



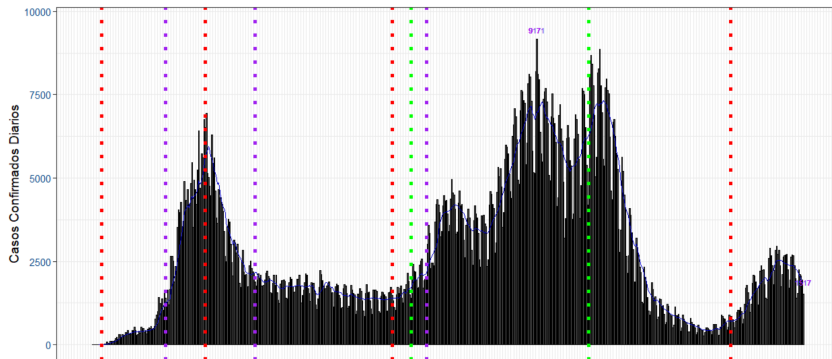
Universidad de Santiago de Chile
Facultad de Ciencia
Departamento de Matemática y Ciencias de la
Computación
Ingeniería Estadística

Sistema de Monitoreo de la pandemia de COVID-19 en Chile mediante análisis de Change-Point

Pedro Osorio Meza

7 de Diciembre de 2021

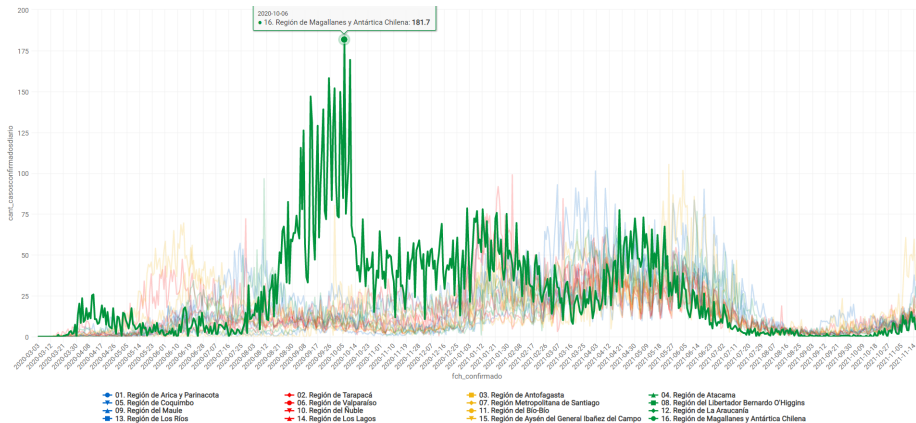
Introducción



Fechas Importantes

- Marzo: 3 Primer caso, 11 WHO declara Pandemia ,26 Cuarentena
- 8 Mayo Cuarentena Región Metropolitana
- Dic: 16 Vacunación, 30 Permiso de Vacaciones
- 26 Mayo Pase de movilidad

Incidencia para las regiones del país



Objetivos

Objetivo General

Desarrollar un sistema de Monitoreo de la pandemia de COVID-19 en Chile, el cual permita comunicar a las regiones y comunas de posibles brotes de ésta enfermedad, incorporará resultados de métodos de Change-Point a una clasificación de riesgo con Alerta con Clusters.

Objetivos Específicos

- Implementar dos métodos de Change-Point a las series de contagios de las distintas regiones y comunas del país.
- Evaluar y comparar los resultados obtenidos de los métodos aplicados.
- Determinar cual método de Change-Point es apropiado para agregar los resultados al análisis de Clusters.

¿Qué es un Change-Point?

Change-Points también se conocen como: breakpoints, segmentation, structural breaks, regime switching, detecting disorder etc.

Para los datos y_1, \dots, y_n , si existe un punto de cambio en τ , entonces y_1, \dots, y_τ difieren de $y_{\tau+1}, \dots, y_n$ de alguna manera.

Así, un modelo de Change-Point para un cambio de media tiene la siguiente forma:

$$y_\tau = \begin{cases} \mu_1 & \text{Si } 1 \leq t \leq \tau_1 \\ \mu_2 & \text{Si } \tau_1 < t \leq \tau_2 \\ \cdot & \cdot \\ \mu_{m+1} & \text{Si } \tau_m < t \leq \tau_{m+1} = n \end{cases}$$

A raíz de la pandemia de Covid-19 en el mundo, se han propuesto métodos para el control de los brotes basados en Change-Point:

Métodos propuestos

- Inkelas et al. **Using control charts to understand community variation in COVID-19.**
Arthur Wakefield Baker, Duke University, EE.UU. (30 Abril 2021)
- **Automatic Selection of Models Outlier DETection for Epidemics (ASMODEE).**
Método contenido en la librería "trendbreaker" de R, propuesta por la OMS. (31 mayo 2021) .

