

Universidad Santiago de Chile
Facultad de Ciencia
Departamento de Matemática y
Ciencia de la Computación



SISTEMA DE MONITOREO DE LA PANDEMIA DE COVID-19 EN CHILE MEDIANTE ANÁLISIS DE CHANGE-POINT

Pedro Ángel Osorio Meza

Profesor Guía:

Felipe Elorrieta López

SANTIAGO - CHILE
2021

Resumen

El **G**rupo **EM**atemático para la **V**igilancia de **EP**andemias de la Universidad de Santiago (GEMVEP-USACH) busca crear un sistema de monitoreo en Chile para epidemias, utilizando los datos del Covid-19, con la intención de entregar de manera oportuna y eficaz la información estadística necesaria que permita dar conocimiento a las autoridades encargadas de la toma de decisiones como también a la población.

Se dispone de un sistema de alerta utilizando Clusters con las variables latentes de casos nuevos de Covid-19 reportados por el ministerio de salud realizando una clasificación de riesgo al contagio en tres niveles distintos para las regiones de Chile la cual es monitorizada por integrantes del grupo.

Los Clusters Regionales son obtenidos a través de componentes principales utilizando las variables: Tasa de Positividad, Tasa de Reproducción, Tasa de Incidencia y variaciones semanales.

Éste trabajo propone incorporar dos métodos distintos de 'Change-Point' para incluir los resultados como nuevas variables al sistema de alerta con clusters, además de otros indicadores que permitan hacer más robusto éste análisis.

Palabras clave: Change-Point, Clusters, Sistema de alerta, Covid-19.

Dedicatoria

Me gustaría agradecer a la Universidad de Santiago que fue donde pasé por mis mejores años de juventud, también aquí fue donde conocí a Nayra Muñoz quién me dio fuerzas para volver a estudiar y juntos tuvimos una hija Martina, que a ella no la detuvo para terminar como Bioquímica y para mí fue mi motivación diaria.

También quisiera agradecer a los profesores que apoyaron y dieron vida a ésta tesis Gloria Icaza, Claudio Vargas y Felipe Elorrieta profesor guía, quién me dío la seguridad de seguir estudiando cuando vi todo cuesta arriba, y que cuando obtuve la Licenciatura me acogió en el GEMVEP-USACH para realizar práctica, donde pude explotar mis conocimientos.

Agradezco a mis padres y hermanos que me apoyaron, también a las personas que ya no están conmigo y que creyeron en mí.

Pasaron años para poder terminar éste ciclo, donde hubieron momentos difíciles, pero estoy muy contento y agradecido por llegar a este punto.

Nunca es tarde para cumplir los sueños y cambiar las estrellas.

Tabla de contenido

Índice de figuras	v
Índice de cuadros	vii
1. Introducción	1
2. Objetivos	3
1. Objetivo General	3
2. Objetivos específicos	3
3. Metodología	5
4. Marco Teórico	6
1. Conceptos Relevantes	6
1.1. Definiciones	6
1.2. Tasa de Positividad	8
1.3. Tasa de Incidencia	8
1.4. Tasa de Reproducción Instantánea	8
1.5. Componentes Principales	9
1.6. Clusters y validación	10
1.7. Conceptos generales de vigilancia estadística	11
1.8. Change-Point	12
1.9. Modelo Lineal Generalizado	13
1.10. Regresión Poisson	13
1.11. GLM Binomiales Negativos	15

1.12.	AIC	15
1.13.	Deviance	15
1.14.	Cartas de Control de Walter Shewhart	16
5.	Análisis Descriptivo	17
6.	Métodos de Change-Point	22
1.	Using control charts to understand community variation in COVID-19	22
1.1.	Procedimiento	23
1.2.	Aplicación para los datos de Chile.	27
2.	Automatic Selection of Models Outlier Detection for Epidemics (ASMODEE)	34
2.1.	Procedimiento	34
2.2.	Aplicación en los datos de Chile	40
7.	Sistema de Alerta para las regiones de Chile	46
1.	Clusters de Alerta	51
8.	Clonclusión	54
Referencias		55
1.	Gráficos de control primer método	58
2.	Gráficos de control segundo método	61

Índice de figuras

4.1.	Gráfico de control de Shewhart	16
5.1.	Cantidad de población por región	18
5.2.	Tasa de incidencia por región	19
5.3.	Calendario con tasa de incidencia reportada diariamente	20
5.4.	Incidencia y positividad reportada diariamente	20
5.5.	Calendario con exámenes PCR	21
6.1.	Épocas asociadas con muertes diarias reportadas por COVID-19	23
6.2.	Cambio de gráfica de control C-Chart a I-Chart.	24
6.3.	Épocas asociadas con casos diarias en Chile reportadas por COVID-19 . .	28
6.4.	Épocas de casos diarios y fallecidos en Chile reportadas por COVID-19 . .	28
6.5.	Grafico de Control de muertes a nivel nacional	29
6.6.	Casos reportados a nivel Nacional y la Región Metropolitana	30
6.7.	Región de Atacama y Valparaíso	31
6.8.	Regiones de Bío-bío y Aysén	31
6.9.	Grafico de Control de casos nuevos para la comuna de Caldera	32
6.10.	Grafico de Control de casos nuevos para la comuna de Pinto	32
6.11.	Grafico de Control de casos nuevos para las comunas de la Región Metro- politana	33
6.12.	Ejemplo ASMODEE	37
6.13.	Ejemplo ASMODEE con incidence2	38
6.14.	ASMODEE con casos a nivel Nacional	40
6.15.	Gráfico de ASMODEE para las Zonas de Chile	41
6.16.	Ejemplo ASMODEE Regiones de Chile	42

6.17. Ejemplo utilizando shiny	42
6.18. Comunas de la Región Metropolitana	43
6.19. Clasificación en el tiempo a nivel Nacional	44
6.20. Cambios de tendencia Incidencia RM	45
6.21. Cambios de tendencia en Incidencia Región del Maule	45
7.1. Gráfico de seguimiento de casos de Covid-19 en Chile actualizado al 2 de Noviembre del 2021	47
7.2. Grupos de Alerta regional en los días 6 de octubre de 2020 y 9 de abril de 2021	48
7.3. Resumen de indicadores diarios	49
7.4. La primera figura corresponde Variaciones semanales Positividad v/s Incidencia y la segunda figura corresponde a la interacción Incidencia v/s R efectivo	50
7.5. Correlación entre las variables	51
7.6. Cluster-Point	53
8.1. Cambios de tendencia nivel Nacional ajustado	58
8.2. Gráficos de control, región de Arica	58
8.3. Gráficos de control, región de Ñuble	59
8.4. Cambios de tendencia con gráficos de control	59
8.5. Cambios de tendencia con gráficos de control para todas las regiones	60
8.6. ASMODEE completo con todos los datos	61
8.7. Cambios de tendencia	62
8.8. Cambios de tendencia	62
8.9. Cambios de tendencia	64

Índice de cuadros

5.1. Resumen de variables para las zonas de Chile	17
5.2. Resumen de variables para las regiones de Chile	18
6.1. Resultados de ajuste para las zonas del país.	41
7.1. Correlación con la variable potencia	51
7.2. Resultados de los cambios de tendencia	52
7.3. Cluster-Point	53
8.1. Resumen de indicadores del 23 de noviembre	63

1

Introducción

Durante la última década, diversos estudios de genómica ambiental han establecido que las entidades biológicas más comunes y abundantes en la biosfera son los virus [1]. Estos son agentes biológicos microscópicos e infecciosos de naturaleza particulada y organización acelular, carentes de metabolismo propio, por lo que requieren de una célula hospedadora para poder replicarse, siendo un parásito intracelular obligado [2].

De esta manera, diversas enfermedades virales han sido relacionadas a conceptos epidemiológicos como brote, epidemia o pandemia. La Organización Mundial de la Salud (OMS), define un brote como la aparición repentina e inesperada de una enfermedad en un foco o segmento limitado de la población, mientras que epidemia es la propagación activa de la enfermedad producto de un descontrol en un determinado brote, aumentando el número de casos en un área geográfica concreta, lo cual se mantiene en el tiempo. Por otra parte, para que se declare estado de pandemia se tienen que cumplir dos criterios: que el brote epidémico afecte a más de un continente y que los casos de cada país ya no sean importados sino provocados por transmisión comunitaria, existiendo así una 'propagación mundial' de dicha enfermedad [3].

Desde la antigüedad, las pandemias provocadas por virus de la influenza han representado la mayor amenaza de calamidad universal causada por una enfermedad infecciosa, siendo una de las pandemias más importantes y con más alta mortalidad del siglo XX, la denominada gripe española, desarrollada entre 1918 y 1920 . Aún no se sabe cuál fue el verdadero origen de ésta, algunos investigadores afirman que empezó en Francia en 1916 o en China en 1917, sin embargo, muchos estudios sitúan los primeros casos en la base militar de Fort Riley (EE.UU.) el 4 de marzo de 1918 [4]. Por otra parte, se sabe que causó la muerte de entre 20 y 100 millones de personas en el mundo, cifra que supera fácilmente el número de muertos asociados a la Primera Guerra Mundial. Cien años después, esta

pandemia no solo es recordada como uno de los desastres más devastadores en la historia de la salud humana, sino que también representa uno de los estudios de caso más importantes disponibles en la actualidad, permitiendo grandes avances en la comprensión de la dinámica de transmisión de las enfermedades y en el análisis del comportamiento de futuras pandemias [5].

Actualmente nos encontramos enfrentando la Pandemia de Covid-19 en el mundo, que inició en Wuhan, China al término del año 2019 y que tuvo su primer caso en Chile el 3 de Marzo del año 2020. Fue declarado Pandemia por la Organización Mundial de la Salud (OMS) el 11 de marzo del año 2020, el denominado 2019 nCoV, que se propaga rápidamente de persona a persona, produce un cuadro respiratorio febril, con síntomas generales, rinorrea, tos intensa y disnea, pueden existir vómitos y diarreas, y donde entre 10 % a 25 % aproximadamente de los casos, según reportes iniciales, sufren un Síndrome de dificultad respiratoria aguda o grave por neumonía severa que puede llevar a un fallo de órganos multisistémico con una letalidad elevada. Los casos más severos se dan en personas con edades avanzadas y pacientes con comorbilidades [6].

Hasta el momento, producto de la pandemia COVID-19 han habido más de 5 millones de fallecidos en el mundo [7]. En Chile, se han alcanzado ya los 49.567 decesos entre confirmados y sospechosos producto de esta enfermedad [8].

Debido a lo difícil que es entregar una alerta temprana debido al tiempo que demora en ser confirmado un caso como positivo. El siguiente proyecto se enfoca en desarrollar un sistema de alerta temprana que permita llevar a las autoridades y a la comunidad la información necesaria para mantener un seguimiento y un control más óptimo de esta pandemia en el país, y cuidando así de la población y en especial de los grupos de riesgo. Para esto, se utilizarán gráficos de control basados en los datos obtenidos día a día por la pandemia COVID-19 en Chile, aplicando dos métodos distintos de 'Change Point' [9], el cual nos permitirá observar y determinar cambios significativos en las tendencias, identificando así distintos escenarios con respecto a la propagación del virus, ya sea la disminución de casos de COVID-19 o posibles nuevos brotes en lugares específicos.

Éstos resultados serán útiles para incorporar en la vigilancia realizada a través de los clusters que entregan alertas por el **Grupo Epidemiológico Matemático para la Vigilancia de Epidemias y Pandemias de la Universidad de Santiago de Chile (GEMVEP-USACH)** a la comunidad a través de sus redes sociales y su página web [10]. Éstos entregan tres niveles distintos de riesgo para las regiones: Alto, Medio y Bajo.

Por lo tanto éste análisis entrega un sistema de alertas robusto, oportuno y sincero que puede ser implementado para cualquier enfermedad.

2

Objetivos

A continuación se presenta el objetivo principal, así como también los objetivos específicos que se desarrollarán a lo largo de éste trabajo.

1. Objetivo General

- El objetivo principal es desarrollar un sistema de monitoreo y alerta temprana para la pandemia de COVID-19 en Chile, el cual permita comunicar a las comunas y regiones del país de posibles nuevos brotes de ésta enfermedad, en dónde se incorporarán resultados de métodos de 'Change-Point' a una clasificación de riesgo con análisis de Clusters.

2. Objetivos específicos

Para cumplir con el objetivo principal que se plantea se llevarán a cabo los siguientes objetivos específicos:

- Realizar una revisión bibliográfica de los conceptos involucrados.
- Hacer un análisis descriptivo de la base de datos, considerando las distintas zonas, regiones y comunas del país.

- Implementar dos métodos de Change-Point.^a las series de contagios de las distintas regiones y comunas del país con enfoque en los casos nuevos reportados con el fin de entregar la información lo más rápido posible.
- Comparar los resultados obtenidos de los métodos de 'Change-Point' aplicados.
- Identificar cuál de los métodos resulta más efectivo para su implementación y desarrollo o bien complementar ambos resultados.
- Determinar variables que resulten clave en el desarrollo de la pandemia para incorporar a la vigilancia a través de una clasificación mediante clustering para las regiones.
- Con los resultados de 'Change-Point' más la clasificación de Clusters, diseñar un sistema que de alerta del riesgo al contagio para las diferentes regiones y comunas del país.

3

Metodología

Como el objetivo pretende encontrar alertas para las regiones y comunas del país, se requiere tener la base de datos con el número de pacientes contagiados día a día, conociendo así el comportamiento de las frecuencias de casos y respectivas tasas involucradas. Por lo tanto es importante explicar los conceptos que estarán involucrados previo al análisis logrando mayor compresión al respecto.

Se realizarán dos análisis de 'Change-Point' a las series de tiempo involucradas que indiquen en tiempo real si existe riesgo de crecimiento de casos o no, para así complementar con el análisis de Cluster que basado en componentes principales entrega alertas según riesgo al contagio para cada región y comuna del país utilizando indicadores latentes de la pandemia (Incidencia, Refectivo, Positividad y variaciones semanales).

Una de las dos metodologías basadas en 'Change-point' se encuentra expuesta en el artículo: 'Using control charts to understand community variation in COVID-19' [11] en donde utilizan un gráfico de control híbrido desarrollado por Perla et al.[12] aplicado a las series de descensos, mientras que la otra metodología fue propuesta por la Organización Mundial de la Salud, denominada como ASMODEE: 'Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection'.[13], éstas se aplicarán con el propósito de encontrar quiebres de tendencia, disminución significativa de casos o brotes epidémicos.

Éstas dos se compararán con el fin de seleccionar o bien complementar para incluir en la entrega de alertas de Clusters de riesgo para las regiones y comunas del país que son divididas en tres tipos de riesgo. Por otra parte, los procedimientos de realizarán usando el software estadístico R [14]. Los detalles técnicos de las tasas y métodos se incluyen a continuación.

Marco Teórico

Para ésta sección se ha recopilado la información bibliográfica necesaria, para la comprensión total del problema ya que es importante para la correcta implementación de un sistema de alerta que estará enfocada a la enfermedad del Covid-19.

1. Conceptos Relevantes

Resulta importante detallar los conceptos epidemiológicos que se abordarán en el desarrollo de éste proyecto como lo son las tasas involucradas, definiciones, entre otros conceptos.

1.1. Definiciones

- Epidemiología: Estudio de la ocurrencia, distribución y control de las enfermedades (epi= sobre, demos= pueblo; población, logos=estudio).
- Virus: Agentes biológicos más abundantes en la biosfera, requieren de una célula hospedadora para poder replicarse.
- Brote Epidémico: Aparición repentina e inesperada de una enfermedad infecciosa en un lugar específico y en un momento determinado.
- Rebrote: Al seguir los comportamientos de las olas que tiene el país, luego de ocurrido un brote en la caída y creencia de disminución de casos nuevos se produce una nueva alza que provoca un nuevo peak de casos.
- Epidemia: Se cataloga como epidemia a una enfermedad que se propaga rápida y activamente con lo que el número de casos aumenta significativamente, aunque se mantiene en un área geográfica concreta.

- Pandemia: Cuando un brote epidémico afecta a más de un continente y los casos de cada país ya no son importados sino provocados por transmisión comunitaria.
- Sindemia: Se habla de este término cuando las causas de muertes en la población son causadas por más de una enfermedad, ya que combina los conceptos de sinergia y endemias. El Covid-19 fue nominado de esta manera por los científicos investigadores ya que afectaron a las personas más vulnerables que poseían enfermedades de base, aún más grave combinada con factores climáticos y la pobreza.
- Criterio de ola: Cuando en un lugar se sobrepasan los límites descritos por la OMS por un período prolongado.
- Serie Temporal: En muchos fenómenos naturales o sociales podemos medir ciertas variables y asignar un valor numérico a cada observación. Cuando hacemos esto, podemos hablar de una serie de tiempo y_t ordenadas en el tiempo. Llamaremos serie de tiempo a cualquier conjunto de observaciones $y_t : t \in T$ ordenadas en el tiempo, donde el tiempo t es estrictamente creciente.
- Potencia: Se le llama potencia o momento a la combinación entre dos variables clave de una enfermedad en propagación (Incidencia * R efectivo, conceptos que serán explicados más adelante).
- Inmunidad de Rebaño: es la resistencia de un grupo a infecciones debido a la inmunidad de una alta proporción de los miembros del grupo, si la proporción de individuos inmunes es alta, la población completa va a estar protegida, la proporción necesaria dependerá de cuán virulento es el agente (Ej., un 70 % de inmunidad a C. diphtheriae es suficiente; un 90 % es necesario para la peste cristal).
- Zoonosis: Enfermedades infecciosas que ocurren en humanos y animales; enfermedades animales que se pueden transmitir a los seres humanos

1.2. Tasa de Positividad

La tasa de positividad se define como la razón entre el número de casos confirmados y el número de exámenes realizados en un intervalo de tiempo dado. Para hacer la estimación de la tasa de positividad se usará los modelos GAM (Generalized Additive Models) Poisson definidos por:

$$\log(\mu_t) = \log(pcr) + s_0 + \sum_{i=1}^k s_i(t)$$

donde $s_i(.)$ es una función spline cúbica natural con 7 grados de libertad y t es el tiempo diario correlativo. Además, μ_t es la tasa de exámenes positivos y pcr es el número de exámenes realizados.

Un modelo aditivo generalizado es un modelo lineal generalizado en el que el predictor lineal depende linealmente de funciones suavizadas de las variables predictoras. [15].

1.3. Tasa de Incidencia

La incidencia es un indicador de ocurrencia, el cual se utiliza para medir el número de casos nuevos de una enfermedad.

$$I\text{ncidencia}(t) = \frac{\text{suma de casos nuevos durante el periodo(t)}}{\text{cantidad de población(t)}} * 100,000$$

Esta medida se identifica por incidencia e indica el número de casos nuevos por cada 100,000 personas durante un periodo de tiempo. Esta también puede ser calculada para cada grupo de edad, por lo que se refiere como incidencia específica por edad, es decir, se determinan los riesgos de desarrollar una enfermedad en cada grupo de edad. [16].

1.4. Tasa de Reproducción Instantánea

El número de reproducción, R, es el número medio de casos secundarios de enfermedad causada por un solo individuo infectado durante su período infeccioso.

El seguimiento del R a lo largo del tiempo es una variable muy importante ya que proporciona información sobre la eficacia de las intervenciones y sirve para identificar lugares con descontrol, lo principal es reducir por debajo del umbral del 1 y lo más cercano a 0, poniendo así una epidemia bajo control.

Clasificación:

- $Re < 1$ significa que las personas enfermas cada vez dejan menos contagiados.
- $Re=1$ significa que, en promedio cada persona enferma contagia a una persona más.
- $Re > 1$ quiere decir que la persona contagiada va dejando más de 1 enfermo.

$$R_e(t) = I(t) \left[\int_0^t I(t-\tau) \omega(\tau) d\tau \right]^{-1}$$

El resultado es en base a la incidencia I en el tiempo t dada por la integral sobre todas las infecciones anteriores, multiplicado por la infectividad respectiva ω multiplicada colectivamente por R_e en el tiempo t [17], este método puede ser implementado con el paquete EpiEstim del software R [13].

