

Estrujemos los datos para vencer el COVID-19

Usando Data Science para aprender sobre qué palancas usar en la batalla contra la pandemia

Hace poco un artículo de la BBC resaltaba que Chile estaba manejando muy bien la crisis del Covid-19, tanto en relación a otros países de la región, como cuando se comparaba con Corea del Sur o Alemania, países reconocidos internacionalmente por sus estrategias para combatir la pandemia.

La noticia reciente del alza en los contagios en Chile, y la decisión de la autoridad de decretar cuarentena para varios millones de habitantes, levantan dudas sobre si realmente lo estamos haciendo tan bien, más aún cuando los medios de comunicación informan constantemente sobre personas imprudentes que se ponen en riesgo a sí mismos y al resto por razones pueriles.

Los avances tecnológicos y el acceso casi universal a la información, hacen que estemos en una situación privilegiada para abordar esta interrogante. A continuación, describo brevemente los resultados del análisis que realicé utilizando las estadísticas públicas del Covid-19¹, más antecedentes disponibles en el Banco Mundial² y Transparencia Internacional, para un total de 170 países. El análisis incorpora variables demográficas, geográficas, económicas y conductuales. El objetivo era intentar entender mejor qué aspectos influyen a nivel de país, tanto en la tasa de contagios como la tasa de muertes. Utilicé técnicas de Machine Learning para identificar las variables relevantes, y sistemas de ecuaciones estructurales para estimar los efectos en las variables de interés. En los cuadros de resultados sólo muestro variables explicativas³ que son estadísticamente significativas, es decir aquellas cuyo impacto es distinto de cero.

Partamos con la tasa de contagios por cada millón de habitantes. Es mucho mayor para los países situados en Europa. El solo hecho de estar allí multiplica la tasa de contagios base por 4. Por ejemplo, en el caso de Nigeria, cuya tasa de contagios es de 24, si estuviese físicamente en Europa, su tasa sería de 97.

El nivel de desempleo también tiene un efecto positivo. Por cada 10% que sube, los contagios aumentan en 3.8%. En muchos países los desempleados se dedican al comercio ambulante, lo que aumenta los puntos de contagio posibles.

Es interesante que tanto la proporción de mujeres como de niños de un país, reducen la tasa de contagios, lo que levanta una pregunta interesante. ¿Se debe esto a motivos biológicos, o a patrones de conducta? Por ejemplo, es bien sabido que el sistema inmune de las mujeres es más potente que el de los hombres, y quizás simplemente se enferman menos. Pero además las mujeres tienden a ser más prudente que los hombres, y esa prudencia la ejercen también sobre sus hijos. Si esta es la razón detrás de este efecto, entonces tenemos que aprender de las mujeres y animarlas a que ejerzan con mucha más convicción y protagonismo ese rol protector benéfico.

También tenemos que, tanto el porcentaje de gente que vive en ciudades, como la ausencia de una cultura de prevención aumentan la tasa de contagios. Lo primero es consistente con que el virus se transmite por proximidad y que las grandes ciudades son “zonas calientes”. Y lo segundo es de sentido común. Menos prevención implica más contagios, y le da la razón a la autoridad sanitaria sobre la relevancia de tomar todas las medidas que estén a nuestro alcance.

Tasa de Contagios	Efecto	Comentario
Europa	4,04	Veces x estar en Europa
Desempleo	3,8%	Por cada 10 % de variación.
% Población femenina	-12,8%	Por cada punto adicional
% Población bajo 15 años	-9,7%	Por cada punto adicional
% Población Urbana	2,8%	Por cada punto adicional
Ausencia cultura de prevención	0,9%	Por cada punto adicional

Siguiendo con los tests por cada millón de habitantes, los análisis indican que son una reacción a la tasa de contagios, ya sea por temor en la población, o por instructivo de la autoridad. Además, la tasa de tests tiende a disminuir en los países con mayor población, pero aumenta en la medida que lo hace el producto per cápita. Por cada 10% que sube el producto, la tasa de test aumenta en más de un 4%, pero cae en más de un 2% por cada 10% que aumenta el nivel de desempleo.

La cantidad de tests también disminuye cuando la cultura de prevención en la población está menos desarrollada, y lo mismo sucede cuando aumenta el nivel de corrupción⁴. En el caso de Alemania, cuyo índice de corrupción es 20, si el nivel de corrupción fuese 21, la tasa de tests caería en un 1.2%. Es algo bien sabido que la corrupción reduce la eficiencia y efectividad

¹ Datos de <https://www.worldometers.info/coronavirus/>, con fecha 13 de mayo de 2020.

