UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA

FACULTAD DE ZOOTECNIA Y ECOLOGÍA

SECRETARÍA DE INVESTIGACIÓN Y POSGRADO

UNIVERSIDAD AUTÓNOMA DE
CHIHUAHUA

ÍNDICE DE VULNERABILIDAD COVID-19

POR:

INGENIERO MARIO ALBERTO MATA LICÓN

TESINA PRESENTADA COMO REQUISITO PARA OBTENER EL GRADO DE:

MAESTRÍA PROFESIONAL EN ESTADÍSTICA APLICADA.

CHIHUAHUA, CHIH., MÉXICO **NOVIEMBRE DE 2025**

Índice de vulnerabilidad COVID-19. Tesina de maestría presentada por Mario Alberto Mata Licón como requisito parcial para obtener el grado de Maestría Profesional en Estadística Aplicada, ha sido aprobada y aceptada por:

D. Ph. Alfredo Pinedo Álvarez
Director de la Facultad de Zootecnia y Ecología

Dra. Rosalía Sánchez Basualdo
Secretaria de Investigación y posgrado

Dr. Lauro Manuel Espino Enríquez
Coordinador Académico

D.Ph. Joel Domínguez Viveros
Presidente del Comité

21 de noviembre del 2025

Fecha

UNIVERSIDAD AUTÓNOMA DE CHIHUAHUA

Fecha: _____

Después de haber reunido la revisión final del trabajo de tesina a nivel Maestría, bajo el nombre: Índice de vulnerabilidad COVID-19

Elaborado por: Mario Alberto Mata Licón

Los integrantes del comité están de acuerdo en que dicho trabajo satisface los requisitos para su publicación (indicándolo con una "X"), desde el punto de vista de:

Puesto en comité	Nombre y firma	Contenido Técnico	Análisis de la Información	Formato (estilo y forma)
Asesor:	Dr. Joel Domínguez Viveros	✓	✓	✓
Representante área relacionada:	Dra. Rosalía Sánchez Bañuelos	✓	✓	✓
Representante área estadística:	Dr. Nelson Guadalupe Aguilar Palma	✓	✓	✓
Representante de la coordinación:	M.P.A. Amanda Castro Ochoa	✓	✓	✓

Dr. Lauro Manuel Espino Enríquez
Unidad/Coordinación Académica
Vo. Bo.

FACULTAD DE ZOOTECNIA Y ECOLOGÍA
Periférico Francisco R. Almada km 1
C.P. 31453 Chihuahua, Chih.
Tel. 52 (614) 434-0363, 434-0304, 434-0458 fax 434-0345
www.fz.uach.mx

AGRADECIMIENTOS

Agradezco a mi compañera y dentista la Maestra Mayra Ramírez, a la comunidad de la Facultad de Zootecnia y Ecología de la UACH quienes me brindaron todo su apoyo y en especial a mi asesor el D. Ph. Joel Domínguez quien me brindó su tiempo y conocimiento para lograrlo.

DEDICATORIA

A mi esposa Eva Mayela que nunca me dejó abandonar el reto.

CURRICULUM VITAE

El autor nació el 15 de mayo de 1989 en la ciudad de Delicias Chihuahua México

2008-2012	Estudió la licenciatura en el Instituto Tecnológico de Estudios Superiores de Monterrey en Chihuahua, Chihuahua.
Junio-agosto 2012	Integrante del equipo de manufactura esbelta en su división transporte General Electric.
2012-2017	Gerente de producción en Comercial de muebles Mata S.A. de C.V.
2013-2016	Profesor de matemáticas y estadística en la Universidad Vizcaya de las Américas.
2017-2021	Director del CONALEP plantel Delicias
2021-2022	Profesor de estadística facultad de Contaduría y Administración en la Universidad Autónoma de Chihuahua
2021-2022	Director de servicios públicos Municipales ciudad Delicias Chihuahua

Miembro de la Asociación única de fabricantes de muebles de Delicias

Miembro del club Activo 20-30 ciudad Delicias

Miembro de CANACINTRA Delicias

RESUMEN

ÍNDICE DE VULNERABILIDAD COVID-19

POR:

ING. MARIO ALBERTO MATA LICÓN

Maestría profesional en estadística aplicada

Secretaría de investigación y posgrado

Facultad de Zootecnia y ecología

Universidad Autónoma de Chihuahua

Presidente: D. Ph. Joel Domínguez Viveros

Él trabajó recopiló y analizó de 51 países, la información establecida por las organizaciones médicas como necesarias para el control y cuidado en los esfuerzos por detener el avance de la enfermedad COVID-19 dadas las condiciones particulares de los países se pudo establecer un índice de vulnerabilidad con un análisis de componentes principales, siendo este un métrico en la búsqueda por categorizar a los países en fortaleza o debilidad para el control de la enfermedad en su territorio y población, se utilizaron en el estudio 4 ejes principales para el análisis: población, sector salud, economía y las enfermedades padecidas comúnmente en cada país, en el modelo final se utilizaron 8 variables las cuales actúan de manera negativa o positiva en el control de la enfermedad, se analizó de manera individual estas variables así como su

correlación, se concluye el trabajo con un análisis de grupos para una subdivisión de la vulnerabilidad y mejor toma de decisiones.

ABSTRACT

COVID-19 VULNERABILITY RATE

BY:

MARIO ALBERTO MATA LICÓN

The study collected and analyzed information from fifty-one countries, focusing on data deemed necessary by medical organizations for the control and management of efforts to stop the spread of COVID-19. Given the specific conditions of each country, a vulnerability index was established by the principal component analysis. This vulnerability rate serves as a metric aimed at categorizing countries based on their strengths or weaknesses in controlling the disease within their territories and populations. The study utilized four main axes for the analysis: population, healthcare sector, economy, and the diseases commonly found in each country. The final model incorporated eight variables, which have either positive or negative effects on disease control. These variables were individually analyzed, as well as their correlations. The study concludes with a cluster analysis to subdivide vulnerability and improve decision-making processes.

CONTENIDO

Página

LISTA DE CUADROS

LISTA DE GRÁFICAS

LISTA DE FIGURAS

LISTA DE ABREVIACIONES

Concepto	Término
OMS	Organización Mundial de la Salud
TM	Tasa de mortalidad
IV	Índice de vulnerabilidad
USD	Dólares americanos

INTRODUCCIÓN

A finales del año 2019 se comenzaron a presentar fallecimientos por neumonía atípica en la comunidad de Wuhan China, el gobierno de este país informó que el broté se pudo haber ocasionado en un mercado de dicha ciudad; en enero de 2020 se contabilizaban 41 personas confirmadas, el mercado fue clausurado y se descartó que la enfermedad fuera síndrome respiratorio agudo grave, gripe o gripe aviar; los casos se habían extendido en Japón y Tailandia, para finales de este mismo mes la Organización Mundial de la Salud (OMS) declaró emergencia sanitaria de preocupación mundial con casos en 30 países, en marzo el virus se había detectado en 100 países, fue reconocida por la OMS como pandemia y ya se tenían 500 casos confirmados. En China, de 44672 casos confirmados en la provincia de Hubei el porcentaje de fallecimientos fue del 2.3 %; además, el grupo de mayor riesgo correspondió a las personas mayores de 80 años, con una tasa de mortalidad del 14.8% (Team, 2020).

En el esfuerzo por sobrellevar la pandemia lo mejor posible, instituciones nacionales e internacionales colaboraron para crear métricos, índices e indicadores de vulnerabilidad; uno de esos estudios fue presentado por el instituto de geografía de la Universidad Autónoma de México (Lastra, 2021), ese métrico permitió identificar la distribución espacial de los diferentes factores que generan mayor susceptibilidad al daño o las consecuencias adversas que pueden tener las personas, el estudio ayudó a los municipios con información *a priori* de los picos de la pandemia.

Otras instituciones mundiales aportaron índices para la evaluación de la respuesta de cada país a la pandemia; la empresa financiera y de software Bloomberg creó el: Ranking de Resiliencia COVID, en este (Roa, 2020) señalan que gracias a estos indicadores las administraciones han manejado el virus de manera más efectiva con la menor cantidad de interrupciones para los negocios y la sociedad; el estudio fue realizado a 53 naciones e integrando 10 indicadores y generó una medida de análisis objetiva; sin embargo, la información mostrada fue sobre el procedimiento de respuesta de los países, no sobre la capacidad para afrontar una pandemia, es decir, evalúa el desempeño del país sobrellevando la pandemia, no el riesgo que tiene un país a no poder controlar la pandemia antes de que ocurra, con esta información las diferentes organizaciones pueden evaluar a los países, clasificarlos y destinar los recursos necesarios a donde más se requieran.

El objetivo del presente estudio fue aplicar las técnicas de análisis multivariado para crear un índice de vulnerabilidad COVID-19 por país; la información permite evaluar la preparación para afrontar una pandemia y generar una medida objetiva para la clasificación por medio de sus capacidades.

REVISIÓN DE LITERATURA

Análisis de correlación

Una aplicación estadística para estudiar la variación de las características físicas en poblaciones es el coeficiente de correlación, este se define como una medida numérica de la fuerza de la relación lineal entre dos variables. Este coeficiente se denota con la literal *r*.

Sean $(x_1, y_1), \ldots, (x_n, y_n)$ los *n* puntos del diagrama de dispersión. Para calcular la correlación, primero se deducen las medias y las desviaciones estándar de las *x* y de las *y*, que se representan mediante $\bar{x}, \bar{y}, s_x$ y s_y. Después se convierte cada *x* y cada *y* a las unidades estándar; en otras palabras, se calculan los puntajes z: $(x_i - \bar{x})/s_x, (y_i - \bar{y})/s_y$. El coeficiente de correlación se obtiene con la siguiente fórmula (Navidi, 2006):

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

y con las siguientes hipótesis:

H_o: $r_{xy} = 0 \Rightarrow$ El coeficiente de correlación ($\rho = 0$)

H_1: $r_{xy} \neq 0 \Rightarrow$ El coeficiente de correlación ($\rho \neq 0$)

Análisis multivariante

Cuadras (2014) afirmó que en el análisis multivariante se encuentran aquellos métodos matemáticos y estadísticos que tienen como finalidad la

interpretación de los datos observados desde la interacción de sus propias variables estudiadas conjuntamente, estas no tienen alguna variable dominante, son homogéneas y correlacionadas, en resumen, el análisis multivariante es el desarrollo y cálculo de métodos estadísticos en el estudio de los elementos o individuos desde el punto de vista de varias variables.

Conceptos básicos del análisis multivariante

En análisis multivariado, el primer paso es buscar y eliminar aquellas variables que se consideren atípicas para el estudio, el agregar variables en esta situación pueden distorsionar las medias y desviaciones típicas de las variables destruyendo las relaciones existentes, también es importante una visualización gráfica para una mejor representación de sus valore individuales y relaciones (Peña, 2002). Dentro de la decisión en el número de variables que integren el modelo, Moroy (2007) señala que, en un análisis multivariante, así como en el análisis de regresión, el procedimiento es encontrar el menor número de variables predictoras que mejor contribuyan a la discriminación (principio de parsimonia). Refiriéndose a la problemática del error de medida se analizan las variables que por concepto propio pudieran dar un parámetro erróneo o poco comparable entre las observaciones, por ejemplo, la densidad de cada país, este tipo de medidas puede añadir ruido al análisis y en específico en la relación de dependencias (Hair, 1999). Por otra parte, se tiene el problema del error de especificación este hace referencia a la inclusión de variables irrelevantes o a la omisión de variables relevantes del conjunto de variables independientes (Hair, 1999).

Método de componentes principales

El análisis de componentes es una aproximación estadística que puede usarse para analizar interrelaciones entre un gran número de variables y explicar estas variables en términos de sus dimensiones subyacentes comunes (factores), el objetivo es encontrar un modo de condensar la información contenida en un número de variables originales en un conjunto más pequeño de variables con una pérdida mínima de información (Hair, 1999). Un estudio de componentes principales tiene como objetivo transformar las variables originales, en general correlacionadas, en nuevas variables no correlacionadas, facilitando la interpretación de los datos (Peña, 2002). El análisis de componentes principales es una metodología de tipo matemático para la cual no es necesario asumir distribución probabilística alguna (Díaz Monroy, 2007).

Determinación de los componentes principales

Se determina la primera componente principal Y_1 la cual sintetiza la mayor cantidad de variabilidad total contenida en los datos con el siguiente modelo:

$$Y_1 = \gamma_{11}X_1 + \gamma_{12}X_2 + \cdots + \gamma_{1p}X_p$$

Donde las ponderaciones $\gamma_{11}, \ldots, \gamma_{1p}$ se escogen de tal forma que maximicen la razón de la varianza de Y_1 a la variación total; con la restricción:

$$\sum_{j=1}^{p} \gamma_{1j}^2 = 1$$

La segunda componente principal Y_2 es una combinación lineal ponderada de las variables observadas, la cual no está correlacionada con la primera

componente principal y reúne la máxima variabilidad restante de la variación total contenida en la primera componente principal Y_1. De manera general, la k-ésima componente es una combinación lineal de las variables observadas X_j, para $J = 1, \ldots, p$, la cual tiene la varianza más grande entre todas las siguientes. De otra manera, los Y_k sintetizan en forma decreciente la varianza del conjunto original de datos.

$$Y_k = \gamma_{k1}X_1 + \gamma_{k2}X_2 + \cdots + \gamma_{kp}X_p,$$

A continuación, se muestra cómo generar las componentes principales. Supóngase que el vector aleatorio $X' = (X_1, \ldots, X_p)$ tiene matriz de varianzas y covarianzas $\sum$. Sin pérdida de generalidad asúmase que la media de los X_i es cero, para todos los $i = 1, \ldots, p$;

Para encontrar la primera componente principal, se examina el vector de coeficientes $\Gamma' = (\gamma_{11}, \ldots, \gamma_{1p})$, tal que la varianza $\Gamma'X$ sea un máximo sobre la clase de todas las combinaciones lineales $\Gamma'X$, con la restricción $\Gamma'\Gamma = 1$.

De esta manera se determina la combinación lineal

$$Y = \sum_{j=1}^{p} \gamma_{1j}X_j \ ,$$

Tal que

$$var(Y) = var\left(\sum_{j=1}^{p} \gamma_{1j}X_j\right) \ ,$$

Sea la máxima, donde $\sum_{j=1}^{p} \gamma_{ij}^2 = 1$

La restricción que Γ sea un vector unitario, se hace para evitar el incremento de la varianza de manera arbitraria; de lo contrario, ésta se incrementaría tan sólo aumentando cualquiera de las componentes de Γ . Ahora se maximiza $var(\Gamma' X) = \Gamma' \Sigma \Gamma$ con respecto a Γ, sujeto a $\Gamma' \Gamma = 1$, lo cual, por multiplicadores de Lagrange, equivale a resolver

$$(\Sigma - \lambda_1 I)\Gamma_1 = 0$$

Para que la solución anterior sea diferente de la trivial, el vector Γ_1 debe ser escogido de tal manera que

$$|\Sigma - \lambda_1 I| = 0$$

Esta ecuación corresponde a la ecuación característica, su solución es el valor propio más grande de Σ y Γ_1 el correspondiente vector propio.

Así el primer componente principal puede describirse de la siguiente forma

$$Y_1 = \Gamma_1' X.$$

El segundo componente principal se determina encontrando un segundo vector normalizado Γ_2, ortogonal a Γ_1, tal que $Y_2 = \Gamma_2' X$ tenga la segunda varianza más grande entre todos los vectores que satisfacen:

$$\Gamma_1' \Gamma_2 = 0 \qquad \text{y} \qquad \Gamma_2' \Gamma_2 = 1$$

Se demuestra que Γ_2 es el vector propio correspondiente al segundo valor propio más grande de Σ. El proceso se desarrolla hasta encontrar los p-vectores; donde Γ_r es ortonormal a $\Gamma_1, \dots, \Gamma_{r-1}$ con $r = 2, \dots, p$. En la mayor parte de los análisis se asume que las raíces de Σ son distintas, esto implica que sus vectores

propios asociados son mutuamente ortogonales; si, además, se asume que Σ es definida positiva, entonces todas las raíces son positivas. En este caso, el rango corresponde al número de valores propios no nulos (Díaz Monroy, 2007).