1.5. Componentes Principales

El análisis de componentes principales tiene como objetivo: dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales.

La técnica de componentes principales es debida a Hotelling (1933), aunque sus orígenes se encuentran en los ajustes ortogonales por mínimos cuadrados introducidos por K. Pearson (1901). Su utilidad es doble:

1. Permite representar óptimamente en un espacio de dimensión pequeña, observaciones de un espacio general p -dimensional. En este sentido componentes principales es el primer paso para identificar posibles variables "latentes." no observadas, que estén generando la variabilidad de los datos.
2. Permite transformar las variables originales, en general correladas, en nuevas variables incorreladas, facilitando la interpretación de los datos.[18]

1.6. Clusters y validación

La premisa en el análisis de conglomerados es: dado un número de individuos, cada uno de los cuales se describe mediante un conjunto de medidas numéricas, idear un esquema de clasificación para agrupar los objetos en un de clases, de forma que los objetos de cada clase sean similares en algún aspecto y se diferencien de los de otras clases. Estas clases deducidas son los clusters. La validación interna y externa son las dos categorías más importantes para la validación de Clustering. I Las técnicas de validación internas miden el Clustering únicamente basada en información de los datos

Validación

- **Cohesión:** Cada miembro perteneciente a un Cluster debe ser lo más cercano posible a los otros mientras del mismo Cluster

$$SSW = \sum_{i=1}^k y_i \sum_{x \in C_i} d^2(m_i, x)$$

- donde k es el número de clusters, x un punto del cluster y m_i el centro del cluster C_i
- **Separación:** Los clusters deben estar ampliamente separados entre ellos,

$$SSB = \sum_{j=1}^k n_j d^2(c_j, \bar{x})$$

- donde k es el número de clusters, n_j el número de elementos en el cluster j y c_j el centro del cluster j y \bar{x} es la media del conjunto de datos

[19]

Paquete ClusterCrit

ClusterCrit es una librería del software R que se usa para la validación de Clusters.

Silhouette 1987

Considerando para cada punto M_i , su distancia media a cada grupo. Se define la distancia media dentro del cluster $a(i)$ como la distancia media al punto M_i a los otros puntos del cluster al que pertenece: si $M_i \in C_k$, tenemos:

$$a(i) = \frac{1}{n_k - 1}$$

Calinski Harabasz 1974

Usando las notaciones de dispersión dentro de del grupo descrita en el paquete, el indice de Calinski-Harabasz se define:

$$C = \frac{BGSS/(K - 1)}{WGSS/(N - K)}$$

Dunnv 1974

Se denota d_{min} la distancia mínima entre puntos de diferentes clusters y d_{max} la mayor distacia dentro del cluster. La distancia entre clusters C_k y $C_{k'}$ se mide por la distancia entre sus puntos más cercanos:

$$d_{kk'} = \min_{\substack{i \in I_k \\ j \in I_{k'}}} \|M_i^{\{k\}} - M_j^{\{k'\}}\|$$

y d_{min} es la más pequeña de las distancias $d_{kk'}$:

$$d_{min} = \min_{k \neq k'} d_{kk'}$$

Por cada cluster C_k se denorta por D_k la distancia más grande que separa dos puntos distintos en el custer (usualmente llamado diametro del cluster).

$$D_k = \max_{\substack{i,j \in I_k \\ i \neq j}} \|M_i^{\{k\}} - M_j^{\{k'\}}\|$$

Luego d_{max} es la mayor de estas distancias D_k :

$$d_{max} = \max_{1 \leq k \leq K} D_k$$

El índice de Dunn se define como el cuociente de d_{min} y d_{max} :

$$C = \frac{d_{min}}{d_{max}}$$

1.7. Conceptos generales de vigilancia estadística

Por vigilancia estadística entendemos la supervisión en línea de un proceso estocástico $X = \{X(t); t = 1, 2, \dots\}$ con el objetivo de detectar un cambio importante en el proceso en un momento desconocido τ , de la forma más rápida y precisa posible. En cada momento de decisión, queremos discriminar entre dos estados del sistema monitoreado: los estados

de control y fuera de control, se denotará por $D(s)$ y $C(s)$ respectivamente.

Para esto usamos las observaciones acumuladasas $X = \{X(t); t \geq s\}$ para formar conjuntos de alarmas $A(s)$, ya que si $X_s \in A(s)$, entonces esto es una indicación que el proceso esta en estado $C(s)$ y se produce una alarma. Por lo general, esto se hace mediante una función de alarma $p(X_s)$ y un límite de control $g(s)$ donde el tiempo de una alarma, t_A , está escrito como:

$$t_A = \min; p(X_s) > g(s)$$

Dependiendo de la aplicación, se utilizan diferentes tipos de estados de control y fuera de control. El caso más estudiado es cuando $D(s) = \{\tau > s\}$ y $C(s) = \{\tau \leq s\}$. A menudo, un cambio en un parámetro en la distribución de X será de interés. Por ejemplo, un cambio en un parámetro puede corresponder a un cambio de nivel, un cambio de variación o posiblemente a un cambio combinado de nivel y variación al mismo tiempo.

La mayoría de los estudios en la literatura se refieren al step change, donde un parámetro cambia de un nivel constante a otro nivel constante. Los diferentes tipos de cambio son de interés en diferentes aplicaciones. Otros tipos incluyen cambios graduales, lineales o un aumento exponencial. Se han estudiado muchas situaciones diferentes en salud pública y se ha sugerido una amplia gama de métodos para la vigilancia, incluida la suma acumulativa de Poisson (CUSUM), el CUSUM exponencial, el método de conjuntos y varios métodos de ventana.[20]

1.8. Change-Point

El concepto de 'Change-Point' es un método aplicable a series temporales que resulta ser eficaz para detectar cambios significativos en tendencia.

Se aplicará a las series de contagios para detectar cambios significativos en la tendencia con especial enfoque en los datos más recientes generados por la pandemia. Cabe mencionar que éste método ofrece numerosos beneficios sobre las gráficas de control convencionales: es robusto para los valores atípicos (es aún más robusto mediante la realización del análisis para rangos), funciona para todos los tipos de datos (los datos de atributos, recuentos, promedios y las desviaciones estándar) y es mucho más fácil de usar si se han producido varios cambios, en comparación con las gráficas de control tradicionales que requieren que se establezcan nuevos límites después de cada cambio, no obstante, ambos se utilizarán de manera complementaria.

Éste análisis permite detectar pequeños cambios sostenidos, entregando niveles de confianza, y proporcionando intervalos de confianza y llevar un control de la tasa de error

para cada cambio detectado con el objetivo de determinar la causa raíz del cambio, normalmente se cree que el cambio ocurre entre el primer punto fuera de los límites de control y el punto anterior. Sin embargo, este no es con frecuencia el caso. En realidad, el cambio puede haberse producido varios puntos antes. Realizar el análisis de 'Change-Point' ayuda a aislar mejor el momento exacto y conocer la verdadera naturaleza del cambio.[21]

1.9. Modelo Lineal Generalizado

El modelo lineal clásico ha sido utilizado extensivamente y con mucho éxito en múltiples situaciones. En el análisis de regresión estamos interesados en predecir la media de una variable, llamada la respuesta, basados en un conjunto de variables, llamadas los predictores. La regresión clásica asume que la respuesta es continua y distribuida normalmente.

El modelo lineal clásico cae en una clase mayor de modelos que se conoce como modelo lineal generalizado, M.L.G., la cual tiene tres componentes básicas:

- Un conjunto de variables aleatorias independientes que pertenecen a la Familia Exponencial, es decir, su función de densidad ha de poder escribirse como:

$$f(y; \theta, \phi) = \exp \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)$$

donde, en cada caso, $a(\cdot)$, $b(\cdot)$ y $c(\cdot)$ serán funciones específicas. El parámetro θ es lo que se llama parámetro canónico de localización y ϕ es un parámetro de dispersión. Pertenece a la familia exponencial la distribución Normal, Bernouilli, Binomial, Poisson, Exponencial, Gamma, entre otras.

- Una matriz de diseño y un vector de parámetros.
- Una función link (enlace, conexión) que relaciona las medias del modelo lineal.

Dentro de la clase de modelos lineales generalizados tenemos el modelo lineal clásico, el modelo loglineal, la regresión Poisson, la regresión logística, etc.

[22]

1.10. Regresión Poisson

La distribución Poisson juega un papel de fundamental importancia en el trabajo aplicado para modelar problemas de conteo en muchas áreas. Los problemas de regresión donde la variable dependiente es un conteo, ocurre con bastante frecuencia.

Como ejemplo tenemos el número de muertos por una cierta enfermedad extraña que

puede explicarse por un número grande de factores, como el clima, salubridad, educación, etc. Otro ejemplo es el número de defectos que aparece en cierto rollo de tela, que depende de la longitud del rollo, época de elaboración. Es común asumir una respuesta poissoniana, que perteneciendo a la familia exponencial puede resolverse con la metodología que estamos desarrollando [22].

$$\begin{aligned}
 Y_i &\sim Poisson (\lambda_i) \\
 f(y_i; \lambda_i) &= \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!} \\
 &= \exp(y_i \log \lambda_i - \lambda_i - \log(y_i!)) \\
 &= \exp(y_i \theta_i - \lambda_i - \log(y_i!)), \\
 &\text{donde,} \\
 \theta_i &= \log(\lambda_i),
 \end{aligned}$$

el cual es el parámetro natural.

$$\begin{aligned}
 E[y_i] &= \lambda_i \\
 Var[y_i] &= \lambda_i,
 \end{aligned}$$

ya que $g(\lambda_i) = \theta_i$ cuando g es la función logaritmo. El link canónico es el link log

$$\log(\lambda_i) = x_i' \beta,$$

ya que $\lambda_i = \exp(\eta_i)$ se tiene que,

$$\frac{\partial \lambda_i}{\partial \eta_i} = \exp(\eta_i) = \lambda_i,$$

las ecuaciones de verosimilitud,

$$\sum_{i=1}^n \frac{(y_i - \lambda_i)}{var(y_i)} x_{ij} \frac{\partial \lambda_i}{\partial \eta_i} = 0 \quad j = 1, \dots, p,$$

se reduce a

$$\sum_{i=1}^n (y_i - \lambda_i) x_{ij} = 0,$$

ya que,

$$\omega_i = \left(\frac{\partial \lambda_i}{\partial \eta_i}\right)^2 \frac{1}{var(y_i)} = \lambda_i,$$

la matriz de covarianza estimada de $\hat{\beta}$ es $(X'\hat{W}X)^{-1}$ donde \hat{W} es la matriz diagonal con elemntos de $\hat{\lambda}$ en la diagonal principal.

1.11. GLM Binomiales Negativos

- La densidad de la distribución binomial negativa es:

$$f(y; k; \mu) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1 - \frac{k}{\mu+k}\right)^y$$

con $y = 0, 1, 2, \dots$ donde k y μ son los parámetros.

- Se tiene que. $E(Y) = \mu$, $var(Y) = \mu + \mu^2/k$
- El parámetro $1/k$ es un parámetro de dispersión.
- Si $1/k \rightarrow 0$, entonces $var(Y) \rightarrow \mu$ y la distribución binomial negativa converge a una distribución de Poisson.
- Con k fijo esta densidad está en la familia exponencial natural y podríamos hablar de un GLM binomial negativo.[23]

1.12. AIC

El criterio de información de Akaike esta basado en la teoría de la información que contiene la verosimilitud del modelo. Mide la calidad del ajuste de una regresión y se calcula con la siguiente expresión:

$$AIC = 2(p+1) - 2\log(L),$$

Donde L es la verosimilitud. Mientras más pequeño sea el AIC, mejor será la bondad de ajuste de modelo [24].

1.13. Deviance

La medida de bondad de ajuste más común en GLM es la deviance. Sean $\ell(\hat{\mu}, \phi, y)$ el log-likelihood del modelo de interés y $\ell(y, \phi, y)$ el log-likelihood del modelo saturado. La deviance o devianza se define como:

$$D = 2(\ell(y, \phi, y) - \ell(\hat{\mu}, \phi, y)) = -2(\ell(\hat{\mu}, \phi, y) - \ell(y, \phi, y))$$

Notemos que hallar el EMV equivale a minimizar la deviance.

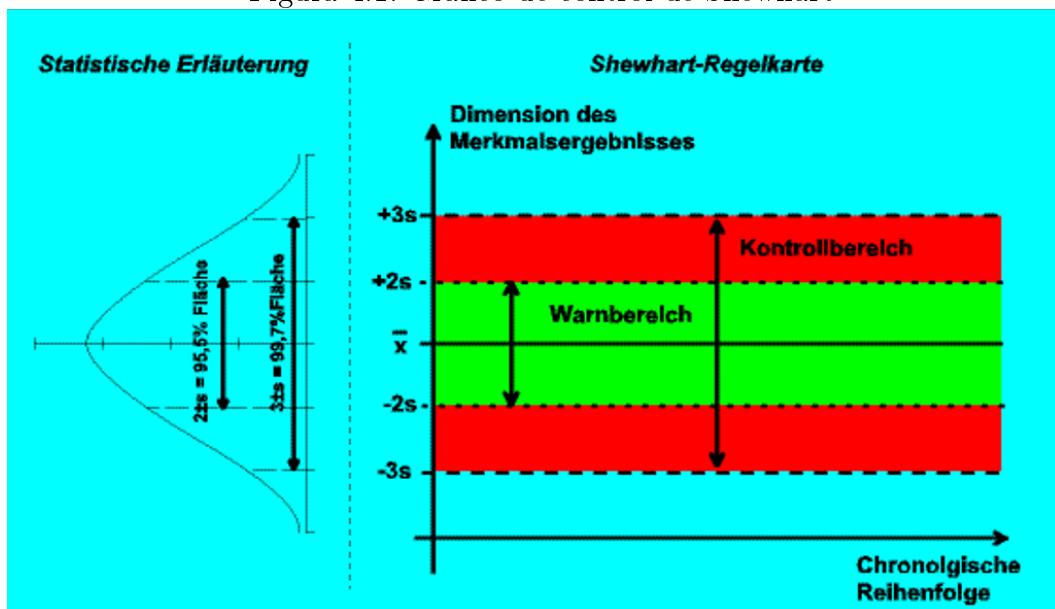
En las distribuciones, como la normal por ejemplo, que tienen un parámetro de escala ϕ , se define la deviance o devianza escalada como $D^* = D/\phi$, que es la medida que se usa para inferencia.

En las distribuciones como la binomial o la Poisson, $\phi = 1$, entonces las dos medidas deviance coinciden.

1.14. Cartas de Control de Walter Shewhart

Fue Walter Shewhart quién creó los gráficos de control para aplicar en Control Industrial. La línea central es el promedio de los puntos de datos individuales y los límites se basan en cálculos estadísticos de variación de causa común que establecen los límites superior e inferior del rendimiento del sistema [25].

Figura 4.1: Gráfico de control de Shewhart



Fuente: <https://www.faes.de/Basis/Basis-Statistik>

5

Análisis Descriptivo

En éste capítulo se realiza una descripción breve de las variables, se mostrarán algunas tablas y gráficos con información descriptiva de la base de datos, con cifras actualizadas a noviembre de 2021. La actualización de los métodos y gráficos se puede encontrar dentro del repositorio GitHub utilizado para éste trabajo [26].

El GEMVEP-USACH realiza una actualización diaria de los datos de Covid-19 en Chile que almacena en el repositorio GitHub [27], ésta actualización utiliza los productos del Minsal[28].

Chile es un país tricontinental lo cual hace que sea muy diverso y ya que el coronavirus ha tenido impactos diferentes en distintos tiempos y lugares es conveniente revisar la realidad que se vive en distintas zonas y regiones.

Cuadro 5.1: Resumen de variables para las zonas de Chile

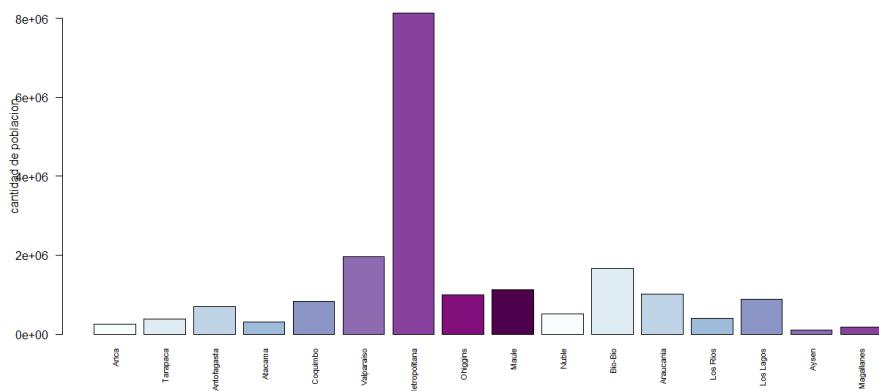
	Region	Población	Confirmados	PCR	Fallecidos	Recuperados
1	Nacional	19.458.310	1.740.545	23.488.691	38.016	1.672.514
18	Norte Grande	1.326.737	136.519	2.341.583	2.615	131.763
19	Norte Chico	1.150.805	80.689	1.295.442	1.408	77.291
20	Zona Centro	14.383.491	1.240.414	15.932.729	29.776	1.189.179
21	Zona Sur	2.311.618	244.074	3.328.645	3.639	236.721
22	Zona Austral	285.659	38.849	525.023	578	37.560
23	Regiones sin RM	11.333.238	1.005.216	14.628.924	18.508	968.131

Cuadro 5.2: Resumen de variables para las regiones de Chile

	Region	Población	Confirmados	PCR	Fallecidos	Recuperados
1	Arica	252.110	28.398	487.123	541	27.467
2	Tarapaca	382.773	43.534	639.877	812	41.972
3	Antofagasta	691.854	64.587	1.214.583	1.262	62.324
4	Atacama	314.709	28.866	505.536	362	27.720
5	Coquimbo	836.096	51.823	789.906	1046	49.571
6	Valparaiso	1.960.170	131.343	1.900.154	3.507	124.608
7	Metropolitana	8.125.072	735.329	8.794.498	19.508	704.383
8	Ohiggins	991.063	71.588	895.356	1.590	69.123
9	Maule	1.131.939	106.532	1.254.582	1.806	103.066
10	Nuble	511.551	37.528	567.076	652	36.226
11	Bio-Bio	1.663.696	158.094	2.521.063	2.713	151.773
12	Araucania	1.014.343	100.630	989.202	1.519	98.066
13	Los Rios	405.835	51.158	557.089	698	49.542
14	Los Lagos	891.440	92.286	1.782.354	1.422	89.113
15	Aysen	107.297	9.297	190.640	79	8.666
16	Magallanes	178.362	29.552	334.383	499	28.894

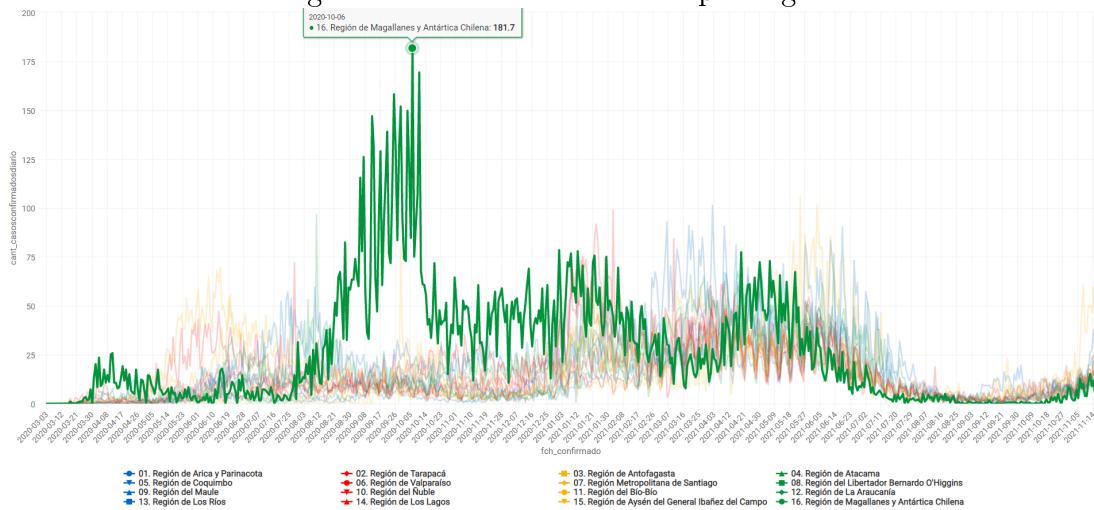
Una de las diferencias que presenta Chile es la cantidad de población por región, donde la Región Metropolitana concentra la mayor cantidad de habitantes (8.125.072), por otro lado es Aysén quién tiene la menor población (107.297). Por lo que es importante utilizar tasas para la comparación entre las regiones. A continuación gráfico que muestra ésta diferencia.

