

# Propuesta de implementación de factores demográficos para el análisis de COVID-19 en Honduras

Luis Felipe Flores

4 de junio de 2020

## 1. Construcción de una métrica de similitud entre países

Se pretende diseñar una métrica para categorizar sistemáticamente las similitudes entre países del mundo respecto a Honduras para hacer más rigurosa la comparación de su realidad con la nuestra. Esto se llevará a cabo en los siguientes tres aspectos:

1. **Similitud socioeconómica:** Se refiere a una medida de productividad, fertilidad y estabilidad económica general, así como sistema político y rubros sociales existentes. Como indicadores específicos y numéricos a considerar tenemos:
  - Índices de pobreza
  - Coeficiente de Gini
  - Indicadores de eficiencia/rendimiento del sistema sanitario
  - Diversos indicadores macroeconómicos
2. **Similitud Cultural:** Se refiere a una similitud en aspectos de costumbres, hábitos y creencias de los individuos del país que pueden tener efectos importantes en la efectividad de ciertas medidas de prevención de la esparcidad del virus.

Este se puede medir mediante cercanía geográfica, que usualmente implica interacción migratoria y turística entre países, así como similar trasfondo histórico, aunque estos indicadores podrían utilizarse de manera independiente en caso de no ser suficientemente homogéneos dentro de una zona. No obstante, en el caso de Honduras, centroamérica es suficientemente homogéneo en este aspecto para utilizar la cercanía geográfica como el indicador global<sup>1</sup>.

Otro indicador importante puede ser la estructura de gobierno y de leyes que rigen al país, así como aspectos de uso de transporte público, contaminación ambiental, etc. No obstante, el único indicador a utilizar (a menos que no se vea de ser suficiente) va a ser el de la cercanía geográfica.

Para este basta con dividir el mundo en regiones como las usuales: Centroamérica, suramérica, Noramérica, etc.

3. **Similitud en estrategias para enfrentar el COVID-19:** Esto se refiere a las políticas y decisiones que fueron tomadas por ley por el gobierno de cada país para hacer frente a la crisis. Es decir, la intensidad de la cuarentena (y la prontitud en que se aplicó), accesibilidad mercantil de productos para bioseguridad (mascarillas, gel, etc.).

Para países con comportamientos parecidos (y correspondientes similitudes culturales y socioeconómicas altas) se espera ver parámetros de modelo (especialmente robustos, como quizá  $R_0$ , el índice de esparcidad local del virus) similares y así poder realizar una prueba de hipótesis con una muestra que corresponda mejor a nuestra realidad.

---

<sup>1</sup>Revisar.

## 2. Tareas a realizar

Las tareas a realizar de manera inmediata entonces son, para cada los país:

1. Encontrar:
  - Índices de pobreza, de rendimiento del sistema sanitario, macroeconómicos y coeficiente de gini
  - Si es válida la asunción de simplificar los aspectos culturales en solo cercanía geográfica, y si no, qué países sí se considerarían culturalmente similares a Honduras, y cuales no tanto. En dicho caso, realizar una escala para pesos.
  - Número de pruebas.
2. Realizar una linea de tiempo para los siguientes eventos:
  - Primer caso de COVID-19 confirmado pero sin haber tomado medidas de cuarentena (que define una etapa de **crisis inicial sin cuarentena**)
  - Confinamiento débil
  - Confinamiento total (términos que deben definirse bien como parte de esta tarea)<sup>2</sup>
  - Reactivación posterior de la economía.
3. Adecuar el código para introducir que aplique el modelo SIR localmente. Es decir, un modelo SIR independiente para cada una de las etapas mencionadas anteriormente, que pueda ser puesto en un sólo gráfico.
4. Diseño de la prueba de hipótesis dada toda la información anterior. Esto incluye determinar cuál será el parámetro que servirá para comparar Honduras con los demás países, siendo una propuesta inicial el parámetro  $R_0$ .

Se entrará en mayor detalle de estas tareas en las siguientes secciones.