Generación de los componentes principales

Supóngase que los valores de *p*-variables $X_1, \dots, X_p$ se obtienen sobre una muestra de *n*-individuos, la matriz $\mathbb{X}$ de tamaño $(n \times p)$ representa tales datos. Sean $\bar{X}_j$ la media muestral de la variable X_j ; s_{jk} la covarianza muestral entre las variables X_j y X_k y la matriz $S = (s_{jk})$, corresponde a la matriz de varianzas y covarianzas muestral de las *p*-variables. Paralelamente se encuentra el primer componente principal que tenga la máxima varianza muestral a través de la combinación lineal

$$Y_1 = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p = \sum_{j=1}^{p} a_{1j}X_j \ .$$

La varianza muestral de esta combinación lineal es igual a $a_1' S a_1$, donde el vector $a_1' = (a_{11}, \dots, a_{1p})$, es tal que $\|a_1\| = 1$. Los componentes de a_1 deben satisfacer

$$[S - l_1 I]a_1 = 0 \ ,$$

Con l_1 el multiplicador de Lagrange. Entonces l_1 es tal que

$$|S - l_1 I| = 0 \ ;$$

La cantidad l_1 es el valor propio más grande de S y a_1 su correspondiente vector propio.

Para la extracción de los demás componentes principales las m componentes principales $(m \leq p)$ que contienen, de manera decreciente, fracciones de la varianza total, se generan a partir de los respectivos a_k ; con $l_1 \geq l_2 \geq \ldots \geq l_m \ldots \geq l_p$ los valores propios de S.

La contribución de la *i*-ésima variable a la *k*-ésima componente principal está dada por la magnitud del coeficiente a_{ki}. La covarianza entre la variable X_i y el componente principal Y_k es:

$$cov(X_i, Y_k) = a_{ki} l_k.$$

La varianza muestral de las observaciones con respecto a la *k*-ésima componente principal es:

$$var(Y_k) = a_k' S a_k = l_k,$$

Con l_k el *k*-ésimo valor propio ordenado descendentemente.

La varianza total de las *p*-variables es:

$$VT = \text{tra}(S) = \sum_{j=1}^{p} l_j$$

La correlación se obtiene con la siguiente división:

$$rX_i Y_k = \frac{a_{ki} l_k}{\left(\sqrt{S_{ii}}\right)\left(\sqrt{l_k}\right)} = \frac{a_{ki}\sqrt{l_k}}{\sqrt{S_{ii}}}$$

Esta última expresión suministra la ponderación (o grado de importancia) de la *i*-ésima variable sobre la *k*-ésima componente principal. (Díaz Monroy, 2007).

Análisis de conglomerados

El análisis de conglomerados busca particionar un conjunto de objetos en grupos, de tal forma que los objetos de un mismo grupo sean similares y los objetos de grupos diferentes sean disímiles (Díaz, 2007). Así, el análisis de conglomerados tiene como objetivo principal definir la estructura de los datos colocando las observaciones más parecidas en grupos. Para alcanzar lo anterior se deben considerar los siguientes aspectos: Similitud, se requiere de un "dispositivo" que permita comparar los objetos en términos de las variables medidas sobre ellos. Tal dispositivo debe registrar la proximidad entre pares de objetos de tal forma que la distancia entre las observaciones (atributos del objeto) indique la similitud.

Reconocer objetos como similares o disímiles es fundamental para el proceso de clasificación. Aparte de su simplicidad, el concepto de similitud para aspectos cuantitativos se presenta ligado al concepto de métrica. Las medidas de similitud se pueden clasificar en dos tipos; en una parte estén las que reúnen las propiedades de métrica, como la distancia; en otra, se pueden ubicar los coeficientes de asociación, estos últimos empleados para datos en escala nominal.

Una métrica *d* (;) es una función (o regla) que asigna un número a cada par de objetos de un conjunto Ω, es decir,

$$\Omega \times \Omega : \overset{d}{\rightarrow} \mathrm{R}$$

$$(x, y) \longrightarrow d(x, y)$$

La cual satisface, sobre los objetos x, y y z de Ω, las siguientes condiciones:

1. No negatividad, $d(x,y) \geq 0$, y $d(x,y) = 0$, si sólo si, $x = y$.
2. Simetría. Dados los objetos x y y, la distancia d, entre ellos satisface $d(x,y) = d(y,x)$.
3. Desigualdad triangular. Para tres objetos x, y y z las distancias entre ellos satisfacen la expresión $d(x,y) \leq d(x,z) + d(z,y)$.
4. Identificación de no Identidad. Dados los objetos x y y si $d(x,y) \neq 0$, entonces $x \neq y$
5. Identidad. Para dos elementos idénticos, x y $\acute{x}$, se tiene que $d(x,\acute{x}) = 0$ es decir, si los objetos son idénticos la distancia entre ellos es cero.

Como observación se tiene que, a cambio de la desigualdad triangular, se satisface $d(x,y) \leq \max\{d(x,z), d(z,y)\}$, para todo x, y y z, a este tipo de distancia se le denomina ultra métrica. Esta distancia juega un papel importante en los métodos de clasificación automática. Las medidas de similitud más frecuentes son: medidas de distancia, coeficiente de correlación, coeficiente de asociación y medidas probabilísticas de similitud.

Formación de los conglomerados.

Este es el procedimiento mediante el cual se agrupan las observaciones que son más similares dentro de un determinado conglomerado, para determinar

la pertenencia al grupo de cada observación, Díaz (2007) propone los siguientes tipos de métodos:

1. Métodos aglomertativos; tienen como primera característica que buscan una matriz de similitudes de tamaño $(n \times n)$, (n número de objetos) desde la cual, secuencialmente, se mezclan los casos más cercanos; aunque cada uno tiene su propia forma de medir las distancias entre grupos o clases. Un segundo aspecto es que cada paso o etapa en la conformación de grupos puede representarse visualmente por un dendrograma. En tercer lugar, se requieren $(n-1)$ pasos para la con formación de los conglomerados de acuerdo con la matriz de similitudes. En el primer paso cada objeto es tratado como un grupo; es decir, se inicia con n conglomerados, y, en el paso final, se tienen todos los objetos en un solo conglomerado. Finalmente, los métodos jerárquicos aglomerativos son conceptualmente simples, los métodos más usados son los siguientes:

    a. Enlace simple o del “vecino más cercano”

    b. Enlace completo o del “vecino más lejano”

    c. Método de *Ward*

Este último método trabaja directamente con los elementos utilizando una medida global de heterogeneidad de una agrupación de observaciones en grupos esta medida es descrita por $\mathbf{W}$ en la siguiente fórmula donde $\overline{\mathrm{X}}_g$ es la media del grupo *g*:

$$\mathbf{W} = \sum_{g} \sum_{i \in g} (\mathrm{X}_{ig} - \overline{\mathrm{X}}_{g})'(\mathrm{X}_{ig} - \overline{\mathrm{X}}_{g})$$

El algoritmo de decisión fue sustraído del libro: Análisis de datos multivariantes de Daniel Peña y dice lo siguientes:

El criterio comienza suponiendo que cada dato forma un grupo, *g* = *n* y por tanto **W** es cero; A continuación se unen los elementos que produzcan el incremento mínimo de **W**; esto implica tomar los más próximos con la distancia euclídea, en la siguiente etapa tenemos n−1 grupos, n−2 de un elemento y uno de dos elementos, se decide de nuevo que dos grupos unir para que **W** crezca lo menos posible, con lo que pasamos a n−2 grupos y así sucesivamente hasta tener un único grupo, Los valores de **W** van indicando el crecimiento del criterio al formar grupos y pueden utilizarse para decidir cuantos grupos naturales contienen nuestros datos, puede demostrarse que, en cada etapa, los grupos que debe unirse para minimizar W son aquellos tales que:

$$\min \frac{n_a n_b}{n_a + n_b} (\overline{\mathrm{X}}_a - \overline{\mathrm{X}}_b)'(\overline{\mathrm{X}}_a - \overline{\mathrm{X}}_b)$$

2. Métodos de partición, se comienza con una partición del conjunto de objetos en algún número específico de grupos y a cada uno de estos grupos se le calcula en centroide, se ubica cada caso u objeto en el conglomerado cuyo centroide esté más cercano a este, se calcula el nuevo centroide de los conglomerados; éstos no son actualizados hasta tanto no se comparen sus centroides con todos los casos, se continúa con el segundo y tercer paso

hasta que los casos resulten irremovibles, los métodos de partición más comunes son los siguientes:

a. Método de las *K*-medias

b. Métodos basados en la traza

c. Nubes dinámicas

Número de conglomerados

Díaz (2007) afirma que, aunque se dispone de un amplio número de estrategias para decidir sobre el número de conglomerados a construir, el criterio decisivo es la homogeneidad "media" alcanzada dentro de los conglomerados. Una estructura simple debe corresponder a un número pequeño de conglomerados, no obstante, a medida que el número de conglomerados disminuye, la homogeneidad dentro de los conglomerados necesariamente disminuye. En consecuencia, se debe llegar a un punto de equilibrio entre el número de conglomerados y la homogeneidad de estos.

Análisis discriminante

Díaz (2007) señala que son dos los objetivos principales abordados por el análisis discriminante, de una parte, está la separación o discriminación de grupos, y de otra, la predicción o asignación de un objeto en uno de entre varios grupos previamente definidos, con base en los valores de las variables que lo identifican. Para este proceso Cuadras (2014) cita que sean Ω_1, Ω_2 dos poblaciones, $X_1, \dots, X_P$ variables observables y $\mathrm{x} = (x_1, \dots, x_p)$ las observaciones

de las variables sobre un individuo ω. Se trata de asignar ω a una de las dos poblaciones, una regla discriminante es un criterio que permite asignar ω conocido $(x_1, \dots, x_p)$ y que a menudo es planteado mediante una función discriminante $D(x_1, \dots, x_p)$. Entonces la regla de clasificación es

Si $D(x_1, \dots, x_p) \geq 0$ asignamos ω a Ω_1

En caso contrario asignamos ω a Ω_2

Esta regla divide a R^p en dos regiones

$$R_1 = \{\mathrm{x} | D(\mathrm{x}) > 0\}, \quad R_2 = \{\mathrm{x} | D(\mathrm{x}) < 0\}$$

En la decisión se identifica ω, nos equivocaremos si asignamos ω a una población a la que no pertenece, esto es la probabilidad de clasificación errónea con sus siglas (pce) donde:

$$pce = P(R_2/\Omega_1)P(\Omega_1) + P(R_1/\Omega_2)P(\Omega_2)$$

En el caso que el individuo ω puede provenir de k poblaciones $\Omega_1, \Omega_2, \dots, \Omega_k$ donde $k \geq 3$, es necesario establecer una regla que permita asignar ω a una de las k poblaciones sobre la base de las observaciones $\mathrm{x} = (x_1, x_2, \dots, x_p)'$ de p variables, para esto se proponen discriminadores lineales y regla de Bayes.

Discriminadores lineales

Con la suposición que la media de las variables en Ω_i es $\boldsymbol{\mu}_i$, y que la matriz de covarianzas $\boldsymbol{\Sigma}$ es común. Si consideramos las distancias de Mahalanobis de ω a las poblaciones

$$M^2(\mathrm{x}, \boldsymbol{\mu}_i) = (\mathrm{x} - \boldsymbol{\mu}_i)'\Sigma^{-1}(\mathrm{x} - \boldsymbol{\mu}_i), i = 1, \dots, k,$$

un criterio de clasificación consiste en asignar ω a la población más próxima:

Si $M^2(\mathrm{x}, \boldsymbol{\mu}_i) = mín\{M^2(\mathrm{x}, \boldsymbol{\mu}_1), \dots, M^2(\mathrm{x}, \boldsymbol{\mu}_k)\}$, asignamos ω a Ω_i

Esto anterior equivale a: Si $L_{ij}(\mathrm{x}) > 0$ para todo $j \neq i$, asignamos ω a Ω_i, introduciendo las funciones discriminantes lineales

$$L_{ij}(\mathrm{x}) = (\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\,\Sigma^{-1}\mathrm{x} - \frac{1}{2}(\boldsymbol{\mu}_i - \boldsymbol{\mu}_j)'\,\Sigma^{-1}(\boldsymbol{\mu}_i + \boldsymbol{\mu}_j)$$

Además, las funciones $L_{ij}(\mathrm{x})$ verifican:

1. $L_{ij}(\mathrm{x}) = \frac{1}{2}\left[M^2(\mathrm{x}, \mu_j) - M^2(\mathrm{x}, \mu_i)\right]$
2. $L_{ij}(\mathrm{x}) = -L_{ji}(\mathrm{x})$
3. $L_{rs}(\mathrm{x}) = L_{is}(\mathrm{x}) - L_{ir}(\mathrm{x})$

Es decir, sólo necesitamos conocer $k - 1$ funciones discriminantes.

Regla de la máxima verosimilitud

Sea $f_i(x)$ la función de densidad de x en la población Ω_i. Podemos obtener una regla de clasificación asignando ω a la población donde la verosimilitud es más grande:

Si $f_i(x) = máx\{f_i(x), \dots, f_k(x)\}$, asignamos ω a Ω_i

Este criterio es más general que el geométrico y está asociado a las funciones discriminantes

$$V_{ij}(\mathrm{x}) = \log f_i(\mathrm{x}) - \log f_j(\mathrm{x})$$

En el caso de normalidad multivariante y matriz de covarianzas común, se verifica $V_{ij}(\mathrm{x}) = L_{ij}(\mathrm{x})$, y los discriminadores máximo-verosímiles coinciden con los lineales. Pero si las matrices de covarianzas son diferentes $\Sigma_1, \dots, \Sigma_k$, entonces este criterio dará lugar a los discriminadores cuadráticos. (Cuadras, 2014)

$$Q_{ij}(\mathrm{x}) = \tfrac{1}{2}\mathrm{x}'(\Sigma_j^{-1} - \Sigma_i^{-1})\mathrm{x} + \mathrm{x}'(\Sigma_j^{-1}\mu_1 - \Sigma_i^{-1}\mu_2) + \tfrac{1}{2}\mu_j' \Sigma_j^{-1}\mu_j - \tfrac{1}{2}\mu_i' \Sigma_i^{-1}\mu_i + \tfrac{1}{2}\log|\Sigma_j| - \tfrac{1}{2}\log|\Sigma_i|.$$

Regla de Bayes

Si además de las funciones de densidad $f_i(\mathrm{x})$, se conoce las probabilidades *a priori*

$$q_1 = P(\Omega_1), \dots, q_k = P(\Omega_k),$$

La regla de Bayes que asigna ω a la población tal que la probabilidad *a posteriori* es máxima

Si $q_i f_i(\mathrm{x}) = máx\{q_1 f_1(\mathrm{x}), \dots, q_k f_k(\mathrm{x})\}$, asignamos ω a Ω_i, está asociada a las funciones discriminantes

$$B_{ij}(\mathrm{x}) = \log f_i(\mathrm{x}) - \log f_j(\mathrm{x}) + \log(q_i/q_j).$$

Finalmente, si $P(j/i)$ es la probabilidad de asignar ω a Ω_j cuando en realidad es de Ω_i, la probabilidad de clasificación errónea es:

$$pce = \sum_{i=1}^{k} q_i \left(\sum_{j \neq i}^{k} P(j/i) \right)$$

Y se demuestra que la regla de Bayes minimiza esta pce. (Cuadras, 2014)

Tasa de error de clasificación por restitución

Díaz (2007) afirma que al tener una regla de clasificación es necesario evaluar la confiabilidad de esta regla, uno de los métodos de validación es el cálculo de la tasa de error por clasificación incorrecta, esta tasa se puede calcular por el método de restitución el cual funciona clasificando a los elementos en el conjunto al que fueron designados, a cada elemento X_i se le aplicará la función de clasificación para asignarlo a uno de los grupos, con esto se tiene el porcentaje de elementos correctos e incorrectos, este porcentaje se le denomina tasa de error aparente el cual se calcula de la siguiente forma:

$$Tasa\ de\ error\ aparente = \frac{n_{12} + n_{21}}{n_1 + n_2} = \frac{n_{12} + n_{21}}{n_{11} + n_{12} + n_{21} + n_{22}}$$

Donde las n_1 observaciones de G_1, n_{11} son clasificadas correctamente en G_1 y n_{12} son clasificadas incorrectamente en G_2, con $n_1 = n_{11} + n_{12}$. Análogamente, de las n_2 observaciones de G_2, n_{21} son asignadas incorrectamente a G_1 y n_{22} son correctamente asignadas a G_2,con $n_2 = n_{21} + n_{22}$.

