajo la hipótesis nula de que no hay brotes de la enfermedad, entonces:

$$C_2(t) \sim N(0,1)$$

Se producirá una alarma si

$$C_2(t) \geq z_{(1-\alpha)}$$

Siendo $z_{(1-\alpha)}$ el cuantil $(1 - \alpha)$ de la distribución normal estándar.

El límite superior $U_2(t)$ se define como:

$$U_2(t) = \bar{Y}_2(t) + z_{(1-\alpha)} S_2(t)$$

Al igual que en el caso anterior, podemos construir un gráfico utilizando los datos de COVIDBENS. En este caso, a la hora de utilizar la función `earsC`, indicaremos que estamos utilizando el método C2:

```
monitoriz_casos2 <- earsC(casos_sts, control = list(method = "C2"))
```

El gráfico construido con este método lo podemos ver en la **Figura 34**. Este gráfico es muy similar al realizado con el método C1 (**Figura 32**), pero si nos fijamos en las señales detectadas en la **Figura 34**, vemos que este gráfico es más sensible a los cambios, al

detectarse un mayor número de señales de alarma. Como vemos, se están detectando todos aquellos incrementos en el número de casos, es decir, se detectan nuevos brotes de la enfermedad. Al igual que con el método C1, este gráfico lo podemos considerar como un gráfico de control de Shewhart. Las señales de alarma se están obteniendo, también, justo antes de que comience el brote.

Figura 34: Gráfico de detección temprana de anomalías (EARS) utilizando el método C2 con la variable “Número de casos activos”

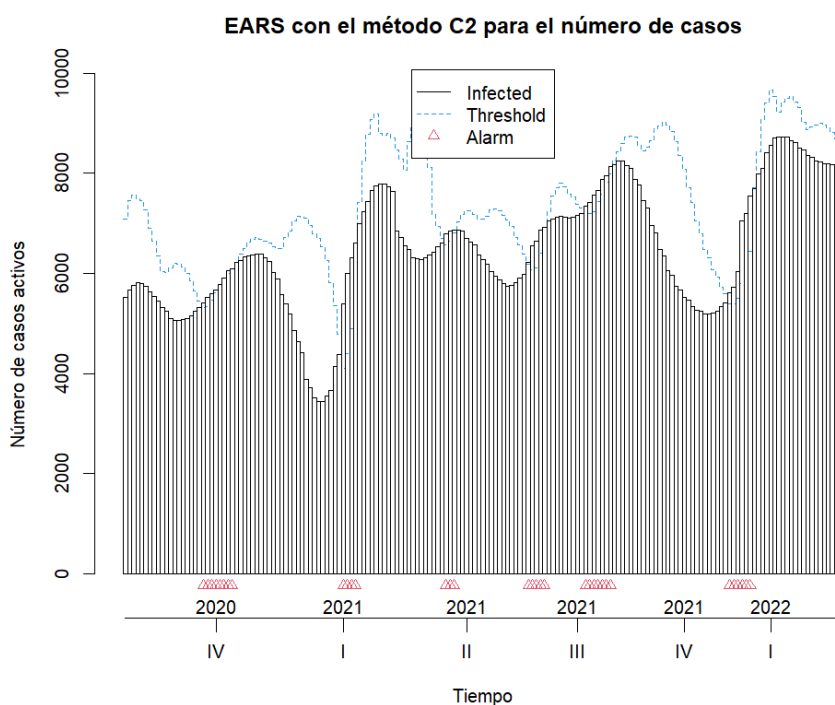


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C2. En negro tenemos representado el número de casos activos de coronavirus, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “earsC” del paquete “surveillance” de R y con el conjunto de datos cedido por COVIDBENS.

El gráfico con los datos de la carga vírica medida en las aguas residuales lo podemos observar en la **Figura 35**.

Figura 35: Gráfico de detección temprana de anomalías (EARS) utilizando el método C2 con la variable “Carga vírica”

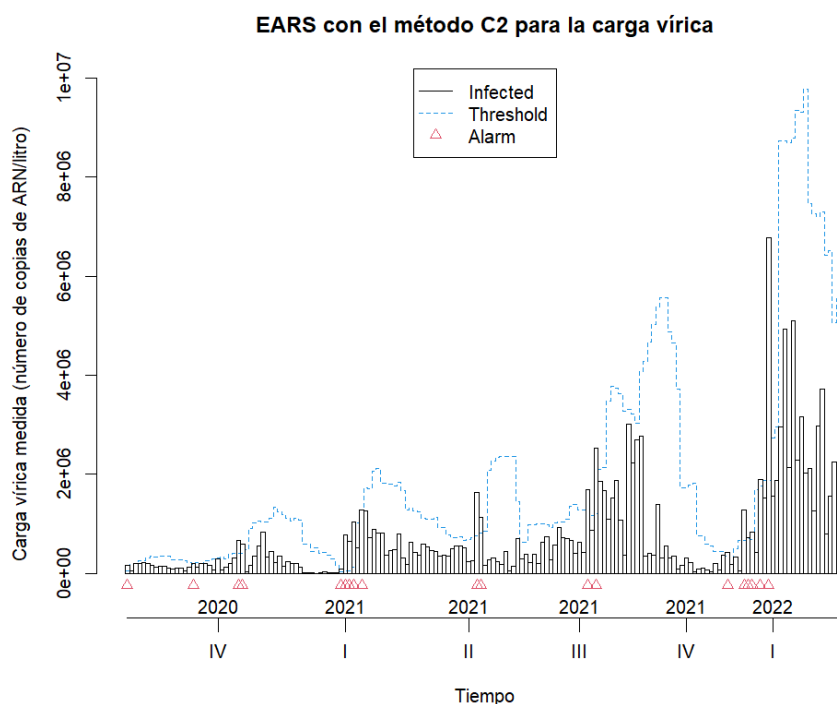


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C2. En negro tenemos representado la carga vírica medida en aguas residuales, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “earsC” del paquete “surveillance” de R y con el conjunto de datos cedido por COVIDBENS.