## 3. Indicadores demográficos para el modelo

Datos provenientes de la OMS, del banco mundial y de un informe de desarrollo sostenible de la ONU que se utilizarán en el modelo:

1. Densidad de población
2. Índice de pobreza
3. Ingreso per cápita
4. Población en área urbana
5. Edad promedio (índice índole biológico)
6. Índice de Gini
7. Conectividad (Pasajeros transportados)
8. Escolaridad años promedio,
9. Expectativa de vida al nacer

---

<sup>2</sup>Además, puede ser importante determinar qué empleos cada país considera como “empleos o funciones **esenciales**” y los mantiene activos como tal, aunque esto puede ser sobre-detallar información para los fines que se buscan. Revisar.

## 4. Línea de tiempo

las etapas son

- Pendiente
- Pendiente
- Pendiente
- Pendiente

## 5. Prueba de hipótesis

**ASUNCIÓN FUNDAMENTAL:** La evolución de la epidemia en Honduras puede ser estimada conociendo la evolución en otros países, difiando las cualidades de dicha evolución **solamente** aspectos que los indicadores anteriormente mencionados contemplan.

Sea  $\mathbb{P}$  el conjunto de todos los países a considerar (todos los que se utilizarán en el modelo de estimación) y sea  $\{\tilde{X}_i(t)\}_{i \in \mathbb{P}}$  el conjunto de series de tiempo que describen el número de casos en un intervalo  $[T_{\text{Inicial},i}, T_{\text{Final},i}]$  donde  $T_{\text{Inicial},i}$  es el día del primer caso detectado en el país  $i$ -ésimo.

Siendo más precisos, para el  $i$ -ésimo país:

$$\begin{aligned} \tilde{X}_i : [T_{\text{Inicial},i}, T_{\text{Final},i}] &\longrightarrow \mathbb{Z}^+ \\ t &\longmapsto \# \text{ de casos en el tiempo } t \end{aligned}$$

donde  $\mathbb{Z}^+ := \{1, 2, 3, \dots\}$ . Para poder comparar estas series de tiempo, podemos utilizar el día del primer caso,  $T_{\text{Inicial},i}$ , desplazarlas todas mediante la transformación  $X_i(t) := \tilde{X}_i(t - T_{\text{ref}})$  y definir un tiempo final común,  $T$ , justo para los datos que los tendremos al momento del análisis. Específicamente:

$$T = \min\{T_{\text{Final},i} - T_{\text{Inicial},i} \mid i \in \mathbb{P}\}$$

De esta manera, obtenemos un nuevo conjunto de series de tiempo,  $\{X_i(t)\}_{i \in \mathbb{P}}$ , bien definidas todas en  $[0, T]$ . La elección de tiempo de referencia  $T_{\text{Inicial},i}$  puede que no contemple el hecho que algunos países fueron menos eficientes en detectar su primer caso respecto al tiempo en que *realmente* ocurrió dicho primer contagio. Esto implica que puede necesitar contemplarse algún ajuste para dichos países en caso de poder obtener evidencia de qué países han sido más o menos eficientes en ese aspecto.

Además, para algunos países la fecha es difícil de determinar, como ejemplo tomamos China donde desde diciembre del año pasado hubieron reportes de múltiples casos de neumonía<sup>3</sup> sin haberse reportado oficialmente hasta el 22 de enero un total de 580 casos.

Para establecer realmente un día inicial  $T_{\text{Inicial},i}$  todos los países para los que no tengamos la data, se puede realizar un ajuste exponencial que nos permita hacer una regresión hacia el día en que *pudieron haber tenido* un solo caso. Definamos entonces a  $T_{\text{Inicial},i}$  como ese día estimado mediante regresión, de manera que ahora tenemos bien definida las series de tiempo  $\{X_i(t)\}_{i \in \mathbb{P}}$ .

La forma propuesta del estimador,  $\hat{X}_H(t)$ , para la serie de tiempo de Honduras (y para cualquier otro, no obstante, centrando por motivos claridad nuestra atención ya al caso de Honduras),  $X_H(t)$ , es la de una combinación convexa de las demás series de tiempo:

$$\hat{X}_H(t) = \frac{\sum_{i \in \mathbb{P}} \omega_i X_i(t)}{\sum_{i \in \mathbb{P}} \omega_i}$$

<sup>3</sup>Ver: <https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19>.