MATERIALES Y MÉTODOS

Definición de las variables de estudio

La Organización Mundial de la salud (OMS) presentó en sus medidas de protección para combatir el COVID-19 que mantener una distancia de un metro entre personas ayuda a disminuir el contagio, por lo que los países a los que se les dificulte lo anterior, es decir, países con densidad de población alta tendrán un mayor riesgo de contagio (OMS, 2020).

Según las recomendaciones de la OMS el sistema de salud es clave para la buena respuesta, control y la no propagación del virus, la atención clínica de calidad para los pacientes de COVID-19 requiere un diagnóstico y pruebas precoces, y ha de ir acompañada de intervenciones de atención clínica adecuadas. El tratamiento con intervenciones clínicas apropiadas reduce el riesgo de que los pacientes desarrollen enfermedad grave y requieran hospitalización.

A su vez existen enfermedades previas que aumentan la tasa de mortalidad (TM) en los pacientes con COVID-19; esto fue analizado por el Centro de Control de China para enfermedades (CCDC por sus siglas en inglés) en su estudio: características epidemiológicas del brote 2019, donde reconocieron que hay hasta el momento 5 enfermedades con la mayor tasa de mortalidad: enfermedades cardiovasculares con una mortalidad de 10.5%; diabetes con 7.3 %; enfermedad pulmonar obstructiva crónica con 6.3 %; hipertensión con 6 %; y, cáncer con 5.6 % (Team, 2020) con base en lo anterior 15,536 pacientes que no

contaban con alguna de estas enfermedades previas, solamente se presentaron 133 fallecimientos siendo un 0.9 % (Médica, 2020).

Por otra parte, el Banco mundial en su informe sobre el desarrollo mundial 2022, finanzas al servicio de la recuperación equitativa, reportaron que los impactos económicos de la pandemia fueron especialmente graves en las economías emergentes, las pérdidas de ingresos pusieron de manifiesto y exacerbaron ciertos factores de fragilidad económica preexistentes; a medida que avanzó la pandemia en 2020, se vio con claridad que muchos hogares y empresas no estaban preparados para soportar una alteración de semejante duración y escala en sus ingresos, diversos estudios basados en datos anteriores a la crisis indicaron que, más del 50 % de los hogares de las economías emergentes y avanzadas no podrían sostener el consumo básico durante más de tres meses en caso de perder sus ingresos (Grupo Banco mundial, 2022).

Por otro lado, el ejercicio presentado por la unidad de investigaciones geográficas de la UNAM en su estudio, índice de vulnerabilidad ante el COVID-19 en México, los investigadores identifican tres dimensiones relevantes para la configuración de la vulnerabilidad: Demográfica, Salud y Socioeconómica; “cada dimensión está construida a partir de una serie de indicadores cuya relevancia se basa en la revisión de trabajos publicados en cada una de las áreas de conocimiento” (Lastra, 2021). Otro punto clave en el control de pandemia es el sector salud de cada país, Ranzani (2022) concluye que la desigualdad en el acceso a la atención de salud ha tenido una influencia evidente en las elevadas tasas de mortalidad por COVID-19 de los grupos vulnerables, lo que agrega

veracidad a la dimensión del sector salud como un factor importante en el análisis de la vulnerabilidad de los países, sumado a lo anterior mencionó que la pandemia de COVID-19 fue particularmente dura en América Latina, donde la disparidad social y económica derivaron en una serie de crisis económicas y de salud sin precedentes.

Base de datos analizada

Los datos fueron recopilados de las páginas oficiales de la Organización mundial de la salud (https://www.who.int/data/gho/info/gho-odata-api) y del departamento de economía de las Naciones Unidas, (https://population.un.org/wpp/Download/Standard/MostUsed/); se eligieron 51 países, tomando en consideración: la importancia del país, la veracidad de sus datos individuales y la existencia total de los registros en cada una de las variables a estudiar. La formación de la base de se dividió en 4 ejes principales como se muestra en la Figura 1: población, sector salud, enfermedades y sector económico, estos ejes fueron comparados con el ejercicio propuesto por Santos et al. (2021) donde señalan que para analizar la capacidad de exposición al COVID-19 deben considerarse los aspectos socioeconómicos, políticos y la infraestructura de salud del país en cuestión.

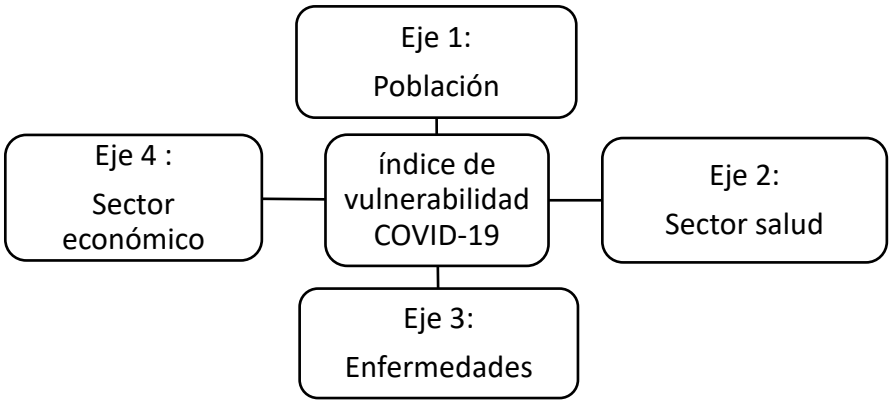


Figura 1. Ejes considerados en el índice de vulnerabilidad

1. Eje población.

La influencia de las variables relacionadas con la población en el índice de vulnerabilidad propuesto se confirmó con el estudio realizado por Md Iderus *et al.* (2022) donde señalaron que el análisis de correlación demostró que tanto la población absoluta como la densidad de población están significativamente correlacionadas con los casos de COVID-19 ($P < 0.001$), con un $r = 0.871$ para la población y un $r = 0.778$ para la densidad de población, por lo tanto para medir el riesgo latente que tiene cada país en la propagación del virus, así como el daño que pudiera ocasionar a su población se agregaron las siguientes variables presentadas en el Cuadro 1.

2. Sector salud

La importancia de las variables en este sector se comprobó con el estudio realizado por Molla y Hiilamo (2023) donde afirma que al aplicar el análisis de regresión en las variables del gasto en salud por país y sus resultados ante el COVID-19, demostraron una relación estadísticamente significativa entre las características de financiamiento en salud y los casos de mortalidad del virus ($P < 0.05$); donde estas características explican alrededor del 15 al 30 % de la variación en el exceso de mortalidad y tasa de letalidad, por lo anterior y dado que el sector salud es el principal responsable de controlar la pandemia en términos médicos, así como de la recuperación de los infectados, se agregaron las siguientes variables en el eje de salud presentadas en el Cuadro 2.

Cuadro 1. Eje Población

Nombre de la variable	Unidades
Población total	Miles de personas
Población de 0 a 14 años	Miles de personas
Población de 70 o más	Miles de personas
Población de 80 o más	Miles de personas
Densidad de población	Personas por Km^2
Ciudades con más de 1 Millón de habitantes	Ciudades

Cuadro 2. Eje salud

Nombre de la variable	Unidades	Notas
Médicos por 10,000 habitantes	Densidad	Densidad de médicos por 10,000 habitantes
Gasto per cápita en salud	Dólares por persona	-
Porcentaje de gasto del país en salud	Porcentaje	Porcentaje del producto interno bruto gastado en salud
Camas por 10,000 habitantes	Camas	Número de camas de hospital por 10,000 habitantes

3. Enfermedades

Marton *et al.* (2023) señalaron que al analizar 1,986,871 casos en Asia, Europa y USA, confirmaron que el factor que da más severidad a la enfermedad es la edad del paciente, sin embargo, es estadísticamente significativo las personas que tienen; Enfermedad pulmonar obstructiva crónica ($P < 0.05$), esta es la enfermedad comórbida con mayor riesgo según el *odds ratio* de 14.06, también señalaron que al estudiar 191 pacientes hospitalizados con COVID-19 en Wuhan, 48 % tenía comorbilidades, (67 % de ellos murieron), 30 % de los pacientes tenían hipertensión (48% de ellos murieron), 19 % de los pacientes tenían diabetes (31 % de ellos murieron), por lo anterior al modelo se introdujeron las variables esperanza de vida saludable al nacer y una variable utilizada por la organización mundial de la salud donde se conjuntan estas enfermedades (probabilidad de morir de enfermedades); estas se explican en el Cuadro 3.

Cuadro 3. Eje enfermedades

Nombre de la variable	Unidades	Notas
Esperanza de vida saludable al nacer	Años	-
Probabilidad de morir por enfermedades	Probabilidad	Probabilidad de morir por enfermedades: cardiovasculares, cáncer, diabetes, respiratorias crónicas dentro, de los 30 y los 70 años

4. Sector económico

El eje económico de cada país forma un ángulo clave en el cuidado de la propagación y mortalidad del virus, tanto como los recursos que pueda cada país invertir en frenar la pandemia como los recursos propios de la ciudadanía para el cuidado y posibilidad de frenar su movilidad, es esta última cuestión la que mayormente afecta a la propagación del virus, la riqueza de las naciones y la primera ola de COVID-19, indicaron que el movimiento de los ciudadanos en la frontera y el servicio local de transporte afectaban más a la propagación del virus que el movimiento de personas por sistemas aéreos (Antionietti *et al.*, 2021) por lo que la posibilidad de las personas a mantenerse en aislamiento el mayor tiempo posible generaría una menor propagación, la variable de interés de este eje fue aquellas personas con un ingreso menor a $3.20 USD al día, este eje se presenta en el Cuadro 4.

Cuadro 4. Eje sector económico

Nombre de la variable	Unidades	Notas
Porcentaje población con un salario menor a $3.20 por día	Porcentaje	Porcentaje de la población con un ingreso menor a $3.20 dólares

La base de datos quedó conformada por 51 países y con las siguientes variables y unidades:

- Población total en miles de personas.
- Población de 0 a 14 años en miles de personas.
- Población de 70 años o más en miles de personas.
- Población de 80 años o más miles de personas.
- Densidad de población en personas por km^2.
- Ciudades con más de 1 millón de habitantes en ciudades.
- Esperanza de vida saludable al nacer en años.
- Probabilidad de morir por enfermedades.
- Médicos por 10,000 habitantes en densidad.
- Gasto per cápita en salud en dólares por persona.
- Porcentaje de gasto del país en salud en porcentaje.
- Camas por 10,000 habitantes en densidad.
- Porcentaje población con un salario menor a $3.20 USD por día en porcentaje.

Las unidades fueron homologadas para los 51 países y se encuentran con sus valores individuales en el Anexo I.

Análisis estadístico

Para el cálculo de los estadísticos básicos (media, desviación estándar, coeficiente de variación, máximos y mínimos) se utilizó el software Excel y Power BI para los apoyos visuales en tablas y gráficas, en el análisis de datos atípicos, errores de medida, variables no comparables y errores de especificación, se le dio un enfoque especial en la verificación en no tener la inclusión de variables irrelevantes o bien en variables que estuvieran dentro de otras variables.

Análisis de correlación

Se realizó un análisis de correlación en el Software SAS ®, dado que en su totalidad las variables son cuantitativas y continuas; se utilizó la prueba de correlación de Pearson (r_{xy}), oscila de -1 a 1, siendo uno el máximo grado de correlación en cualquiera de sus signos y cero el mínimo grado de correlación, el ejercicio se realizó para la combinación de todas las variables, con un valor de significancia del 5 %, los resultados de las correlaciones se compararon entre sí con un gráfico de cuerdas creado en el software Power BI, bajo el nombre del apoyo visual: Chord, este presenta de manera gráfica las relaciones a escala según su valor de correlación.

Análisis multivariado con componentes principales

Los análisis se realizaron en el software SAS ® con la función PRINCOMP de manera iterativa buscando una mejor explicación de la varianza en cada cambio, se comenzó con las 13 variables expuestas en el anexo I, se continuó con las iteraciones siguiendo los resultados de exclusión de variables analizados

en el apartado de datos atípicos y selección de variables. Dada la variabilidad del estudio en cuestión se tuvo como objetivo probar por lo menos el 60 % de la variabilidad y dado que los valores originales de las variables no fueron estandarizados y las escalas originales son bastante diferentes entre sí, el proceso de generación de los componentes principales expuesto en la revisión de literatura fue realizado sustituyendo la matriz de covarianzas S por la siguiente matriz de correlaciones muestral R.

$$R = \begin{pmatrix} 1 & -0.238 & -0.190 & -0.263 & -0.229 & 0.326 & -0.085 & 0.243 \\ -0.238 & 1 & 0.629 & 0.674 & 0.731 & -0.825 & 0.509 & -0.873 \\ -0.190 & 0.629 & 1 & 0.812 & 0.742 & -0.611 & 0.399 & -0.478 \\ -0.263 & 0.674 & 0.812 & 1 & 0.660 & -0.647 & 0.414 & -0.592 \\ -0.229 & 0.731 & 0.742 & 0.660 & 1 & -0.581 & 0.543 & -0.668 \\ 0.326 & -0.825 & -0.611 & -0.647 & -0.581 & 1 & -0.358 & 0.584 \\ -0.085 & 0.509 & 0.399 & 0.414 & 0.543 & -0.358 & 1 & -0.474 \\ 0.243 & -0.873 & -0.478 & -0.592 & -0.668 & 0.584 & -0.474 & 1 \end{pmatrix}$$

estas dos matrices se relacionan con la siguiente expresión

$$R = D^{-\frac{1}{2}} S D^{-\frac{1}{2}}$$

Donde $D^{-\frac{1}{2}} = \text{Diag}\left({}^{1}/_{s_i}\right)$

El modelo para el índice fue el siguiente:

$$Y_1 = \gamma_{11}X_{i1} + \gamma_{12}X_{i2} + \gamma_{13}X_{i3} + \gamma_{14}X_{i4} + \gamma_{15}X_{i5} + \gamma_{16}X_{i6} + \gamma_{17}X_{i7} + \gamma_{18}X_{i8}$$

Donde:

Y_1 Indicador de la vulnerabilidad de un país contra el COVID-19

γ_{1j} Vectores propios en Y_1 de X_{ij}

X_{i1} Población total

X_{i2} Esperanza de vida saludable al nacer

X_{i3} Gasto per cápita en salud

X_{i4} Porcentaje del PIB en gasto de salud

X_{i5} Número de médicos por 10,000 habitantes

X_{i6} Probabilidad de morir de enfermedades cardiovasculares, cáncer, diabetes, enfermedades crónicas respiratorias.

X_{i7} Camas de hospital por 10,000 habitantes

X_{i8} Porcentaje de la población con salario menor a $3.20 USD por día

Para $j = 1, \dots, 8$ y para $i = 1, \dots, 51$

Cabe señalar que, para el cálculo de cada uno de los índices, las variables individuales fueron normalizadas, restando la media de la variable, dividiéndola por la desviación estándar y estos fueron realizados con el componente principal uno obtenido.

Análisis de conglomerados

Se realizó un estudio de conglomerados en el software SAS ® con el comando CLUSTER y con el método de Ward descrito en la revisión de literatura,

con esto se encontraron los países con alta homogeneidad dentro del grupo y heterogeneidad entre grupos, a su vez se identificaron los países más parecidos y el país más contrastante, para visualizar de manera gráfica lo anterior se procedió con el software MiniTab®. Para la decisión del número de grupos óptimo se utilizó el método cuantitativo pseudo estadístico de t cuadrado y el método de análisis gráfico en el software Minitab ® con medición de distancia de Pearson.