En el caso del gráfico construido con la carga vírica (**Figura 35**) nos ocurre lo mismo que en el caso anterior, el gráfico realizado a través del método C2 es más sensible a los cambios que el realizado con el método C1 (**Figura 33**) ya que se detectan más señales de alarma. Estas alarmas se detectan justo antes del comienzo del brote de coronavirus. Lo vemos muy bien en la ola de enero de 2021, en donde se detectan las señales de alarma en el gráfico de la carga vírica antes de que comience a subir el número de casos activos. Si nos fijamos en el número de señales de alarma, el gráfico del número de casos (**Figura 34**), detecta un mayor número de alarmas, pero todas estas alarmas se agrupan dando lugar a varios brotes. En el caso de la carga vírica, se detectan un mayor número de brotes, por lo que es más sensible a los cambios que el

gráfico del número de casos. Además, este gráfico detecta un mayor número de olas reales. Nos ocurre lo mismo que con el método C1, el gráfico del número de casos activo (**Figura 34**) no llega a detectar la primera ola, correspondiente al verano de 2020, en el comienzo del gráfico.

▪ Método C3

El método C3 es bastante diferente, pero se basa en el método C2. De hecho, utiliza $C_2(t)$ en el momento del tiempo t y los dos momentos de tiempo anteriores. Por lo tanto, la línea base consistirá en los momentos de tiempo t (*baseline* + 4) en $t - 3$. El estadístico $C_3(t)$ será la suma de las diferencias entre las observaciones y las predicciones, por lo que podemos compararlo con el método CUSUM que ya conocemos.

$$C_3(t) = \sum_{i=t}^{t-2} \max(0, C_2(i) - 1)$$

Bajo la hipótesis nula de ausencia de brotes, entonces:

$$C_3(t) \sim N(0,1)$$

Se producirá una alarma si

$$C_3(t) \geq z_{(1-\alpha)}$$

Siendo $z_{(1-\alpha)}$ el cuantil $(1 - \alpha)$ de la distribución normal estándar.

El límite superior $U_3(t)$ se define como:

$$U_3(t) = \bar{Y}_2(t) + S_2(t) \left(z_{(1-\alpha)} - \sum_{i=t-1}^{t-2} \max(0, C_2(i) - 1) \right)$$

El gráfico obtenido con el método C3 se realiza con la misma función que los gráficos anteriores, modificando el método e indicando que utilizaremos el C3.

```
monitoriz_casos3 <- earsC(casos_sts, control = list(method = "C3"))
```

Este gráfico lo vemos en la **Figura 36**. El gráfico es más sensible a los cambios y detecta un mayor número de señales de alarma que los otros gráficos (**Figura 32 y 34**). Además, este gráfico detecta antes las señales de alarma que los otros dos gráficos. Si lo

comparamos con el obtenido con el método C2 (**Figura 34**), es muy similar en el sentido de que detecta el mismo número de olas, pero este último (**Figura 36**), detecta un mayor número de señales de alarma, al adelantarse a las de la **Figura 34**.

Figura 36: Gráfico de detección temprana de anomalías (EARS) utilizando el método C3 con la variable “Número de casos activos”

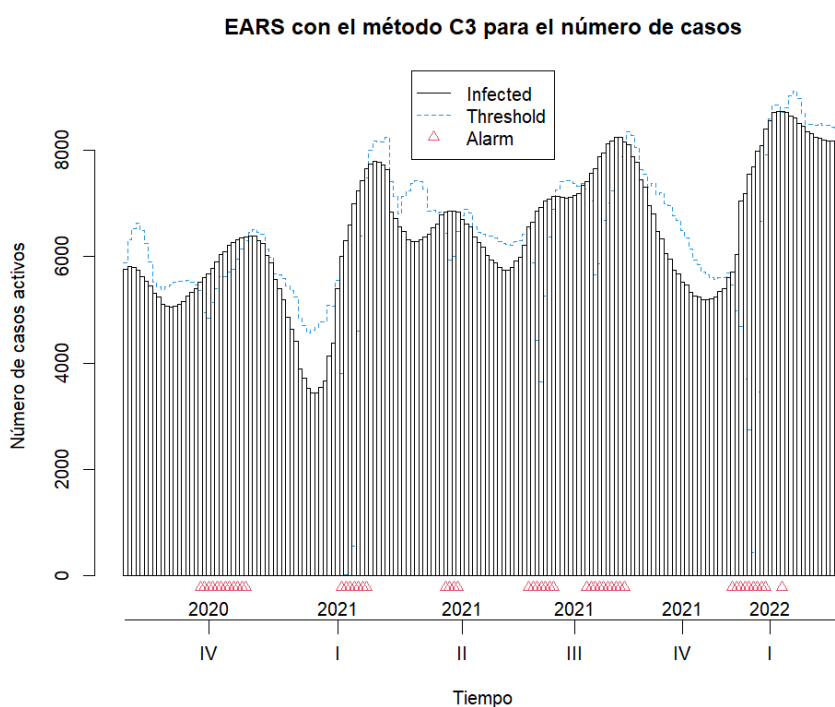


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C3. En negro tenemos representado el número de casos activos de coronavirus, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “earsC” del paquete “surveillance” de R y con el conjunto de datos cedido por COVIDBENS.

El gráfico obtenido con los datos de la carga vírica lo vemos en la **Figura 37**. Este gráfico también es muy similar al obtenido mediante el método C2 para los datos de la carga vírica, pero es más sensible a los cambios pequeños ya que detecta más señales de alarma en todos los incrementos en la carga vírica medida en las aguas residuales. Es decir, está detectando más brotes reales que los otros dos gráficos. Además, también es más sensible a los cambios que el gráfico para el número de casos activos, al detectarse un mayor número de brotes. En este caso, el gráfico de la carga vírica detecta

más brotes y estos se detectan en instantes antes que el gráfico del número de casos activos (**Figura 36**). Esto se ve más claramente en grandes olas, como en la que tenemos a comienzos de 2022.