Los pesos  $\omega_i$  deben depender de parámetros encontrados en la sección anterior, enunciados a continuación por categoría:

- |   |   |                       |
|---|---|-----------------------|
| <ul style="list-style-type: none"> <li>■ Culturales/demográfico/geográficos               <ol style="list-style-type: none"> <li>1. Densidad de población</li> <li>2. Conectividad (Pasajeros transportados hacia el país) REVISAR: debe ser por <math>x</math> cantidad de persona.</li> <li>3. Población en área urbana %</li> </ol> </li> <li>■ Socioeconómico               <ol style="list-style-type: none"> <li>1. Índice de pobreza</li> <li>2. Ingreso per cápita</li> <li>3. Índice de Gini</li> <li>4. Escolaridad años promedio</li> </ol> </li> <li>■ Biológicos               <ol style="list-style-type: none"> <li>1. Edad promedio</li> <li>2. Expectativa de vida al nacer</li> </ol> </li> <li>■ Otros               <ol style="list-style-type: none"> <li>1. ¿Índice de corrupción? CPI</li> </ol> </li> </ul> | } | Factores predictores  |
| <ul style="list-style-type: none"> <li>■ Estratégicos frente a la pandemia               <ol style="list-style-type: none"> <li>1. Número de pruebas realizadas por millón</li> <li>2. proporción de fallecidos</li> <li>3. ¿Días sin cuarentena?</li> </ol> </li> </ul>  | } | Factores de respuesta |

Denotemos como  $M$  al número total de parámetros que tenemos y a  $M_C, M_S, M_B, M_E$  respectivamente el número total de parámetros culturales, socioeconómicos, biológicos y estratégicos. Estos deben cumplir la obvia relación  $M = M_C + M_S + M_B + M_E$ . Por facilidad de escritura escribamos

$$M = \sum_{\mu \in \mathcal{C}} M_\mu$$

donde  $\mathcal{C} := \{C, S, B, E\}$  es el conjunto de categorías de los parámetros. Además, si denotamos el  $j$ -ésimo parámetro de la categoría  $\mu$  del  $i$ -ésimo país como  $P_{\mu,j}^{(i)}$ , podemos escribir el conjunto de todos los parámetros de un país  $i$  fijo como:

$$\mathcal{P}^{(i)} = \underbrace{\{ \underbrace{P_{C,1}^{(i)}, P_{C,2}^{(i)}, \dots, P_{C,M_C}^{(i)}}_{\substack{\text{Parámetros culturales} \\ =: \mathcal{P}_C^{(i)}}}, \underbrace{P_{S,1}^{(i)}, P_{S,2}^{(i)}, \dots, P_{S,M_S}^{(i)}}_{\substack{\text{Parámetros socioeconómicos} \\ =: \mathcal{P}_S^{(i)}}}, \underbrace{P_{B,1}^{(i)}, P_{B,2}^{(i)}, \dots, P_{B,M_B}^{(i)}}_{\substack{\text{Parámetros biológicos} \\ =: \mathcal{P}_B^{(i)}}}, \underbrace{P_{E,1}^{(i)}, P_{E,2}^{(i)}, \dots, P_{E,M_E}^{(i)}}_{\substack{\text{Parámetros de estrategia} \\ =: \mathcal{P}_E^{(i)}}} \}_{}}_{\text{Parámetros de predicción} =: \mathcal{P}_P^{(i)}}$$

De manera que  $\mathcal{P}^{(i)}$  es una unión disjunta de  $\mathcal{P}_C^{(i)}, \mathcal{P}_S^{(i)}, \mathcal{P}_B^{(i)}$  y  $\mathcal{P}_E^{(i)}$ . También se ha definido a  $\mathcal{P}_P^{(i)}$  para denotar a todos los parámetros predictores. Esto nos provee el lenguaje para definir con claridad la propuesta de forma funcional para  $\omega_i$ .