Análisis discriminante y error de clasificación

Con la finalidad de clasificar a otros países que estén fuera de este estudio y tener una variable cuantitativa clasificatoria, se realizó un estudio discriminante. Se tomó como base el índice de vulnerabilidad propuesto en este ensayo, el criterio clasificatorio se dividió en 4 ejes, en función de las posibilidades que tiene cada país de manejar la pandemia. Con las variables propuestas y con el índice de vulnerabilidad generado se creó un método de clasificación direccionado en las posibilidades que tiene cada país de combatir la pandemia; el índice de vulnerabilidad (IV) se describe en el Cuadro 5.

Cuadro 5. Criterio de corte

Rango de índice de vulnerabilidad	Nivel de manejo	Abreviación
1.5 < IV < 3.5	Buen manejo	BM
0 < IV ≤ 1.5	Manejo medio	MM
-1.5 < IV ≤ 0	Problemas de manejo	PM
-4.5 < IV ≤ -1.5	Extremo manejo	EM

Este análisis se realizó en el software SAS ® con el comando DISCRIM, con probabilidades *a priori* iguales para todos los niveles de manejo, en este caso 0.25 cada uno con el comando PRIORS EQUAL, el método utilizado fue con el comando METHOD=NORMAL, el cual con un método paramétrico crea una función discriminante lineal para cada nivel propuesto asumiendo una distribución normal multivariable, también se especificó el comando POOL=YES para utilizar la matriz de covarianzas en el cálculo de estas funciones discriminantes, esté análisis se probó con el estadístico Wilks´Lamda λ con las siguientes pruebas de hipótesis:

H_0: No existe diferencias multivariadas en general entre grupos $\lambda \geq 0.05$

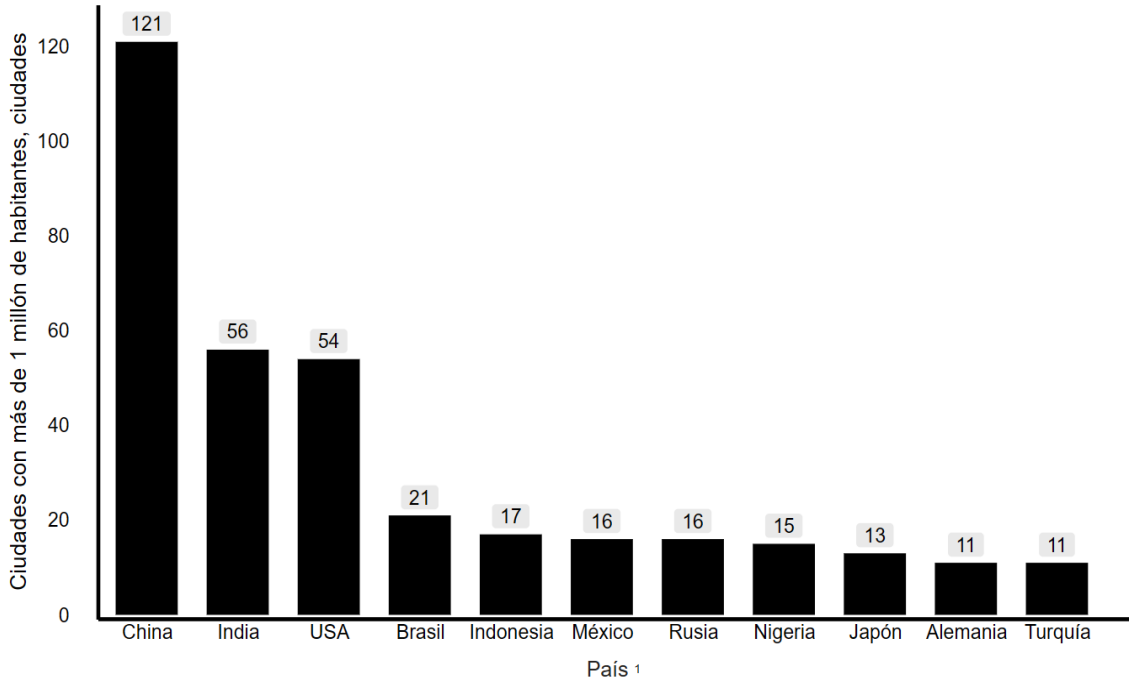
H_1: Existe diferencia en por lo menos una de las variables explicativas $\lambda < 0.05$

Se evaluó el error de clasificación de las funciones con el software SAS® por el método de restitución previsto en la revisión de literatura bajo el comando de CROSSLISTER.

RESULTADOS Y DISCUSIÓN

En el análisis de datos atípicos se concluyó que la variable de ciudades con más de 1 millón de habitantes no era confiable, ya que China tiene 121 ciudades con esta característica, lo cual es muy cercano al doble que el país inmediato anterior, esto se puede ver de mejor manera en la Gráfica 1. En el análisis de error de especificación se concluyó que la densidad de población tampoco era una medida confiable para el modelo, tomando como ejemplo Brasil, Chile y Perú estos tienen una diferencia de densidad de menos de una persona por kilómetro cuadrado, pero, el tamaño de los países es considerablemente diferente entre sí, esto se puede apreciar en el Cuadro 6. Países con ciudades densamente pobladas como Brasil y China no resultan con una densidad de población alta dado su tamaño total en kilómetros cuadrados.

En el análisis de error de especificación, las variables de población de 0 a 14 años, población de 70 o más, población de 80 o más y esperanza de vida saludable al nacer, se observa que las variables de 70 y 80 se abarcan entre sí, es decir, una persona que está en la categoría de 80 y más eventualmente también está en 70 y más, también cabe señalar que la variable de esperanza de vida abarca a las otras 3 y suma información más relevante al modelo en el análisis de vulnerabilidad.



Gráfica 1. Cantidad de ciudades de cada país con más de 1 millón de habitantes

[1] se seleccionaron los 11 países con mayor número de ciudades para su visualización

Cuadro 6. Análisis de densidad

País	Densidad de población	Tamaño
Brasil	25.4	8,515,770 km^2
Perú	25.8	1,285,215 km^2
Chile	25.7	756,102 km^2
Suecia	24.6	450,295 km^2

Como resultado del análisis de datos atípicos y selección de variables, se elimina del modelo la variable de ciudades con más de un millón de habitantes y continuando con el principio de parsimonia el cual prioriza las explicaciones más sencillas, se eliminó del modelo las variables de densidad de población, población de 0 a 14 años, población con más de 70 años y población con más de 80 años.

En el análisis de estadísticas básicas (Cuadro 7) se observó que la variable con más variación es población total, seguido de porcentaje de la población con un ingreso menor a $3.20 dólares por día, estas dos variables se analizaron de manera conjunta en un mapa, el cual se puede visualizar de una mejor manera en la Figura 2. En este mapa, el tamaño del círculo representa la variable X_1 (población total en miles de habitantes), mientras más grande sea el círculo mayor será la población total en comparación con el resto los países estudiados, de aquí que los países de China y la India tienen el círculo más grande, también, en escala de grises se tiene la variable X_8 (porcentaje de la población con un salario menor a $3.20 dólares diarios), entre más oscuro se encuentre el círculo mayor será su porcentaje de personas viviendo con este ingreso, aquí fueron notorios países de África central como: El Congo, Nigeria y Tanzania con porcentaje de 91, 77.6 y 76.6 % respectivamente con menos de $3.20 dólares diarios, se observó también que en La India se tiene el conjunto de estos 2 problemas, es decir, se tiene un problema económico y un alto número de habitantes, ambos aspectos negativos.

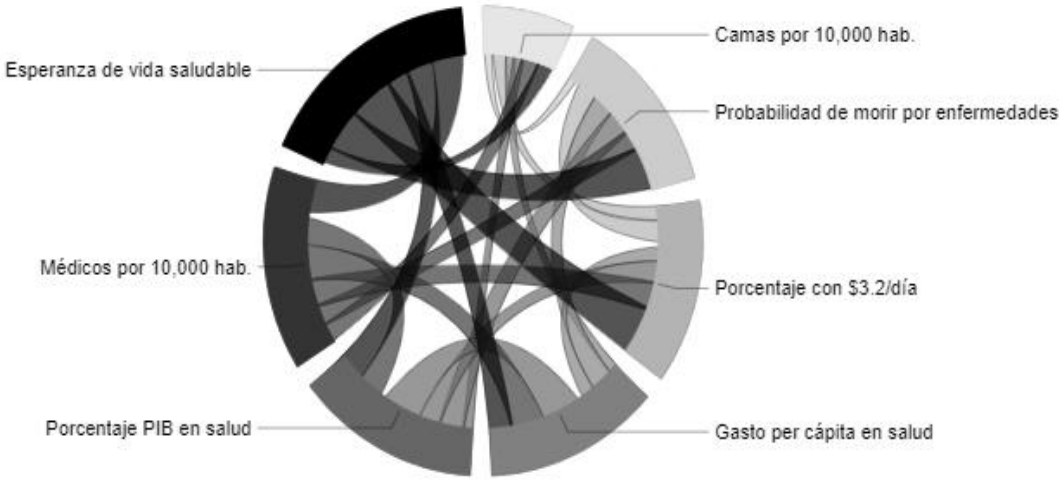
Cuadro 7. Estadísticas básicas[1]

Variable	Promedio	Desviación estándar	Coeficiente de variación	Mínimo	Máximo
Población total	121325	273397	225.3%	4315	1,439,324
Porcentaje con $3.2/día	13.96	23.53	168.6%	0.1	91
Gasto per cápita en salud	2245	2561	114.1%	21	9870
Camas por 10,000 hab.	32.05	27.71	86.5%	3.12	134
Médicos por 10,000 hab.	22.7	14.68	64.7%	0.4	54
Porcentaje PIB en salud	7.663	3.05	39.8%	2.4	17.1
Probabilidad de morir por enfermedades	14.947	5.683	38.0%	7.8	27.7
Esperanza de vida saludable	67.437	6.088	9.0%	48.9	74.8

[1] Ordenados de mayor a menor según el coeficiente de variación de cada variable.

Figura 2. Mapa relación ingreso y población

En el análisis de correlación en 7 de 8 variables se rechazó la hipótesis nula ($P < 0.05$) concluyendo que las variables están correlacionadas, en el caso de la variable X_1, únicamente fue significativa con X_6 (probabilidad de morir por enfermedades), sin embargo, dadas las circunstancias que el virus COVID-19 arrojó en la revisión de literatura, háblese de densidad poblacional, ciudades grandes y las implicaciones de propagación que esto conlleva, se realizó un gráfico de cuerdas general (Gráfica 2); donde la variable con mayor relación es la esperanza de vida saludable al nacer (Gráfica 3) y por el contrario la variable con menor relación es camas de hospital por 10,000 habitantes (Gráfica 4), se observó que las dos variables que se encuentran con mayor correlación positiva son: X_3 (Gasto per cápita en salud) y X_4 (Porcentaje del PIB en gasto de salud) con un coeficiente de correlación de $r = 0.812$ la tabla con los resultados individuales se aprecia en el Cuadro 8.



Gráfica 2. Gráfico de cuerdas de correlación

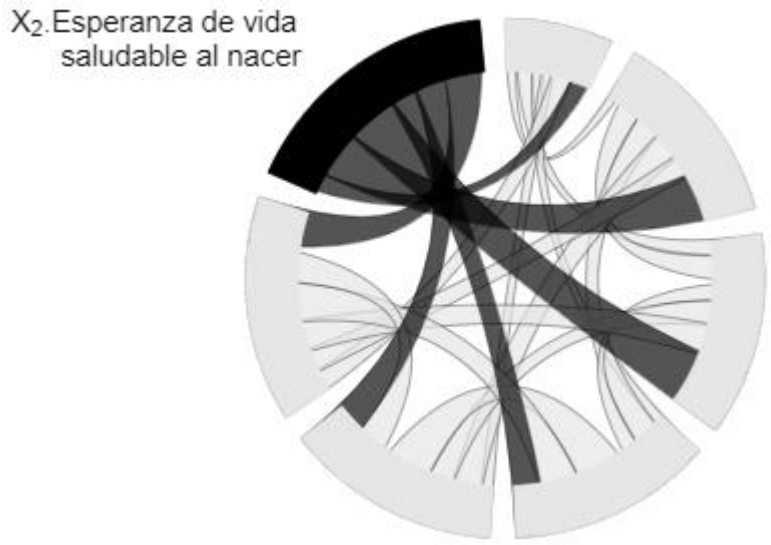


Gráfico 3. Correlación de esperanza de vida saludable al nacer

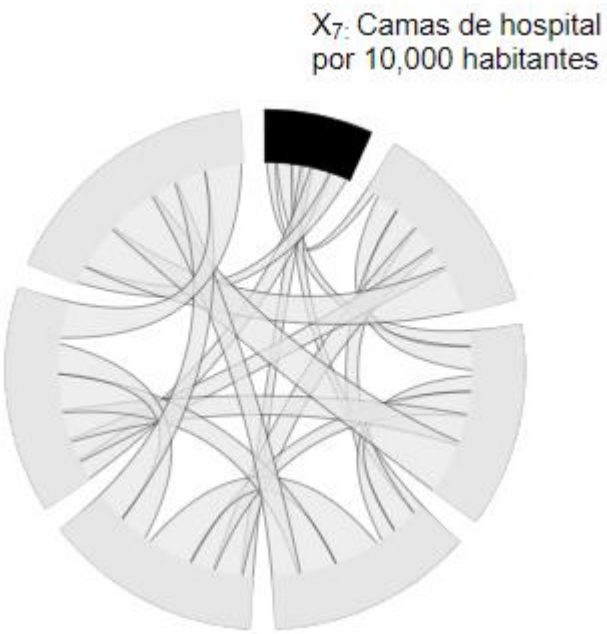


Gráfico 4. Correlación de camas de hospital por 10,000 habitantes

Cuadro 8. Correlación de Pearson [1] *

		X_1	X_2	X_3	X_4	X_5	X_6	X_7
X_2	Pearson	-0.23						
	p-value	0.09						
X_3	Pearson	-0.19	0.62					
	p-value	0.18	0					
X_4	Pearson	-0.26	0.67	0.81				
	p-value	0.06	0	0				
X_5	Pearson	-0.22	0.73	0.74	0.66			
	p-value	0.10	0	0	0			
X_6	Pearson	0.32	-0.82	-0.61	-0.64	-0.58		
	p-value	0.02	0	0	0	0		
X_7	Pearson	-0.08	0.50	0.39	0.41	0.54	-0.35	
	p-value	0.55	0	0	0	0	0.01	
X_8	Pearson	0.24	-0.87	-0.47	-0.59	-0.66	0.58	-0.47
	p-value	0.08	0	0	0	0	0	0

[1] Donde:

X_1. Población total en miles de habitantes

X_2. Esperanza de vida saludable al nacer en años

X_3. Gasto per cápita en salud

X_4. Porcentaje del PIB en gasto de salud

X_5. Número de médicos por 10,000 habitantes

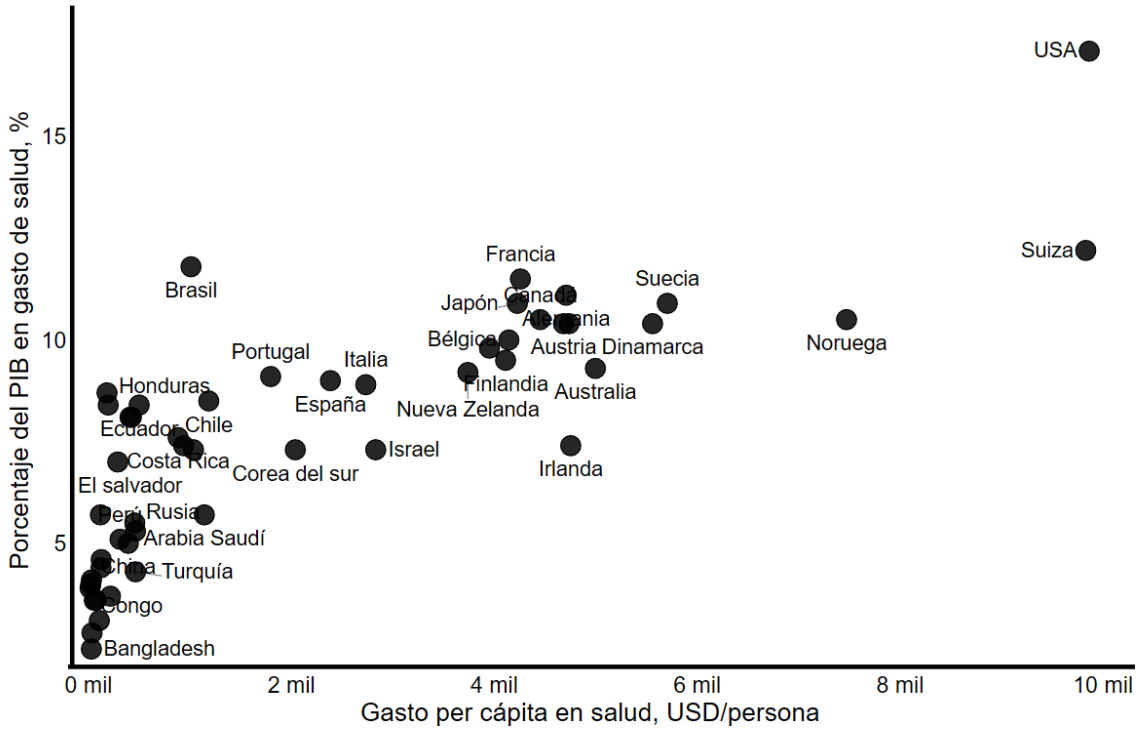
X_6. Probabilidad de morir de enfermedades cardiovasculares, cáncer, diabetes, enfermedades crónicas respiratorias.