Figura 37: Gráfico de detección temprana de anomalías (EARS) utilizando el método C3 con la variable “Carga vírica”

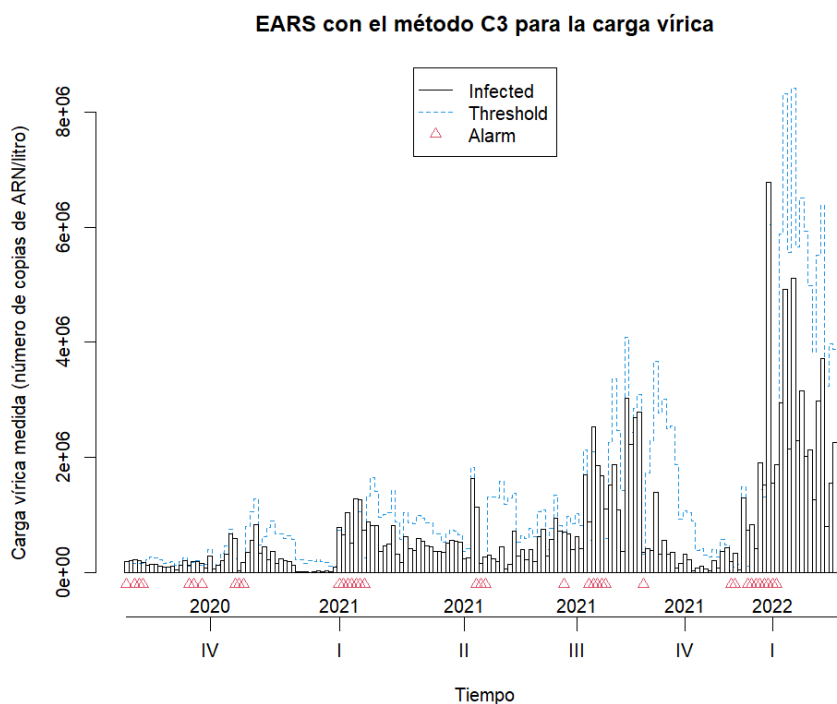


Gráfico construido mediante el Sistema de Detección Temprana de Anomalías (EARS) y el método C3. En negro tenemos representado la carga vírica medida en aguas residuales, en la línea azul discontinua se representa el umbral o límite de control y con unos triángulos rojos se representan las que serían las señales de alarma. El gráfico se elaboró con la función “earsC” del paquete “surveillance” de R y con el conjunto de datos cedido por COVIDBENS.

Conclusiones y líneas futuras

En este trabajo se ha llevado a cabo un análisis de la bibliografía acerca del Control Estadístico de Procesos y, más concretamente, de los Gráficos de Control. Se ha estudiado su uso en el ámbito de la salud pública y se han aplicado las técnicas a datos de las aguas residuales del área metropolitana de A Coruña.

En este análisis, hemos visto que se pueden aplicar los gráficos de control a datos de tipo epidemiológico, siempre y cuando tengamos cuidado con la autocorrelación de los mismos. Las variables epidemiológicas, dada su naturaleza, suelen depender de los datos correspondientes a momentos anteriores en el tiempo. En el ejemplo del número de casos, hemos observado que estos dependen, al menos, de la observación anterior. Por lo tanto, antes de construir los gráficos de control, ante una variable que presente autocorrelación, debemos asegurarnos de corregir la misma, como hemos realizado en este trabajo, a través del uso de los residuos del modelo ARIMA que mejor se ajuste a los datos.

Se observa que cada ola o brote es diferente, ya sea en tamaño o forma, además de sucederse continuamente sin apenas etapas en las que se pueda considerar que el proceso está controlado. Esto hace que la aplicación de gráficos de control estándar pueda no ser tan eficiente como precisaríamos. De hecho, se observa que es crítica la elección de la muestra de calibrado para estimar los límites de control, que serán fijos y que no se recalcularán hasta que el proceso cambie. Por tanto, la aplicación eficiente de este tipo de gráficos requeriría que hubiese un intervalo de tiempo en el que el proceso o situación sanitaria se considerase bajo control, además del sucesivo recálculo de los límites de control siempre y cuando la situación se estabilizase a un nivel diferente. A todas estas dificultades se une la dificultad añadida para interpretar unos gráficos de control contruidos a partir de los residuos de un modelo de ajuste, no a partir de las variables originales, siendo por tanto menos intuitivo e interpretable para profesionales y autoridades sanitarias y políticas. Por tanto, se observa la necesidad del empleo de nuevas alternativas de gráficos aplicadas a las variables originales y que presenten límites de control variables para cada instante de tiempo.

Para evitar estos problemas, entre otros, se ha visto que existen gráficos específicos para datos epidemiológicos, basados en el Control Estadístico de Procesos. Hemos aplicado tres de estos gráficos específicos a nuestros datos: dos gráficos basados en los gráficos de control de Shewhart y uno basado en los gráficos CUSUM. De estos tres

gráficos, se ha visto que el basado en el gráfico CUSUM (EARS con el método C3) detecta un mayor número de señales de alarma, al ser más sensible a los cambios pequeños, y las llega a detectar antes de que el número de casos comience a aumentar. Desde el punto de vista epidemiológico, este gráfico es mejor que los otros dos métodos ya que cualquier pequeño aumento en el número de casos, muertes, u otra variable de interés, se detectará como una alarma antes de que comiencen a aumentar de forma preocupante para la población. Además, también es el que detecta, tanto para el caso del estudio de la carga vírica como en el número de casos activos, un mayor número de alarmas reales. Es decir, al compararse con las olas y brotes reales de coronavirus, este es el gráfico que más olas ha conseguido detectar.

Los datos que hemos utilizado han sido cedidos por el proyecto COVIDBENS en el que utilizaban los datos de la carga vírica en el área metropolitana de A Coruña para detectar nuevas olas de coronavirus y poder alertar a la población antes de que el número de casos en la población aumentara. En estos últimos gráficos del paquete “surveillance” también hemos visto cómo, en su mayoría, el aumento de la carga vírica provoca señales de alarma antes de que se vea reflejado el aumento del número de casos activos, lo que justifica la monitorización de esta variable, por lo menos, para el caso de estudio del coronavirus.