Figura 5.1: Cantidad de población por región



También se debe considerar que las regiones de Chile han experimentado sus propios períodos de alzas de casos de coronavirus, que se puede apreciar a continuación en el gráfico de tasa de incidencia cada 100.000 habitantes por región actualizado hasta el lunes 15 de noviembre de 2021, dónde destaca la región de Magallanes con la mayor incidencia observada el 2020-10-06 con 181,7 casos/100.000hab.

Figura 5.2: Tasa de incidencia por región



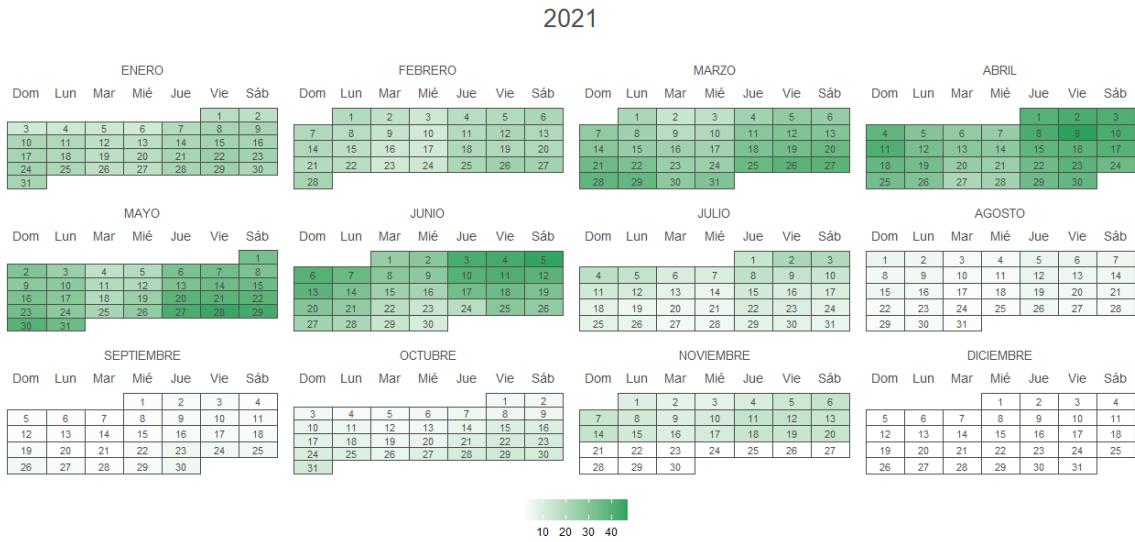
Efecto días de la semana en las cifras reportadas

Los casos que se informan diariamente tienen un efecto por día de la semana, por lo que es importante notar para los métodos de 'Change-Point' que se aplicarán.

A continuación se muestra un calendario con la incidencia que se reportó desde el inicio del año 2021 hasta el 20-11-2021.

Éste mapa de calor muestra como aumentan progresivamente los casos desde marzo en adelante y que los días que se informan la menor incidencia de casos serían los días martes y miércoles, correspondientes a los exámenes realizados el fin de semana, por lo que los casos se reportarían al menos con 3 días de desfase . Ésto se debe al proceso de tiempo que existe en la notificación de los casos de Covid-19, por lo que es importante notar que siempre nos encontraremos detrás curva y los resultados vistos serán muestra del pasado.

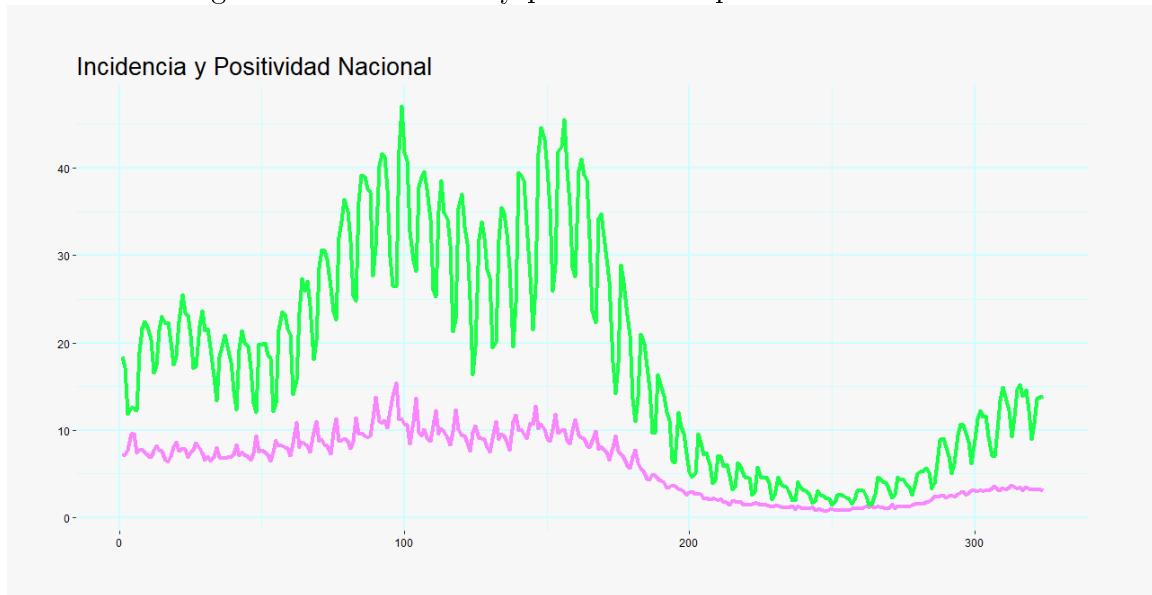
Figura 5.3: Calendario con tasa de incidencia reportada diariamente



Ocupando el mismo período te tiempo a continuación la gráfica con la tasa de incidencia más la tasa de positividad.

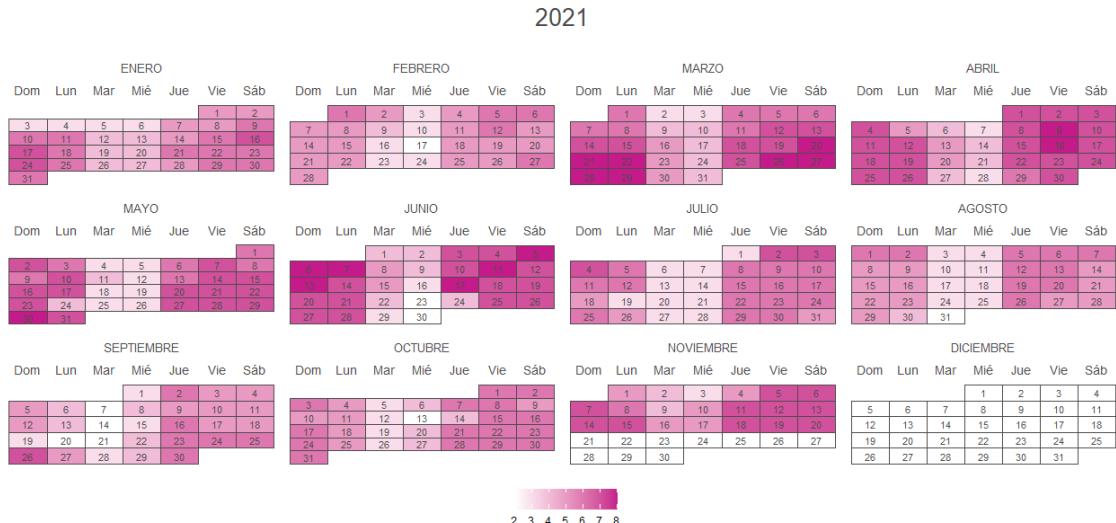
Note que la tasa de Incidencia y tasas de Positividad van de la mano.

Figura 5.4: Incidencia y positividad reportada diariamente



Respecto a la tasa de Positividad es posible reconocer cómo afectan los días sábados y domingos, que corresponderían a los exámenes PCR reportados los martes y miércoles como se muestra a continuación en el siguiente mapa de calor.

Figura 5.5: Calendario con exámenes PCR



6

Métodos de Change-Point

A raíz de la pandemia en el mundo se han originado múltiples trabajos para detectar brotes de Covid-19, ante ésta emergencia sanitaria nacen los métodos de 'Change-Point' en dos aplicaciones distintas.

En ésta sección se abordarán los contenidos para la implementación mediante el software estadístico R para dos métodos de 'Change-Point' propuestos, además de los resultados obtenidos con los datos de Chile, que se incorporarán en la clasificación de Clusters que realiza el GEMVEP-USACH.

1. Using control charts to understand community variation in COVID-19

El siguiente procedimiento implementa un método basado en los gráficos de control de Shewhart para ver las fases de los datos notificadas de COVID-19. El método fue desarrollado por Lloyd Provost, Shannon Provost, Rocco Perla, Gareth Parry y Kevin Little, con un enfoque inicial en las series de muertes, para éste trabajo se aplicarán en las distintas series con enfoque en los casos nuevos reportados, con el fin de generar una alerta temprana antes que se manifiesten en hospitalizaciones y muertes.

Gareth Parry utilizó SPSS para desarrollar la presentación inicial de IHI en la primavera y el verano de 2020. Su código SPSS creó tablas de datos para países y estados y territorios de estados unidos. Gareth creó un script PowerBI para leer estas tablas y creó las pantallas de datos.[\[12\]](#) En septiembre, Kevin Little asesorado por Lloyd Provost reemplazó el código SPSS con el código R que requiere menos intervención humana diaria. Este documento describe la función de código R y las limitaciones. El proceso para su desarrollo está contenido en el repositorio de datos de Gibhub con los códigos y ejemplos para el desarrollo.[\[29\]](#)

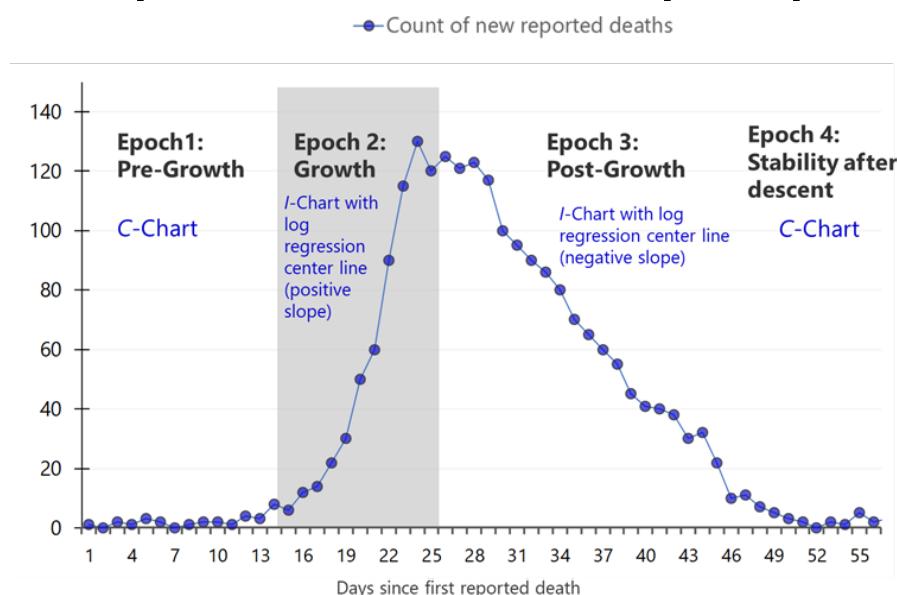
1.1. Procedimiento

Para realizar el proceso se debe conocer de ciertos criterios que se mostrarán y ejemplificarán a continuación. Primero se utilizan épocas y fases para describir los patrones de las series de Covid-19.

Descripción de épocas:

- 1 Crecimiento preexponencial: gráfico en la escala original
- 2 Crecimiento exponencial: gráfico individual ajustado al logaritmo 10 de la serie, transformado a la escala original.
- 3 Crecimiento postexponencial: trayectoria plana o descenso exponencial: gráfico de individuos ajustado al logaritmo 10 de la serie, transformado a la escala original.
- 4 Estabilidad tras el descenso.

Figura 6.1: Épocas asociadas con muertes diarias reportadas por COVID-19



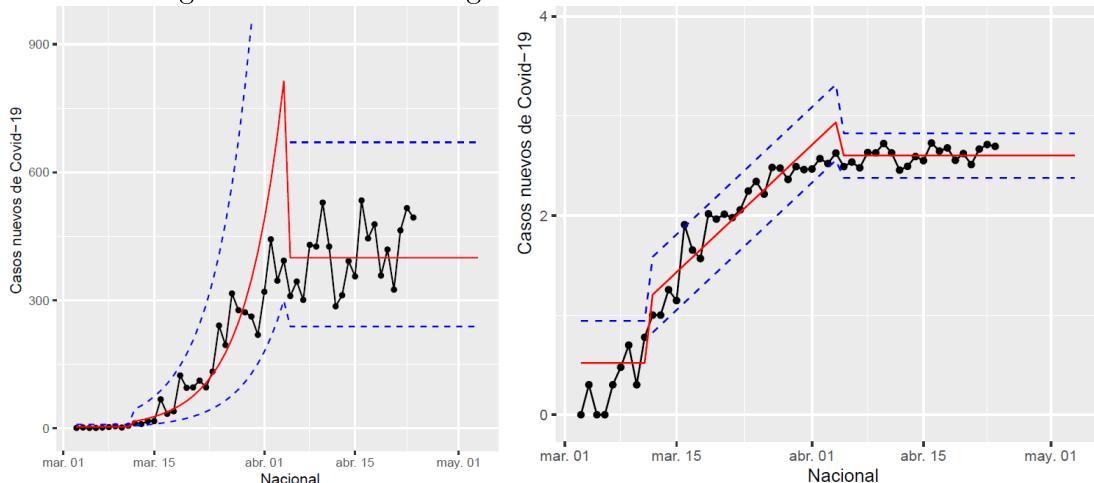
Fuente: International Society for Quality in Health Care

Note que las épocas se muestran con la forma de una ola, cuando el algoritmo detecta una causa especial éste se cambia de época a un crecimiento exponencial, así hasta que llega al peak y detecta una disminución para pasar a la siguiente época, que en su disminución vuelve a la primera época.

Cartas de control

Para la mejor comprensión de cómo se calcula la línea central y los límites de control en el cambio de época de una carta de control C-Chart a una carta de control I-Chart, se muestra un ejemplo utilizando los casos nuevos reportados en Chile, desde marzo hasta mayo del 2020, en donde los resultados de las bases de datos generadas por el algoritmo fueron transformadas a una escala logarítmica para su posterior representación gráfica.

Figura 6.2: Cambio de gráfica de control C-Chart a I-Chart.



Es importante considerar que dentro de cualquier época, se necesita al menos una fase. Por ejemplo, dentro de la Época 1, si el algoritmo no detecta un crecimiento exponencial pero muestra un aumento en el promedio, las fases adicionales mostrarán gráficos con medias más altas que la primera fase.

Funciones en el software R

Los scripts contenidos en el repositorio que se realizan para el desarrollo del método son:

1 generate-data-files.R.

- Este archivo carga los datos de sitios web externos para los casos diarios de COVID de EE.UU y de sus estados, éstos datos serán reemplazados con los casos que se reportaron en Chile.
- Realiza una edición mínima de los marcos de datos para asegurar nombres comunes.
- Para el archivo de estado/territorio de los Estados Unidos o en éste caso para las regiones de Chile, convierte las muertes o casos acumulados en muertes reportadas diariamente.

- Cuando se ejecuta en modo interactivo, crea gráficos del país de entrada y los archivos de estado y territorio de EE. UU. y guarda los gráficos como archivo pdf en una carpeta denominada samples.
- Crea cinco archivos de valores separados por comas que se usan como entrada en PowerBI, escribiendo los archivos en una carpeta denominada output.

2 functions.R.

Este archivo contiene las funciones principales, además de varias pequeñas funciones auxiliares, algunas de éstas son:

- **detect.outlier.dates**, una función que identifica las fechas con registros considerados inusualmente grandes que se excluirán de la creación de los gráficos de control.
Inputs: marco de datos de entrada; umbral para declarar un valor atípico.
Outputs : una columna añadida al marco de datos con valor = TRUE si el valor de las muertes para un día determinado se considera un valor atípico.
- **forcemonotonicity**, una función que obliga a las series de datos a ser monótonas y no decrecientes.
Input: vector de muertes por estado o territorio de la fuente
Output: vector de muertes para el estado o territorio específico con valores negativos contabilizados
- **findphasedates**, una función que hace el 'trabajo pesado'; esta función comproba el comienzo y el final de las fases y genera los parámetros del gráfico de control para cada fase. También ajusta la estacionalidad dentro de la semana para las fases con al menos 21 registros.
Inputs: un marco de datos con las series de casos/defunciones y los indicadores de datos fantasma (ghosted values), por ubicación; una variable lógica para ajustar los datos a la estacionalidad dentro de la semana.
Outputs: un marco de datos que añade nuevas columnas al marco de datos de entrada: indicadores de épocas y fases dentro de las épocas, fechas de inicio de las fases; parámetros del gráfico Shewhart (línea media y límites superior e inferior) para cada fase.

Parámetros clave

Estos parámetros están actualmente 'codificados', pero deberían expresarse como parámetros para la generalización y las pruebas de sensibilidad.

Longitud mínima del gráfico: al menos cinco registros en las épocas 1 y 4 para estimar los parámetros. En las épocas 2 y 3, necesitamos 21 registros para estimar los parámetros. Si el ajuste exponencial en las épocas 2 o 3 se está probando con los registros más actuales, requerimos un mínimo de cinco registros no nulos para un ajuste preliminar.

longitud máxima de la serie para establecer los límites del gráfico: 21 registros, correspondientes a tres semanas naturales.

Número inicial de casos: se requieren al menos ocho casos en la primera fase de las épocas 1 y 4 para estimar los parámetros del gráfico. Este parámetro tiene en cuenta el número potencialmente elevado de días con cero casos en la primera fase de las épocas 1 y 4.

Longitud del desplazamiento: ocho valores consecutivos por encima o por debajo de la línea media de una fase señalan una causa especial y el inicio de una nueva fase.

Límite de transición de la época 4: un valor límite inferior a 2 es un requisito para la transición de la época 3 a la 4. Este parámetro ajusta la sensibilidad de la transición de la época 3. Para eventos distintos a los casos, por ejemplo, las hospitalizaciones, el límite puede ser más alto.

Número de días para extender el gráfico más allá del registro más reciente: 10 días.

Si la serie termina con una fase que tiene suficientes registros para calcular la línea media y los límites, se extienden por el número de días.

Marcar y dejar de lado valores inusualmente grandes: ghosting.

La función `detect_outlier` examina cada registro diario para evaluar si el registro es inusualmente grande en comparación con los días anteriores y posteriores al registro (siguieron los informes durante la primavera de 2020, de noticias de '**data dumps**' y marcaron estos eventos manualmente). Tales '**data dumps**' son un ejemplo simple y claro de una causa especial de variación en la serie de datos. Si adapta el código R para su propio uso, recomiendan que permita a los usuarios identificar los registros que deben excluirse de los cálculos utilizando su conocimiento de las series de casos notificadas. Éste equipo se referió a los registros marcados por la función `detect_outlier` como '**ghosting**' porque

el script original de PowerBI trazaba dichos valores con un punto pálido.

Monotonidad.