X_7. Camas de hospital por 10,000 habitantes

X_8. Porcentaje de la población con un salario menor a $3.20 dólares por día

* ($P < 0.05$)

Se observó que países con una población alta, no pueden tener un buen gasto per cápita en salud aún y cuando su porcentaje de producto interno bruto se designe para ello (Gráfica 5), como ejemplo se tiene a Brasil y Suiza, ambos se encuentran en la misma región del eje Y (que es el porcentaje del PIB destinado a salud), sin embargo, al momento de seguir con el eje X (Gasto per cápita en salud) es apreciable como existe un amplio margen de diferencia entre ambos países, esto hace que aún y cuando el país destine un buen porcentaje en su esfuerzo por tener un sistema de salud robusto, este se puede ver rebasado por su tamaño de población, en el Cuadro 9 se muestra como la diferencia de porcentaje entre estos dos países es de sólo .4 % pero la diferencia en gasto per cápita dada la desigualdad de población es de más de 9 veces.



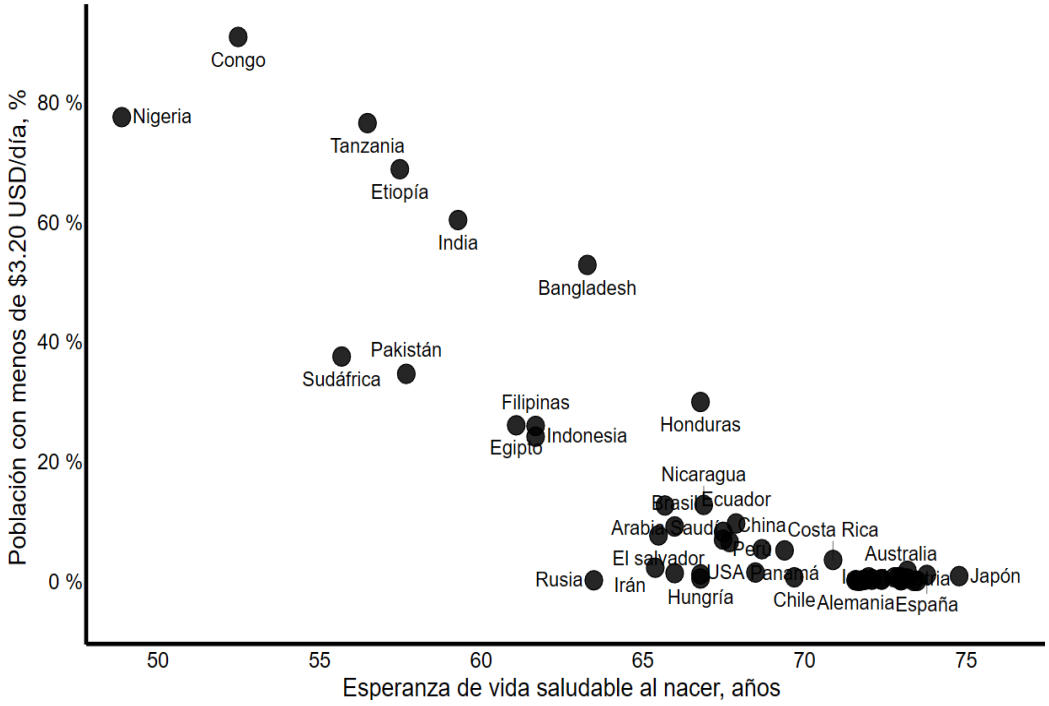
Gráfica 5. Correlación positiva en X_3: Gasto per cápita en salud y X_4: Porcentaje del PIB en gasto de salud

Cuadro 9. Comparación de Brasil y Suiza

País	X_3: Gasto per cápita en salud, USD$	X_4: Porcentaje del PIB en gasto de salud	Población total, Millones
Suiza	9,836	12.2	8,655
Brasil	1,016	11.8	212,559

Se analizaron las dos variables más correlacionadas de manera negativa (Gráfica 6), X_2 (Esperanza de vida saludable al nacer en años) y X_8 (Porcentaje de la población con un salario menor a $3.20 dólares por día), se observó como los países que tienden a tener un mejor ingreso, es decir menos porcentaje de personas viviendo con menos de $3.20 dólares por día (eje Y) tienden a una esperanza de vida mayor, lo cual demostró que a una mejor estabilidad económica, una mejor calidad de vida, esto es notorio en la esquina inferior derecha de la gráfica, donde países como Japón, España, Alemania, Francia e Italia comparten una esperanza de vida alta y un bajo porcentaje de personas en pobreza.

En el análisis multivariado por componentes principales se obtuvieron los eigenvectores mostrados en el Cuadro 10; donde, los eigenvectores del componente principal 1 (Cuadro 11), las variables fueron agrupadas con el mismo signo dependiendo su interacción con la enfermedad, por ejemplo, la esperanza de vida y los médicos por 10,000 habitantes están con el mismo signo (positivo), por el contrario, la probabilidad de morir por enfermedades y el porcentaje de población con menos de $3.20 dólares por día se encuentran con el mismo signo (negativo), es decir, los dos grupos son correctos con las variables que lo integran, esto concuerda con lo fundamentado por Peña (2002) donde afirma que los componentes tienden a contraponerse en dos grupos con distinto signo y se pueden asignar como factores de forma, dentro de la explicación de la varianza se obtuvo un 60 % con el componente principal uno (Cuadro 12), lo cual fue aceptable para el trabajo en cuestión.



Gráfica 6 Correlación de X_2 y X_8

Cuadro 10. Eigenvectores por cada componente principal

Componente principal	CP1	CP2	CP3	CP4
Población total	-0.157	0.889	0.248	-0.331
Esperanza de vida saludable al nacer	0.420	0.035	-0.173	-0.393
Gasto per cápita en salud	0.373	0.067	0.545	0.273
Porcentaje de gasto del país en salud	0.385	-0.021	0.404	0.143
Médicos por 10,000 habitantes	0.393	0.127	0.034	0.192
Probabilidad de morir por enfermedades	-0.374	0.180	-0.085	0.354
Camas por 10,000 habitantes	0.279	0.390	-0.549	0.571
Porcentaje población con un salario menor a $3.20 por día	-0.374	-0.046	0.372	0.390

Componente principal	CP5	CP6	CP7	CP8
Población total	-0.070	0.003	0.070	-0.057
Esperanza de vida saludable al nacer	-0.056	0.041	-0.173	0.777
Gasto per cápita en salud	0.052	0.086	-0.682	-0.097
Porcentaje de gasto del país en salud	0.004	-0.644	0.493	0.100
Médicos por 10,000 habitantes	0.432	0.637	0.447	-0.017
Probabilidad de morir por enfermedades	0.679	-0.254	-0.148	0.383
Camas por 10,000 habitantes	-0.353	-0.115	-0.056	-0.033
Porcentaje población con un salario menor a $3.20 por día	-0.465	0.305	0.180	0.475

Cuadro 11. Eigenvectores componente principal 1

Nombre de la variable	Eigenvectores[1]
Esperanza de vida saludable al nacer	0.419749
Médicos por 10,000 habitantes	0.392626
Porcentaje de gasto del país en salud	0.385355
Gasto per cápita en salud	0.37259
Camas por 10,000 habitantes	0.278684
Población total	-0.157357
Probabilidad de morir por enfermedades	-0.373868
Porcentaje población con un salario menor a $3.20 por día	-0.374341

[1] Las variables fueron ordenadas de mayor a menor según el valor del eigenvetor

Cuadro 12. Explicación de la varianza por cada eigenvalor

Componente principal	Valores propios	Proporción de la varianza	Varianza acumulada
1	4.80	60.0%	60.0%
2	0.97	12.1%	72.1%
3	0.73	9.1%	81.2%
4	0.63	7.9%	89.1%
5	0.40	5.0%	94.1%
6	0.29	3.7%	97.8%
7	0.13	1.7%	99.4%
8	0.05	0.6%	100.0%

Los indicadores de la vulnerabilidad de un país contra el COVID-19 fueron calculados utilizando el componente principal 1 y el modelo para el cálculo fue el siguiente.

$$Y_1 = -0.1573X_{i1} + .4197X_{i2} + 0.3725X_{i3} + 0.3853X_{i4} + 0.3926X_{i5} + -0.3738X_{i6} + 0.2786X_{i7} + -0.3743X_{i8}$$

Donde:

Y_1 Indicador de la vulnerabilidad de un país contra el COVID-19

X_{i1} Población total

X_{i2} Esperanza de vida saludable al nacer

X_{i3} Gasto per cápita en salud

X_{i4} Porcentaje del PIB en gasto de salud

X_{i5} Número de médicos por 10,000 habitantes

X_{i6} Probabilidad de morir de enfermedades cardiovasculares, cáncer, diabetes, enfermedades crónicas respiratorias.

X_{i7} Camas de hospital por 10,000 habitantes

X_{i8} Porcentaje de la población con salario menor a $3.20 USD por día

Para $i = 1, \ldots, 51$

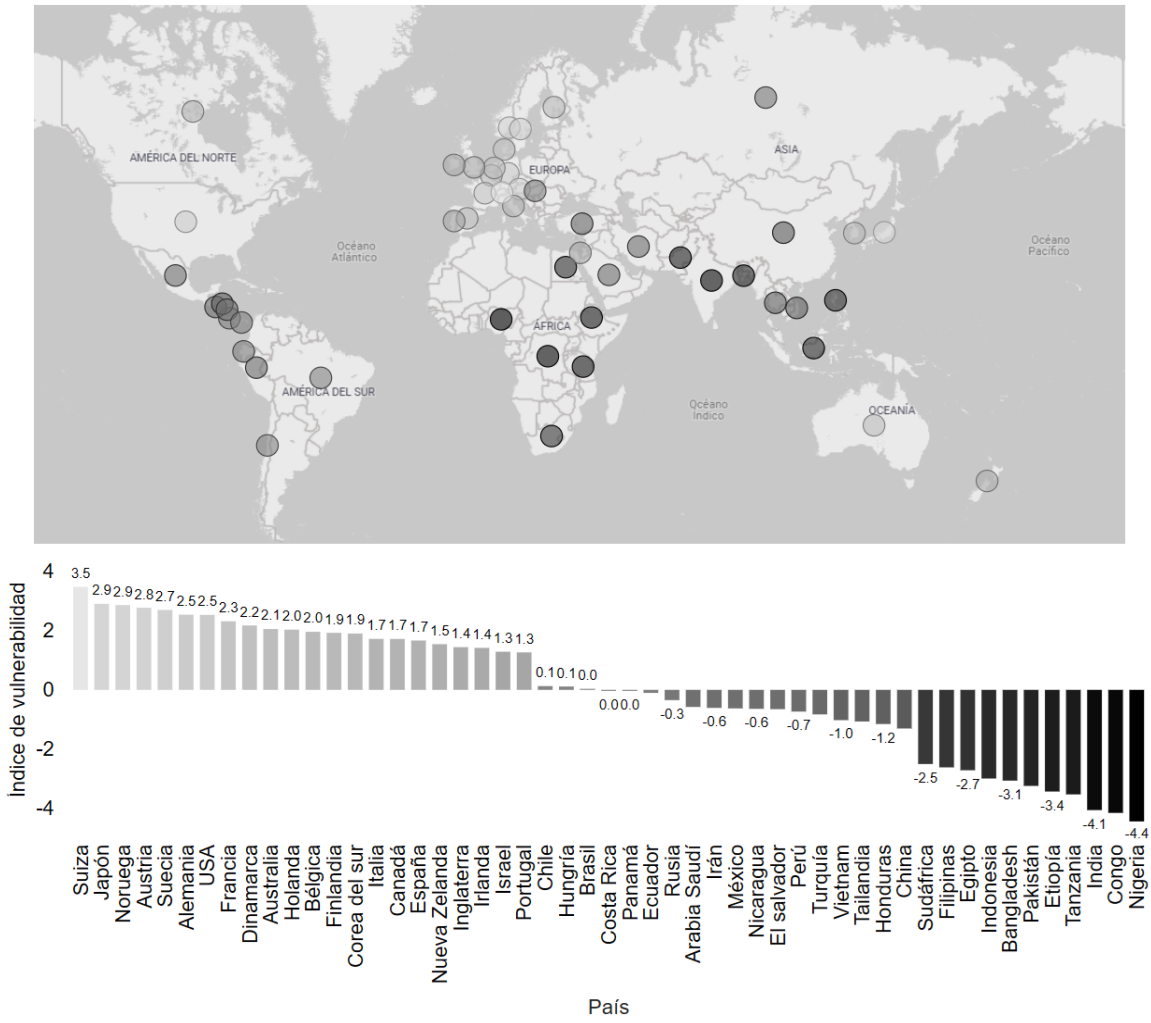
El total de los indicadores se muestra a continuación en el Cuadro 13.