En el futuro, podría ser interesante, dado que se ha justificado la monitorización de la carga vírica, reactivar el proyecto de COVIDBENS y el estudio de la misma. Dada la importancia que ha tenido este proyecto, sería interesante poder actualizarlo a otras enfermedades u otras zonas geográficas.

Se ha visto que existen nuevos métodos de estudio de variables epidemiológicas y algún gráfico surgido a raíz de la crisis de coronavirus en el mundo. En futuros trabajos, sería interesante implementar en R el gráfico híbrido de Shewhart que hemos estudiado en el **Capítulo 2** y su uso para el estudio del número de casos y la carga vírica del área metropolitana de A Coruña que hemos utilizado en el presente trabajo, ya que se ha visto que podría ser muy útil para el estudio del coronavirus y otras epidemias que puedan surgir en el futuro.

Apéndice A

Código de R utilizado

```
## -- EJEMPLO CON DATOS DE CASOS DIAGNOSTICADOS EN GALICIA PUNTO 2.3.3

diagnosticados <- read.csv2("diagnosticados_GA.csv")
attach(diagnosticados)

diagnosticados$FECHA <- as.Date(diagnosticados$FECHA)
diagnosticados$CASOS <- as.numeric(diagnosticados$CASOS)


plot(diagnosticados[53:145,], type="o", ylab="Número de casos",
xlab="2020", main="Número de casos de COVID-19 diagnosticados en
Galicia", lwd = 2, col = 2)

abline(v = diagnosticados[59,], lwd = 2, col = 4)
abline(v = diagnosticados[80,], lwd = 2, col = 4)
abline(v = diagnosticados[137,], lwd = 2, col = 4)

#####

## -- EJEMPLO CON DATOS DE COVIDBENS

##-- Lectura de datos

covidbens <- read.csv2("Casos_estimados_20220316.csv")
covidbens$Fecha <- as.Date(covidbens$Fecha, format = "%d/%m/%Y")
#Nos interesarán Carga_medida y Casos

attach(covidbens)

#####

#### PAQUETE QCC ####

###--- Gráficos de control para la variable CASOS ---###

library(qcc)
```



```

#- Gráfico de medidas individuales
qcc(Casos, type="xbar.one", restore.par=F,
    title = "Gráfico xbar.one para el número de casos activos por día",
    xlab = "Día", ylab = "Número de casos activos")

#- Gráfico EWMA
ewma(Casos, lambda = 0.2, nsigmas = 3, restore.par=F,
    title = "Gráfico EWMA para el número de casos activos por día",
    xlab = "Día", ylab = "Número de casos activos") #Revisar parámetros

#- Gráfico CUSUM
cusum(Casos, decision.interval = 5 ,restore.par=F,
    title = "Gráfico CUSUM para el número de casos activos por día",
    xlab = "Día", ylab = "Sumas acumuladas") #Revisar parámetros

#- Test de Ljung-Box (Contraste de independencia)
Box.test(Casos, lag = 1, type = "Ljung-Box") #Se rechaza independencia

##-- Modelo ARIMA --##
library(fpp2)
library(tseries)

casos_ts <- ts(Casos, start=c(2020,7), frequency=365)

```

```
auto.arima(casos_ts, d=NA, D=NA, lambda="auto", seasonal=FALSE,
stepwise=FALSE, approximation=FALSE, trace=TRUE) #Nos sugiere un
ARIMA(2,1,2)
```

```
ajuste_casos <- Arima(casos_ts, order=c(2,1,2), lambda=1,
include.constant=FALSE)
```

```
ajuste_casos #Sin constante porque no es significativamente distinta de
0
```

```
#Diagnosis de los residuos para ver si el modelo se ajusta a los datos
```

```
checkresiduals(ajuste_casos)
```

```
tsdiag(ajuste_casos)
```

```
t.test(ajuste_casos$residuals, mu=0) #Contraste de media nula
```

```
qqnorm(ajuste_casos$residuals)
```

```
qqline(ajuste_casos$residuals) #No son gaussianos
```

```
jarque.bera.test(ajuste_casos$residuals)
```

```
shapiro.test(ajuste_casos$residuals)
```

```
#El modelo escogido hace un buen ajuste de los datos, podemos utilizar
#los residuos para construir los gráficos de control.
```

```
###--- Repetimos los GRÁFICOS DE CONTROL, pero ahora con los RESIDUOS
```

```
casos_res <- ajuste_casos$residuals
```

```
#- Gráfico de medias para datos de medidas individuales
```

```
xbarone_casos_res <- qcc(casos_res, type="xbar.one", restore.par=F,
title = "Gráfico xbar.one para los residuos del número de casos", xlab
= "Día", ylab = "Residuos")
```

```
xbarone_casos_res$violations
```

```
#- Gráfico EWMA
```

```
ewma_casos_res <- ewma(casos_res, lambda = 0.2, nsigmas = 3,
restore.par=F, title = "Gráfico EWMA para los residuos del número de
casos", xlab = "Día", ylab = "Residuos")
```

```
ewma_casos_res$violations
```

```
#- Gráfico CUSUM
```

```
cusum_casos_res <- cusum(casos_res, decision.interval = 5,
restore.par=F, title = "Gráfico CUSUM para los residuos del número de
casos", xlab = "Día", ylab = "Sumas acumuladas")
```

```
cusum_casos_res$violations
```

```
## Representación de los residuos fuera de los límites en el gráfico de
## la variable "Número de casos"
```

```
plot(x = Fecha[1:184], y = Casos[1:184], col = 2,
type = "b", xlab="Fecha",
ylab = "Número de casos",
main = "Número de casos activos del COVID-19 en Galicia")
```

```
casos_fuera <- Casos[c(2,4,54,63,64,65,66,68,75,76,160,163,165,171)]
```

```
fecha_fuera <- Fecha[c(2,4,54,63,64,65,66,68,75,76,160,163,165,171)]
```

```

points(x = fecha_fuera, y = casos_fuera, col = 4, pch = 16)

###--- Calibramos una muestra inicial y monitorizamos las siguientes

#- Gráfico de medias para medidas individuales

xbarone_casos_cali <- qcc(casos_res[1:60], newdata = casos_res[61:184],
type="xbar.one", restore.par=F, title = "Gráfico xbar.one para los
residuos del número de casos", xlab = "Día", ylab = "Residuos")

xbarone_casos_cali$violations

#- Gráfico EWMA

ewma_casos_cali <- ewma(casos_res[1:60], newdata = casos_res[60:184],
lambda = 0.2, nsigmas = 3, restore.par=F, title = "Gráfico EWMA para
los residuos del número de casos", xlab = "Día", ylab = "Residuos")

ewma_casos_cali$violations

#- Gráfico CUSUM

cusum_casos_cali <- cusum(casos_res[1:60], newdata = casos_res[60:184],
decision.interval = 5 , restore.par=F, title = "Gráfico CUSUM para los
residuos del número de casos", xlab = "Día", ylab = "Sumas acumuladas")

cusum_casos_cali$violations