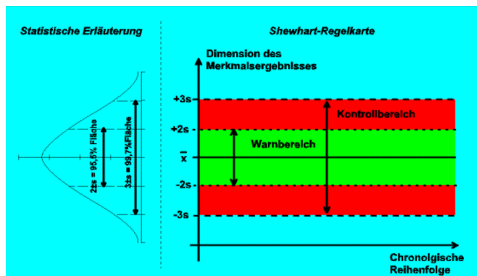
Cartas de Control de Shewhart

Walter Andrew Shewhart (1891-1967)

En 1931, teorizó que la variación en una medida tiene dos orígenes potenciales: causas comunes y causas especiales.



Figura: Gráfico de control de Shewhart

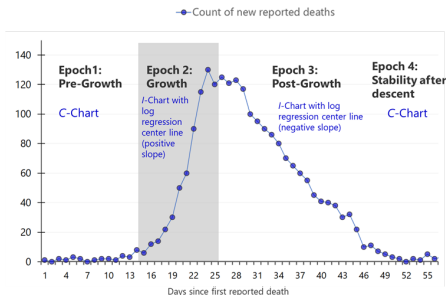


Fuente: <https://www.faes.de/Basis/Basis-Statistik>

Using control charts to understand community variation in COVID-19

- Utiliza un gráfico de control híbrido para los datos de recuento y el crecimiento o disminución exponencial desarrollado por Perla et al.(26 de junio de 2020): para ver las fases o épocas en una pandemia

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



Fuente: <https://isqua.org/latest-blog/shewhart-charts-for-covid-19-reported-deaths.html>

Época-1 (pre-growth)

Casos reportados son bajos y estables Incluye linea central (media) y límites de control superior e inferior 3 sigma por encima y debajo de la línea central . Un total de al menos 8 observaciones y al menos 5 distintos de 0.

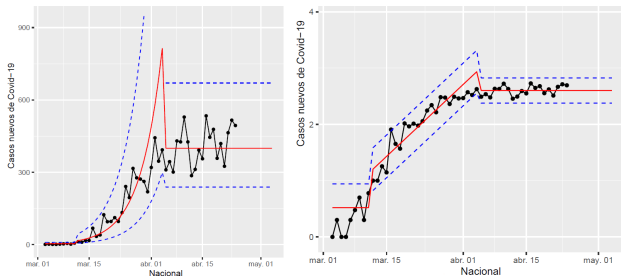
cambia de fase con 2 puntos consecutivos por encima del limite superior, 8 puntos por encima o debajo de la linea central.

Época-2 (growth)

Comienza con una señal de causa especial en el gráfico C que indica un aumento significativo. Una pendiente de regresión no negativa significativa después de al menos 5 días de observaciones incluidas en los cálculos de regresión. Ajustado al \log_{10} de la serie y transformado de nuevo a la escala original, Termina después de 21 días de crecimiento para volver a calcular la linea central y los límites.

Metodología cartas de control de Shewhart

Figura: Épocas asociadas con casos reportados por COVID-19



Época-3 (post-growth)

No hay crecimiento exponencial (meseta)(descenso)

Época-4 Estabilidad después del descenso

Comienza a causa especial que marcan el final de una fase en la Época 3,

Cartas de control de Shewhart Fase 3 Para Chile

Fechas importantes de cambio de tendencia

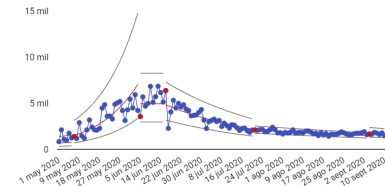
Casos nuevos: 5 Junio 3563 , 16 Junio 6392, 21 Julio 2109, 4 Sept 1667

Fallecidos: 28 Junio 143, 27 Julio 69, 23 Agosto 55

International Society for Quality in Health Care (ISQua)

Figura: Épocas asociadas con casos diarias y fallecidos en Chile reportadas por COVID-19

Chile : Adjusted Daily Covid-19 Reported Cases

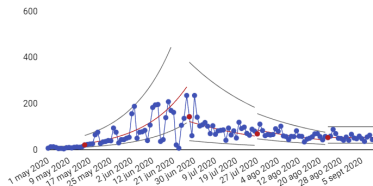


Chile is in Epoch 3: Plateau or Descent in Daily Reported Cases

A_Z | :

Chile : Adjusted Daily Covid-19 Reported Deaths

A_Z | :