Cuadro 13. índices principales

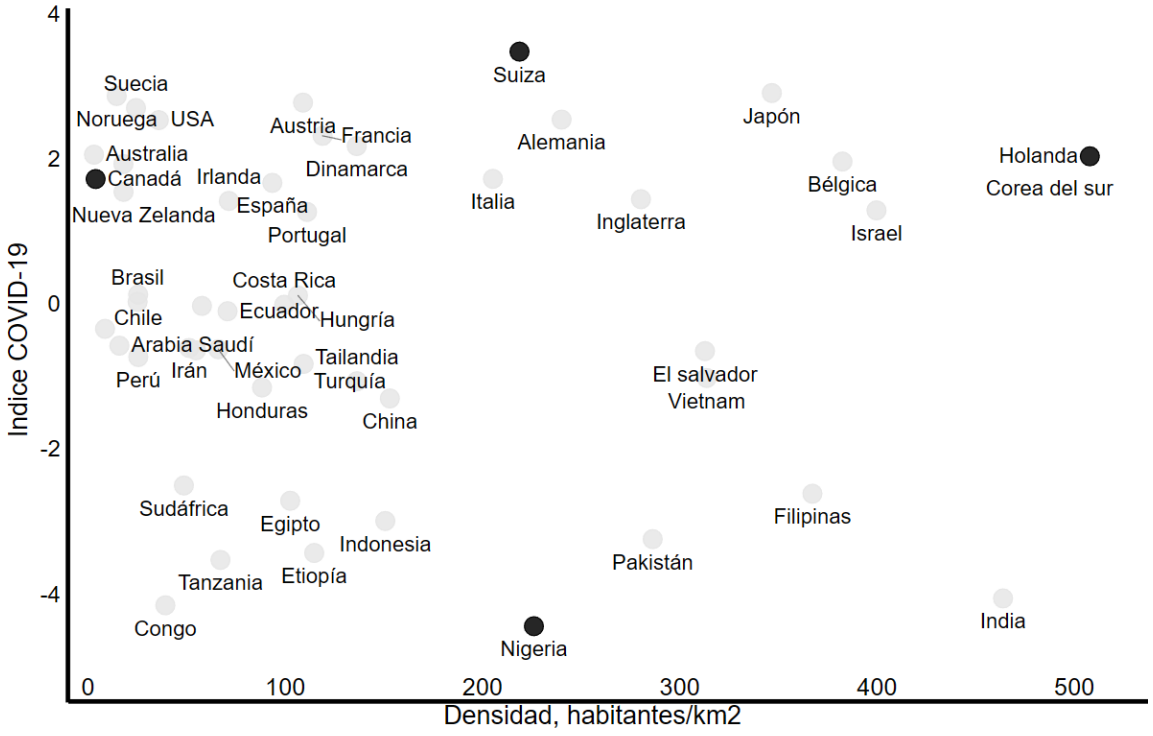
País	Índice COVID-19 [1]	País	Índice COVID-19
Suiza	3.47381	Panamá	-0.02927
Japón	2.90374	Ecuador	-0.10296
Noruega	2.86251	Rusia	-0.34588
Austria	2.77164	Arabia Saudí	-0.57712
Suecia	2.69676	Irán	-0.60864
Alemania	2.53739	México	-0.62632
USA	2.53039	Nicaragua	-0.64054
Francia	2.31424	El salvador	-0.65209
Dinamarca	2.17299	Perú	-0.73386
Australia	2.05406	Turquía	-0.82853
Holanda	2.03227	Vietnam	-1.02238
Bélgica	1.95962	Tailandia	-1.0676
Finlandia	1.92475	Honduras	-1.15326
Corea del sur	1.90142	China	-1.30411
Italia	1.72064	Sudáfrica	-2.50357
Canadá	1.71869	Filipinas	-2.6171
España	1.66599	Egipto	-2.71403
Nueva Zelanda	1.54153	Indonesia	-2.99262
Inglaterra	1.44086	Bangladesh	-3.06432
Irlanda	1.41693	Pakistán	-3.24251
Israel	1.28688	Etiopía	-3.43296
Portugal	1.26525	Tanzania	-3.52946
Chile	0.12628	India	-4.0592
Hungría	0.11313	Congo	-4.15402
Brasil	0.02609	Nigeria	-4.44217
Costa Rica	-0.01334		

[1] Los países fueron ordenados de mayor a menor según su indicador de vulnerabilidad

Se analizó el índice de vulnerabilidad por área geográfica distinguiendo 3 áreas principales (Gráfica 7), por un lado, los países europeos son los mejores preparados para el manejo de la pandemia, siguiendo con los países de medio oriente y américa latina, los países africanos y asiáticos tienen el peor índice. Japón es el mejor preparado del continente asiático, USA es el mejor preparado de América y Suiza el mejor preparado de Europa, el país con el mayor riesgo expresado por el índice es Nigeria, este último en África. Se contrastaron algunas de las variables retiradas del modelo y el índice de vulnerabilidad propuesto, se apreció que existen países que si bien sus indicadores son muy parecidos, al graficar con la segunda variable, en este caso Densidad (Gráfica 8), los diferencia en su totalidad, por ejemplo, Nigeria y Suiza tiene una densidad muy parecida, sin embargo, dadas las condiciones de cada país su índice es muy diferente, caso contrario con Canadá y Holanda, sus índices son muy parecidos y sin embargo, su densidad muy distinta, esto indica que hay países que aún y con un sistema de salud robusto pueden tener aspectos geográficos que disminuyan su capacidad de respuesta.



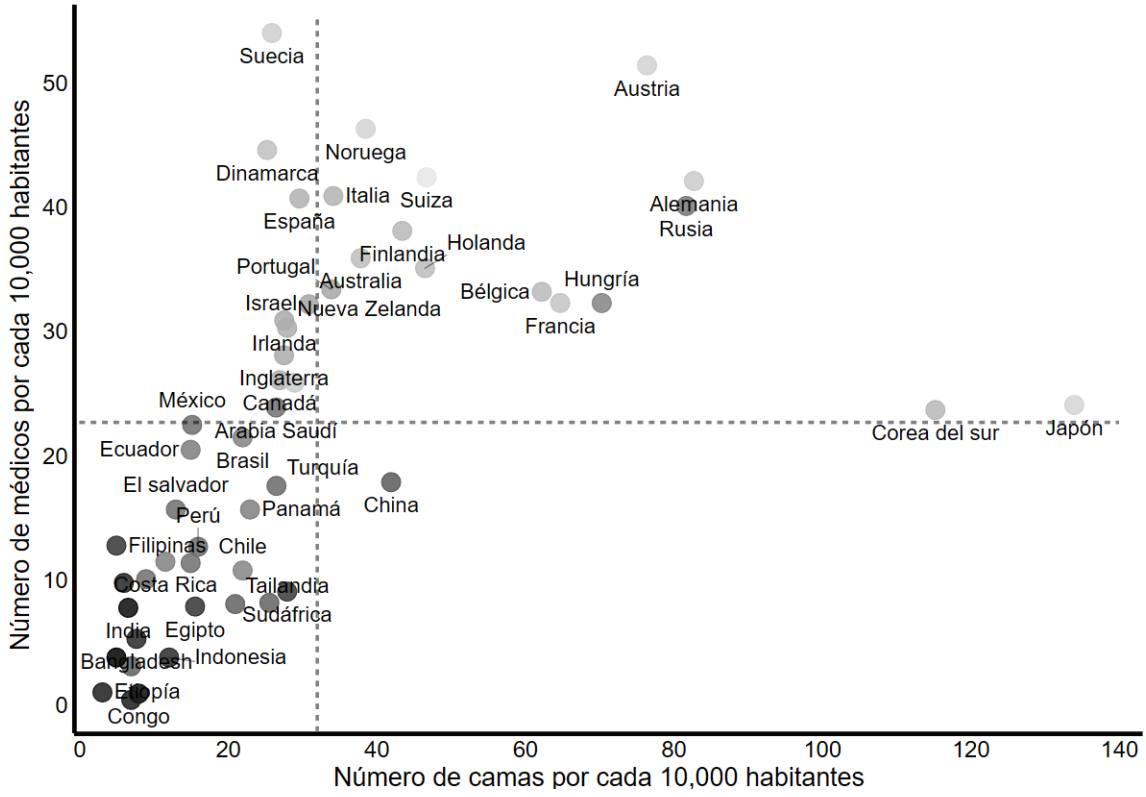
Gráfica 7. índice de vulnerabilidad por región



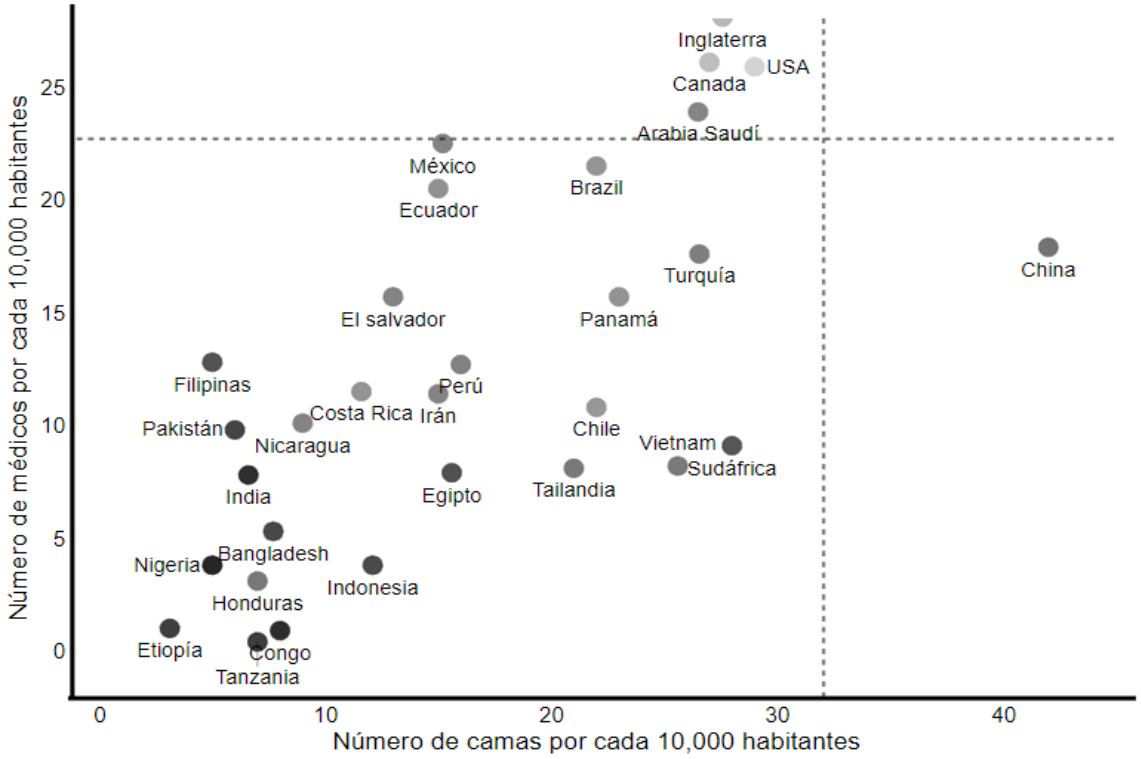
Gráfica 8. índice de vulnerabilidad y densidad

Se compararon las variables del sector salud y el indicador de vulnerabilidad propuesto, En la gráfica 9 se tiene en el eje Y el número de médicos por 10,000 habitantes con su línea promedio punteada, en el eje X se tiene el número camas por 10,000 habitantes también con su línea promedio punteada y por último el color de la burbuja representa el nivel de índice de vulnerabilidad yendo en escala de grises, siendo el más oscuro el menor índice.

Los países en el cuadrante inferior izquierdo tienden a ser de color más oscuro dado que sus 2 variables médicas son bajas. Difirieron de manera significativa Japón y Suecia quienes se encuentran muy por encima de alguna de las dos variables, pero cerca de la línea promedio de la variable contraria, esto resulta igual de peligroso al momento de controlar la pandemia ya que el paciente no contaría con médico o bien no contaría con cama de hospital, los países mejor preparados en este rubro fueron Austria, Rusia y Alemania, quienes tienen un balance entre las 3 variables. Analizando el cuadrante inferior izquierdo (Gráfica 10), se observa que los países de Etiopía, Tanzania, El Congo y Honduras tienen carencias en las 3 variables y son los más propensos a ser superados en capacidad médica.



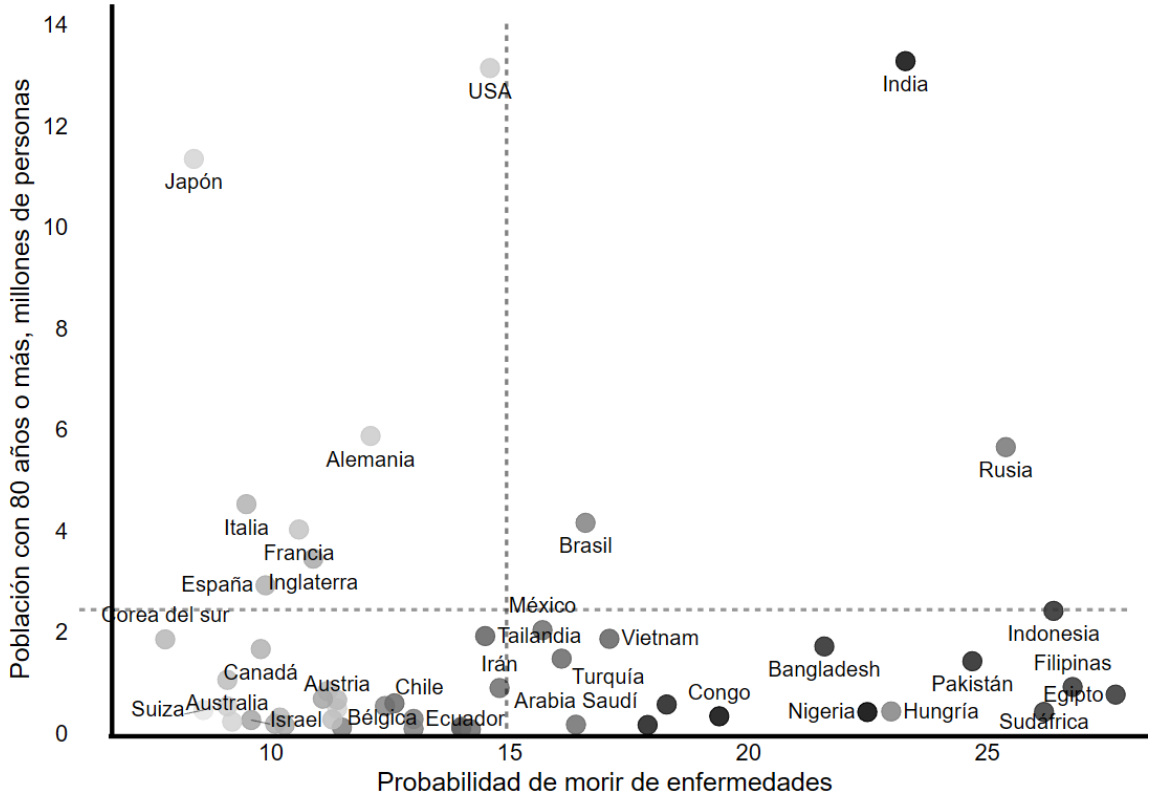
Gráfica 9. Camas y médicos por cada 10,000 habitantes



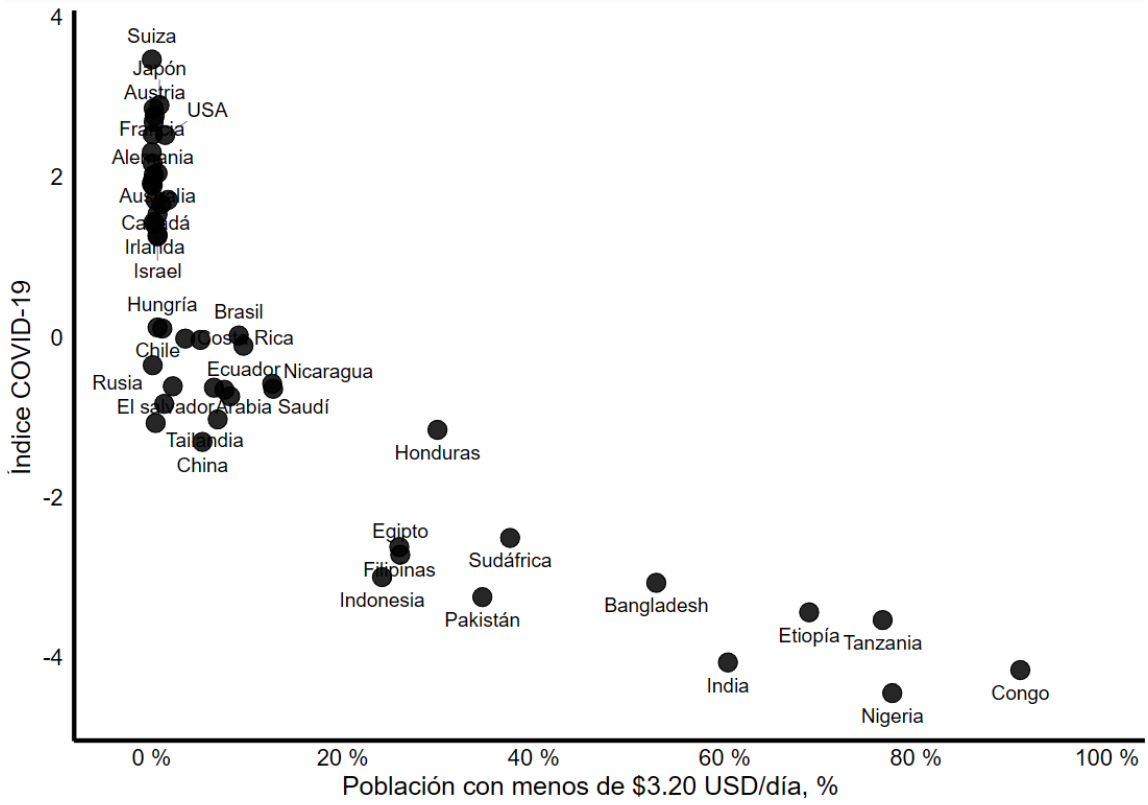
Gráfica 10. Camas y médicos por cada 10,000 habitantes cuadrante inferior izquierdo

Se analizaron las variables de probabilidad de morir de enfermedades y el porcentaje de población con más de 80 años, que en el caso de esta pandemia es el sector con más riesgo de presentar complicaciones, se agregaron las 2 líneas promedio de cada variable de forma punteada y el color de la burbuja representa el índice de vulnerabilidad COVID-19, entre más oscura sea la burbuja mayor es su vulnerabilidad a la pandemia (Gráfica 11), resaltaron los países de India y Rusia los cuales su probabilidad de morir por enfermedades está muy por encima de la media, su porcentaje con personas de más de 80 años no es bajo y su índice de vulnerabilidad no es bueno. En los resultados del sector económico se contrastaron las variables de porcentaje de población con menos de $3.20 dólares al día y el índice de vulnerabilidad propuesto (Gráfica 12), aquí los países de El Congo, Nigeria y Tanzania tendrían las mayores dificultades de permanecer en cuarentena ya que el porcentaje de población con un ingreso bajo es muy por encima de la media, por otro lado países como Suiza, Austria y Francia tienen un porcentaje de pobreza bajo lo cual teóricamente ayudaría a permanecer más tiempo en cuarentena.