La tabla de datos del New York Times proporciona recuentos acumulativos de muertes para cada estado o territorio de los Estados Unidos. El código diferencia la serie de muertes acumuladas para obtener muertes diarias. La serie de conteo acumulativo de muertes muestra ajustes para 27 estados y territorios al 8 de noviembre que hacen que la serie no sea monótona aumentando: 52 registros son menores que los registros anteriores, dentro del estado o territorio. Esto significa que las series diferenciadas tendrán valores negativos. Para eliminar las muertes negativas en las series diferenciadas, la función asigna los valores negativos a los registros anteriores de modo que la serie revisada solo tenga valores no negativos.

Ajuste.

En junio, éste equipo de desarrollo vio que algunos lugares reportaron muertes de manera sistemática: uno o dos días a la semana tendían a ser más bajos que los otros seis o cinco días en cada semana. Como resultado, la aplicación de IHM se ajustó para un efecto de día de la semana. Desarrollamos el código R para imitar la aplicación IHM: el código R genera datos ajustados para cada ubicación. La serie de datos ajustados se utiliza como entrada al algoritmo de épocas y fases.

1.2. Aplicación para los datos de Chile.

Los dos métodos de 'Change-Point' propuestos son aplicables las las distintas series de tiempo, que pueden ser : la frecuencia absoluta, tasa de positividad, tasa de incidencia, tasa de mortalidad, ocupación de camas UCI etc. Ésto permite que se pueda evaluar el estado de los distintos escenarios de una enfermedad.

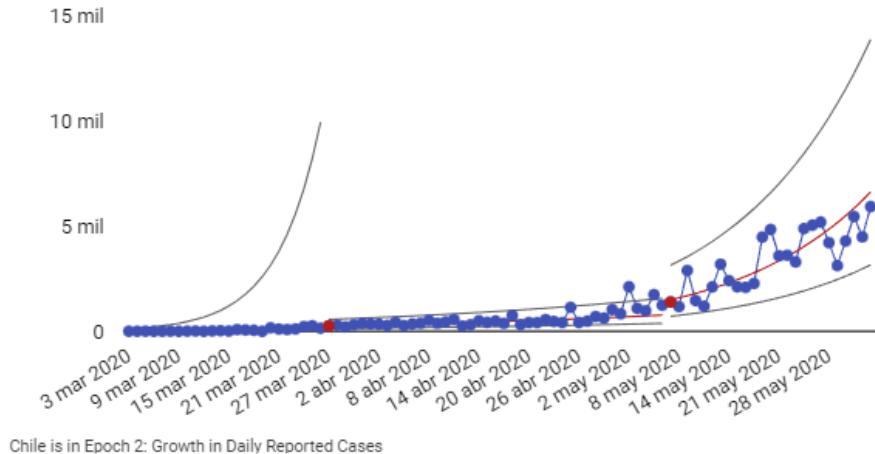
International Society for Quality in Health Care implementó los resultados del método para los países del mundo, donde los gráficos se actualizan diariamente desde el sitio Our World [30] in Data y el repositorio Github del New York Times[31].

Los casos reportados se ajustan para tener en cuenta la variación del día de la semana en la notificación.

Para el comienzo de la pandemia en Chile éste método identifica 2 días de cambio en la tendencia.

- Causa especial 1 : 27 de marzo con 257 casos reportados.
- Causa especial 2 : 7 de mayo 1391 casos reportados.

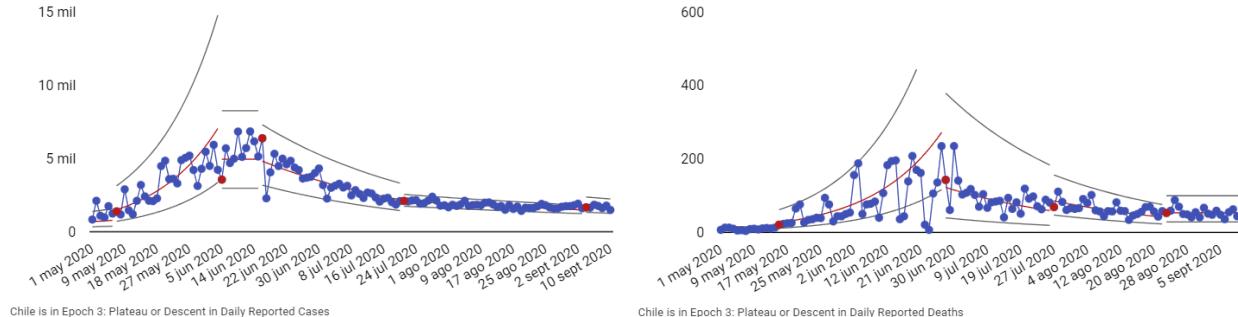
Figura 6.3: Épocas asociadas con casos diarias en Chile reportadas por COVID-19
Chile : Adjusted Daily Covid-19 Reported Cases



Si se utiliza una ventana desde la segunda causa especial el 7 de mayo, se puede apreciar cuando se presentan los siguientes cambios de épocas. También se muestra como luego del cambio de época de los casos nuevos reportados a una escala exponencial, afecta a los fallecidos reportados, presentando las siguientes causas especiales de puntos de cambio:
Casos nuevos: 5 Junio 3563 , 16 Junio 6392, 21 Julio 2109, 4 Sept 1667.
Fallecidos: 28 Junio 143, 27 Julio 69, 23 Agosto 55.

Figura 6.4: Épocas de casos diarios y fallecidos en Chile reportadas por COVID-19
Chile : Adjusted Daily Covid-19 Reported Cases

Chile : Adjusted Daily Covid-19 Reported Deaths



Fuente: International Society for Quality in Health Care

Nivel Nacional

A continuación se mostrarán resultados a nivel nacional, por regiones y zonas de las muertes por COVID-19 y la evolución de casos nuevos utilizando éste método.

En primera instancia se evaluó el método utilizando las muertes a nivel Nacional. Note que existe un punto rojo que sobresale muy por encima de los demás, un outlier, día en que se reportaron 1.057 fallecidos el 17 de julio. Éste es un buen ejemplo de 'ghosting' que se recomienda eliminar, sin embargo no es menor mencionar lo que se vivió en la primera ola de Covid-19 en Chile. Y que en éste punto, sin tener el desarrollo de una vacuna la solución inmediata fueron las cuarentenas para cortar la circulación del virus.

Los gráficos contemplan los decesos desde el inicio de la pandemia hasta el 26 de Agosto del año 2021. Se puede notar que existe una disminución significativa para los últimos días

Figura 6.5: Grafico de Control de muertes a nivel nacional

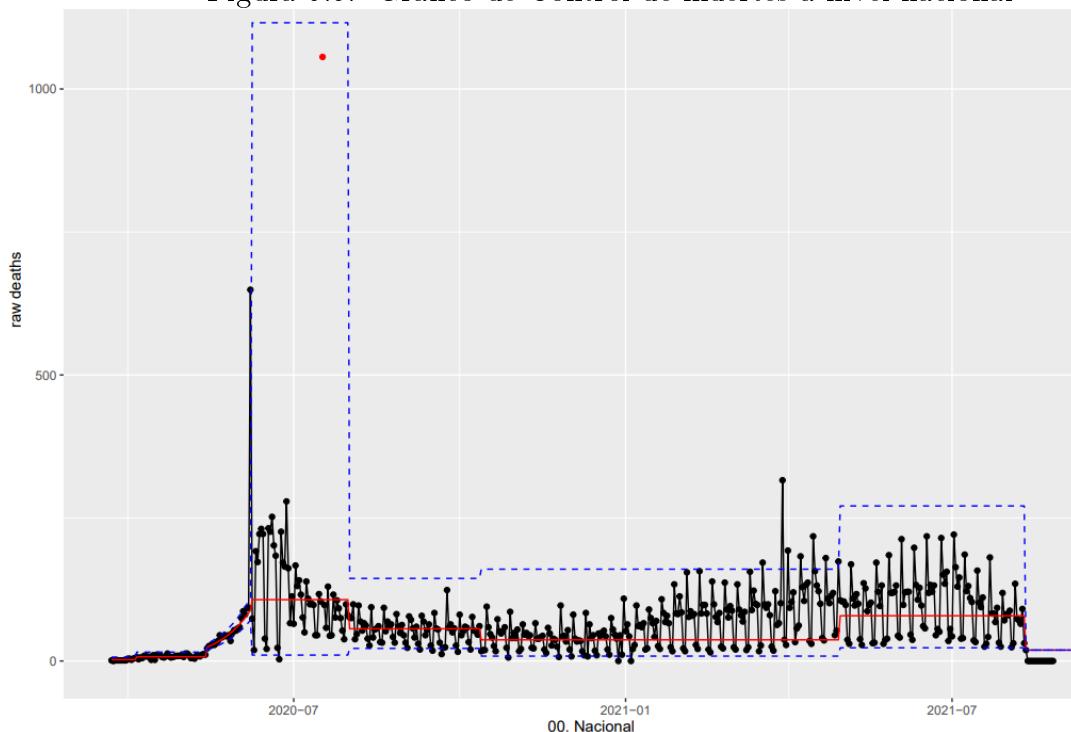
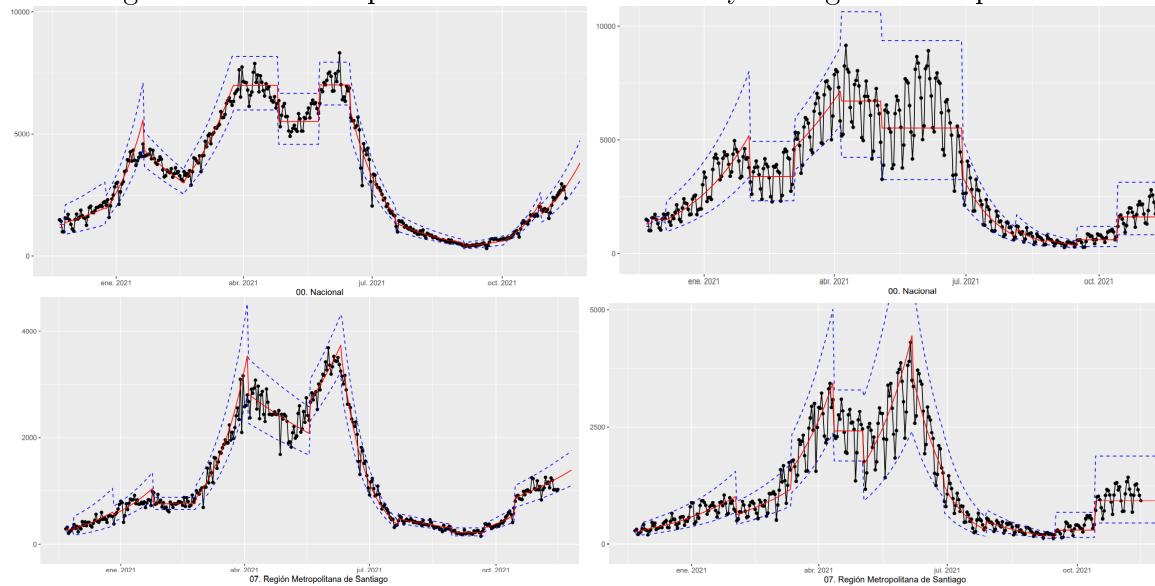


Figura 6.6: Casos reportados a nivel Nacional y la Región Metropolitana



De llevar una estabilidad al rededor de los 1.000 casos diarios después de la primera ola, con un relajo en las restricciones y con la salida de vacunas contra el Covid-19 comenzó un nuevo aumento de casos lo que marcó el inicio de la segunda ola antes del inicio del proceso de vacunación en Chile (16 diciembre del 2020). Para el siguiente análisis se considerará una ventana desde el 20-11-2020 hasta el 15-11-2021 con los casos reportados a nivel Nacional y la Región Metropolitana. Note que los casos a nivel nacional serían explíciamente por la región Metropolitana. Al lado izquierdo se muestran los gráficos ajustados por día de la semana y a la derecha los casos brutos.

El método también libera lo que sería el comienzo de una tercera ola a principios de octubre del 2021.

Nivel Regional

A continuación los resultados de las regiones del país que muestran presentan un crecimiento exponencial para los últimos casos reportados y más el gráfico ajustado por día de la semana.

Más resultados para las regiones son expuestos en la sección de Anexos.

Figura 6.7: Región de Atacama y Valparaíso

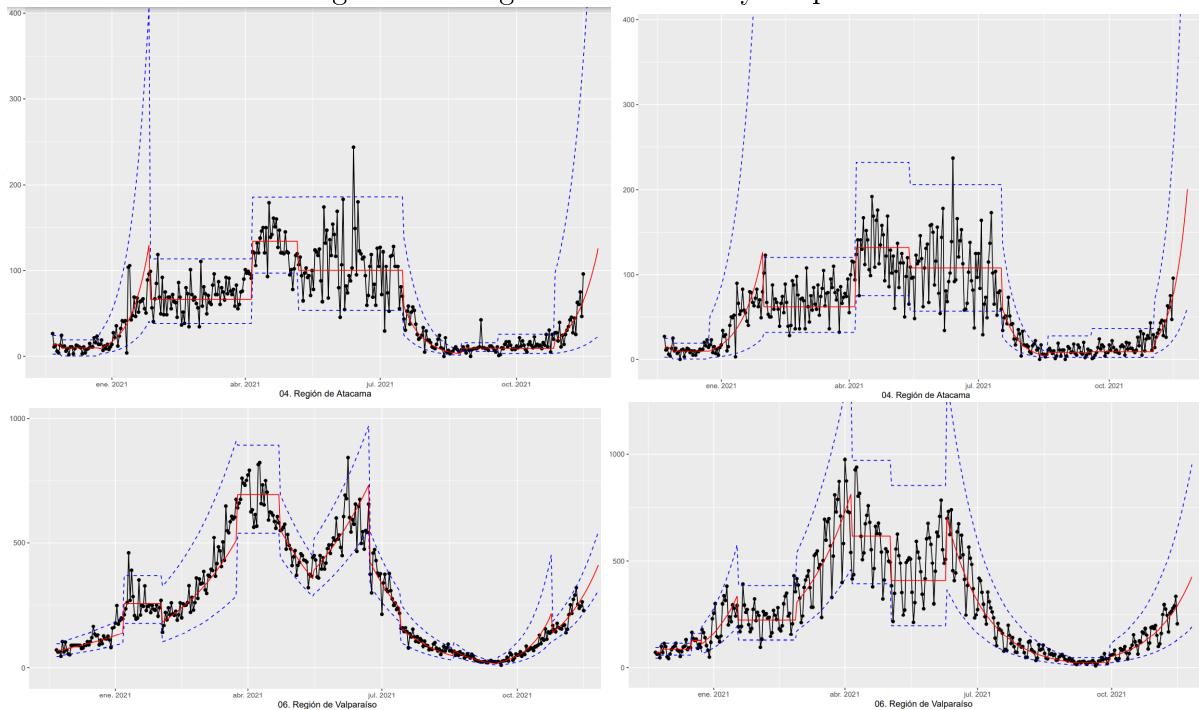
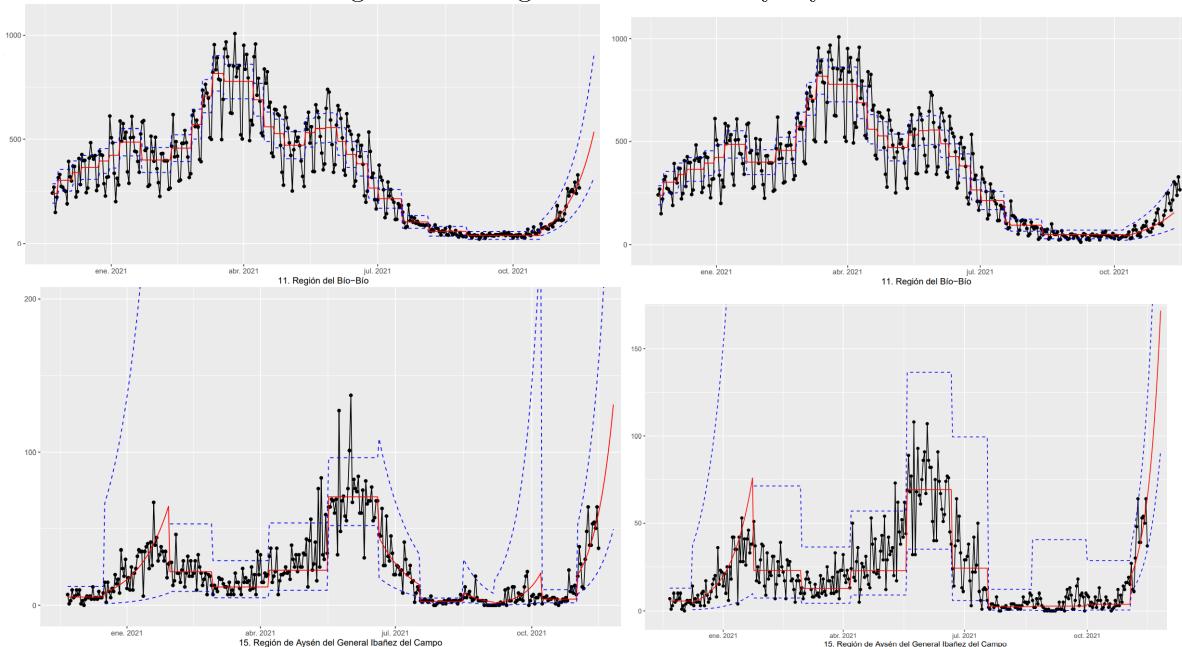


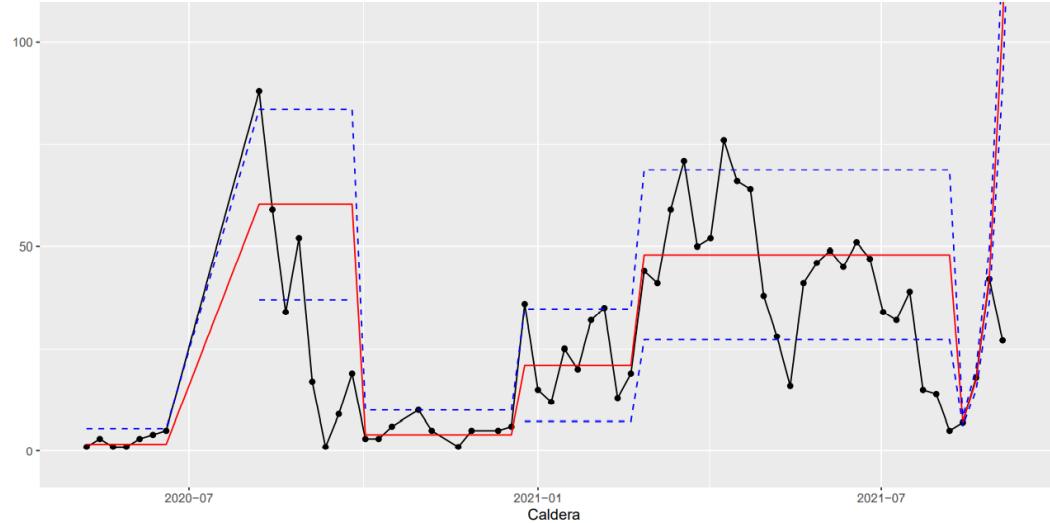
Figura 6.8: Regiones de Bío-bío y Aysén



Nivel Comunal

Para la simulación a nivel comunal se considerarán los casos semanales, ya que éstos son reportados dos veces a la semana y en ocasiones una sola vez.

Figura 6.9: Grafico de Control de casos nuevos para la comuna de Caldera



El método resulta efectivo de realizar a pesar de que los casos fueron entregados semanalmente.