Chile is in Epoch 3: Plateau or Descent in Daily Reported Deaths

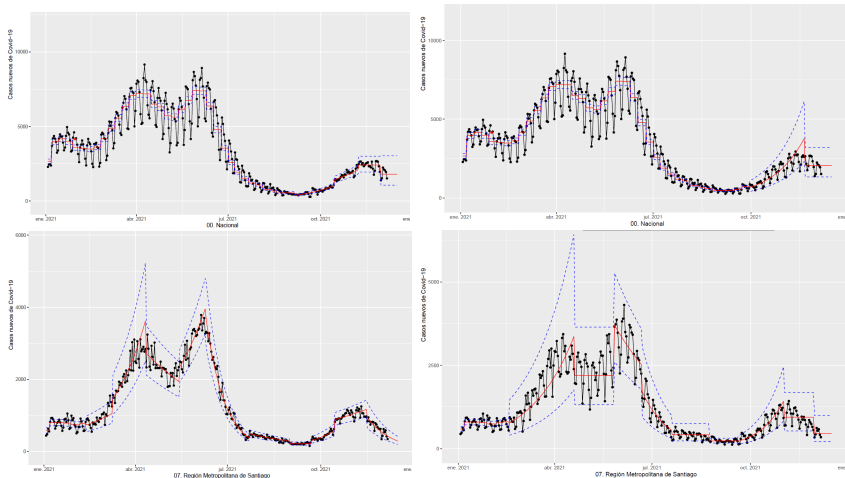
Fuente: <https://isqua.org/covid19-research-page.html>

- La aplicación se ajustó para un efecto de día de la semana.
- La serie de datos ajustados se utiliza como entrada al algoritmo de épocas y fases.

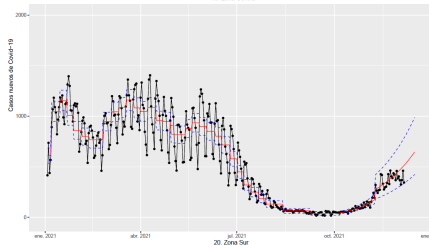
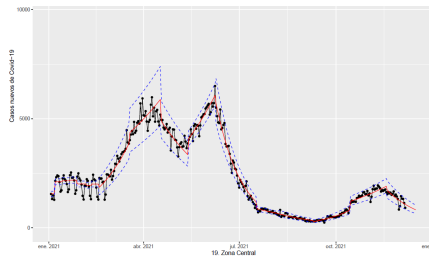
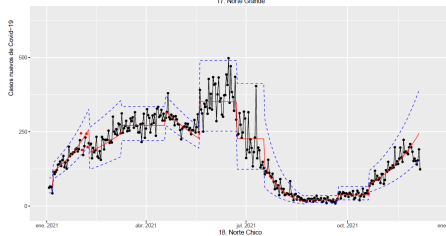
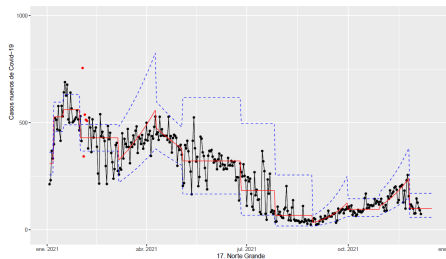
Epoca	Cálculo
2	$p < 0,05$ pendiente positiva: crecimiento exponencial
3	$p < 0,05$ pendiente negativa: decaimiento exponencial
3	$p \geq 0,05$: ni crecimiento ni decaimiento