² Utilicé los datos del año 2017 disponibles en el Banco Mundial. Para los años posteriores hay menos información y eso reduce el tamaño de la muestra.

³ Los resultados deben interpretarse bajo el criterio “ceteris paribus”, es decir, es el efecto de la variable manteniendo todo el resto constante.

⁴ El índice de Transparencia Organizacional va de 0 a 100, con los valores más altos indicando una percepción más baja de la corrupción, es decir integridad. Usé el valor inverso para medir corrupción. En el caso de Chile, cuyo índice de integridad es 67, el índice de corrupción sería 33.

Estrujemos los datos para vencer el COVID-19

Usando Data Science para aprender sobre qué palancas usar en la batalla contra la pandemia

de los países, y por lo tanto tiene consecuencias económicas. Estos resultados indican que también tiene consecuencias en la salud pública, reduciendo la capacidad de los países de identificar el Covid-19 en forma oportuna.

Tasa de Test	Efecto	Comentario
Tasa de contagios	3,7%	Por cada 10 % de variación.
Total Población	-1,5%	Por cada 10 % de variación.
PGB per cápita	4,2%	Por cada 10 % de variación.
Desempleo	-2,1%	Por cada 10 % de variación.
Ausencia cultura de prevención	-1,0%	Por cada punto adicional
Corrupción	-1,2%	Por cada punto adicional

Llegamos ahora a lo más relevante, la tasa de muertes por cada millón de habitantes. El sólo hecho de estar en Europa multiplica por 1.84 veces la tasa base. Por lo tanto, la evaluación auspiciosa de Chile realizada por la BBC podría ser inadecuada, pues no consideran el efecto de estar fuera del epicentro de la pandemia, que, aunque partió en China, ha azotado Europa con más fuerza. Los alemanes pueden ser muy ordenados y responsables, pero están expuestos a la insensatez e improvisación de sus vecinos, y por ese solo hecho el Covid-19 los está golpeando fuerte. Estando a miles de kilómetros de distancia, nuestro problema es menor de lo que podría ser estando allá.

Continuando, y como era de esperarse, una mayor cantidad de tests reduce la tasa de muertes, probablemente por la identificación temprana que corta la cadena de contagios, y que permite un tratamiento oportuno del afectado. Es un efecto importante: por cada 10% que aumenta la tasa de tests, la tasa de muertes cae en un 3%.

Por otro lado, un incremento en la tasa de contagios tiene un incremento casi a la par en la tasa de muertes y por ello hace todo el sentido del mundo la preocupación de la autoridad por reducir los contagios.

Si bien el desempleo no tiene un efecto directo en la tasa de muertes, sí tiene un efecto indirecto a través del incremento en la tasa de contagios, y la disminución en la tasa de tests. El modelo indica que, si el nivel de desempleo aumenta, por ejemplo, del 7% al 8%, la tasa de muertes crece en 5.4%. Por lo tanto, es falso que debamos escoger entre la Salud Pública y la Economía ya que ambas están íntimamente relacionadas. Con una economía por el suelo, los más vulnerables y afectados harán lo que sea necesario para que sus familias no pasen hambre, aunque esto implique exponerse al contagio y sus consecuencias.

Es interesante ver que la densidad poblacional reduce las muertes. Mi interpretación es que esto puede relacionarse con la mayor facilidad de acceso a servicios de salud en lugares más densos, aunque podría deberse a otro motivo. Este es un hallazgo que sería interesante explorar con mayor profundidad.

Siguen dos efectos que son coherentes con lo que sabemos. La gente mayor es población de riesgo, y por lo tanto la tasa de muertes aumenta en países con una mayor esperanza de vida, y una mayor proporción de la población sobre los 65 años.

No es tan claro por qué crece la tasa de muertes en países con más niños. Una hipótesis a explorar, es el potencial rol de los niños como vectores del virus, al menos en las etapas primeras de la pandemia. Es bien sabido que las guarderías infantiles son un caldo de cultivo de resfríos durante el invierno, y bien podría haber sucedido algo similar con el Covid-19⁵.

Es interesante la caída en la tasa de muertes en países con mayor tasa de natalidad, lo que habla de poblaciones donde hay una mayor porción de madres y bebés. Nuevamente, ¿será un efecto conductual o biológico? Perfectamente pueden ser ambos. Desde el punto de vista conductual, las madres con bebés tienden a ser precavidas y prudentes, generando un ambiente propicio para las medidas preventivas, y una rápida reacción ante señales preocupantes.