Figura 6.10: Grafico de Control de casos nuevos para la comuna de Pinto

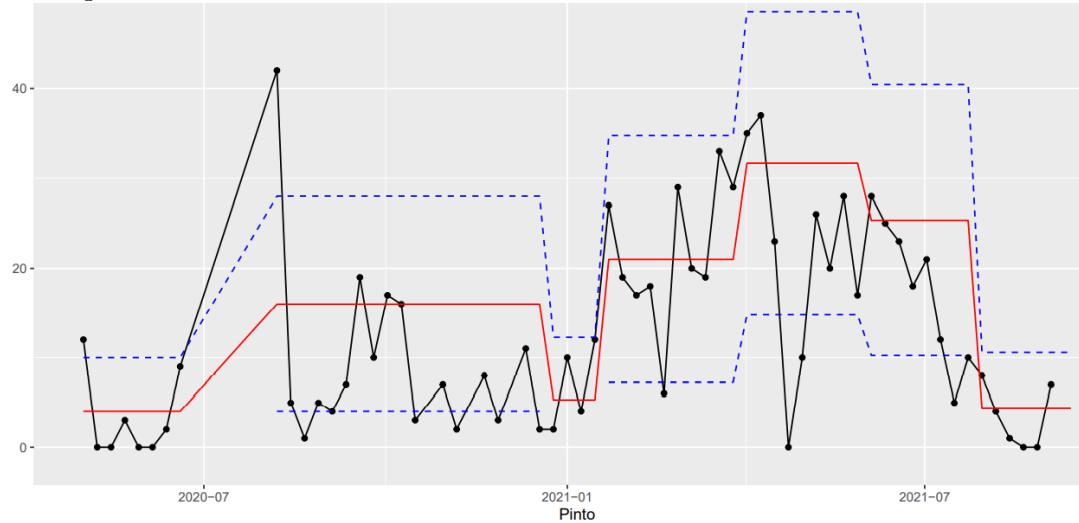
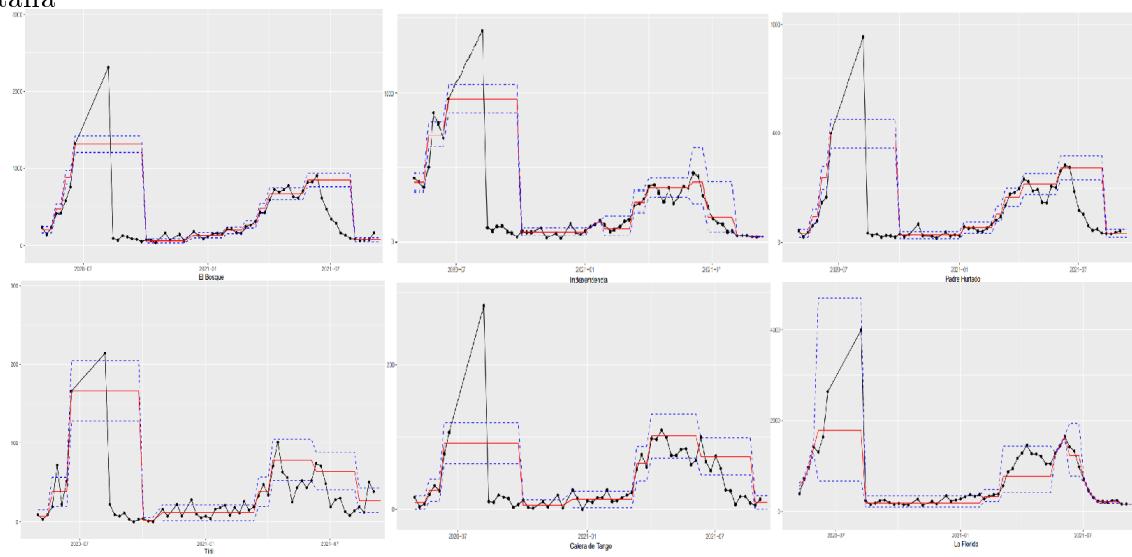


Figura 6.11: Grafico de Control de casos nuevos para las comunas de la Región Metropolitana



2. Automatic Selection of Models Outlier Detection for Epidemics (ASMODEE)

A raíz de la pandemia provocada por el Coronavirus, la Organización Mundial De La Salud creó una serie de algoritmos orientada para el análisis de series temporales epidémicas, la cual está contenida en la librería 'trendbreaker' del software estadístico R, ésta se encuentra libre para descarga en el repositorio de datos de Github.[32]

2.1. Procedimiento

Los pasos a seguir para aplicar éste método son:

1. Dividir la serie de tiempo en dos ventanas de tiempo complementarias: una ventana de calibración, excluyendo los más recientes k puntos; y una ventana de predicción, con los últimos k puntos de datos. El valor k es definido con la finalidad de encontrar la mejor ventana de tiempo requerida para las diferentes series de tiempo.
2. Ajustar modelos de series de tiempo definidos por el usuario al conjunto de calibración, se conserva el modelo de mejor predicción o de mejor ajuste; aquí describimos implementaciones con regresiones lineales, varios GLM (incluidos los modelos de Poisson y Binomial Negativo) y modelos de regresión bayesiana.
3. Establecer intervalos de predicción en un determinado alfa umbral ($\alpha = 5\%$), para todos los puntos de datos, incluidos en el conjunto de predicciones. Este umbral es igual a la proporción esperada de puntos de datos que se clasificarán como valores atípicos si todos los puntos siguieron la misma tendencia temporal.
4. Identificar valores atípicos como puntos de datos que se encuentran fuera del intervalo de predicción; el número de puntos de datos que muestran un incremento marcado de las tendencias pasadas en el conjunto de predicciones se pueden utilizar para diseñar un sistema de alarma. Similarmente, una disminución en la tendencia se puede ver en los puntos de datos que se encuentran por debajo del límite inferior del intervalo de predicción. Los puntos dentro del intervalo se pueden considerar como normal, es decir perteneciente a la tendencia reciente.[13]

Librería

El paquete **trendbreaker** implementa herramientas para detectar cambios en las tendencias temporales de una única variable de respuesta. Implementa la elección Automatica de Modelos y Deteción de Outliers para Epidemias (ASMODEE), un algoritmo diseñado originalmente para detectar cambios en la incidencia de casos de COVID-19.

De manera simple ASMODEE se utiliza:

1. definir un conjunto de entrenamiento excluyendo los últimos puntos de datos k.
2. identificar la tendencia temporal en el conjunto de entrenamiento ajustando un rango de modelos (especificados por el usuario) a los datos y conservando el mejor modelo de predicción / ajuste.
3. calcular el intervalo de predicción Intervalo de Predicción de la tendencia temporal.
4. clasificar cualquier punto de datos fuera del PI como valor atípico.

El algoritmo se puede aplicar con un valor fijo especificado por el usuario de k, para monitorear los posibles cambios en este período de tiempo reciente. Alternativamente, el valor óptimo de k se puede determinar automáticamente.

Una vez que sea incorporado en CRAN se podrá instalar:

```
install.packages('trendbreaker')
```

Por el momento la versión de desarrollo se puede instalar desde GitHub con:

```
if(!require(remotes)) {
  install.packages('remotes')
}
remotes::install_github('reconhub/trendbreaker')
```

Para la mejor comprensión para usar éste paquete es leer la documentación de la función y ejecutar su ejemplo:asmodee

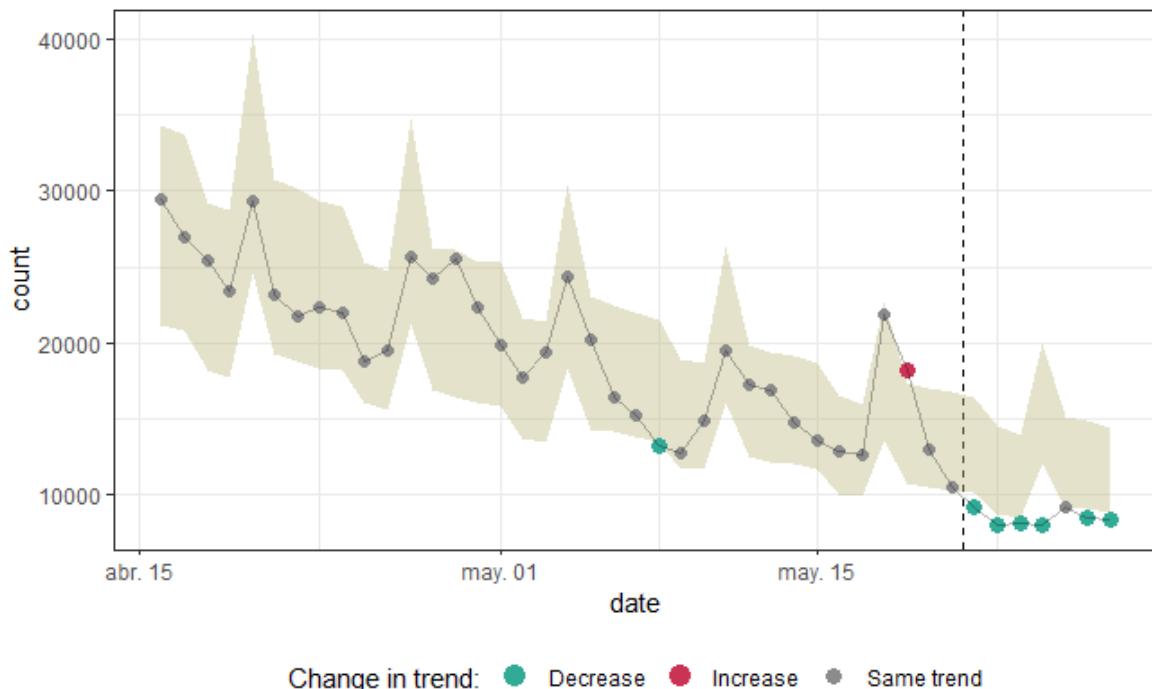
```
library(trendbreaker)
?asmodee
example(asmodee)
```

Características principales de ASMODEE

Se ilustra ASMODEE utilizando datos de vías del NHS disponibles públicamente, que registran el autoinforme de posibles casos de COVID-19 en Inglaterra (consulte para tener más información).?nhs_pathways_covid19

```
library(trendbreaker) # ASMODEE
library(dplyr) # para manipular los datos
library(future)
plan('multisession')
# datos
data(nhs_pathways_covid19)
# seleccionar las últimas 6 semanas de los datos
first_date <- max(nhs_pathways_covid19$date, na.rm = TRUE) - 6*7
pathways_recent <- nhs_pathways_covid19 %>%
filter(date >= first_date)
# definir modelos candidatos
models <- list(
  regression = lm_model(count ~ day),
  poisson_constant = glm_model(count ~ 1), family='poisson'),
  negbin_time = glm_nb_model(count ~ day),
  negbin_time_weekday = glm_nb_model(count ~ day +weekday),
)
# análisis de los datos
counts_overall <- pathways_recent %>%
group_by(date, day, weekday) %>%
summarise(count=sum(count))
# resultados con un 'k'=7
res<-asmodee(
  counts_overall,
  models,
  k =7,
  date_index='date',
  method= evaluate_aic,
  simulate_pi=TRUE)
plot(res, 'date')
```

Figura 6.12: Ejemplo ASMODEE



ASMODEE normalmente sería más útil para investigar los cambios en las tendencias de un gran número de series temporales (por ejemplo, a una escala geográfica fina). Para facilitar este tipo de análisis, **trendbreaker** también funciona con la librería [incidence2](#). Para ilustrar esto, podemos considerar las tendencias sobre las regiones del NHS.

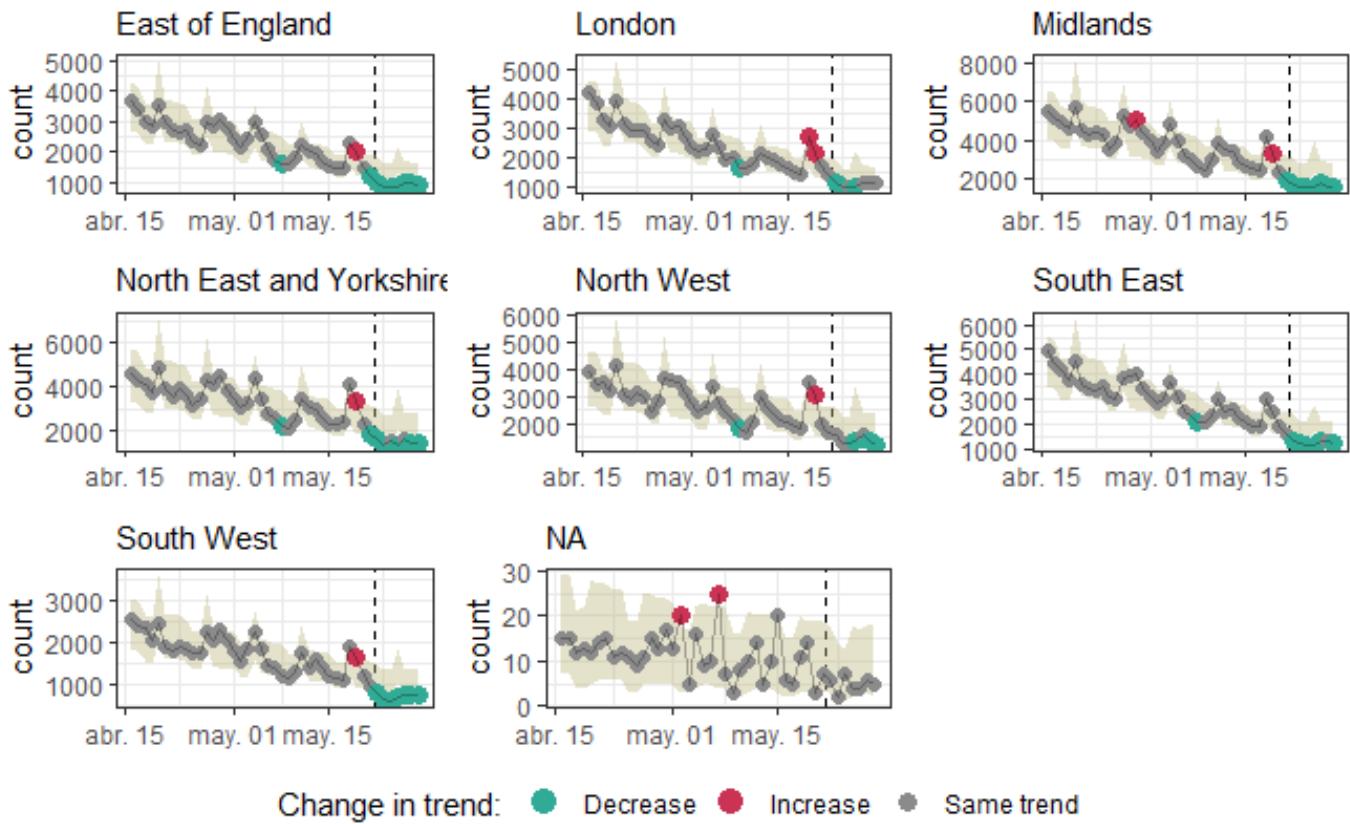
```
library/incidence2)
# seleccionar las últimas 6 semanas de los datos
first_date <- max(nhs_pathways_covid19$date, na.rm = TRUE) - 6*7
pathways_recent <- filter(nhs_pathways_covid19, date >= first_date)
# crear un objeto de incidencia con variables adicionales
lookup <- select(pathways_recent, date, day, weekday) %>% distinct()
dat <-
pathways_recent %>%
incidence(date_index=date, groups=nhs_region, count=count) %>%
left_join(lookup, by=c('date_index'= 'date'))
# definir modelos candidatos
models <- list(
  regression = lm_model(count ~ day),
```

```

poisson_constant = glm_model(count ~ 1), family='poisson'),
negbin_time = glm_nb_model(count ~ day),
negbin_time_weekday = glm_nb_model(count ~ day +weekday),
)
# analyses on all data
res<-asmodee(dat, models, method = evaluate_aic, k=7)
plot(res)

```

Figura 6.13: Ejemplo ASMODEE con incidence2



Podemos notar que esta función implementa un algoritmo para el análisis de series temporales epidémicas con el objetivo de detectar la desviación reciente de la tendencia seguida por los datos.

Los datos se dividen primero en datos 'recientes', utilizando las últimas k observaciones como individuales suplementarios, y los datos más antiguos se utilizan para ajustar la tendencia. El ajuste de la tendencia se realiza ajustando una serie de modelos especificados por el usuario para la serie temporal, con diferentes métodos para seleccionar el mejor ajuste.

A continuación, se calcula el intervalo de predicción para el mejor modelo, y todos los puntos de datos (incluyendo el conjunto de entrenamiento y los individuos suplementarios) que caen fuera se clasifican como 'valores atípicos'.

Es importante mencionar algunos de los elementos de la función como lo es el método para evaluar el ajuste del modelo. Las opciones actuales son **evaluate_aic** (por defecto) y **evaluate_resampling**. **evaluate_aic** utiliza el Criterio de Información de Akaike, que es más rápido pero posiblemente menos bueno a la hora de seleccionar modelos con el mejor poder predictivo. **evaluate_resampling** utiliza la validación cruzada y, por defecto, el RMSE para evaluar el ajuste del modelo.

alpha: Otro argumento es el umbral alfa que se utilizará para el cálculo del intervalo de predicción; por defecto es 0,05, es decir, se calculan intervalos de predicción del 95 %.

2.2. Aplicación en los datos de Chile

Para los datos de casos nuevos de Covid-19 reportados por el Minsal se utiliza la frecuencia absoluta y una ventana de tiempo de 6 semanas como propone el ejemplo en el punto anterior para mostrar los resultados con las fechas entre el día el 27 de septiembre al 8 de noviembre del 2021. Donde el valor de k calibrado resulta ser 7, y además se ocupan los mismos argumentos del método en `evaluate_resampling` y con un umbral alpha igual al 95 % .

```
res<-asmodee(Nacional, models, method= evaluate_resampling, alpha =0,05
k=7)
```

El resultado con estos datos, la función selecciona como mejor modelo de ajuste al binomial negativo `negbin_time` con los siguientes parámetros:

```
theta= 19,87
SE.theta= 4,7
twologlink= -504,08
```

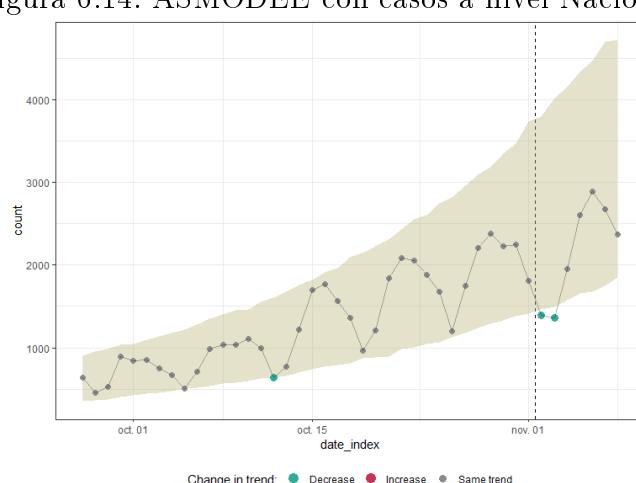
También entrega los resultados de evaluación del modelo:

```
deviance= 36,33
aic= 510,08
null.deviance= 149,59
```

Y el gráfico que muestra es el siguiente:

```
plot(res)
```

Figura 6.14: ASMODEE con casos a nivel Nacional



Se puede apreciar que en la ventana de tiempo de los datos otorgados a la función la tendencia resulta ser creciente y para los datos de predicción encontramos que 2 días fueron clasificados como Decrease y también uno ubicado en la ventana de calibración.