Ejemplos de Resultados desde el 20-11-2020

Figura: Casos reportados Nacional y Región Metropolitana

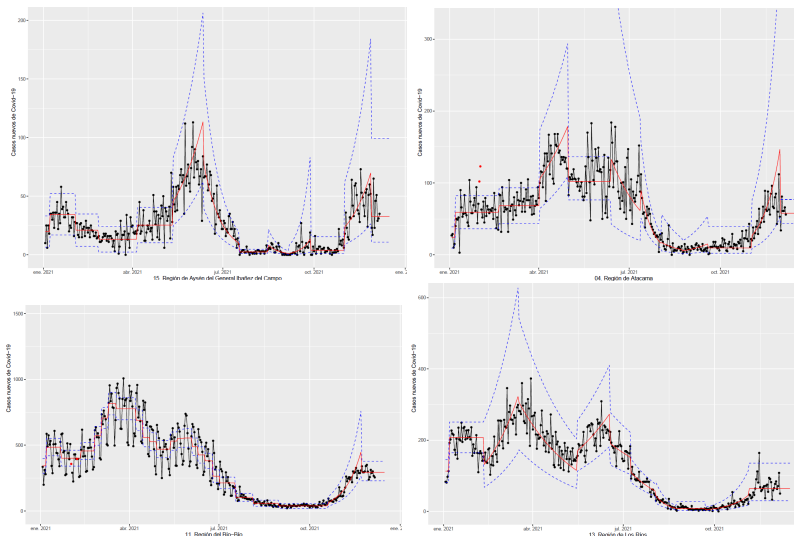


Casos nuevos para las zonas del país



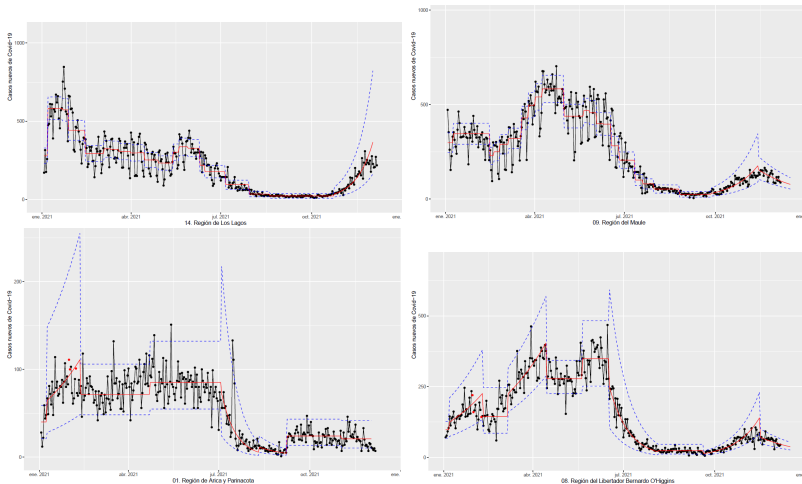
Casos nuevos en regiones del país

Figura: Ejemplo del del método para las regiones de Chile



Casos para las regiones

Figura: Ejemplo del método para las regiones de Chile



Resultados

Resumen de de las fases y epocas para las zonas y regiones del país para el día 6 de diciembre de 2021

state	Casos	phase	epoch	midline	lcl	ucl
Nacional	1517	38	3	1798.88	1066.53	3034
Norte Grande	74	13	3	99.35	57.87	171
Norte Chico	124/167	10	2	243.2	119.8	493.7
Zona Central	914	17	3	1077	854	1359
Zona Sur	344	34	2	513.53	336	785
Zona Austral	53	13	3	50.86	20	127
R Sin Metro	1174	37	3	1341	799	2249

Resumen

state	Casos	phase	epoch	midline	lcl	ucl
Arica	7/10/8/12	9	3	22.49	7.43	68.1
Tarapacá	22	15	1	24.40	9.58	39.22
Antofagasta	45	16	1	67.36	42.74	91.98
Atacama	56	12	3	57.95	43.59	77.04
Coquimbo	81/101	9	3	103.7	77	139.7
Valparaíso	165	13	3	212.82	116.70	388.14
Metropo	343	15	3	424.90	293.55	615.03
O'Higgins	45	10	3	46.60	32.68	66.46
Maule	87	24	3	95.93	65.53	140.43
Ñuble	59	12	3	65.22	47.62	89.32
Bío-Bío	256	25	3	294.44	229.41	377.90
Araucanía	76	13	3	93.61	64.37	136.14
Los Ríos	50	11	3	63.66	29.87	135.65
Los Lagos	218/292	20	2	445.96	166	1198
Aysén	35	13	3	32.73	10.81	99.14
Magallanes	13	13	3	10.46	2.96	36.89

Automatic Selection of Models Outlier DEtection for Epidemics (ASMODEE)

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 R.

Está diseñado para detectar desviaciones, aberraciones recientes de tendencias temporales pasadas, la variable respuesta representa el recuento de casos, pero también se puede utilizar en tasas de incidencia, mortalidad, entre otras. Interesa clasificar valores atípicos en los puntos temporales más recientes.

Implementa un marco de series de tiempo flexible utilizando una variedad de modelos que incluyen regresión lineal, modelos lineales generalizados(GLM).

K-Doble

Los datos se dividen aleatoriamente en k particiones de tamaños aproximadamente iguales (plieges). Cada pliege se utiliza una vez para una ronda de validación cruzada: el $k-1$ se utiliza como entrenamiento, para adaptarse al modelo, mientras que los restantes se utilizan como conjunto de testeo, para evaluar la calidad de las predicciones. y los datos de prueba se miden utilizando el RMSE. Computacionalmente intensivo. **evaluate resampling**

AIC

Selección de modelos utilizando AIC es un estándar para comparar la bondad de ajuste de modelos teniendo en cuenta su respectiva complejidad.