Finalmente, aquellos países en que la población tiene un mayor acceso a las tecnologías de comunicación, tienen una menor tasa de muertes. Este efecto podría relacionarse con la mayor disponibilidad de información oportuna en dichos países, que ayudaría a que se tomen las medidas adecuadas a tiempo, pero puede también deberse a otros motivos, aún por identificar

Tasa de Muertes	Efecto	Comentario
Europa	1,82	Veces x estar en Europa
Tasa de Tests	-3,0%	Por cada 10 % de variación.
Tasa de contagios	9,4%	Por cada 10 % de variación.
Densidad poblacional	-1,0%	Por cada 10 % de variación.
% Población sobre 65 años	10,7%	Por cada punto adicional
Experanza de vida	3,3%	Por cada punto adicional
% Población bajo 15 años	12,4%	Por cada punto adicional
Tasa de natalidad	-9,5%	Por cada punto adicional
Masificación tecn. comunic.	-0,4%	Por cada punto adicional

⁵ En Europa los abuelos tienen un rol relevante en el cuidado de sus nietos, es decir habría una conexión directa entre vectores y población de riesgo.

Estrujemos los datos para vencer el COVID-19

Usando Data Science para aprender sobre qué palancas usar en la batalla contra la pandemia

Pasemos ahora a ver cómo estamos respecto a Alemania y Corea, con los que ya nos comparó la BBC, y también respecto a Polonia, que tiene un tamaño más o menos similar al nuestro. El cuadro muestra el ratio entre los datos reales y los que proyecta el modelo, para cada caso⁶. Mientras más cercano a “1”, más exacta es la cuadratura entre el dato real y el del modelo. Las diferencias entre el modelo y la realidad, las explican factores adicionales no incorporados, pues se trata de un modelo simple. En el caso alemán, la cuadratura es casi perfecta.

En el caso chileno, tenemos una tasa mayor de contagios, pero una tasa menor de muertes. Es decir, hasta ahora efectivamente lo estamos haciendo mejor en términos de muertes de lo que se podría esperar dadas nuestras circunstancias, entre ellas nuestra ubicación geográfica, PGB per cápita, y el tamaño de población, por nombrar algunas. Probablemente, la identificación oportuna de casos más un sistema de salud que aún no colapsa, tienen algo que ver en esta explicación. Y también podría deberse a la entrega y buen trabajo de los profesionales de la salud, y una buena gestión de la autoridad sanitaria. La gran duda es cómo podría empeorar la tasa de muertes si la capacidad hospitalaria colapsa.

Por otro lado, tenemos una tasa mayor de contagios, y me atrevo a plantear que la aparente poca consciencia de parte de la población explica eso, al menos en parte. Además, está la forma de contar los contagios. El efecto de incorporar a los asintomáticos también podría explicar algo de esa diferencia, pero sólo una parte por ser un ajuste reciente que afecta poco a los datos acumulados.

Nótese que, si Chile estuviese físicamente en Europa, la situación sería mucho peor. Crece la tasa de contagios, y se dispara la de muertes. Por lo tanto, cualquier comparación alegre con lo que está sucediendo en Europa, es simple y llanamente ceguera o soberbia, ambas muy dañinas para salir bien parados de esta pandemia.

Pais	Contagios	Tests	Muertes
Alemania	1,02	0,98	1,02
Chile	1,15	0,97	0,73
Polonia	0,88	0,97	0,98
Corea del S.	0,86	1,01	1,06
Chile EUR (*)	1,20	1,00	1,48

(*) Los indicadores equivalen a dividir la tasa logarítmica real por la estimada por el modelo, salvo en el caso Chile EUR, en que se toma la tasa real corregida, es decir la que daría si Chile estuviese en Europa, y se divide por la tasa real.

A partir de estos datos, podemos concluir que en Chile lo estamos haciendo razonablemente bien, pero que aún lo podríamos hacer mucho mejor. Primero, podríamos fomentar con mayor énfasis la detección temprana usando tests. Además, cambiar el tono de la comunicación a la población, siendo mucho más claros sobre lo que está en juego. La idea no es generar histeria colectiva, sino tender a una comunicación más cruda y directa para abordar la dimensión conductual. También, que empecemos a usar más los números en la comunicación, pues son herramientas sumamente efectivas para cambiar patrones de comportamiento. No es lo mismo decir “*Cuidémonos todos*” y que la gente lo interprete como se le venga en gana, que decir con claridad “*Cada vez que los contagios suben un 10%, los muertos suben un 9.4%. Si no te cuidas, puedes matarte a ti mismo, y a otros*”. Y en la misma línea, ser más estrictos con las personas que no respetan las normas. Que no quede espacio a dudas.

