



Facultad de Matemática y Computación, Universidad de la Habana

Centro de Estudios de Población y Desarrollo

Oficina Nacional de Estadísticas e Información

Modelación Bayesiana de los parámetros de la Epidemia de SARS-CoV-2 en Cuba

Autor: Yan Carlos Leyva Labrador Lic.

Tutor: Vivian Sistachs Vega, Dr.C

Tesis Para optar por el Título de Master en Matemática

Mención: Estadística y Probabilidades

La Habana, 2022

Agradecimientos

A mi esposa por su apoyo y amor incondicional.

A mi hermana por estar siempre ahí.

A mi madre por todo su cariño y apoyo en los momentos difíciles.

A mi padre quien siempre me brindó su apoyo.

A mi tutora Vivian por su motivación y apoyo en el desarrollo de esta investigación.

A mis estimados compañeros Ernesto y Mercedes por su apoyo incondicional.

A mis oponentes por sus valiosos comentarios para la culminación de esta tesis.

A todo aquel que haya olvidado nombrar explícitamente.

Muchas gracias

Resumen

A finales del 2019 en Wuhan, China, se originó a una pandemia sin precedentes en la historia de la humanidad, que ha sido atribuida a un coronavirus llamado SARS-CoV-2. Esta pandemia cobró hasta la fecha la vida de millones de personas en todo el mundo. Además, impactó significativamente en la economía de los países y la vida cotidiana de sus habitantes. Cuba no estuvo exenta de esta situación, se produjeron grandes oleadas de contagios y de muertes atribuidas fundamentalmente, a la incorporación de nuevas variantes del virus. Con el objetivo de mitigar el impacto y avance de la epidemia en Cuba, se desarrollaron masivas campañas de vacunación a partir de fármacos producidos propiamente por el país. Este trabajo estima mediante un modelo jerárquico Bayesiano simulado en el software Stan a través del software R, los parámetros fundamentales de la epidemia de SARS-CoV-2 en Cuba. Estos parámetros son explicados a partir de las nuevas variantes del virus, así como la cantidad de personas vacunadas, permitiendo evaluar el impacto que produjo la campaña de vacunación y la introducción de nuevas variantes del virus sobre el avance de la epidemia en Cuba.

Los principales impactos que presentó la campaña de vacunación en Cuba fue la disminución de los casos susceptibles desde un 99% hasta un 49%, revelando que las vacunas poseen una efectividad del 50%. Así mismo, el índice de mortalidad por infección (IFR) descendió desde 1.4% [1.29%-1.49%] hasta 0.62% [0.499%-0.751] alcanzando un mínimo de 0.24% [0.190%-0.308%]. Con respecto a la variante de Wuhan, las variantes delta, beta y omicron resultaron ser más virulentas siendo esta última la más transmisible. Así mismo, las variantes épsilon y alfa mostraron ser menos transmisibles con respecto a la variante de Wuhan. La vacunación de la población contribuyó a disminuir significativamente el IFR, mientras que la variante omicron, incrementa el IFR en comparación al resto de las variantes que afectaron a la población cubana.

Índice

Introducción	8
Problema	13
Objetivo General	13
Objetivos específicos	13
Estructura de la tesis	14
I Modelos empleados para describir la epidemia de Covid 19	16
I.1 Aspectos generales	16
I.2 Modelos deterministas	18
I.2.1 Modelo SIRM	18
I.2.2 Modelo Logístico de Crecimiento de Richards	19
I.2.3 Consideraciones sobre los modelos deterministas	20
I.3 Modelos estocásticos	20
I.3.1 Estimación del índice de mortalidad por infección	20
I.3.2 Estimación del número reproductivo básico (R_0) y el número variable en el tiempo (R_t)	22
I.3.3 Modelo del Colegio Imperial de Londres (CIL) Reporte 26	26
I.3.4 Modelo del Colegio Imperial de Londres Reportes 13, 20, 21 y 23	27
I.3.5 Consideraciones sobre los modelos estocásticos	28
I.4 Comparación entre modelos	29
II Enfoque bayesiano y simulación del modelo	35
II.1 Enfoque Bayesiano.....	35
II.1.1 Modelos Jerárquicos	35
II.1.2 Paradigma Bayesiano.....	36
II.1.3 Métodos Markov Chain Monte Carlo (MCMC)	39

II.2	Simulación del Modelo	42
II.2.1	R	42
II.2.2	RStan.....	44
III	Modelación para Cuba	46
III.1	Descripción del modelo.....	47
III.1.1	Variables independientes:.....	48
III.1.2	Susceptibles dependencias:.....	48
III.1.3	Número reproductivo en el tiempo	49
III.1.4	IFR	49
III.1.5	Contagiados.....	49
III.1.6	Activos	50
III.1.7	Fallecidos	50
III.2	Red Bayesiana	51
III.3	Evaluación del modelo.....	56
III.3.1	R-hat.....	57
III.3.2	Logaritmo de la densidad posterior lp_{\cdot}	58
III.3.3	Tamaño efectivo de la muestra.....	59
III.3.4	Autocorrelación y transiciones divergentes en R_0 e IFR_0	60
III.3.5	El coeficiente de determinación	62
III.3.6	Pruebas de normalidad.....	62
III.4	Análisis de los parámetros epidemiológicos	63
III.4.1	Análisis descriptivo	63
III.4.2	Susceptibles y parámetros asociados	67
III.4.3	Número reproductivo en el tiempo y parámetros asociados	69

III.4.4	Índice de mortalidad por infección y parámetros asociados	73
III.4.5	Nuevos contagios y nuevas muertes	77
Conclusiones		80
Recomendaciones		82
Referencias bibliográficas.....		84