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

Donde L es la verosimilitud. L es la probabilidad del modelo es la desviación del modelo. Mientras más pequeño sea el AIC, mejor será la bondad de ajuste de modelo. Este procedimiento es rápido pero no

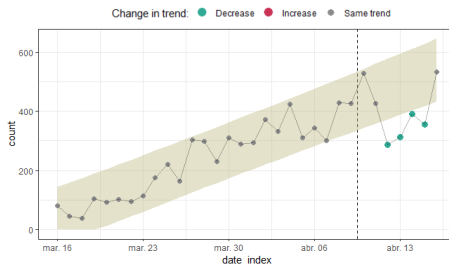
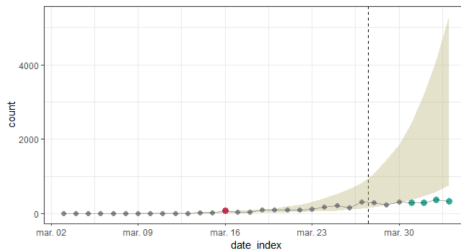
ASMODEE se utiliza:

- Conjunto de entrenamiento excluyendo los últimos k puntos de datos. **calibración y predicciones**
- Entrena modelos especificados por el usuario y conserva el de mejor ajuste o predicción.
- Intervalos de predicción $\alpha = 5\%$
- Clasifica los puntos en: Increase, Normal, Decrease.

```
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)  
)  
res <- asmodee(dat, models, method = evaluate_aic, alpha=0.05, k = 4)
```

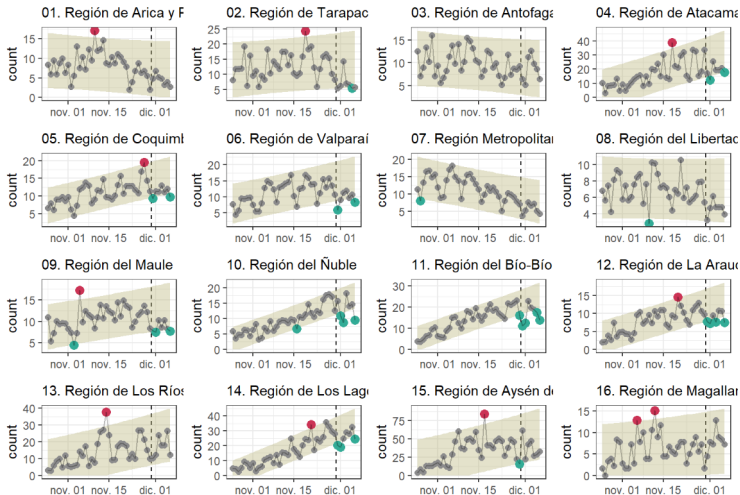
Resultados primera ola del 3 marzo al 3 de abril del 2020

Con una ventana de 4 semanas y un $k=7$ y tomando de referencia el 16 de marzo para el segundo gráfico.



Resultados de tasa de incidencia en regiones

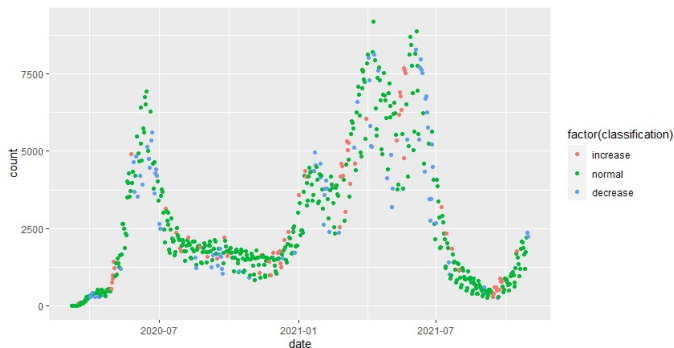
Tomando de las últimas 6 semanas



Change in trend: ● Decrease ● Increase ● Same trend

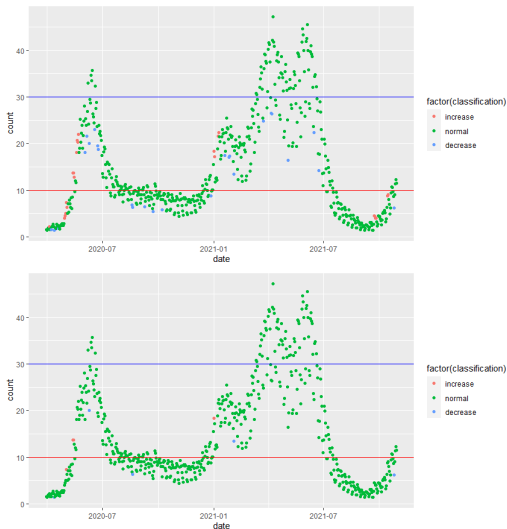
Resultado casos nuevos a nivel Nacional.