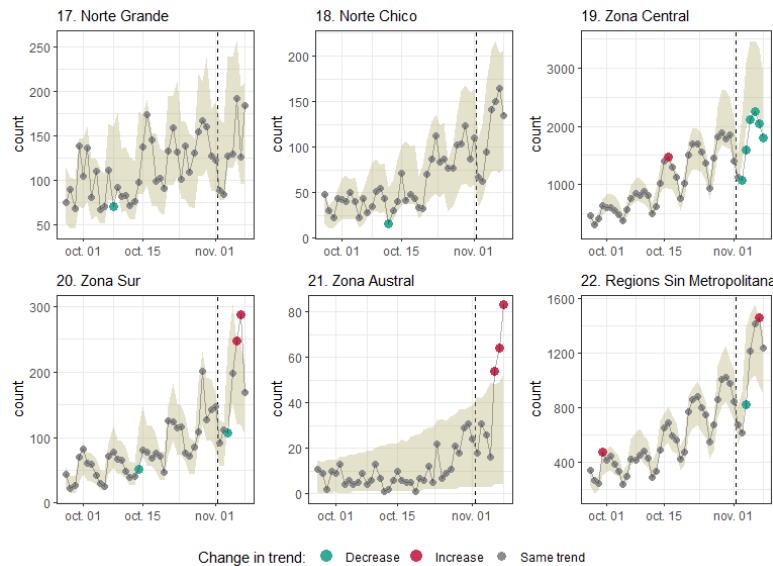
A continuación se aplicará el método para las zonas, regiones y comunas del país:

- **Zonas del país:**

```
res<-asmodee(Zonas, models, method= evaluate_aic, alpha =0.05 k=7)
```

Se aplicó para las 5 zonas del país más las regiones sin la Región Metropolitana, éstas ajustan un modelo Binomial Negativo **negbin_time_weekday** y la zona Austral no aplicó el efecto de los días de la semana. A continuación se muestran los resultados.

Figura 6.15: Gráfico de ASMODEE para las Zonas de Chile



Cuadro 6.1: Resultados de ajuste para las zonas del país.

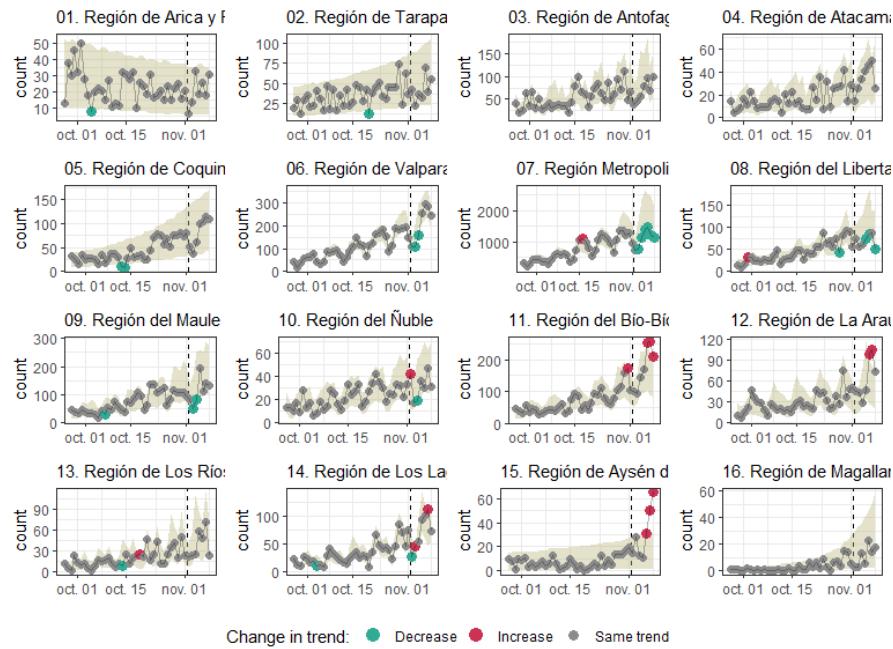
Zonas	N. Grande	N. Chico	Z. Central	Z. Sur	Z. Austral	R. sin M.
Intercepto	-3.77	-15.55	-17.27	-15.77	-17.24	-12.37
day	0.01421730	0.033	0.041	0.034	0.033	0.032
jueves	0.22159405	-0.06859456	0.05388837	0.0755	-	0.045
lunes	0.01051709	0.005500777	-0.1786387	-0.167	-	-0.107
martes	-0.12330461	-0.4625435	-0.561754	-0.5	-	-0.437
miercoles	-0.15757441	-0.3403673	-0.269663	-0.58	-	-0.344
sabado	0.27859792	0.113863	0.07852169	0.03924647	-	0.108
viernes	0.19537283	0.08024719	0.1276279	0.29	-	0.094
deviance	35.98822	37.99	35.37	34.91	37.84	37.686
aic	330.16	305.7	447.87	299.19	224.16	401.71
null.deviance	97.2	146.3	825.37	297.44	52.01	690.46

- **Regiones de Chile**

```
res<-asmodee(Regiones, models, method= evaluate_aic, alpha =0,05 k=7)
```

El gráfico anterior muestra los cambios en la tendencia a simple vista. Ahora para

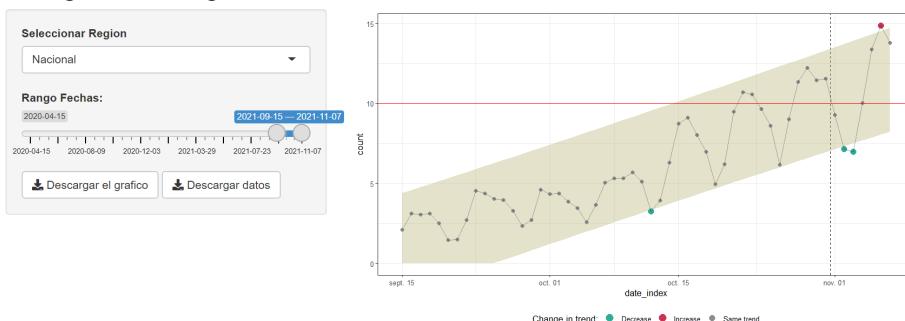
Figura 6.16: Ejemplo ASMODEE Regiones de Chile



una mejor manipulación de las ventanas de tiempo así como la elección de las zonas y regiones de Chile, se implementó un algoritmo utilizando la librería **shiny** de R.

Figura 6.17: Ejemplo utilizando shiny

Change - Point Regional 08 Noviembre



■ Comunas de Chile

Para dar una alerta más focalizada, después de encontrar las regiones que presentan una tendencia creciente, es correcto analizar cuál de las comunas están siendo afectadas con aumentos en los casos de Covid-19.

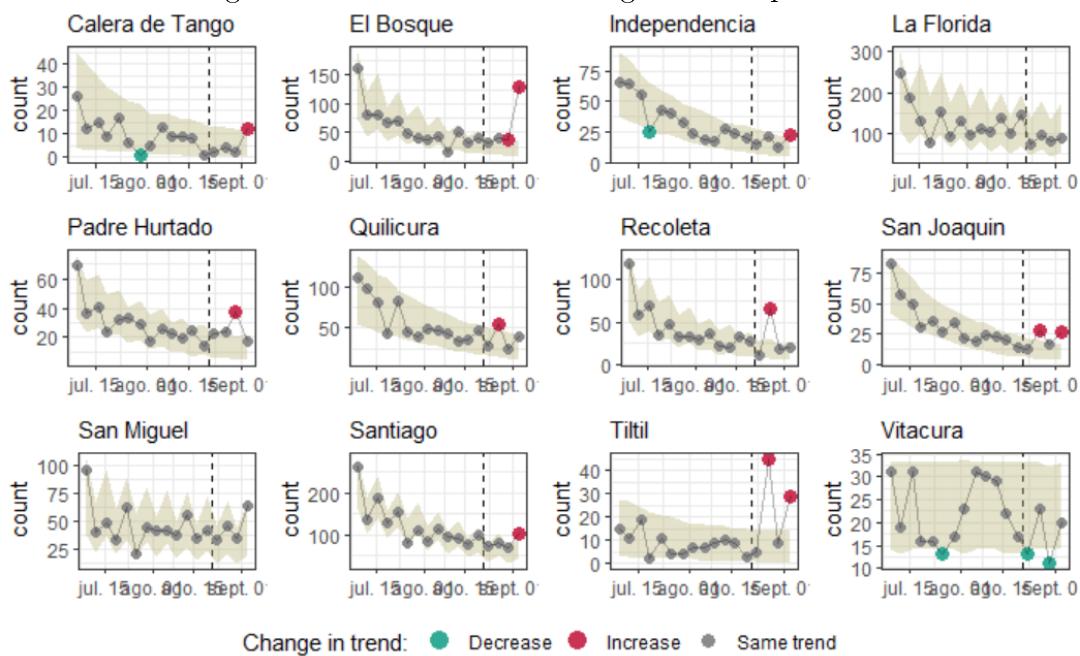
La base de datos con los casos por comunas del país se actualiza menos frecuentemente que los casos regionales, por lo general 2 veces a la semana, sin embargo éste método se puede aplicar de forma sencilla.

En el periodo de tiempo que los casos en Chile seguían una fuerte disminución gracias al proceso de vacunación que alcanzaba ya más del 80 % de la población, se relajaron las medidas y restricciones lo que provocó incrementos en algunas regiones, por lo tanto se aplicó el método para encontrar cuáles de las comunas estaban siendo afectadas.

Como ejemplo se seleccionaron algunas de las comunas de la Región Metropolitana, período entre julio y septiembre con una ventana de 6 semanas y un $k = 4$, que representa a las 2 últimas semanas como ventana de predicción, donde se detectaron comunas con alzas en los casos.

En Anexos se encuentran más resultados en el mismo período de tiempo para las comunas en otras regiones.

Figura 6.18: Comunas de la Región Metropolitana

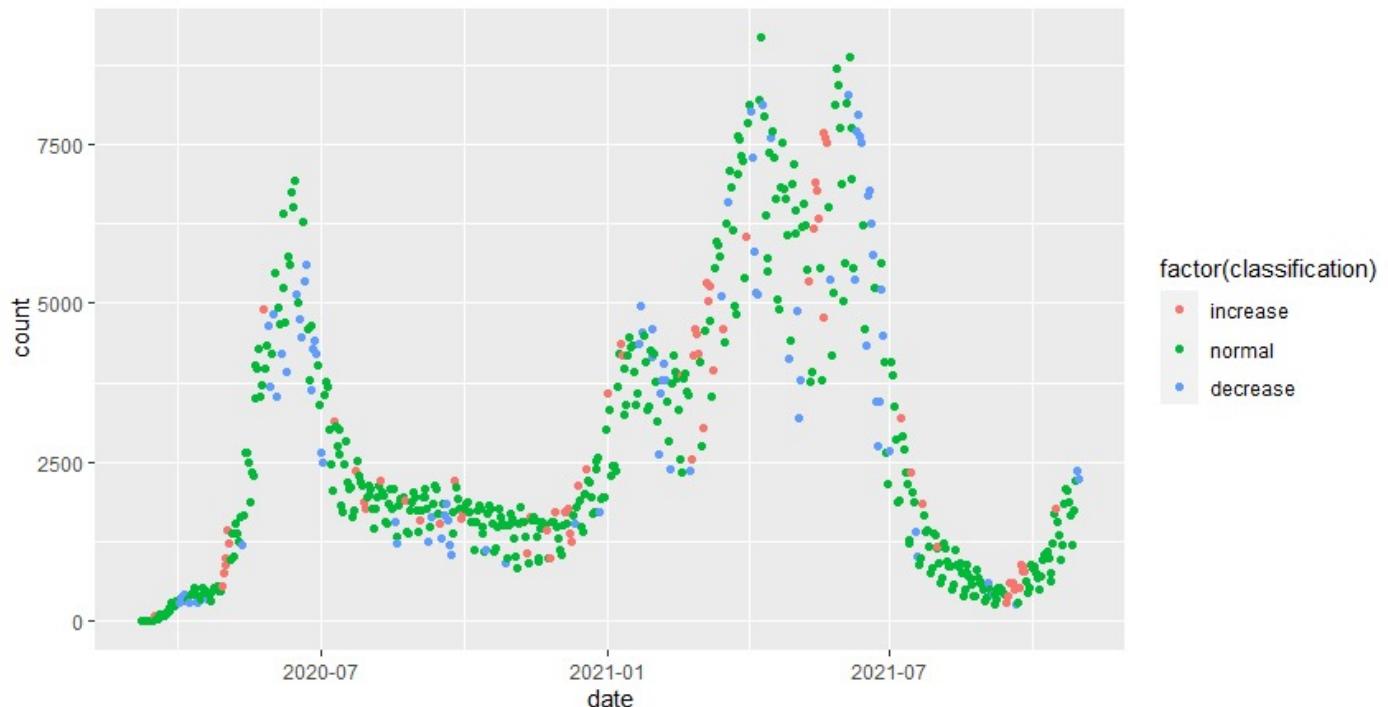


Clasificación en el tiempo

A modo de exploración del método se aplicó el algoritmo con una ventana de 4, 5 y 6 semanas en la que se guardó las clasificaciones de la ventana de predicción de 7 días, secuencialmente, donde con 4 semans se obtienen mejores resultados

Los resultados de las clasificaciones acertaron en detectar los cambios de tendencia que marcaron los inicios de las olas.

Figura 6.19: Clasificación en el tiempo a nivel Nacional



Al utilizar ésta clasificación en el tiempo, se nota que también muestra cuando los casos deberían comenzar a bajar con los puntos de color azul, éste análisis resulta muy interesante de analizar y demuestra que el método es relevante para la investigación. También se realiza el análisis para la Tasa de Incidencia en las Regiones utilizando la misma ventana de tiempo anterior, además de agregar una línea horizontal en los puntos donde la Incidencia alcanza los 10 y 30 casos/100.000hab. Más resultados pueden encontrarse en los Anexos.

Figura 6.20: Cambios de tendencia Incidencia RM

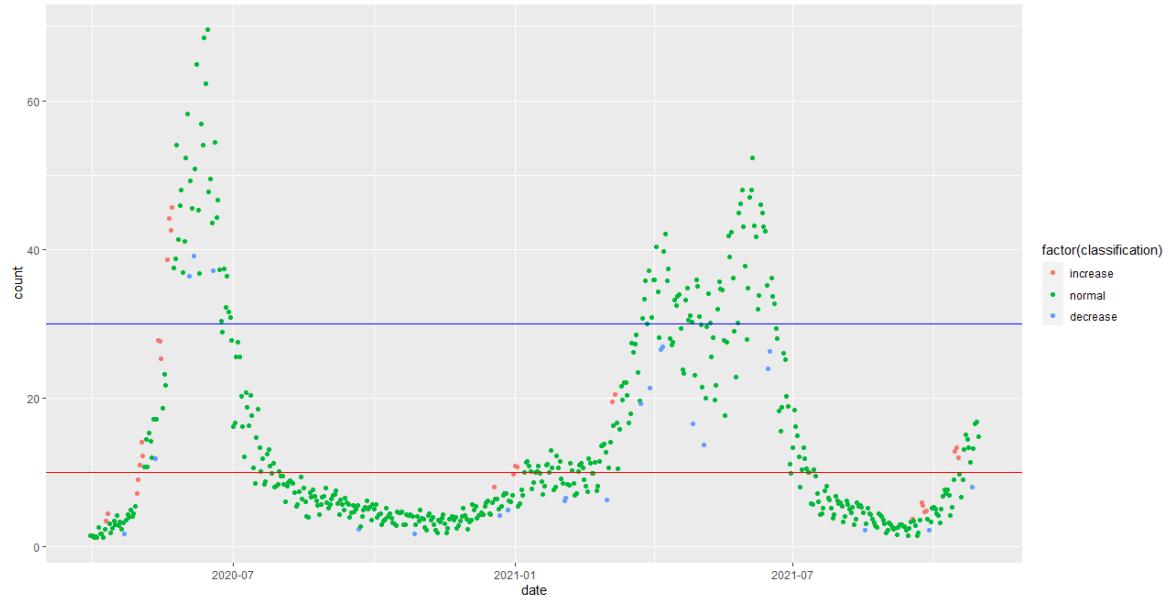
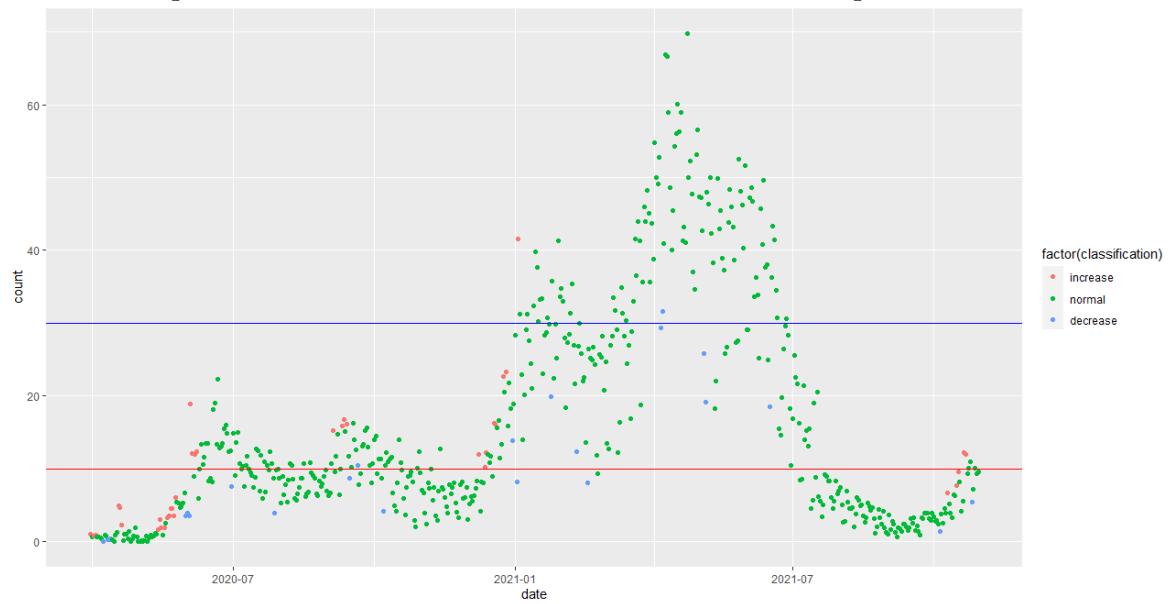


Figura 6.21: Cambios de tendencia en Incidencia Región del Maule



Sistema de Alerta para las regiones de Chile

El Grupo Epidemiológico Matemático para la Vigilancia de Epidemias y Pandemias de la USACH, nacido en Marzo del 2020 ha creado un servicio web analítico que cuenta con dashboards de seguimiento de los datos del COVID-19 proporcionados por el Ministerio de la Ciencia.

Mientras se liberen gradualmente las medidas de distanciamiento, con la creación de vacunas pero en ausencia de una inmunidad generalizada y duradera la detección rápida de nuevos puntos críticos de COVID-19, la intervención inmediata sera clave para evitar el resurgimiento de la transmisión a gran escala.

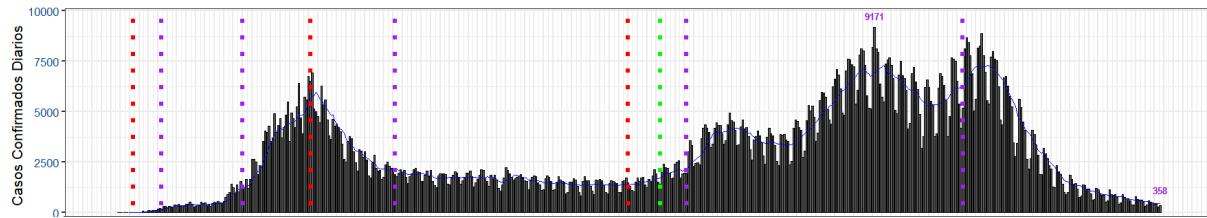
Ante ésta situación, como equipo de investigación, surge la necesidad de crear un Sistema de Monitoreo que permita entregar de manera rápida y eficiente una alerta para las regiones y localidades de Chile sobre la propagación del virus.

A casi 2 años del primer caso de Covid-19 en Chile , podemos notar que se han producido 2 grandes olas en el país con el mayor peak de casos el día viernes 9 de abril del 2021 llegando a 9171 casos.

Ante una enfermedad nueva, donde el desconocimiento de los efectos era absoluto, el comienzo de la primera ola de casos era inminente por lo que fue necesario implementar cuarentenas rápidamente para cortar la circulación del virus ya que los casos seguirían aumentando, se saturan los hospitales y se perderían muchas vidas. Para el año 2021 llegaría una segunda ola de casos, ya había más conocimiento sobre la enfermedad, se contaba con un sistema más robusto de camas críticas y se iniciaba el proceso de vacunación para los grupos de riesgo, sin embargo se subestimó el potencial de la enfermedad lo que provocó una ola de casos aún más grande que la primera. Para lograr una inmunidad generalizada en la población el R efectivo nos indica que el país debiese tener al menos un 80 % de su población vacunada, cifra que se alcanzó en agosto del año 2021, en éste periodo se nota una disminución significativa de casos nuevos como también en hospitalizaciones y fallecidos.