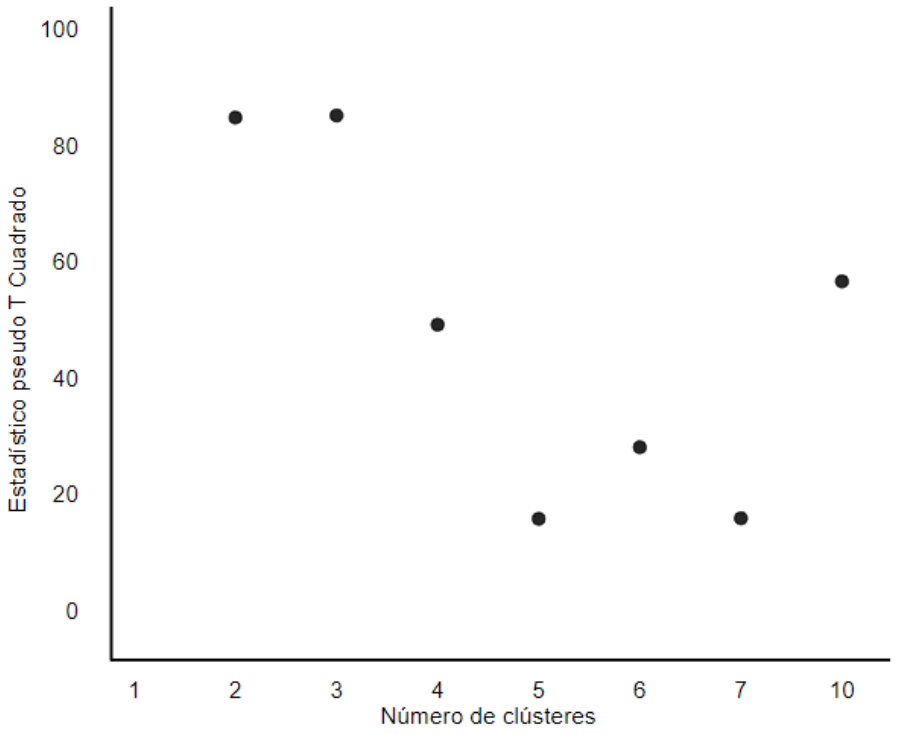
Para el análisis clúster, los resultados en el método de pseudo estadístico de t cuadrado para la formación de grupos arrojó que existen dos posibles resultados: 5 y 7 grupos (Gráfica 13), se realizó también un análisis gráfico del agrupamiento de los países en el software Minitab ® con medición de distancia de Pearson, con 7 grupos y método de vinculación completo (Gráfica 14), la tabla de los 51 países con su grupo correspondiente se encuentra en el Cuadro 14.



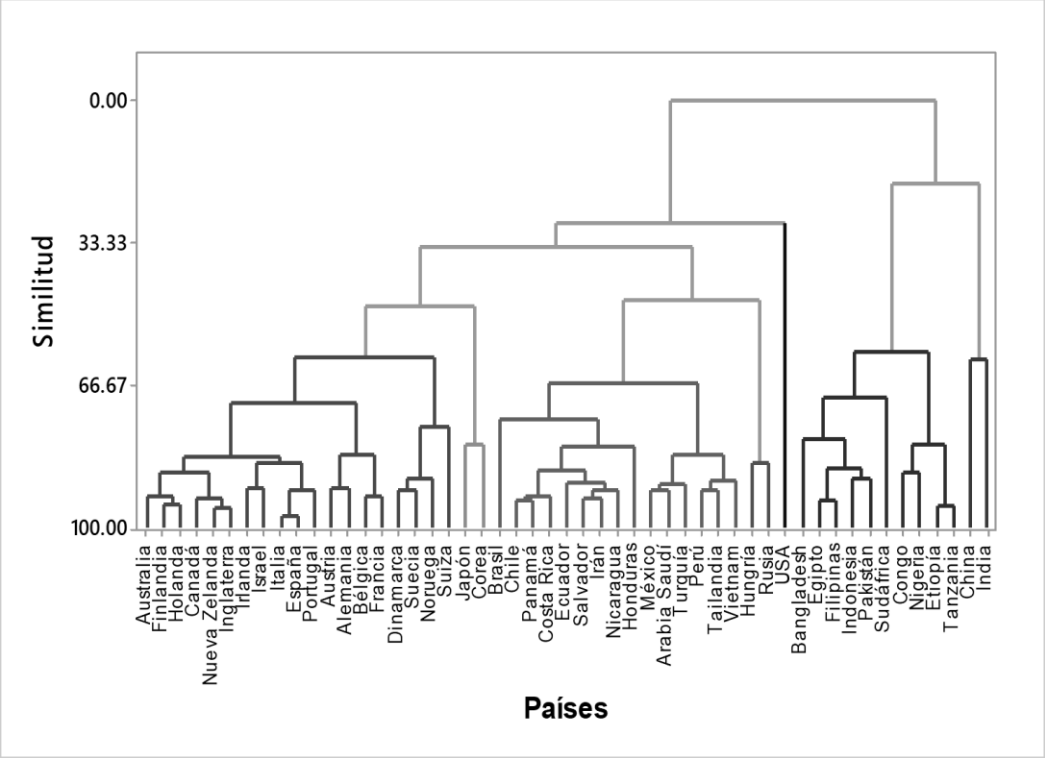
Gráfica 11. Probabilidad de morir por enfermedades y población con más de 80 años



Gráfica 12. Índice y población con menos de $3.20 USD/día



Gráfica 13. Pseudo estadístico t cuadrado.



Gráfica 14. Dendograma de similitud clúster

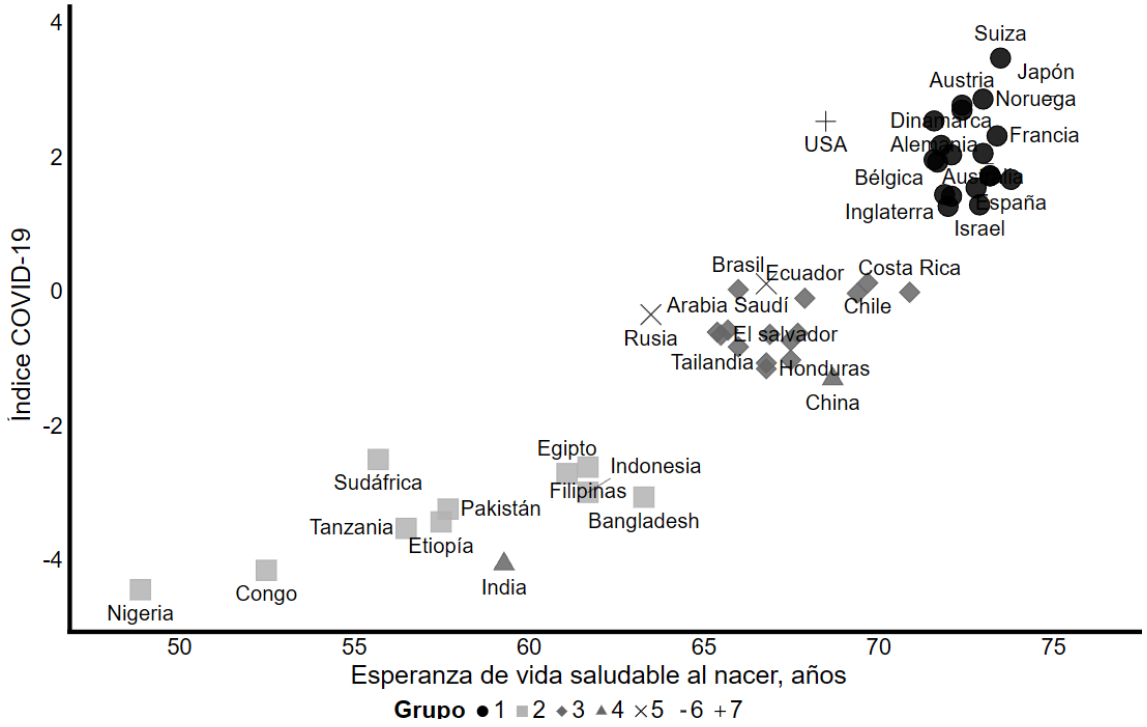
Cuadro 14. Grupos de países

País	Grupo	País	Grupo
Australia	1	Brasil	3
Austria	1	Chile	3
Bélgica	1	Costa Rica	3
Canadá	1	Ecuador	3
Dinamarca	1	El salvador	3
Finlandia	1	Honduras	3
Francia	1	Irán	3
Alemania	1	México	3
Irlanda	1	Nicaragua	3
Israel	1	Panamá	3
Italia	1	Perú	3
Holanda	1	Arabia Saudí	3
Nueva Zelanda	1	Tailandia	3
Noruega	1	Turquía	3
Portugal	1	Vietnam	3
España	1		
Suecia	1	China	4
Suiza	1	India	4
Inglaterra	1		
		Hungría	5
Bangladesh	2	Rusia	5
Congo	2		
Egipto	2	Japón	6
Etiopía	2	Corea del sur	6
Indonesia	2		
Nigeria	2	USA	7
Pakistán	2		
Filipinas	2		
Sudáfrica	2		
Tanzania	2		

Se observó que la totalidad de los países están dentro de un grupo con por lo menos un país en conjunto, con excepción de USA quien es un grupo por sí mismo, también se observa que los dos países más parecidos son El Salvador y Nicaragua y que los últimos 2 países en unirse fueron China y La India.

En la comparación del índice de vulnerabilidad y la esperanza de vida saludable al nacer, se pudo distinguir como los grupos propuestos se comportan de manera similar conforme avanzan las dos variables, los grupos suben casi de manera lineal y rara vez se separan, por el contrario existen dos países que si se separan dentro de su mismo grupo, estos son del grupo 4: India y China y del grupo 6: Japón y Corea del sur, esto indica que aún y cuando los grupos son similares entre sí, existen variables puntuales que crean diferencias entre los grupos (Gráfica 15). En el análisis discriminante se obtuvo que las 8 variables propuestas en el modelo agregan poder discriminativo, también se encontró evidencia estadística suficiente para afirmar que existen diferencias entre los grupos propuestos para el manejo de la pandemia (BM, MM, PM y EM) la clasificación final se encuentra en el Cuadro 15.

En el análisis de varianza multivariada (MANOVA) El software arrojó que efectivamente el estadístico λ es significativo por lo que sí existe diferencia en por lo menos una de las variables explicativas.



Gráfica 15. Esperanza de vida saludable e índice de vulnerabilidad

Cuadro 15. Países con criterio de corte.[1]

País	Índice COVID-19	Nivel de manejo	País	Índice COVID-19	Nivel de manejo
Suiza	3.47381	BM	Ecuador	-0.10296	PM
Japón	2.90374	BM	Rusia	-0.34588	PM
Noruega	2.86251	BM	Arabia Saudí	-0.57712	PM
Austria	2.77164	BM	Irán	-0.60864	PM
Suecia	2.69676	BM	México	-0.62632	PM
Alemania	2.53739	BM	Nicaragua	-0.64054	PM
USA	2.53039	BM	El salvador	-0.65209	PM
Francia	2.31424	BM	Perú	-0.73386	PM
Dinamarca	2.17299	BM	Turquía	-0.82853	PM
Australia	2.05406	BM	Vietnam	-1.02238	PM
Holanda	2.03227	BM	Tailandia	-1.0676	PM
Bélgica	1.95962	BM	Honduras	-1.15326	PM
Finlandia	1.92475	BM	China	-1.30411	PM
Corea del sur	1.90142	BM	Sudáfrica	-2.50357	EM
Italia	1.72064	BM	Filipinas	-2.6171	EM
Canadá	1.71869	BM	Egipto	-2.71403	EM
España	1.66599	BM	Indonesia	-2.99262	EM
Nueva Zelanda	1.54153	BM	Bangladesh	-3.06432	EM
Inglaterra	1.44086	MM	Pakistán	-3.24251	EM
Irlanda	1.41693	MM	Etiopía	-3.43296	EM
Israel	1.28688	MM	Tanzania	-3.52946	EM
Portugal	1.26525	MM	India	-4.0592	EM
Chile	0.12628	MM	Congo	-4.15402	EM
Hungría	0.11313	MM	Nigeria	-4.44217	EM
Brasil	0.02609	MM			
Costa Rica	-0.01334	MM			
Panamá	-0.02927	MM			

[1] Los países fueron ordenados mayor a menor según su índice de vulnerabilidad

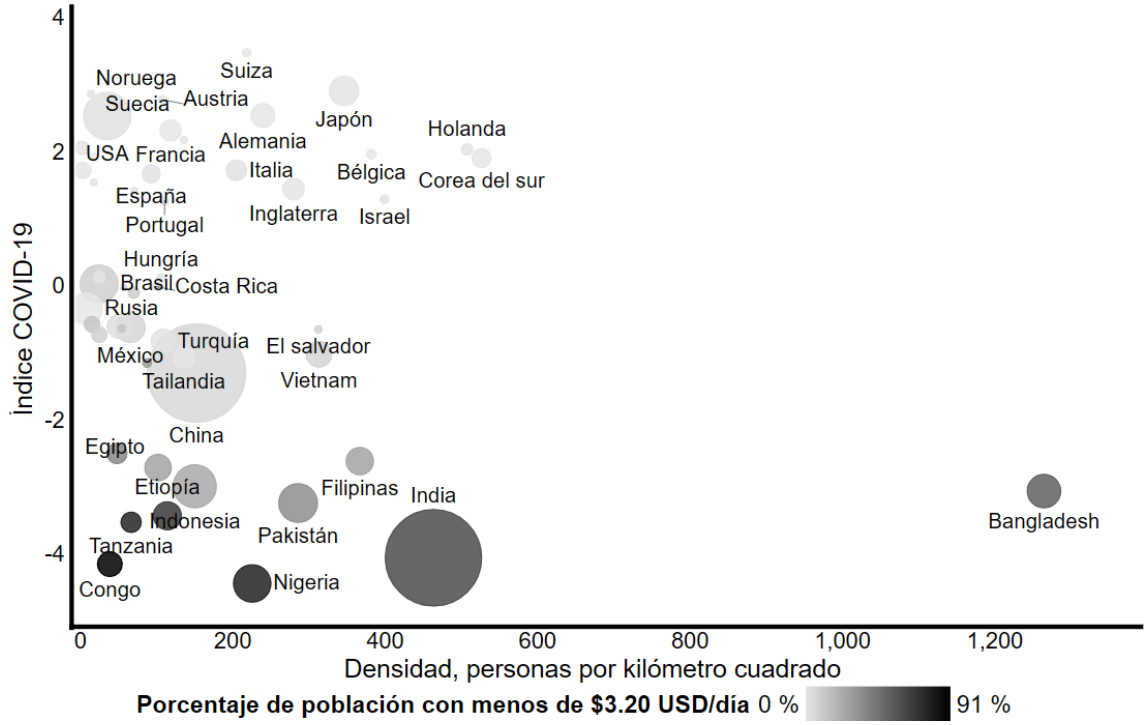
En el análisis de error de clasificación por el método de restitución se encontraron 2 países mal clasificados (Observación 24 y 28) que corresponden a Hungría con un IV=.11313 y en nivel MM el cual debería estar según este estudio en PM y también Ecuador con un IV= -.10296 y un nivel PM que debería estar en MM, en el resumen de este análisis (Cuadro 16), se apreció que en el peor de los niveles se llegó a un 88.89 % de datos bien clasificados, y el total de error por el método de restitución sólo arroja un 4.70 % de error, dando indicio a que el método de clasificación es correcto.