Otro ámbito de acción tiene que ver con el rol muy potente que tienen las madres. Sin su prudencia, precaución y cuidados, muchos no estaríamos aquí. Los datos sugieren que su influencia positiva está reduciendo los contagios y muertes por Covid-19 en el mundo. Aprender de ellas, y darles más protagonismo, puede salvar miles de vidas.

Para cerrar, creo relevante levantar dos puntos. Primero, que usualmente le sacamos poco provecho a los datos, y que, en situaciones de contingencia, cuando hay mucho en juego, es más importante que nunca estrujarlos para aprender más rápido y tomar decisiones en forma más informada. Muchos datos están disponibles a un clic de distancia, y existen herramientas freeware para analizarlos. A pesar de esto, se estima que aprovechamos a menos del 1% de la información que generamos. Lo que falta es que nos demos el tiempo para hacerlo, y que consigamos el know-how, si no lo tenemos en casa.

Segundo que, aunque “*todos los modelos tienen fallas, algunos son útiles*”⁷. Esta es la máxima de quienes intentamos hacer investigación. El modelo que utilicé es sencillo, una simplificación con limitaciones. Además, refleja la situación a la fecha, y tratándose de un fenómeno dinámico, podría darse que el efecto de algunas variables cambie con el tiempo, por ejemplo, el beneficio de estar fuera de Europa podría disminuir. A pesar de eso, el modelo da pistas de qué factores están influyendo en los contagios y muertes por Covid-19. Y más importante, nos dice cuánto influye cada variable. Si bien sería irresponsable

⁶ Las tasas están en escala logarítmica.

⁷ La frase original de George Box, es “All models are wrong but some are useful”.

Estrujemos los datos para vencer el COVID-19

Usando Data Science para aprender sobre qué palancas usar en la batalla contra la pandemia

asumir estos resultados como definitivos, son un punto de partida para explorar con más profundidad los temas que plantea. Está en juego la vida de miles de compatriotas, y el bienestar de millones. Bien vale la pena hacer el esfuerzo.

Anexo1 – Variables explicativas de la Tasa de Muertes

La tabla de abajo muestra parte de los resultados del sistema de ecuaciones estructurales, aquellos que estiman los efectos de diversas variables explicativas en la tasa de muertes cada millón de habitantes. Todas las tasas (Muertes, Tests y Contagios) fueron transformadas logarítmicamente, una práctica frecuente que se aplica a distribuciones asimétricas.

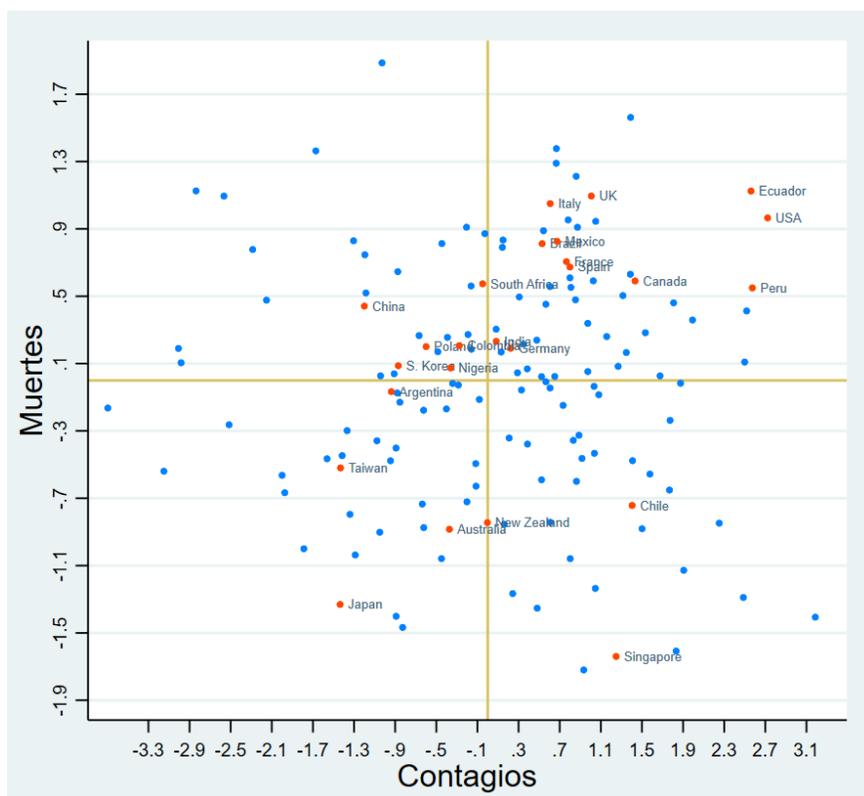
Para recuperar la escala, deben exponenciarse las variables y coeficientes, como se muestra en la ecuación de más abajo.