Considere a  $\omega_i$  como una medida de similitud que el  $i$ -ésimo país tiene respecto al país para el cual se desea estimar la serie de tiempo (Honduras en este caso). Entonces  $\omega$  debe tener la siguiente dependencia funcional:

$$\omega_i = \omega_i \left( \underbrace{\mathcal{P}_C^{(i)}, \mathcal{P}_S^{(i)}, \mathcal{P}_B^{(i)}}_{\substack{\text{Parámetros del} \\ i\text{-ésimo país}}}, \underbrace{\mathcal{P}_C^{(H)}, \mathcal{P}_S^{(H)}, \mathcal{P}_B^{(H)}}_{\substack{\text{Parámetros de} \\ \text{Honduras}}} \right) = \omega_i \left( \mathcal{P}_P^{(i)}, \mathcal{P}_P^{(H)} \right)$$

es decir, debe depender de los parámetros de los países que se están comparando, pero solamente de  $C, S$  y  $B$ . Esto es porque los parámetros de estrategia no reflejan similitudes intrínsecas de los países, si no que acciones específicas a la crisis del Covid-19 y se utilizarán de manera indirecta en un procedimiento de regresión para minimizar el error a la data que tenemos.

Ahora, podemos proponer que  $\omega_i$  depende de la diferencia por pares de dichos parámetros, es decir de las diferencias  $Z_{\mu,j}^{(i)} := \left| P_{\mu,j}^{(i)} - P_{\mu,j}^{(H)} \right|^{-1}$ . La asunción más simple que podemos tener es que la de una dependencia lineal<sup>4</sup> de la forma:

$$\omega_i = \sum_{j=1}^{M_C} \alpha_j Z_{C,j}^{(i)} + \sum_{j=1}^{M_S} \beta_j Z_{S,j}^{(i)} + \sum_{j=1}^{M_B} \gamma_j Z_{B,j}^{(i)}$$

donde los pesos  $\alpha_j, \beta_j$  y  $\gamma_j$  son inicialmente libres, pero deseamos encontrarlos con un ajuste de regresión lineal generalizada con los datos que tenemos. Este procedimiento requiere de pares ordenados de la forma  $(x_i, y_i)$  donde aquí  $Z_{\mu,j}^{(i)}$  juega el rol de las  $x_i$ . las  $y_i$  deben ser propuestas como los  $\omega_i$  “reales” que sea externa al modelo, y para ello se utilizarán los parámetros de estrategia,  $\mathcal{P}_E^{(i)}$ , antes determinados y aun no utilizados.

Utilizar un modelo general lineal asume homocedasticidad de los errores. Esto puede ser asumido o alcanzado bajo el siguiente argumento: Considere primero la siguiente notación para el modelo lineal:

$$\begin{aligned} \omega &:= [\omega_1, \dots, \omega_N]^\top \\ \Theta &:= [\alpha_1, \dots, \alpha_{M_C}, \beta_1, \dots, \beta_{M_S}, \gamma_1, \dots, \gamma_{M_B}]^\top \\ Z &:= \begin{bmatrix} Z_{C,1}^{(1)} & \dots & Z_{B,M_B}^{(1)} \\ \vdots & \ddots & \vdots \\ Z_{C,1}^{(N)} & \dots & Z_{B,M_B}^{(N)} \end{bmatrix} \\ \epsilon &:= [\epsilon_1, \dots, \epsilon_N]^\top \end{aligned}$$

donde se ha asumido que el número de países en el modelo es  $|\mathbb{P}| = N$ . Resumido en forma matricial el modelo lineal sería:

$$\omega = Z\Theta + \epsilon$$

donde  $\epsilon$  es el vector de errores tal que  $\mathbb{E}[\epsilon | Z] = 0$  y  $\text{Cov}[\epsilon | Z] = \Omega$ , matriz la cual se asume conocida. La condición de **homocedasticidad** impone varianzas iguales para todos los errores, es decir, los elementos diagonales de  $\Omega$  serán iguales.

Esto no será siempre cierto, pues la estimación que los **predictores**,  $Z$ , pueden lograr de la respuesta,  $\omega$ , varía por la calidad de ambos datos. Los datos en  $\mathcal{P}_P^{(i)}$  fueron obtenidos de las mismas fuentes las cuales se pueden asumir que tienen un control de calidad homogéneo, no obstante los de  $\mathcal{P}_E^{(i)}$  dependen de reportes individuales de países.

Para medir la diferencia de estos errores, y posteriormente transformarlos para cumplir con la homocedasticidad, se propone utilizar un indicador de transparencia como el índice de corrupción CPI.

---

<sup>4</sup>Se puede considerar mejor a  $Z_{\mu,j}^{(i)} := \exp\left(-\frac{|P_{\mu,j}^{(i)} - P_{\mu,j}^{(H)}|^{-1}}{(2\sigma)}\right)$