En la gráfica 16 se resumieron los mayores problemas a los que se enfrentaría un país al momento de controlar la pandemia, la gráfica se lee de la siguiente manera: en el eje X se tiene la densidad de los países por kilómetro cuadrado, siendo el lado derecho el más denso, en el eje Y se tiene el índice de vulnerabilidad, entre más alto, mejor preparado está el país, el tamaño de la burbuja es la población total del país, siendo la burbuja más grande el país con mayor población (China) y por último la variable del porcentaje de personas en pobreza (viviendo con menos de $3.20 USD/Día) se refleja en el color de la burbuja en escala de grises.

Se observa que los países que se encuentran en un sector vulnerable tanto económico, geográfico y de salud son Nigeria, India y Bangladesh, estos tres tienen la combinación de variables negativas para no poder sobrellevar con éxito esta pandemia.

Cuadro 16. Número de observaciones y porcentajes por restitución

Nivel de manejo	BM	EM	MM	PM	Total
BM	18 100.00%	0 0.00%	0 0.00%	0 0.00%	18 100.00%
EM	0 0.00%	11 100.00%	0 0.00%	0 0.00%	11 100.00%
MM	0 0.00%	0 0.00%	8 88.89%	1 11.11%	9 100.00%
PM	0 0.00%	0 0.00%	1 7.69%	12 92.31%	13 100.00%
Total	18 35.29%	11 21.57%	9 17.65%	13 25.49%	51 100.00%



Gráfica 16. Valores extremos

CONCLUSIONES Y RECOMENDACIONES

Los países con poder económico alto y baja población son según el indicador de la vulnerabilidad de un país contra el COVID-19, los que tienen mayor probabilidad de sobrellevar la pandemia, el país con mayor riesgo fue Bangladesh y el mejor preparado fue Suiza, si bien se encontró evidencia estadística para realizar un índice de vulnerabilidad, agrupaciones clúster y análisis discriminante, existen variables las cuales no es posibles prever como las son: las decisiones de los jefes de gobierno, la obediencia de la población a mantenerse en confinamiento, ciudades altamente turísticas y el nivel de vida del virus debido al clima, estas pueden afectar en ambos lados la propagación y mortandad del virus, todo esto sumado a que los países se reservarán los datos oficiales de las consecuencias que haya dejado la pandemia.

LITERATURA CITADA

Cuadras, C. M. (2014). *Nuevos métodos de análisis multivariante.* Barcelona; España: CMC Editions.

Díaz Monroy, L. G. (2007). *Estadística multivariada: Inferencia y métodos.* Bogotá, Colombia: Universidad Nacional de Colombia.

Grupo Banco mundial. (2022). *Banco mundial*. Recuperado el 5 de Junio de 2023, de https://www.bancomundial.org/es/publication/wdr2022/brief/chapter-1-introduction-the-economic-impacts-of-the-covid-19-crisis#:~:text=Los%20impactos%20econ%C3%B3micos%20de%20la%20pandemia%20y%20los%20nuevos%20riesgos%20para%20la%20recuperaci%C3%B3n&text=La

Hair, J. (1999). *Análisis multivariante* (Quinta ed.). (A. Otero, Ed.) España: Prentice Hall International, Inc.

Lastra, M. S. (20 de Septiembre de 2021). *Instituto de geografía UNAM.* Obtenido de https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-46112021000100101

Md Iderus, N., Lakha Singh, S., Mohd Ghazali, S., Yoon Ling, C., Cia Vei, T., Md Zamri, A., . . . Gill, B. (2022). Correlation between Population Density and COVID-19 Cases during the Third Wave in Malaysia: Effect of the Delta Variant. *Int. J. Environ. Res. Public Health*, 6.

Médica, R. (07 de Abril de 2020). *REDACCIÓN MÉDICA*. Obtenido de https://www.redaccionmedica.com/secciones/sanidad-hoy/las-5-enfermedades-mas-frecuentes-detras-de-las-muertes-por-coronavirus-9239

Navidi, W. (2006). *Estadística para ingenieros y científicos.* México: Mc Graw Hill.

OMS. (07 de 04 de 2020). *Organización mundial de la salud.* Obtenido de https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public

Organización Mundial de la Salud. (14 de Septiembre de 2022). *Organización Mundial de la Salud*. doi:WHO/2019-nCoV/Policy_Brief/Clinical/2022.1

Peña, D. (2002). *Análisis de datos multivariantes.* España: MCGRAW-HILL.

Ranzani, B. H. (17 de 08 de 2022). Enseñanzas de la pandemia de COVID-19 en América Latina: la vulnerabilidad genera más vulnerabilidad. *Rev Panam Salud Publica*(46), 1-2. doi:https://doi.org/10.26633/RPSP.2022.59

Roa, M. M. (1 de Diciembre de 2020). *Statista*. Obtenido de https://es.statista.com/grafico/23659/paises-segun-su-puntuacion-en-el-ranking-de-resiliencia-a-la-covid-19/

Team, T. (2020, 02 08). *China CDC Weekly*. Retrieved from http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51

ANEXO I

Variables iniciales y unidades (Primera parte)

País	Población total	Población de 0 a 14 años	Población de 70 o más	Población de 80 o más	Densidad de población	Ciudades con más de 1 Millón de habitantes
Nombre	Miles de personas	Miles de personas	Miles de personas	Miles de personas	Personas por Km^2	Ciudades
Australia	25 500	4 920	2 902	1 055	3.3	5
Austria	9 006	1 298	1 279	488	109.3	1
Bangladesh	164 689	44 062	5 812	1 716	1265.2	3
Bélgica	11 590	1 974	1 598	658	382.7	2
Brasil	212 559	44 019	12 961	4 159	25.4	21
Canadá	37 742	5 954	4 683	1 664	4.2	6
Chile	19 116	3 678	1 531	538	25.7	2
China	1 439 324	254 930	98 112	26 618	153.3	121
Costa Rica	5 094	1 061	344	113	99.8	1
Congo	89 561	41 015	1 604	335	39.5	5
Dinamarca	5 792	943	859	273	136.5	1
Ecuador	17 643	4 833	853	283	71.0	2
Egipto	102 334	34 713	3 295	764	102.8	2
El salvador	6 486	1 725	376	127	313.0	1
Etiopía	114 964	45 891	2 498	568	115.0	1
Finlandia	5 541	879	894	311	18.2	1
Francia	65 274	11 523	9 755	4 027	119.2	6
Alemania	83 784	11 693	13 347	5 876	240.4	11
Honduras	9 905	3 030	310	104	88.5	2
Hungría	9 660	1 392	1 278	431	106.7	1
India	1 380 004	361 018	52 460	13 284	464.1	56
Indonesia	273 524	70 941	9 993	2 419	151.0	17
Irán	83 993	20 784	3 249	894	51.6	8
Irlanda	4 938	1 029	489	158	71.7	1
Israel	8 656	2 409	721	261	400.0	1
Italia	60 462	7 852	10 557	4 529	205.6	5
Japón	126 476	15 744	27 537	11 351	346.9	13
México	128 933	33 310	6 227	2 038	66.3	16
Holanda	17 135	2 690	2 428	837	508.2	2
Nueva Zelanda	4 822	937	549	187	18.3	1
Nicaragua	6 625	1 954	227	76	55.0	1
Nigeria	206 140	89 645	3 042	419	226.3	15
Noruega	5 421	936	675	229	14.8	1

Pakistán	220 892	76 914	5 983	1 424	286.5	9
Panamá	4 315	1 143	246	88	58.0	1
Perú	32 972	8 141	1 848	592	25.8	2
Filipinas	109 581	32 921	3 554	914	367.5	4
Portugal	10 197	1 331	1 701	682	111.3	2
Corea del sur	51 269	6 431	5 417	1 856	527.3	7
Rusia	145 934	26 797	14 205	5 655	8.9	16
Arabia Saudí	34 814	8 598	663	170	16.2	5
Sudáfrica	59 309	17 082	1 898	422	48.9	4
España	46 755	6 732	6 939	2 924	93.7	5
Suecia	10 099	1 780	1 526	532	24.6	1
Suiza	8 655	1 295	1 211	459	219.0	1
Tailandia	69 800	11 554	5 788	1 921	136.6	2
Turquía	84 339	20 193	4 828	1 475	109.6	11
Inglaterra	67 886	12 000	9 281	3 451	280.6	10
Tanzania	59 734	26 017	905	163	67.4	1
USA	331 003	60 811	37 230	13 147	36.2	54
Vietnam	97 339	22 577	4 645	1 866	313.9	3

Variables iniciales y unidades (segunda parte)

País	Esperanza de vida saludable al nacer	Probabilidad de morir por enfermedades	Médicos por 10,000 hab.	Gasto per cápita en salud	% de gasto del país en salud	Camas por 10,000 hab.	% población con un salario menor a $3.20
Nombre	Años	Probabilidad	Densidad	Dólares por persona	Porcentaje	Camas	Porcentaje
Australia	73	9.1	35.9	5 002	9.3	37.9	0.70
Austria	72.4	11.4	51.4	4 688	10.4	76.47	0.40
Bangladesh	63.3	21.6	5.3	34	2.4	7.7	52.90
Bélgica	71.6	11.4	33.2	4 149	10	62.29	0.20
Brasil	66	16.6	21.5	1 016	11.8	22	9.20
Canadá	73.2	9.8	26.1	4 458	10.5	27	0.50
Chile	69.7	12.4	10.8	1 191	8.5	22	0.70
China	68.7	17	17.9	398	5	42	5.40
Costa Rica	70.9	11.5	11.5	889	7.6	11.6	3.60
Congo	52.5	19.4	0.9	21	3.9	8	77.00
Dinamarca	71.8	11.3	44.6	5 566	10.4	25.3	0.20
Ecuador	67.9	13	20.5	505	8.4	15	9.70
Egipto	61.1	27.7	7.9	131	4.6	15.6	26.10
El salvador	65.5	14	15.7	294	7	13	7.70
Etiopía	57.5	18.3	1	28	4	3.12	68.90
Finlandia	71.7	10.2	38.1	4 117	9.5	43.5	0.10
Francia	73.4	10.6	32.3	4 263	11.5	64.77	0.10
Alemania	71.6	12.1	42.1	4 714	11.1	82.78	0.20
Honduras	66.8	14	3.1	200	8.4	7	30.00
Hungría	66.8	23	32.3	943	7.4	70.37	1.20
India	59.3	23.3	7.8	62	3.6	6.6	60.00
Indonesia	61.7	26.4	3.8	112	3.1	12.1	24.20
Irán	65.4	14.8	11.4	415	8.1	15	2.30
Irlanda	72.1	10.3	30.9	4 759	7.4	27.61	0.40
Israel	72.9	9.6	32.2	2 837	7.3	30.92	0.70
Italia	73.2	9.5	40.9	2 739	8.9	34.22	1.80
Japón	74.8	8.4	24.1	4 233	10.9	134	0.90
México	67.7	15.7	22.5	462	5.5	15.2	6.60
Holanda	72.1	11.2	35.1	4 742	10.4	46.57	0.30
Nueva Zelanda	72.8	10.1	30.3	3 745	9.2	28	0.70
Nicaragua	66.9	14.2	10.1	188	8.7	9	12.80
Nigeria	48.9	22.5	3.8	79	3.6	5	77.60

Noruega	73	9.2	46.3	7 478	10.5	38.58	0.30
Pakistán	57.7	24.7	9.8	40	2.8	6	34.70
Panamá	69.4	13	15.7	1 041	7.3	23	5.20
Perú	67.5	12.6	12.7	316	5.1	16	8.30
Filipinas	61.7	26.8	12.8	129	4.4	5	26.00
Portugal	72	11.1	33.4	1 801	9.1	33.95	0.70
Corea del sur	73	7.8	23.7	2 044	7.3	115.3	0.50
Rusia	63.5	25.4	40.1	469	5.3	81.75	0.20
Arabia Saudí	65.7	16.4	23.9	1 147	5.7	26.5	.
Sudáfrica	55.7	26.2	9.1	428	8.1	28	37.60
España	73.8	9.9	40.7	2 390	9	29.65	1.10
Suecia	72.4	9.1	54	5 711	10.9	25.94	0.30
Suiza	73.5	8.6	42.4	9 836	12.2	46.77	0.10
Tailandia	66.8	14.5	8.1	222	3.7	21	0.50
Turquía	66	16.1	17.6	469	4.3	26.56	1.40
Inglaterra	71.9	10.9	28.1	3 958	9.8	27.58	0.30
Tanzania	56.5	17.9	0.4	35	4.1	7	76.60
USA	68.5	14.6	25.9	9 870	17.1	29	1.50
Vietnam	67.5	17.1	8.2	123	5.7	25.6	7.00

ANEXO II

CUADROS DE ITERACIONES

Variables segunda iteración

Nombre de la variable
Población de 0 a 14 años
Población de 70 o más
Densidad de población
Ciudades con más de 1 Millón de habitantes
Esperanza de vida saludable al nacer
Probabilidad de morir por enfermedades
Médicos por 10,000 habitantes
Gasto per cápita en salud
Porcentaje de gasto del país en salud
Camas por 10,000 habitantes
Porcentaje población con un salario menor a $3.20 por día

Variables tercera iteración

Nombre de la variable
Población total
Esperanza de vida saludable al nacer
Probabilidad de morir por enfermedades
Médicos por 10,000 habitantes
Gasto per cápita en salud
Porcentaje de gasto del país en salud
Camas por 10,000 habitantes
Porcentaje población con un salario menor a $3.20 por día

ANEXO III

Variables finales y fuentes

Variable	Unidades	Fuente
Población total	Miles de personas	https://population.un.org/wpp/Download/Standard/Population/
Esperanza de vida saludable al nacer	Años	https://www.who.int/gho/publications/world_health_statistics/2019/en/
Gasto per cápita en salud	Per cápita dólares	https://www.who.int/gho/publications/world_health_statistics/2019/en/
Porcentaje del PIB en gasto de salud	Porcentaje PIB	https://www.who.int/gho/publications/world_health_statistics/2019/en/
Número de médicos por 10,000 habitantes	Médicos	https://www.who.int/gho/publications/world_health_statistics/2019/en/
Probabilidad de morir de enfermedades	Proporción	https://www.who.int/gho/publications/world_health_statistics/2019/en/
Camas por 10,000 habitantes	Camas	https://www.who.int/data/gho/data/indicators/indicator-details/GHO/hospital-beds-(per-10-000-population)
Población con menos de $3.20 dólares al día	Porcentaje	https://population.un.org/wpp/Download/Standard/Population/