Variable	Coef	Err	Z	Pval	IC 95%	
Tasa Tests (LN)	-0,320	0,06	-4,9	0,00	-0,447	-0,193
Tasa Contagios (LN)	0,945	0,05	17,3	0,00	0,838	1,053
Europa	0,599	0,21	2,9	0,00	0,187	1,012
% Población sobre 65 años	0,102	0,02	5,1	0,00	0,063	0,141
% Población bajo 15 años	0,117	0,04	3,0	0,00	0,041	0,193
Experanza de vida	0,032	0,02	1,8	0,08	-0,004	0,068
Masificación tecn. comunic.	-0,004	0,00	-1,7	0,09	-0,009	0,001
Tasa de natalidad	-0,099	0,04	-2,7	0,01	-0,172	-0,027
Densidad poblacional (LN)	-0,110	0,05	-2,4	0,02	-0,201	-0,019
Constante	-4,248	1,65	-2,6	0,01	-7,475	-1,022

$$\text{Tasa Muertes} = T. \text{Tests}^{-0.303} * T. \text{Contagios}^{0.945} * e^{0.599 * \text{Europa} + 0.102 * \text{Pob65} + 0.117 * \text{Pob15} + 0.032 * \text{Esp.Vida} - 0.004 * \text{TecCom} - 0.99 * \text{Natal} - 0.110 * \text{Dens} - 4.254 + \varepsilon}$$

Anexo 2 – Desempeño relativo de los países

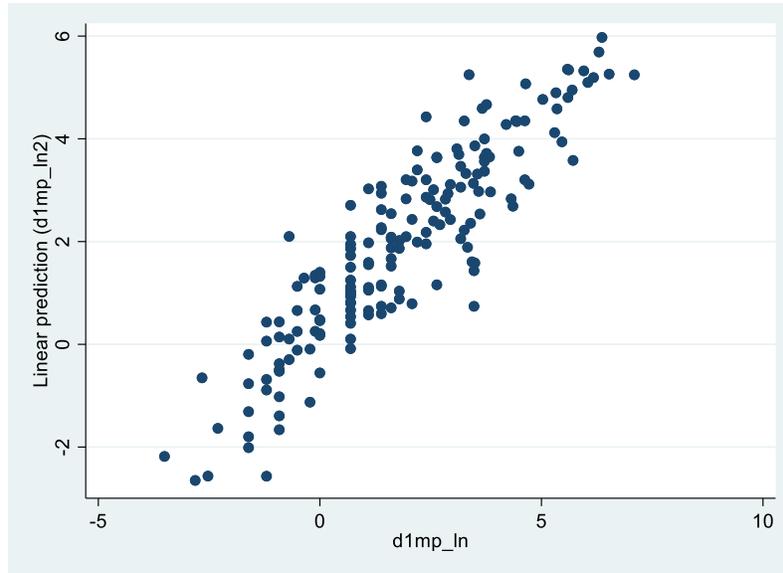
El gráfico muestra el desempeño relativo de los países en términos de Contagios y Muertes por Covid-19. Para cada país, compara el valor real de ambas variables, con el que podría esperarse dadas sus características geográficas, económicas, y culturales, tal cual lo estima el modelo. Si el índice en cualquiera de esas dimensiones es mayor que "0", entonces el valor efectivo de contagios y muertes es superior al esperable, y si es menor es inferior. Por ejemplo, Japón lo está haciendo mejor que lo esperable en ambas dimensiones, China lo hace muy bien en contagios, pero bastante peor de lo esperado en muertes, y Estados Unidos tiene resultados peores a los esperados en ambas dimensiones.



Estrujemos los datos para vencer el COVID-19
Usando Data Science para aprender sobre qué palancas usar en la batalla contra la pandemia

Anexo 3 – Gráfico que muestra tasa de muertes por país estimada por el modelo versus la real

Una relación perfecta sería una recta. El gráfico muestra una relación bastante lineal y esto, junto a que el R² es de 89%, indica que el modelo es razonablemente bueno.



Anexo 4 – Diagrama del Sistema de Ecuaciones Estructurales

Diagrama ilustrativo de las relaciones entre las variables, y sus efectos.