Para ver todos los cambios producidos en el tiempo, se guardan todos los puntos detectados en la predicción



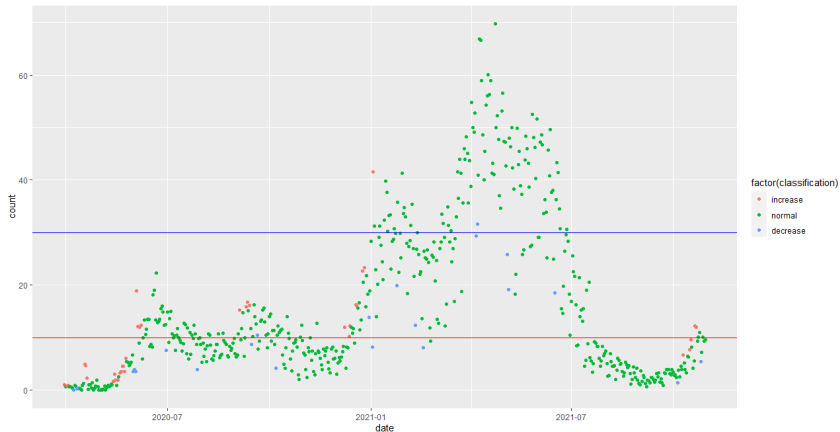
Evaluación con ventanas de 4 y 5 semanas en Incidencia

Con una ventana de 4 semanas se obtienen increase consecutivos el 2020-04-29

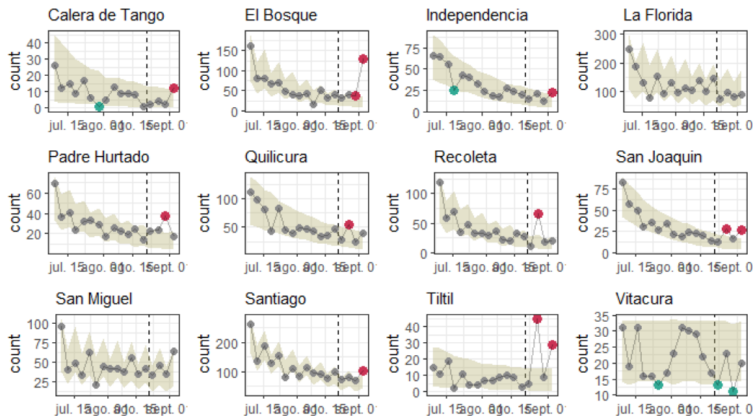


Region del Maule

Con una ventana de 4 semanas, primera ola 13-05-2020
segunda ola 8-12-2020
tercera ola 10-10-2021



Región Metropolitana



Change in trend: ● Decrease ● Increase ● Same trend

Implementación con librería shiny

Aplicación con shiny

Para tener una mejor visualización se creo una aplicación con shiny en donde se pueden manipular fechas y lugares.

Change - Point Regional 07 Diciembre

Selecccionar Region

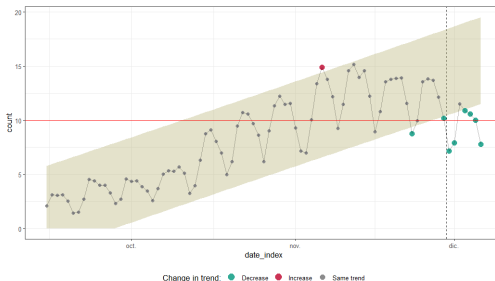
Nacional

Rango Fechas:

2020-03-03 2021-09-15 2021-12-06

2020-03-03 2020-07-11 2020-11-18 2021-03-28 2021-08-05 2021-12-06

Descargar el grafico Descargar datos



Los métodos aplicados resultan ser distintos en algunos aspectos:

- Método 1: Realiza un ajuste considerando los días de la semana. complejo de realizar en las comunas y con datos muy pequeños, necesita de almenos 21 días para entregar resultados coherentes
- Método 2: flexible con los datos entregados
Simple y flexible
Entregó la alerta antes que el otro método y se adapta a los datos por comunas
No realiza ajuste por fines de semana pero se puede adaptar con los modelos que se ocupen.