```

```

## Representación gráfica de estos puntos en el gráfico de dispersión
## del número de casos

plot(x = Fecha[1:184], y = Casos[1:184], col = 2,
     type = "b", xlab="Fecha",
     ylab = "Número de casos",
     main = "Número de casos activos del COVID-19 en Galicia")

casos_fuera1 <-
Casos[c(2,3,4,54,55,61,63,64,65,66,67,68,69,70,71,75,76,77,80,82,160,
161,162,163,164,165,166,167)]

fecha_fuera1 <-
Fecha[c(2,3,4,54,55,61,63,64,65,66,67,68,69,70,71,75,76,77,80,82,
160,161,162,163,164,165,166,167)]

points(x = fecha_fuera1, y = casos_fuera1, col = 4, pch = 16)

#####

##-- Gráficos de control para la variable CARGA VÍRICA MEDIDA

library(qcc)

#- Gráfico de medidas individuales

xbar_carga <- qcc(Carga_medida, type="xbar.one", restore.par=F, title =
"Gráfico xbar.one para la carga vírica medida por muestra (día)",
xlab="Día", ylab = "Carga vírica (número de copias de ARN del virus /
1)")

xbar_carga$violations

```

#- Gráfico EWMA

```
ewma(Carga_medida, lambda = 0.2, nsigmas = 3, restore.par=F, title =
"Gráfico EWMA para la carga vírica medida por muestra (día)",
xlab="Día", ylab = "Carga vírica (número de copias de ARN del virus /
1)")
```

#- Gráfico CUSUM

```
cusum(Carga_medida, restore.par=F, decision.interval = 5, title =
"Gráfico CUSUM para la carga vírica medida por muestra (día)",xlab =
"Día", ylab = "Sumas acumuladas")
```

##- Test de independencia

```
Box.test(Carga_medida, lag = 1, type = "Ljung-Box") #Se rechaza
independencia
```

##-- Modelo ARIMA --##

```
library(fpp2)
```

```
library(tseries)
```

```
carga_ts <- ts(Carga_medida, start=c(2020,7), frequency=365)
```

```
auto.arima(carga_ts, d=NA, D=NA, lambda="auto", seasonal=FALSE,
stepwise=FALSE, approximation=FALSE, trace=TRUE) #Nos sugiere un
ARIMA(2,1,2)
```

```
ajuste_carga <- Arima(carga_ts, order=c(1,1,4), lambda=1,
include.constant=F)
```

```
ajuste_carga #Sin constante porque no es significativamente distinta de
0
```

```

#Diagnosis de los residuos para ver si el modelo se ajusta a los datos
checkresiduals(ajuste_carga)
tsdiag(ajuste_carga)
t.test(ajuste_carga$residuals, mu=0) #Contraste de media nula
qqnorm(ajuste_carga$residuals)
qqline(ajuste_carga$residuals) #No son gaussianos
jarque.bera.test(ajuste_carga$residuals)
shapiro.test(ajuste_carga$residuals)

#El modelo escogido hace un buen ajuste de los datos, podemos utilizar
#los residuos para construir los gráficos de control.

###--- Repetimos los GRÁFICOS DE CONTROL, pero ahora con los RESIDUOS
carga_res <- ajuste_carga$residuals

#- Gráfico de medias para datos de medidas individuales
medias_carga_res <- qcc(carga_res, type="xbar.one", restore.par=F,
title = "Gráfico xbar.one para los residuos del modelo ARIMA ajustado a
la carga vírica", xlab = "Día", ylab = "Residuos")

medias_carga_res$violations

#- Gráfico EWMA
ewma_carga_res <- ewma(carga_res, lambda = 0.2, nsigmas = 3,
restore.par=F, title = "Gráfico EWMA para los residuos del modelo ARIMA
ajustado a la carga vírica", xlab = "Día", ylab = "Residuos")