De referencia se toma la serie con los casos reportados a nivel Nacional desde el inicio de la pandemia hasta 2 de Noviembre del 2021, en éste gráfico se muestran seleccionadas las fechas que destacan distintos momentos importantes, que marcan la importancia de la comunicación de riesgo que hacen las autoridades, ya que de haber contado con la información correcta se podría haber evitado una explosión de casos.

Figura 7.1: Gráfico de seguimiento de casos de Covid-19 en Chile actualizado al 2 de Noviembre del 2021



Fuente: Reportes Covid-19 GEMVEP-USACH

fechas importantes:

- 3 de Marzo: Primer Caso de Covid-19 en Chile.
- 18 de Marzo: Casos locales pasaron a ser mayoritarios en el país.
- 26 de Marzo: Primera cuarentena, sector oriente de Santiago.
- 24 de Abril: Plan Retorno Seguro.
- 8 de Mayo: Todas las comunas de Santiago a cuarentena.
- 13 de Junio: Enrique Paris asume como Ministro de Salud en reemplazo de Jaime Mañalich.
- 17 de Junio: Se integran al conteo oficial de contagios 31412 casos no contabilizados previamente.
- 17 de Julio: Se incorporan 1057 fallecidos al conteo oficial.
- 28 de Julio: Se inicia el Plan Paso a Paso.
- 25 de Octubre: Plebiscito por una Nueva Constitución.
- 16 de Diciembre comienza el proceso de vacunación en Chile
- 30 de Diciembre se anuncia el permiso de vacaciones

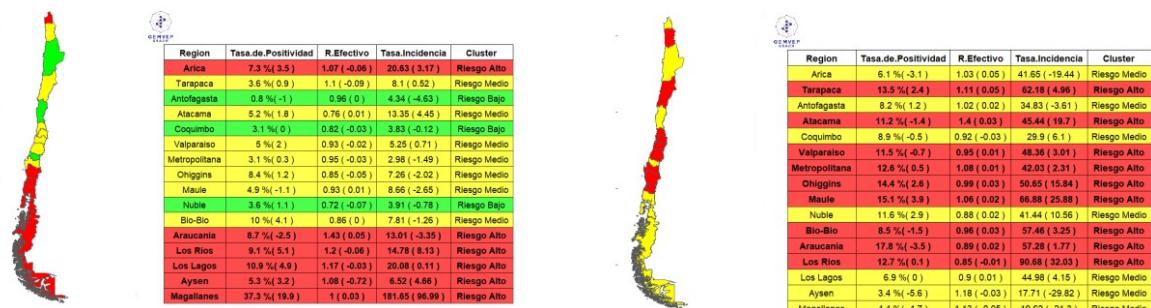
Balance Regional

Se ha observado que la propagación del Covid-19 tiene un comportamiento distinto en diferentes lugares y tiempos. A continuación se presenta la metodología y resultados del sistema de alerta con Clusters que agrupan a las distintas regiones en 3 niveles de riesgo. Ésta clasificación utiliza las variables latentes de la pandemia: tasa de Incidencia, R efectivo y Tasa de positividad, además de sus variaciones semanales.

Para la automatización de esta metodología, se consideran las 6 variables mencionadas previamente. Luego, se realiza una clasificación inicial, mediante componentes principales en la que se usan los dos primeros componentes para separar las zonas de riesgo en cuadrantes. En una segunda etapa, miembros del GEMVEP-USACH hacen revisión de los grupos y verifican que se cumplan los criterios. En caso contrario se realizan las correcciones necesarias. Finalmente, estos grupos son calculados y monitoreados día a día para entregar una alerta lo más temprana posible.

En la Figura a continuación se muestran ejemplos de los clusters de alertas en distintos momentos de la pandemia en Chile. En el gráfico de la izquierda se presentan los grupos obtenidos el día 6 de Octubre de 2020, día en la región de Magallanes alcanzó la mayor tasa de Incidencia desde el inicio de la pandemia en Chile. En la figura de la derecha se presentan los clusters obtenidos el día que se registró la mayor cantidad de casos de covid-19 en Chile (9.171 casos el día viernes 9 de abril de 2021).

Figura 7.2: Grupos de Alerta regional en los días 6 de octubre de 2020 y 9 de abril de 2021



Fuente: Reportes Covid-19, Balance Regional GEMVEP-USACH

Los tres grupos de riesgo se definen a continuación :

- De Riesgo bajo o contención sostenida: En ésta clasificación se encontrarán las regiones con los más bajos indicadores, para estar en esta clasificación las regiones

deben cumplir con los criterios mínimos que exige la OMS (Organización Mundial de la Salud, 2020) y en especial se considera que la región tenga un R efectivo inferior a uno, que presente una tasa de Incidencia menor a 5 casos/100.000 habitantes y una tasa de positividad menor al 5 %.

- De Riesgo medio o de Probable Crecimiento: Para esta clasificación se consideran las regiones que tengan indicadores que se mantengan estables. En este grupo se espera que los indicadores presenten una incidencia entre 5 y 10 casos/100.000 habitantes y una tasa de positividad menor al 5 %.
- De Riesgo alto o Tendencia en Crecimiento: Esta clasificación sugiere medidas de confinamiento ya que representa altas tasas de contagio. En esta clasificación se encuentran las regiones que tengan una tasa de Incidencia en crecimiento mayor a 10 casos/100.000 habitantes, una tasa de positividad mayor al 5 % y un R efectivo mayor a 1.[33]

A continuación se muestra un ejemplo de la clasificación que tendrían las regiones considerando las variables ya mencionadas y además un resumen de los indicadores del día 26 de Agosto del 2021.

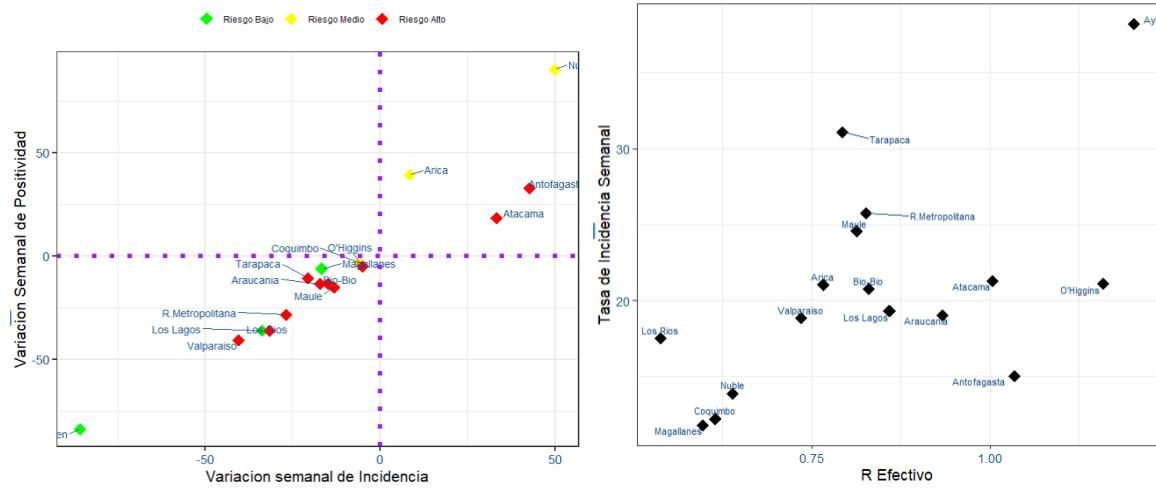
Figura 7.3: Resumen de indicadores diarios

Region	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	Cluster
Arica	0.5 %(0.3)	0.77 (0.07)	3.17 (0.79)	Riesgo Alto
Tarapaca	0.9 %(0.5)	0.79 (0.08)	6.27 (5.23)	Riesgo Alto
Antofagasta	1.1 %(0.7)	1.04 (0.18)	4.63 (3.47)	Riesgo Medio
Atacama	0.8 %(0.5)	1 (0.14)	6.04 (4.77)	Riesgo Medio
Coquimbo	1.1 %(0.2)	0.61 (0.01)	2.75 (1.67)	Riesgo Medio
Valparaiso	1.1 %(-0.2)	0.74 (-0.06)	2.4 (0.56)	Riesgo Alto
Metropolitana	1.7 %(0.5)	0.83 (0)	4.52 (1.9)	Riesgo Medio
Ohiggins	2 %(0.6)	1.16 (0.05)	4.14 (2.83)	Riesgo Alto
Maula	1.4 %(0.3)	0.81 (-0.02)	4.42 (2.39)	Riesgo Alto
Nuble	1 %(0.1)	0.64 (-0.03)	2.74 (1.76)	Riesgo Medio
Bio-Bio	1 %(0.1)	0.83 (-0.07)	4.15 (2.58)	Riesgo Medio
Araucania	1.7 %(0.8)	0.93 (0.03)	3.25 (2.17)	Riesgo Alto
Los Rios	1.3 %(-0.3)	0.54 (0.06)	4.44 (0.99)	Riesgo Alto
Los Lagos	0.8 %(0.2)	0.86 (0.05)	4.82 (3.48)	Riesgo Bajo
Aysen	1.7 %(0.4)	1.2 (-0.04)	10.25 (5.59)	Riesgo Alto
Magallanes	0.2 %(0.2)	0.6 (-0.02)	1.12 (1.12)	Riesgo Medio

Fuente: Clusters con colores asociado al riesgo

Para entregar una correcta clasificación de riesgo se realiza una revisión de los indicadores, que permite clasificar a los grupos de regiones cuando existe incertidumbre con los resultados que entregan los clusters.

Figura 7.4: La primera figura corresponde Variaciones semanales Positividad v/s Incidencia y la segunda figura corresponde a la interacción Incidencia v/s R efectivo



Fuente: Visualizador Regional GEMVEP

1. Clusters de Alerta

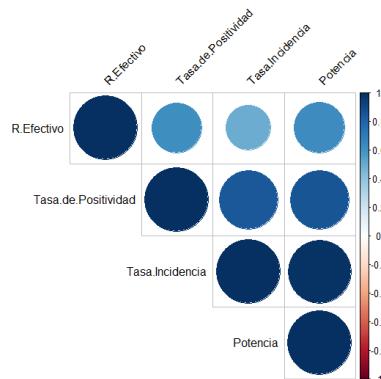
En éste capítulo se investigarán variables a considerar para mejorar el comportamiento de los clusters además de incorporar los resultados de 'Change-Point', dónde se mostrarán los análisis para la validación de los distintos resultados..

Los clusters de alertas entregados se han creado utilizando componentes principales con variables latentes de los casos nuevos y sus variaciones (tasa de positividad, tasa de incidencia, tasa de reproducción). En esta sección se realizará la validación de las variables incorporadas y se incorporarán muevas para luego hacer la comparación de validación con el fin de mejorar la clasificación. Las variables que se incorporan son:

- Potencia

La potencia es el producto entre la tasa de incidencia con el R efectivo. A modo de prueba con los datos del día 23 de noviembre sin las variaciones de los indicadores, se prueba la correlación que existe. Puede ser evidenteLa que la variable potencia esté bien correlacionada por su composición.

Figura 7.5: Correlación entre las variables



Cuadro 7.1: Correlación con la variable potencia

	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	Potencia
Tasa.de.Positividad	1.00	0.62	0.84	0.85
R.Efectivo	0.62	1.00	0.50	0.63
Tasa.Incidencia	0.84	0.50	1.00	0.98
Potencia	0.85	0.63	0.98	1.00

- Clasificación de los métodos de 'Change-Point'.

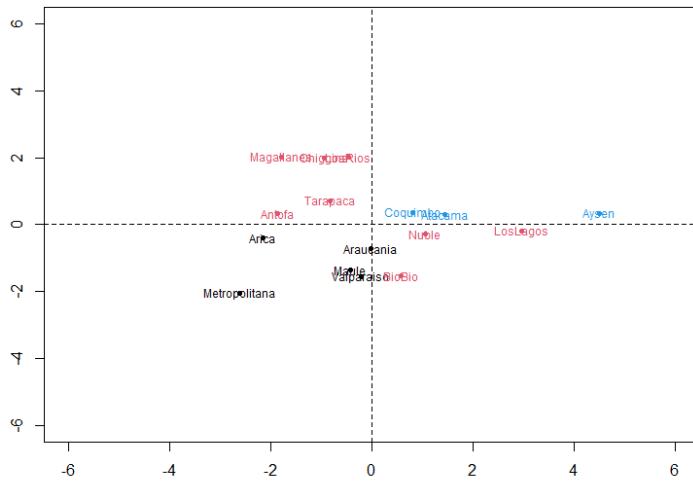
En ésta clasificación se incorporarán resultados obtenidos por los dos métodos de Change-Point investigados.

Del primer método se incorporará la clasificación de épocas y fases, mientras que del segundo método se tomarán en cuenta las clasificaciones de los valores atípicos.

Cuadro 7.2: Resultados de los cambios de tendencia

Region	Posit	R.Efectivo	Incidencia	Potencia	clasif	fase	epoca
Arica	1.40	0.70	3.97	2.77	-2.00	7.00	3.00
Tarapaca	0.80	1.10	6.01	6.61	0.00	10.00	1.00
Antofagasta	1.40	0.82	5.20	4.26	-2.00	13.00	1.00
Atacama	6.50	1.33	15.25	20.27	0.00	6.00	2.00
Coquimbo	8.30	1.15	12.92	14.85	-1.00	12.00	1.00
Valparaiso	5.80	1.06	7.91	8.37	-1.00	15.00	3.00
Metropolitana	1.90	0.78	5.08	3.99	-5.00	17.00	3.00
Ohiggins	4.70	0.91	7.26	6.65	-2.00	0.00	1.00
Maule	2.70	1.05	8.57	9.00	0.00	14.00	3.00
Nuble	4.30	1.33	12.51	16.60	1.00	11.00	2.00
Bio-Bio	5.00	1.16	15.93	18.54	-1.00	14.00	3.00
Araucania	4.60	1.12	8.18	9.20	0.00	15.00	2.00
Los Rios	5.00	0.80	12.81	10.27	-1.00	1.00	1.00
Los Lagos	8.30	1.47	24.45	35.99	0.00	9.00	2.00
Aysen	11.50	1.12	47.53	53.01	0.00	7.00	2.00
Magallanes	1.70	0.77	2.80	2.16	-1.00	1.00	1.00

Figura 7.6: Cluster-Point



Cuadro 7.3: Cluster-Point

	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	clusterreg
Arica	1.40	0.70	3.97	1.00
Tarapaca	0.80	1.10	6.01	2.00
Antofagasta	1.40	0.82	5.20	2.00
Atacama	6.50	1.33	15.25	3.00
Coquimbo	8.30	1.15	12.92	3.00
Valparaiso	5.80	1.06	7.91	1.00
Metropolitana	1.90	0.78	5.08	1.00
O'Higgins	4.70	0.91	7.26	2.00
Maule	2.70	1.05	8.57	1.00
Nuble	4.30	1.33	12.51	2.00
BíoBío	5.00	1.16	15.93	2.00
Araucania	4.60	1.12	8.18	1.00
LosRíos	5.00	0.80	12.81	2.00
LosLagos	8.30	1.47	24.45	2.00
Aysen	11.50	1.12	47.53	3.00
Magallanes	1.70	0.77	2.80	2.00

8

Clonclusión

Resulta fundamental tener un sistema en el país que permita entregar la información de casos lo más rápido posible, ya que al procesar éstos datos se puede alertar de forma correcta y a tiempo.

Es relevante que para dar una alerta a la ciudadanía, esta sea lo más rigurosa posible y el monitoreo de los indicadores resulta ser clave.

Los métodos de Change-Point que se aplicaron resultan ser eficientes en la entrega de información relevante para la vigilancia de la pandemia y que conjuntamente se vuelven una herramienta poderosa a la hora de dar una alerta, tanto para las regiones como dentro de ellas.

Éstos métodos son fáciles de usar y se pueden implementar en distintos escenarios, aunque el enfoque fue alertar lo más rápido posible con el reporte de casos nuevos, los métodos también sirven para monitorear y detectar los incrementos en las hospitalizaciones y muertes por lo que es importante añadir estos métodos en la toma de decisiones.

Como respuesta a la pandemia en Chile se complementaron los resultados de Change-Point y se agregaron a la clasificación de riesgo con clusters para alertar a las distintas regiones del país sobre los aumentos de casos.

Se logró automatizar los métodos con ayuda de la librería shiny con una mayor flexibilidad al momento de identificar cambios significativos en los casos por región y comuna.

La pandemia dejó su huella a nivel mundial y éste trabajo marca la diferencia a la hora de alertar a las regiones y comunidades del país.

Referencias

- [1] Fermin, G. (2018). Host range, host–virus interactions, and virus transmission. *Viruses*, 10(1).
- [2] Madigan, M. T., Martinko, J. M., Barrachina, C., Berlanga, M., Gonzalo, M., Diaz, C., & Ruiz Berraquero, F. (2009). Brock; 2004. Biología de los microorganismos 10ma Edición.
- [3] OMS (2010) : Organización Mundial de la Salud. ¿Qué es una pandemia?.
https://www.who.int/csr/disease/swineflu/frequently_asked_questions/pandemic/es/
- [4] Hays, J. N. (2005). Epidemics and pandemics: their impacts on human history. Abc-clio.
- [5] Yan, P., & Chowell, G. (2019). Quantitative methods for investigating infectious disease outbreaks (Vol. 70). Cham, Switzerland: Springer.
- [6] Serra Valdés, M. Á. (2020). Infección respiratoria aguda por COVID-19: una amenaza evidente. *Revista Habanera de Ciencias Médicas*, 19(1), 1-5.
- [7] WHO COVID-19 Explorer. (2020). Shinyapps.io. Recuperado el 05-12-2021.
<https://worldhealthorg.shinyapps.io/covid/>
- [8] Fallecimientos en Chile-DEIS.
Recuperado el 05-12-2021: <https://gemvepusach.shinyapps.io/FallecimientosDeis/>

-
- [9] Chen, J., & Gupta, A. K. (2000). Parametric statistical change point analysis (Oberwolfach seminars).
 - [10] Elorrieta F, Vargas C, Guerrero C, Maripangui C, Crespo F, Navarro V, Oviedo F., Hernández C., Osorio P., Ayala A. (2021). 'Dashboard e Informes Analíticos para el estudio y seguimiento del COVID-19 GEMVEP-USACH'. Disponible en:<http://www.gemvep.usach.cl/>
 - [11] Inkelas M, Blair C, Furukawa D, Manuel VG, Malenfant JH, Martin E, et al. (2021) Using control charts to understand community variation in COVID-19. PLoS ONE 16(4): e0248500. <https://doi.org/10.1371/journal.pone.0248500>
 - [12] 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.
 - [13] Jombart, T., Ghozzi, S., Schumacher, D., Taylor, T. J., Leclerc, Q. J., Jit, M., ... & Edmunds, W. J. (2021). Real-time monitoring of COVID-19 dynamics using automated trend fitting and anomaly detection. Philosophical Transactions of the Royal Society B, 376(1829), 20200266.
 - [14] R Core Team (2013). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL: <http://www.R-project.org>.
 - [15] Wood, S. N. (2017). Generalized additive models: an introduction with R. CRC press.
 - [16] Figueroa, N. R., Suárez, E., De La Torre, T., Epidemióloga, C. T. R., Torres, M., Pérez, J., & Bioestadístico, C. T. R. INCIDENCIA INCIDENCIA Y MORTALIDAD DE CÁNCER EN VIEQUES CÁNCER EN VIEQUES CÁNCER EN VIEQUES 1990-2004.
 - [17] Cori, A., Ferguson, N. M., Fraser, C., & Cauchemez, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. American journal of epidemiology, 178(9), 1505-1512.
 - [18] Peña, D. (2002). Análisis de datos multivariantes (Vol. 24). Madrid: McGraw-hill.
 - [19] Desgraupes, B., & Desgraupes, M. B. (2018). Package 'clusterCrit'.
 - [20] Sonesson, C., & Bock, D. (2003). A review and discussion of prospective statistical surveillance in public health. Journal of the Royal Statistical Society: Series A (Statistics in Society), 166(1), 5-21.