Componentes Principales

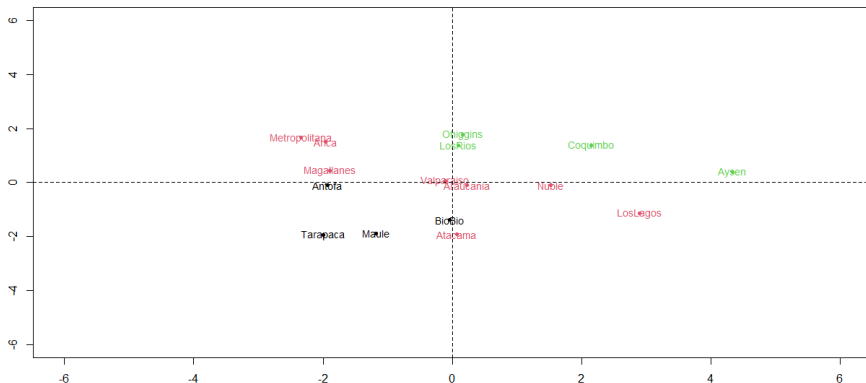
El análisis de componentes principales tiene como objetivo: dadas n observaciones de p variables, se analiza si es posible representar adecuadamente esta información con un número menor de variables construidas como combinaciones lineales de las originales.

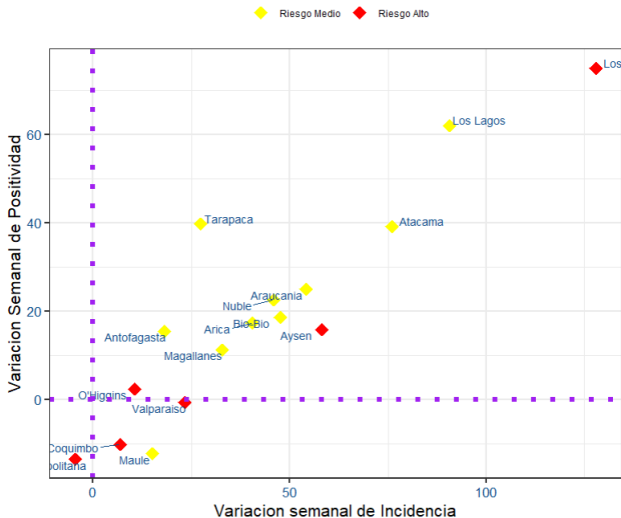
Tasa de Incidencia, R efectivo y Tasa de positividad más variaciones semanales.

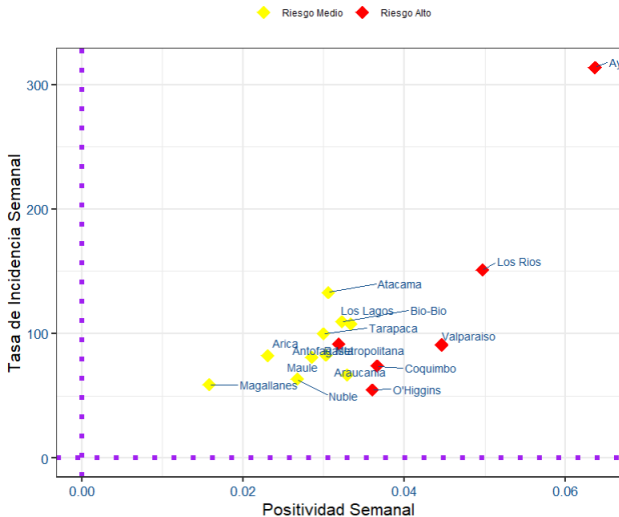
Criterios mínimos que exige la OMS, se considera un R efectivo inferior a uno, presente una incidencia menor a 10 casos/100.000 hab. y una positividad menor al 5.

- Riesgo bajo: Baja cantidad de casos y estable.
- Riesgo medio: Posible crecimiento.
- De Riesgo alto: Tendencia en crecimiento

Gráficos utilizando respuestas de Change-Point





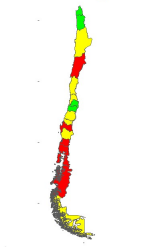


Clusters de alerta

Se muestra el resumen de indicadores con la corrección de los riesgos asociados a cada región del país, publicado por el Gemvop del día miércoles 6 de Diciembre.