```

```
ewma_carga_res$violations
```

```
#- Gráfico CUSUM
```

```
cusum_carga_res <- cusum(carga_res, decision.interval =
5,restore.par=F,title = "Gráfico CUSUM para los residuos del modelo
ARIMA ajustado a la carga vírica",xlab = "Día", ylab = "Sumas acumuladas
de los residuos")
```

```
cusum_carga_res$violations
```

```
## Representación de estos puntos en el gráfico de dispersión de la
carga vírica
```

```
plot(x = Fecha[1:184], y = Carga_medida[1:184], col = 4,
     type = "o", xlab="Fecha",
     ylab = "Carga vírica (número de copias de ARN del virus / l)",
     main = "Carga vírica del COVID-19 en Galicia")
```

```
carga_fuera <-
Carga_medida[c(124,125,132,134,135,136,137,138,139,140,141,142,143,
144,146,147,148,149,150,151,152,153,154,166,167,168,169,170,171,172,
173,174,175,176,177,178,179,180,181,182,183,184)]
```

```
fecha_fuera2 <-
Fecha[c(124,125,132,134,135,136,137,138,139,140,141,142,143,144,
146,147,148,149,150,151,152,153,154,166,167,168,169,170,
171,172,173,174,175,176,177,178,179,180,181,182,183,184)]
```

```
points(x = fecha_fuera2, y = carga_fuera, col = 2, pch = 16)
```



```

###--- Calibramos una muestra inicial y monitorizamos las siguientes

#- Gráfico de medias para medidas individuales

xbar_carga_cali <- qcc(carga_res[1:100], newdata = carga_res[100:184],
type="xbar.one", restore.par=F, title = "Gráfico xbar.one para los
residuos de la carga vírica con una muestra de calibrado [1:100]", xlab
= "Día", ylab = "Residuos")

xbar_carga_cali$violations

#- Gráfico EWMA

ewma_carga_cali <- ewma(carga_res[1:100], newdata = carga_res[100:184],
lambda = 0.2,nsigmas = 3, restore.par=F,title = "Gráfico EWMA para los
residuos de la carga vírica con una muestra de calibrado [1:100]",xlab
= "Día", ylab = "Residuos")

ewma_carga_cali$violations

#- Gráfico CUSUM

cusum_carga_cali <- cusum(carga_res[1:100], newdata =
carga_res[100:184], decision.interval = 5 ,restore.par=F, title =
"Gráfico CUSUM para los residuos de la carga vírica con una muestra de
calibrado [1:100]", xlab = "Día", ylab = "Sumas acumuladas de los
residuos")

cusum_carga_cali$violations

### Representación de los puntos FC en el gráfico de dispersión

```

```
plot(x = Fecha[1:184], y = Carga_medida[1:184], col = 4,
     type = "o", xlab="Fecha",
     ylab = "Carga vírica (número de copias de ARN del virus / l)",
     main = "Carga vírica del COVID-19 en Galicia")
```

```
carga_fuera1 <-
```

```
Carga_medida[c(63,64,65,66,67,68,69,70,71,73,74,75,77,78,95,96,
97,99,100,101,102,103,104,105,106,123,124,125,126,
127,128,129,130,131,132,133,134,135,136,137,138,
139,140,141,142,143,144,145,146,147,148,149,150,
151,152,153,154,155,156,157,158,159,160,161,162,
163,164,165,166,167,168,169,170,171,172,173,174,
175,176,177,178,179,180,181,182,183,184)]
```

```
fecha_fuera3 <-
```

```
Fecha[c(63,64,65,66,67,68,69,70,71,73,74,75,77,78,95,96,
97,99,100,101,102,103,104,105,106,123,124,125,126,
127,128,129,130,131,132,133,134,135,136,137,138,
139,140,141,142,143,144,145,146,147,148,149,150,
151,152,153,154,155,156,157,158,159,160,161,162,
163,164,165,166,167,168,169,170,171,172,173,174,
175,176,177,178,179,180,181,182,183,184)]
```

```
points(x=fecha_fuera3, y = carga_fuera1, col =2, pch = 16)
```

```
#####
```

```
#### PAQUETE SURVEILLANCE ####
```

```
#####
```

```
library(surveillance)
```

```
##-- Gráfico de detección de anomalías (EARS) variable CASOS

casos_sts <- sts(Casos, start = c(2020,7), frequency = 365,
               epochAsDate = TRUE, epoch = as.numeric(Fecha))

#-Método C1 (Gráfico de Shewhart)

monitoriz_casos1 <- earsC(casos_sts, control = list(method = "C1"))

plot(monitoriz_casos1, main = "EARS con el método C1 para el número de
casos", ylab = "Número de casos activos", xlab= "Tiempo")

#-Método C2 (Gráfico de Shewhart)

monitoriz_casos2 <- earsC(casos_sts, control = list(method = "C2"))

plot(monitoriz_casos2, main = "EARS con el método C2 para el número de
casos", ylab = "Número de casos activos", xlab= "Tiempo")

#-Método C3 (CUSUM)

monitoriz_casos3 <- earsC(casos_sts, control = list(method = "C3"))

plot(monitoriz_casos3, main = "EARS con el método C3 para el número de
casos", ylab = "Número de casos activos", xlab= "Tiempo")

##-- Gráfico de detección de brotes variable CARGA VÍRICA

carga_sts <- sts(Carga_medida, start = c(2020,7), frequency = 365,
               epochAsDate = TRUE, epoch = as.numeric(Fecha))

#-Método C1 (Gráfico de Shewhart)

monitoriz_carga1 <- earsC(carga_sts, control = list(method = "C1"))
```

```
plot(monitoriz_carga1, main = "EARS con el método C1 para la carga
vírica", ylab = "Carga vírica medida (número de copias de ARN/litro)",
xlab = "Tiempo")
```

```
#-Método C2 (Gráfico de Shewhart)
```

```
monitoriz_carga2 <- earsC(carga_sts, control = list(method = "C2"))
plot(monitoriz_carga2, main = "EARS con el método C2 para la carga
vírica", ylab = "Carga vírica medida (número de copias de ARN/litro)",
xlab = "Tiempo")
```

```
#-Método C3 (CUSUM)
```

```
monitoriz_carga3 <- earsC(carga_sts, control = list(method = "C3"))
plot(monitoriz_carga3, main = "EARS con el método C3 para la carga
vírica", ylab = "Carga vírica medida (número de copias de ARN/litro)",
xlab = "Tiempo")
```

```
#####
```

