

Datos Covid-19: Caso Chileno

emilianomm

09-11-2020

En el mes de Marzo, año 2020, se registro en Chile el primer caso oficialmente confirmado de Covid-19, una nueva enfermedad viral, relacionada principalmente con problemas respiratorios. El brote inicial fue en China, pocos meses antes de llegar a Chile, cuando ya en otros países como España e Italia, considerados de mayor desarrollo, la expansión del virus era causa de gran mortalidad. Sin tener claridad aún de la envergadura ni los tratamientos adecuados (**fuentes**) ya era clara una cosa, la propagación de la enfermedad era rápida, sugiriendo una alta capacidad de contagio y el grupo de mayor riesgo es la tercera edad (**fuentes**), esta última hipótesis influenciada principalmente por los primeros datos de los países mencionados. Así los gobiernos del mundo se aprestaron a imponer medidas de confinamiento entre otras coerciones a la libertad, como principal medida contra la expansión del contagio.

Algunas de las medidas tomadas no tienen precedentes en la historia contemporánea. También la cantidad de información *almacenada* desde el inicio de la situación hasta la fecha de elaboración del presente informe ha sido sustantiva dadas las capacidades tecnológicas actuales. De lo anterior surge la necesidad de, primero, convertir la información *almacenada* en conocimiento útil para la toma de decisiones y futuros eventos, así como evaluar el *impacto* de aquellas medidas en el manejo de la enfermedad, y también en otros aspectos de la vida humana.

En el caso de Chile el gobierno ha dispuesto una base de datos a través de la cuenta GitHub del Ministerio de Ciencia, en ella se encuentran distintos *productos de datos*, cada uno de los cuales consiste en un archivo en formato *.csv* con datos relacionados a la realidad Chilena en el contexto del Covid-19. Con estos datos pretendemos poner a prueba las principales ideas de la opinión pública acerca de la enfermedad, es decir si los datos suministrados desde la fuente mencionada permiten hacer dichas inferencias. También buscamos y proponemos nuevas alternativas para caracterizar y evaluar el *desempeño* de las distintas regiones del país

Para nuestros propósitos, primero realizamos un análisis descriptivo de las bases de datos que consideraremos (el repositorio mencionado es vasto y aquí se deben concentrar los esfuerzos). Luego, sobre los *productos de datos* escogidos aplicamos la metodología de la *regresión lineal* junto a una validación por *prueba de hipótesis* para realizar inferencia. Para proponer una evaluación del *desempeño*, se sigue la metodología de la *transformación*, usando nuevas variables creadas a partir de combinaciones de las variables existentes. Finalmente, con lo anterior se concluye si es posible apoyar o rechazar ciertas hipótesis de la opinión pública así como entregar nuestra propuesta de caracterización de un grupo de riesgo basado en variables innovadoras.

Resultados

Tengamos en cuenta que consideramos la base de datos actualizada hasta el *16-11-2020*

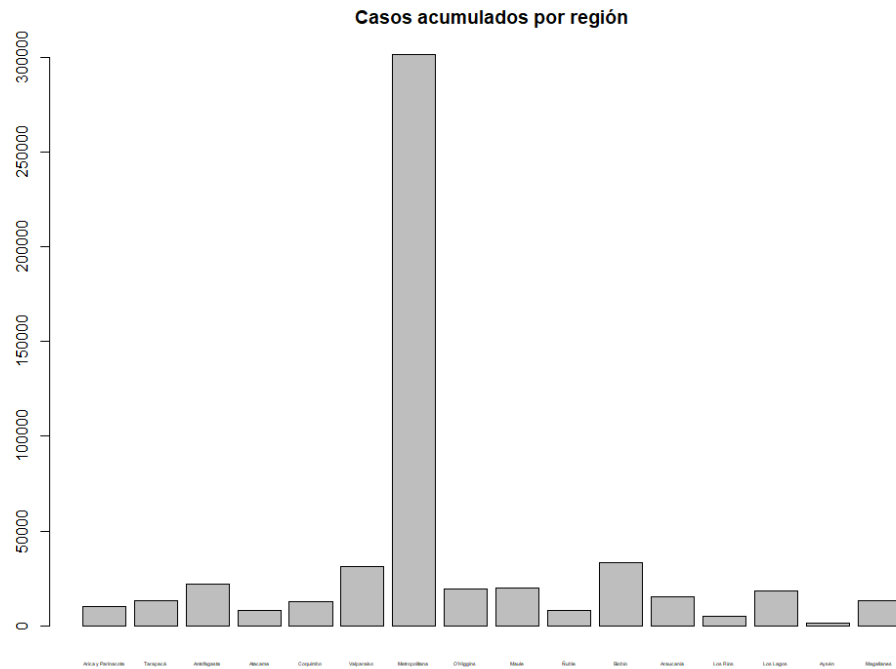
Resumen descriptivo

Análisis descriptivo de las bases de datos consideradas posteriormente

Una mirada al comportamiento de las regiones

Consideramos los productos de datos **03**

Una manera natural de evaluar como se han *comportado* las regiones en relación al control de la enfermedad, podría ser la cantidad de casos totales. Para cada región tenemos la cantidad de casos acumulados a la fecha de corte del estudio, por ello usamos un gráfico de barra, el cual visualizamos a continuación



Como vemos, la región Metropolitana es la que mas casos presenta, sin embargo sabemos que es la región mas poblada. Podemos considerar otras métricas para corregir ese hecho.

Una alternativa para evaluar el comportamiento de las regiones

Consideramos los productos de datos **03**, **07** y **13**

Aca proponemos una *transformación*

Regresión lineal sobre casos por grupo etario

Consideramos los productos de datos **9** y **10**