- [21] What is a Change-Point Analysis? (2021). Variation.com.
Recuperado de: https://variation.com/wp-content/change-point_analyzer_help/hs2.htm
- [22] Correa Morales, J. C., & Barrera Causil, C. J. (2018). Introducción a la estadística bayesiana: notas de clase. Instituto Tecnológico Metropolitano.
- [23] Calahorra Tovar, J. (2017). Procesos puntuales espacio-temporales con aplicación a la modelización de accidentes de tráfico.
- [24] Akaike, H. (1973). Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika*, 60(2), 255-265.
- [25] Shewhart, W. A. (1931). Economic control of quality of manufactured product. Macmillan And Co Ltd, London.
- [26] <https://github.com/0sorio8Pedro/NewFT>
- [27] <https://github.com/GEMVEP-USACH/Datos>
- [28] <https://github.com/MinCiencia/Datos-COVID19/>
- [29] klittle314. (2021). GitHub-klittle314/IHI_Covid_Public: This is the public version of the private repo IHI_Covid_Private used to drive the IHI Covid PowerBI display. GitHub. https://github.com/klittle314/IHI_Covid_Public
- [30] <https://isqua.org/covid19-research-page.html>
- [31] <https://github.com/nytimes/covid-19-data>
- [32] reconhub. (2021, April 23). GitHub - reconhub/trendbreaker: Analysing epidemic trends using AutoML approaches.
GitHub. <https://github.com/reconhub/trendbreaker>
- [33] VOZ, E. P. (2021). Gestión del covid-19 en América Latina. Una respuesta interdisciplinaria para comunicación de riesgo en pandemia. El proyecto del Grupo Epidemiológico Matemático para la Vigilancia de Epidemias y Pandemias de la Universidad de Santiago de Chile. Por Felipe Elorrieta, Claudio Vargas, Fernando Crespo, Cesar Maripangui, Valentina Navarro, Catalina Hernández, Pedro Osorio y Andrés Ayala.

Anexos

1. Gráficos de control primer método

Figura 8.1: Cambios de tendencia nivel Nacional ajustado

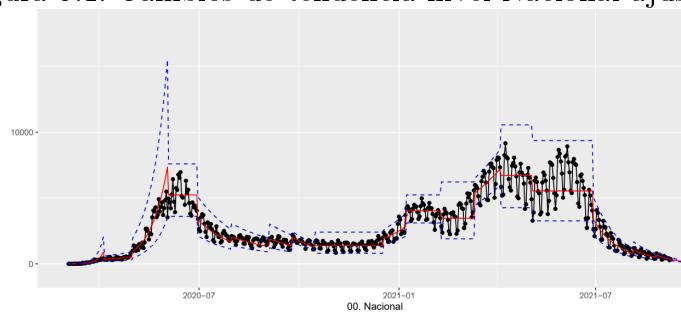


Figura 8.2: Gráficos de control, región de Arica

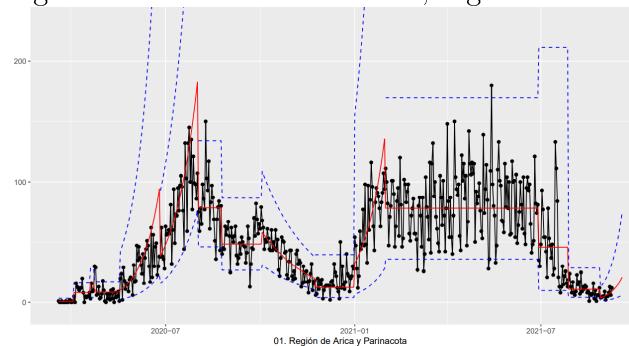


Figura 8.3: Gráficos de control, región de Ñuble

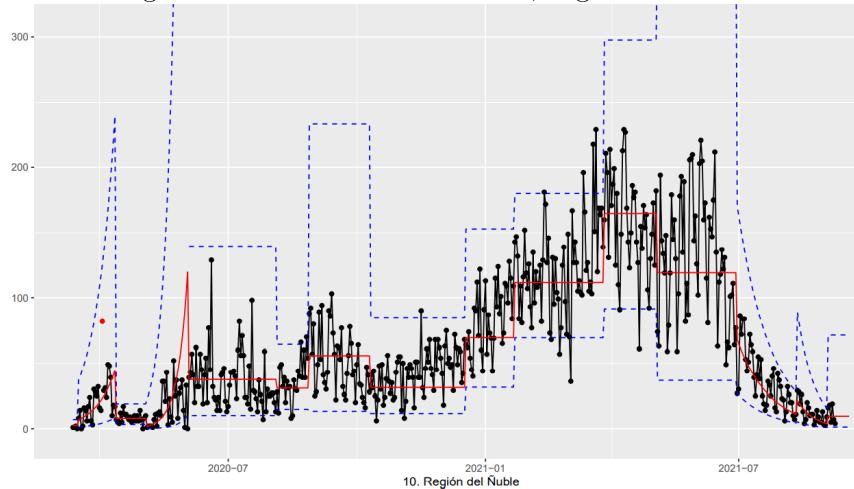


Figura 8.4: Cambios de tendencia con gráficos de control

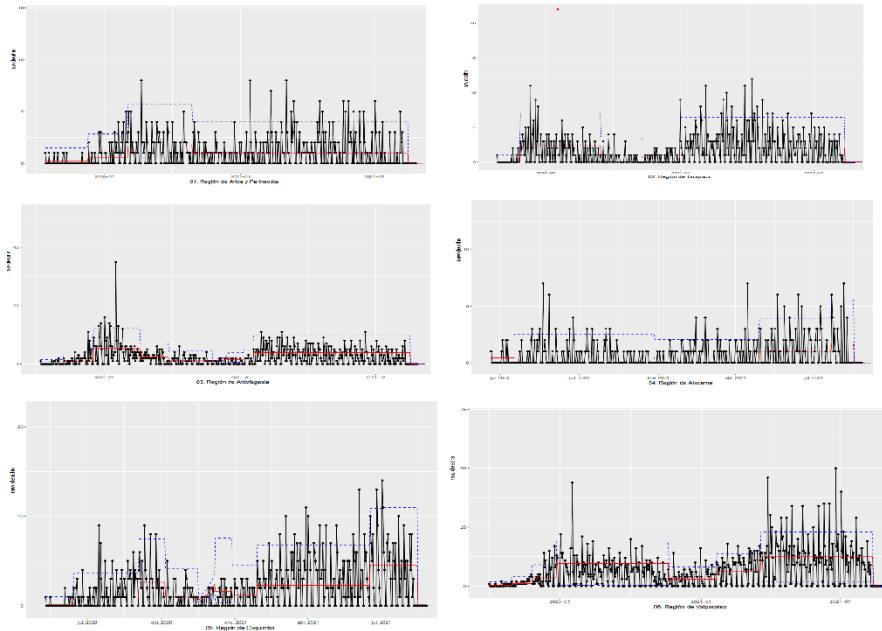
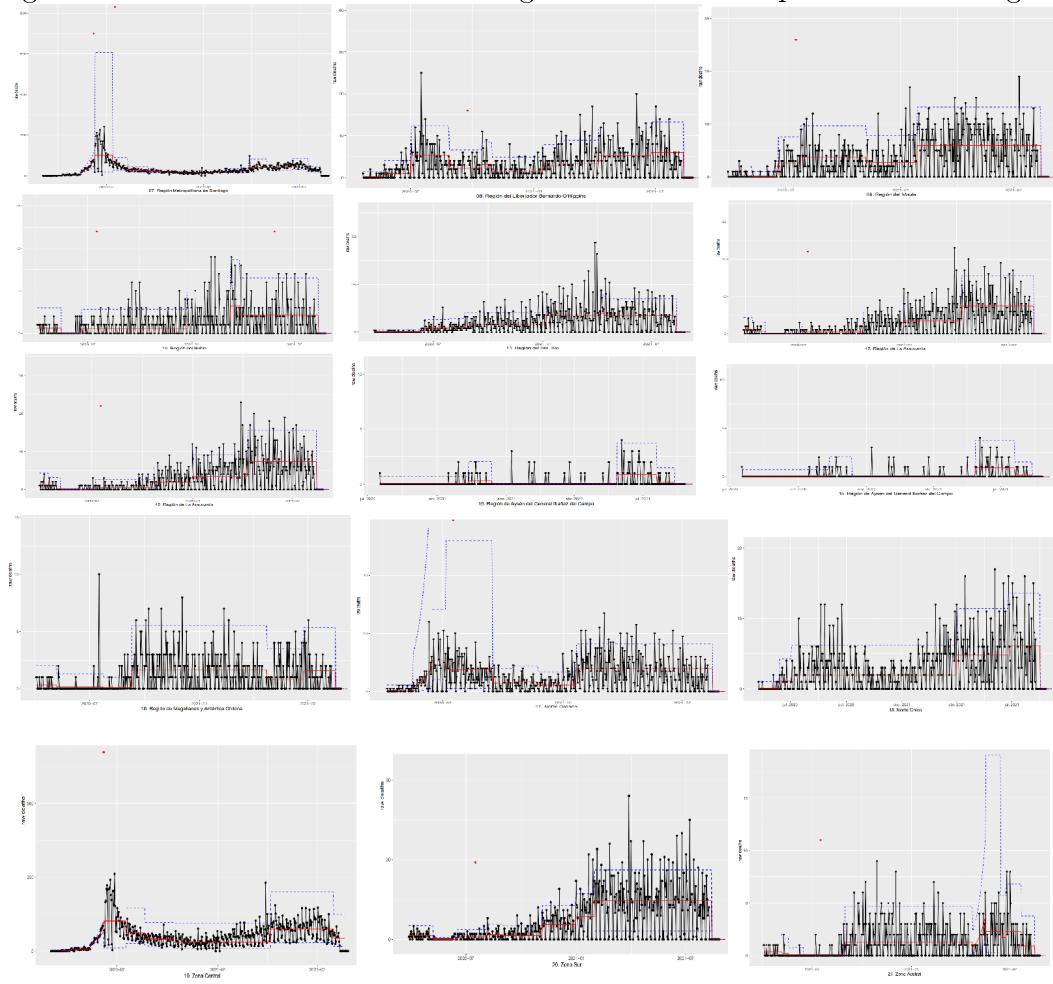


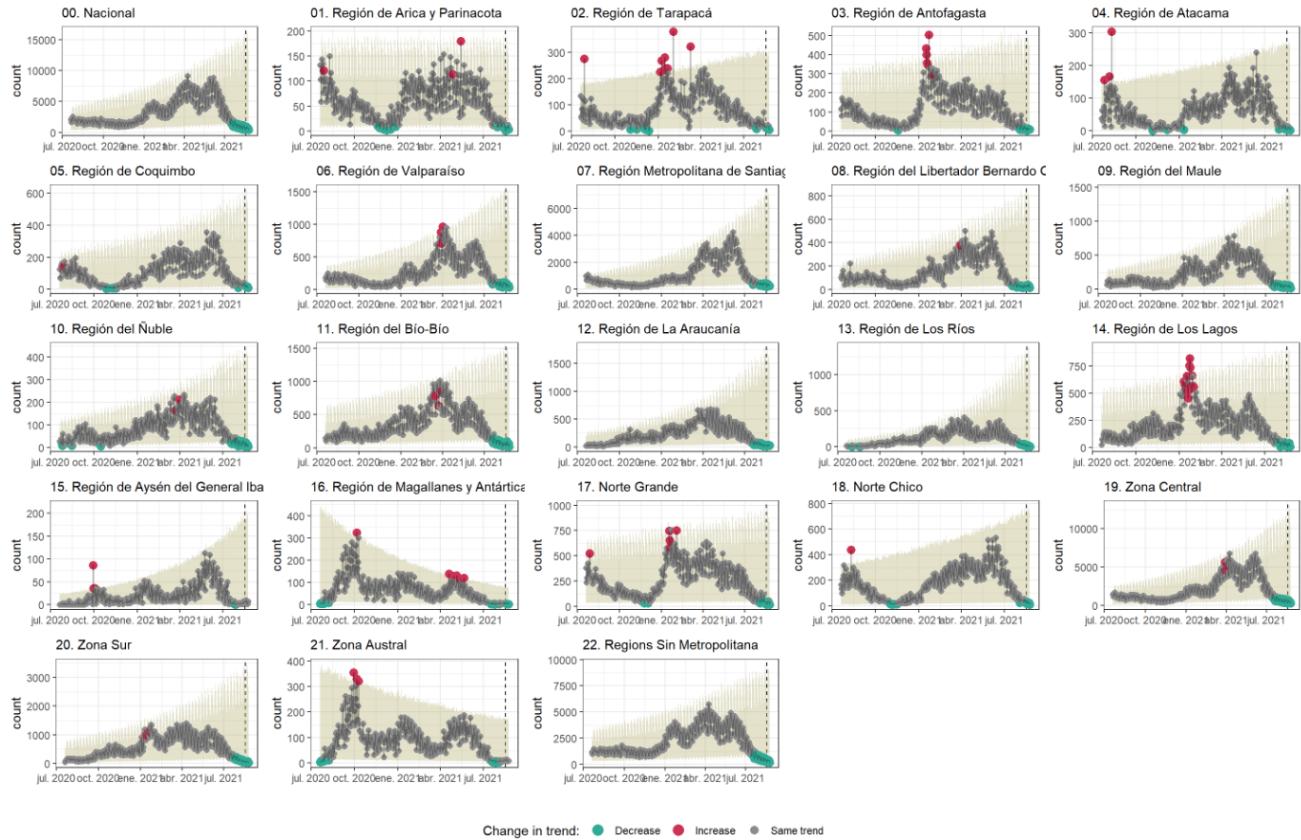
Figura 8.5: Cambios de tendencia con gráficos de control para todas las regiones



2. Gráficos de control segundo método

A continuación la imagen muestra los resultados a nivel nacional, regiones y zonas del país de la cual se puede notar la disminución en los casos para éstos últimos días en donde en ésta se tomó una ventana o valor de $k = 7$ días, desde el 20 de Julio hasta el 26 de Agosto del presente año.

Figura 8.6: ASMODEE completo con todos los datos



Comunas de la Región Metropolitana de Chile

```
res<-asmodee(Regiones, models, method= evaluate_aic, alpha =0,05 k=7)
```

Figura 8.7: Cambios de tendencia

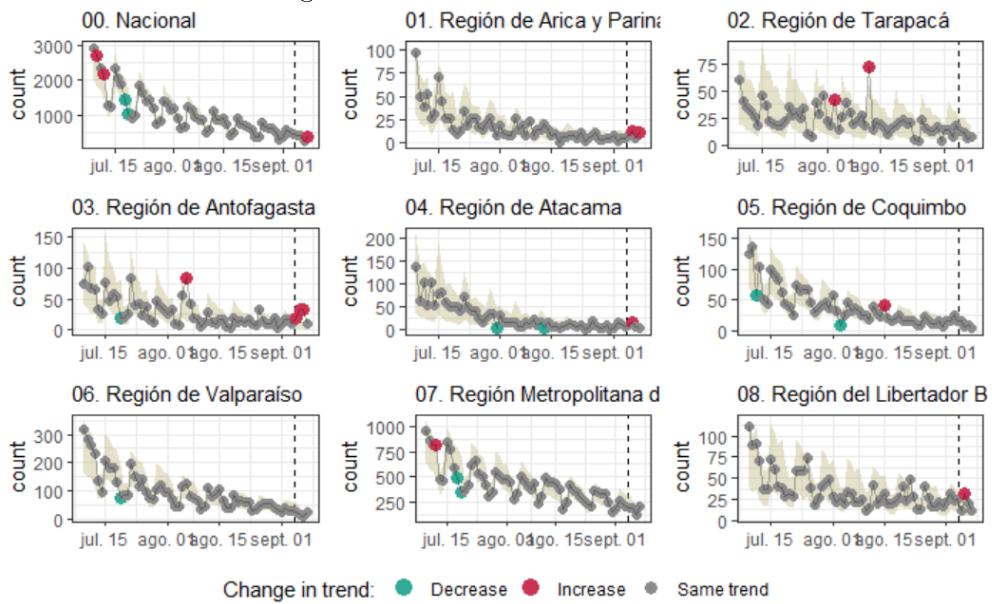
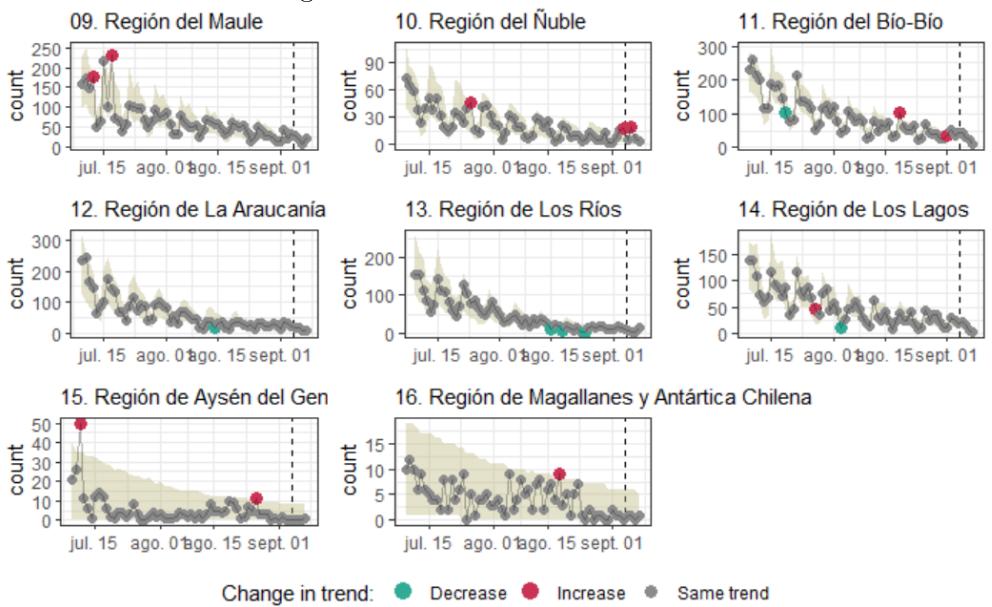


Figura 8.8: Cambios de tendencia



Cuadro 8.1: Resumen de indicadores del 23 de noviembre

Region	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	Potencia
1 Nacional	3.50	0.97	8.74	8.47
2 Tarapacá	0.80	1.10	6.01	6.61
3 Antofagasta	1.40	0.82	5.20	4.26
4 Atacama	6.50	1.33	15.25	20.27
5 Coquimbo	8.30	1.15	12.92	14.85
6 Valparaíso	5.80	1.06	7.91	8.37
7 O'Higgins	4.70	0.91	7.26	6.65
8 Maule	2.70	1.05	8.57	9.00
9 Bío-Bío	5.00	1.16	15.93	18.54
10 Araucanía	4.60	1.12	8.18	9.20
11 Los Lagos	8.30	1.47	24.45	35.99
12 Aysén	11.50	1.12	47.53	53.01
13 Magallanes	1.70	0.77	2.80	2.16
14 Metropolitana de Santiago	1.90	0.78	5.08	3.99
15 Los Ríos	5.00	0.80	12.81	10.27
16 Arica	1.40	0.70	3.97	2.77
17 Ñuble	4.30	1.33	12.51	16.60
18 Norte Grande	1.20	0.89	5.20	4.63
19 Norte Chico	7.70	1.22	13.56	16.52
20 Zona Central	3.00	0.92	7.41	6.79
21 Zona Sur	6.70	1.20	15.27	18.29
22 Zona Austral	7.70	1.03	19.60	20.28
23 Regions Sin Metropolitana	4.60	1.10	11.36	12.53

Por último un ejemplo de clusters obtenido con los datos anteriores.

Figura 8.9: Cambios de tendencia