Bibliografía

- Adab, P., Rouse, A. M., Mohammed, M. A., & Marshall, T. (2002). Performance league tables: The NHS deserves better. *BMJ*, 324(7329), 95-98.
<https://doi.org/10.1136/bmj.324.7329.95>
- Aldama, Z., Barrena, J. C., & Ribera, A. (2020, abril 28). *Coronavirus | ¿Cómo es la desescalada en otros países?* El Comercio. <https://www.elcomercio.es/sociedad/como-es-desescalada-en-otros-paises-20200428001152-ntvo.html>
- Barbeito, I., Zaragoza, S., Tarrío-Saavedra, J., & Naya, S. (2017). Assessing thermal comfort and energy efficiency in buildings by statistical quality control for autocorrelated data. *Applied Energy*, 190, 1-17. <https://doi.org/10.1016/j.apenergy.2016.12.100>
- Benneyan, J. (1998). Statistical quality control methods in infection control and hospital epidemiology, part I: Introduction and basic theory. *Infection control and hospital epidemiology*. <https://doi.org/10.2307/30143442>
- Biobot Analytics · About. (s. f.). Biobot Analytics. Recuperado 11 de agosto de 2022, de <https://biobot.io/about/>
- Bolsin, S., & Colson, M. (2000). The use of the Cusum technique in the assessment of trainee competence in new procedures. *International Journal for Quality in Health Care: Journal of the International Society for Quality in Health Care*, 12(5), 433-438.
<https://doi.org/10.1093/intqhc/12.5.433>
- Casado, I. (2022, enero 20). Diferencias entre Test Antígenos de laboratorio y farmacia | Fly COVID Test. *Fly Covid Test Center*. <https://flycovidtestcenter.com/diferencias-entre-los-test-de-antigenos-de-laboratorio-y-de-autodiagnostico/>
- COVIDBENS. (2022a). COVIDBENS. Detección del COVID-19 en las aguas residuales. *Edar Bens*. <https://edarbens.es/covid19/>
- COVIDBENS. (2022b). Proyecto COVIDBENS: variantes encontradas. *Edar Bens*. <https://edarbens.es/covid19/variantes-encontradas/>

- Fortea-Sanchis, C., & Escrig-Sos, J. (2019). Técnicas de control de calidad en cirugía. Aplicación de las gráficas de control cumulative sum. *Cirugía Española*, 97(2), 65-70. <https://doi.org/10.1016/j.ciresp.2018.11.008>
- Frisén, M. (2009). Optimal Sequential Surveillance for Finance, Public Health, and Other Areas. *Sequential Analysis*, 28(3), 310-337. <https://doi.org/10.1080/07474940903041605>
- García, M. (2020, marzo 12). *El cierre de escuelas para frenar el coronavirus será generalizado en toda España*. La Vanguardia. <https://www.lavanguardia.com/vida/20200312/474101012713/cierre-escuelas-espana.html>
- Goldman, E. (2020). Exaggerated risk of transmission of COVID-19 by fomites. *The Lancet Infectious Diseases*, 20(8), 892-893. [https://doi.org/10.1016/S1473-3099\(20\)30561-2](https://doi.org/10.1016/S1473-3099(20)30561-2)
- Güell, O. (2020, enero 30). La OMS declara la alerta internacional ante la imparable expansión del coronavirus de Wuhan. *El País*. https://elpais.com/sociedad/2020/01/30/actualidad/1580399024_626205.html
- Höhle, M., Meyer, S., Paul, M., Held, L., Burkom, H., Correa, T., Hofmann, M., Lang, C., Manitz, J., Riebler, A., Bové, D. S., Salmon, M., Schumacher, D., Steiner, S., Virtanen, M., Wei, W., Wimmer, V., & R), R. C. T. (A few code segments are modified versions of code from base. (2022). *surveillance: Temporal and Spatio-Temporal Modeling and Monitoring of Epidemic Phenomena* (1.20.0). <https://CRAN.R-project.org/package=surveillance>
- Inkelas, M., Blair, C., Furukawa, D., Manuel, V. G., Malenfant, J. H., Martin, E., Emeruwa, I., Kuo, T., Arangua, L., Robles, B., & Provost, L. P. (2021). Using control charts to understand community variation in COVID-19. *PLOS ONE*, 16(4), e0248500. <https://doi.org/10.1371/journal.pone.0248500>
- Johnson, N. L. (1961). A Simple Theoretical Approach to Cumulative Sum Control Charts. *Journal of the American Statistical Association*, 56(296), 835-840. <https://doi.org/10.2307/2281998>

- Johnson, N. L., & Leone, F. C. (1962). Cumulative Sum Control Charts—Mathematical Principles Applied to Their Construction and Use, Part I, II and III. *Industrial Quality Control*, 19(2).
- La Moncloa. (2020a, abril 28). *La Moncloa. Plan de desescalada [Consejo de Ministros]*.
<https://www.lamoncloa.gob.es/consejodeministros/Paginas/enlaces/280420-enlace-desescalada.aspx>
- La Moncloa. (2020b, mayo 9). *La Moncloa. Estado de Alarma*.
<https://www.lamoncloa.gob.es/covid-19/Paginas/estado-de-alarma.aspx>
- Larraz, I. (2020, marzo 13). *El confinamiento de Igualada, Vilanova del Camí, Santa Margarida de Montbui y Òdena. Preguntas y respuestas*. Newtral. <https://www.newtral.es/el-confinamiento-de-igualada-vilanova-del-cami-santa-margarida-de-montbui-y-odena-preguntas-y-respuestas/20200313/>
- Lema, I. (2020, junio). *Análisis de coronavirus en superficies*. Revista AENOR.
<https://revista.aenor.com/359/analisis-de-coronavirus-en-superficies.html>
- Levey, S., & Jennings, E. R. (1950). The use of control charts in the clinical laboratory. *American Journal of Clinical Pathology*, 20(11), 1059-1066.
https://doi.org/10.1093/ajcp/20.11_ts.1059
- Llaneras, K., Linde, P., & Zafra, M. (2022, enero 6). *Pruebas de antígenos con ómicron: Cuándo son más fiables, cómo interpretarlas y cuándo sospechar de un falso negativo*. El País. <https://elpais.com/sociedad/2022-01-06/pruebas-de-antigenos-con-omicron-cuando-son-mas-fiables-como-interpretarlas-y-cuando-sospechar-de-un-falso-negativo.html>
- MANDAL, S., ROYCHOWDHURY, T., & BHATTACHARYA, A. (2021). Pattern of genomic variation in SARS-CoV-2 (COVID-19) suggests restricted nonrandom changes: Analysis using Shewhart control charts. *Journal of Biosciences*, 46(1), 11.
<https://doi.org/10.1007/s12038-020-00131-5>