Balance Regional Covid-19 06 Diciembre

Region	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	Cluster
Arica	0.6 % (-0.2)	0.74 (0.05)	2.78 (-1.19)	Riesgo Bajo
Tarapaca	0.5 % (-0.4)	0.58 (0)	5.75 (0.26)	Riesgo Bajo
Antofagasta	1.6 % (-0.4)	1.03 (-0.04)	6.5 (-2.17)	Riesgo Medio
Atacama	2.4 % (-1.4)	0.82 (-0.09)	17.79 (-3.18)	Riesgo Alto
Coquimbo	2.9 % (-1.3)	0.84 (-0.02)	9.69 (-2.39)	Riesgo Alto
Valparaíso	4 % (0.6)	0.77 (-0.01)	8.42 (-2.35)	Riesgo Alto
Metropolitana	1.4 % (-0.1)	0.68 (-0.02)	4.22 (-0.9)	Riesgo Bajo
Onigginis	2.3 % (0.1)	0.73 (-0.02)	3.94 (-0.91)	Riesgo Medio
Maule	1.8 % (0)	0.78 (0.01)	7.69 (-0.62)	Riesgo Medio
Nuble	2.8 % (-0.7)	0.92 (-0.05)	9.58 (-5.08)	Riesgo Alto
Bio-Bio	4 % (0.9)	0.9 (-0.02)	13.88 (-3.61)	Riesgo Medio
Araucania	2.3 % (-0.9)	0.86 (-0.03)	7.49 (-3.15)	Riesgo Bajo
Los Rios	3.8 % (-2.1)	1.01 (0.02)	12.32 (-14.29)	Riesgo Medio
Los Lagos	5.1 % (0.4)	0.96 (-0.03)	24.45 (-8.3)	Riesgo Alto
Aysen	7.3 % (3.1)	0.96 (0.08)	32.62 (3.73)	Riesgo Alto
Magallanes	2.4 % (0.9)	1.33 (0.13)	7.29 (-1.12)	Riesgo Medio









Region	Tasa.de.Positividad	R.Efectivo	Tasa.Incidencia	Cluster
Arica	0.6 % (-0.2)	0.74 (0.05)	2.78 (-1.19)	Riesgo Bajo
Tarapaca	0.5 % (-0.4)	0.58 (0)	5.75 (0.26)	Riesgo Bajo
Antofagasta	1.6 % (-0.4)	1.03 (-0.04)	6.5 (-2.17)	Riesgo Medio
Atacama	2.4 % (-1.4)	0.82 (-0.09)	17.79 (-3.18)	Riesgo Alto
Coquimbo	2.9 % (-1.3)	0.84 (-0.02)	9.69 (-2.39)	Riesgo Medio
Valparaíso	4 % (0.6)	0.77 (-0.01)	8.42 (-2.35)	Riesgo Medio
Metropolitana	1.4 % (-0.1)	0.68 (-0.02)	4.22 (-0.9)	Riesgo Bajo
Onigginis	2.3 % (0.1)	0.73 (-0.02)	3.94 (-0.91)	Riesgo Bajo
Maule	1.8 % (0)	0.78 (0.01)	7.69 (-0.62)	Riesgo Medio
Nuble	2.8 % (-0.7)	0.92 (-0.05)	9.58 (-5.08)	Riesgo Medio
Bio-Bio	4 % (0.9)	0.9 (-0.02)	13.88 (-3.61)	Riesgo Alto
Araucania	2.3 % (-0.9)	0.86 (-0.03)	7.49 (-3.15)	Riesgo Medio
Los Rios	3.8 % (-2.1)	1.01 (0.02)	12.32 (-14.29)	Riesgo Alto
Los Lagos	5.1 % (0.4)	0.96 (-0.03)	24.45 (-8.3)	Riesgo Alto
Aysen	7.3 % (3.1)	0.96 (0.08)	32.62 (3.73)	Riesgo Alto
Magallanes	2.4 % (0.9)	1.33 (0.13)	7.29 (-1.12)	Riesgo Medio

¿Es efectivo el análisis?

- Los métodos de Change-Point son aplicables en distintos escenarios.
- Resultan efectivos para la detección de brotes localizados.
- Son importantes en la toma de decisiones.
- Se logra el objetivo de diseñar un sistema de alerta que considere todas las aristas del problema.
- Se logra dar una alerta a las regiones primero para luego encontrar las comunas afectadas.
- La clasificación entregada con clusters utilizando Change-Point no resulta la más efectiva
- Retraso en la notificación de informes y efectos de fin de semana se deben tener en consideración.

Bibliografía

-  Fermin, G. (2018). Host range, host–virus interactions, and virus transmission. *Viruses*, 101.
-  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.
-  OMS (2010) : Organización Mundial de la Salud. ¿Qué es una pandemia?
https://www.who.int/csr/disease/swineflu/frequently_asked_questions/pandemic/es/
-  Hays, J. N. (2005). Epidemics and pandemics: their impacts on human history. *Abc-clio*.
-  Yan, P., & Chowell, G. (2019). Quantitative methods for investigating infectious disease outbreaks (Vol. 70). Cham, Switzerland: Springer.
-  Serra Valdés, M. Á. (2020). Infección respiratoria aguda por COVID-19: una emergencia evidente. *Revista Ubense de Ciencias*