- Medema, G., Heijnen, L., Elsinga, G., Italiaander, R., & Brouwer, A. (2020). Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental Science & Technology Letters*, 7(7), 511-516. <https://doi.org/10.1021/acs.estlett.0c00357>
- Real Decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el COVID-19, Pub. L. No. Real Decreto 463/2020, BOE-A-2020-3692 25390 (2020).
<https://www.boe.es/eli/es/rd/2020/03/14/463>
- Ministerio de Sanidad. (2020a). *Actualización nº13. Neumonía por nuevo coronavirus (2019-nCoV) en Wuhan, provincia de Hubei, (China)*.
https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/Actualizacion_13_2019-nCoV_China.pdf
- Ministerio de Sanidad. (2020b). *Actualización nº 14. Neumonía por nuevo coronavirus 2019 n-CoV en Wuhan, provincia de Hubei, (China)*.
- Ministerio de Sanidad. (2021). *Vigilancia epidemiológica—Ambiental de las aguas residuales: SARS-CoV-2*.
- Montgomery, D. C. (2009). *Introduction to Statistical Quality Control* (6.^a ed.). John Wiley & Sons.
- Noyez, L. (2009). Control charts, Cusum techniques and funnel plots. A review of methods for monitoring performance in healthcare. *Interactive CardioVascular and Thoracic Surgery*, 9(3), 494-499. <https://doi.org/10.1510/icvts.2009.204768>
- Organización Mundial de la Salud. (2020, marzo 11). *Alocución de apertura del Director General de la OMS en la rueda de prensa sobre la COVID-19*.
<https://www.who.int/es/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020>
- Page, E. S. (1954). Continuous Inspection Schemes. *Biometrika*, 41(1/2), 100-115.
<https://doi.org/10.2307/2333009>

- Parry, G., Provost, L. p, Provost, S. m, Little, K., & Perla, R. j. (2021). A hybrid Shewhart chart for visualizing and learning from epidemic data. *International Journal for Quality in Health Care*, 33(4). <https://doi.org/10.1093/intqhc/mzab151>
- Perla, R. J., Provost, S. M., Parry, G. J., Little, K., & Provost, L. P. (2021). Understanding variation in reported covid-19 deaths with a novel Shewhart chart application. *International Journal for Quality in Health Care*, 33(1), mzaa069. <https://doi.org/10.1093/intqhc/mzaa069>
- Prat Bartés, A., Tort-Martorell Llabrés, X., & Grima Cintas, P. (2015). *Métodos estadísticos: Control y mejora de la calidad*. Universitat Politècnica de Catalunya. <https://elibro-net.accedys.udc.es/es/ereader/bibliotecaudc/61421>
- Roberts, S. W. (1959). Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1(3), 239-250. <https://doi.org/10.1080/00401706.1959.10489860>
- Roy, D., Tripathy, S., Kar, S. K., Sharma, N., Verma, S. K., & Kaushal, V. (2020). Study of knowledge, attitude, anxiety & perceived mental healthcare need in Indian population during COVID-19 pandemic. *Asian Journal of Psychiatry*, 51, 102083. <https://doi.org/10.1016/j.ajp.2020.102083>
- Salmon, M., Schumacher, D., & Höhle, M. (2016). Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. *Journal of Statistical Software*, 70 (10), 1-35. <https://doi.org/10.18637/jss.v070.i10>
- Santos, M. (2020, abril 16). *El estudio de las aguas residuales facilitará la vigilancia epidemiológica frente al COVID*. EL ÁGORA DIARIO. <https://www.elagoradiario.com/agua/estudio-aguas-residuales-vigilancia-epidemiologica-covid/>
- Shortell, S. M., Bennett, C. L., & Byck, G. R. (1998). Assessing the Impact of Continuous Quality Improvement on Clinical Practice: What It Will Take to Accelerate Progress. *The Milbank Quarterly*, 76(4), 593-624. <https://doi.org/10.1111/1468-0009.00107>

- Snyder, A., & Cullinane, S. (2020, abril 26). *Some scientists are using sewage to measure the prevalence of coronavirus in their communities*. CNN.
<https://www.cnn.com/2020/04/26/us/covid-19-sewage-testing/index.html>
- Spiegelhalter, D. (2002). Funnel plots for institutional comparison. *Quality and Safety in Health Care*, 11(4), 390-391. <https://doi.org/10.1136/qhc.11.4.390-a>
- Useros, V. (2020, marzo 4). *Sanidad ordena cancelar todos los congresos y jornadas del personal sanitario y médico*. El Mundo. <https://www.elmundo.es/comunidad-valenciana/2020/03/04/5e5ecf2dfdddf97848b45ef.html>
- Vallejo, J. A., Trigo-Tasende, N., Rumbo-Feal, S., Conde-Pérez, K., López-Oriona, Á., Barbeito, I., Vaamonde, M., Tarrío-Saavedra, J., Reif, R., Ladra, S., Rodiño-Janeiro, B. K., Nasser-Ali, M., Cid, Á., Veiga, M., Acevedo, A., Lamora, C., Bou, G., Cao, R., & Poza, M. (2022). Modeling the number of people infected with SARS-COV-2 from wastewater viral load in Northwest Spain. *Science of The Total Environment*, 811, 152334.
<https://doi.org/10.1016/j.scitotenv.2021.152334>
- Vargas, C. M., Acosta, R. G., & Bernilla, A. T. (2020). El nuevo Coronavirus y la pandemia del Covid-19. *Revista Medica Herediana*, 31(2), 125-131.
<https://doi.org/10.20453/rmh.v31i2.3776>
- Vidal Liy, M. (2020, febrero 24). La OMS, tras visitar la 'zona cero' del coronavirus, ve "real" el descenso de contagios en China. *El País*.
https://elpais.com/sociedad/2020/02/24/actualidad/1582558456_859856.html
- Western Electric. (1956). *Statistical Quality Control Handbook* (1.^a ed.). Western Electric Company.
- Woodall, W. (2006). The Use of Control Charts in Health-Care and Public-Health Surveillance. *Journal of Quality Technology*, 38, 89-104.
<https://doi.org/10.1080/00224065.2006.11918593>
- Wortham, A. W., & Ringer, L. J. (1971). *Control via exponential smoothing*.