Introducción

Introducción

A finales del 2019 en Wuhan, China, se originó a una pandemia sin precedentes en la historia de la humanidad, una neumonía viral atípica y altamente contagiosa, COVID-19, que ha sido atribuida a un coronavirus de transmisión zoonótica al que se llamó SARS-CoV-2 [1]. Pocos meses después de su descubrimiento y a pesar de los grandes esfuerzos de los gobiernos, las instituciones y las personas, el virus se encontraba esparcido de manera global [2], su rápida propagación, en un inicio se debió a diferentes factores, como son:

1. La capacidad de transmitirse desde personas asintomáticas o desde personas que aún no presentan síntomas, así como la facilidad de transmisión por las diferentes vías [3]:
 - Transmisión por gotitas de flügge.
 - Transmisión por el aire.
 - Transmisión por superficies.
 - Transmisión fecal-oral.
2. Las características socio culturales de los países.
3. Las conexiones marítimo-aéreo-terrestres que existen en el mundo para el comercio y para la transportación de personas.
4. Los bajos niveles de inmunidad de la población contra este virus emergente.

Así mismo, el rápido crecimiento de casos infectados por SARS-CoV-2 y la propagación global acelerada se puede explicar mediante el número reproductivo básico (R_0).

R_0 describe el contagio y la transmisibilidad de los patógenos, y es una función de los índices de contacto entre individuos. Esto contrasta con el número reproductivo variable en el tiempo (R_t), definido como el número de personas en una población que fueron infectadas por un individuo infeccioso en un momento dado. Este último refleja los niveles cambiantes de inmunidad en la población y el impacto de las medidas de control que limitan la transmisión [4]. Además, R_0 se utiliza como una medida que permite comparar la capacidad que poseen los virus para propagarse en la sociedad. En la siguiente tabla se observa que el R_0

del virus SARS-CoV-2 es el más elevado en comparación con el número reproductivo de las otras epidemias que han afectado a la humanidad en el siglo XXI, evidenciando la alta capacidad de propagación de este virus [5].

Epidemia/ año	Número Reproductivo Básico (R0)
SARS 2003	≈ 2.75
Ébola 2014	De 1.51 a 2.53
H1N1 2009	De 1.46 a 1.48
MERS 2012	≈ 1
SARS-CoV-2 2019	De 1.5 a 3.5

Tabla 1. Número reproductivo básico de las epidemias correspondientes al SARS, Ébola, H1N1, MERS y al SARS-CoV-2.

La rápida propagación del SARS-CoV-2 incidió de manera directa sobre los sistemas de salud de los países afectados, los cuales no estaban preparados para atender simultáneamente los elevados números de personas contagiadas que surgían diariamente, provocando el colapso de sus sistemas sanitarios. A su vez, el colapso sanitario ocasionó un incremento en las muertes por Covid 19 debido a la llegada tardía de la asistencia médica y en algunos casos inexistente para las personas que la necesitaban (Figura 1).

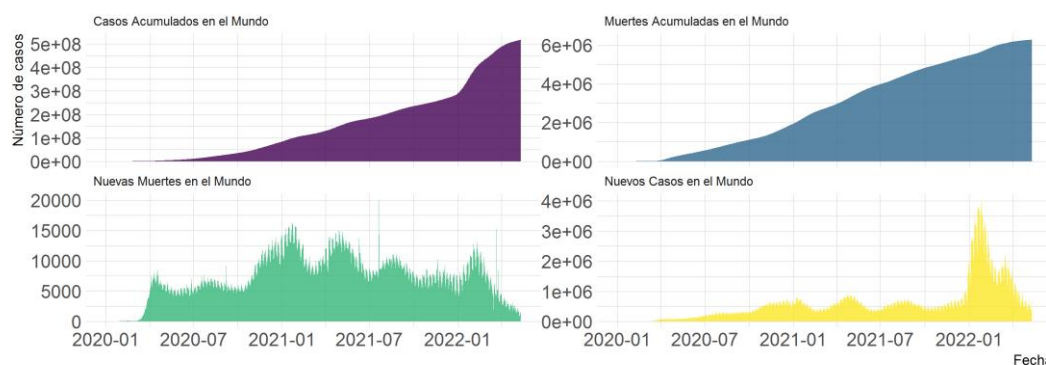


Figura 1. Casos Acumulados, Muertes Acumuladas, Nuevos Casos, Nuevas muertes en el mundo desde enero del 2020 hasta la fecha. Datos tomados de: <https://github.com/owid/covid-19-data/tree/master/public/data>

La Figura 1 muestra la caótica propagación de la epidemia de SARS-Cov-2 a nivel mundial. Hasta el 12 de mayo de 2022 se reportaron aproximadamente 516 922 683 casos de personas infectadas y 6 259 945 decesos producto de la infección por SARS-Cov-2.

Los primeros casos de SARS-CoV-2 en Cuba se reportaron el 11 de marzo del 2020. Desde ese momento y hasta la actualidad, el país ha llevado a cabo una extensa batalla para mitigar la propagación de la epidemia y sus consecuencias. Entre los principales impactos se destacan, de la esfera económica, el decrecimiento del producto interno bruto, mientras que en lo social, se afectó la seguridad alimentaria de la población debido a la reducción en la disponibilidad de los alimentos (Figura 2) [6].

En tal sentido, la difícil situación económica que atraviesa el país impide que se adopten medidas más restrictivas para mitigar el rápido crecimiento de la epidemia, debido a ello se han reportado grandes oleadas de contagios (Figura 3)

La Figura 3 muestra el avance de la epidemia de SARS-CoV-2 en Cuba, hasta el 12 de mayo del 2022 se reportaron aproximadamente 1 104 206 casos de personas infectadas y 8 529 decesos.

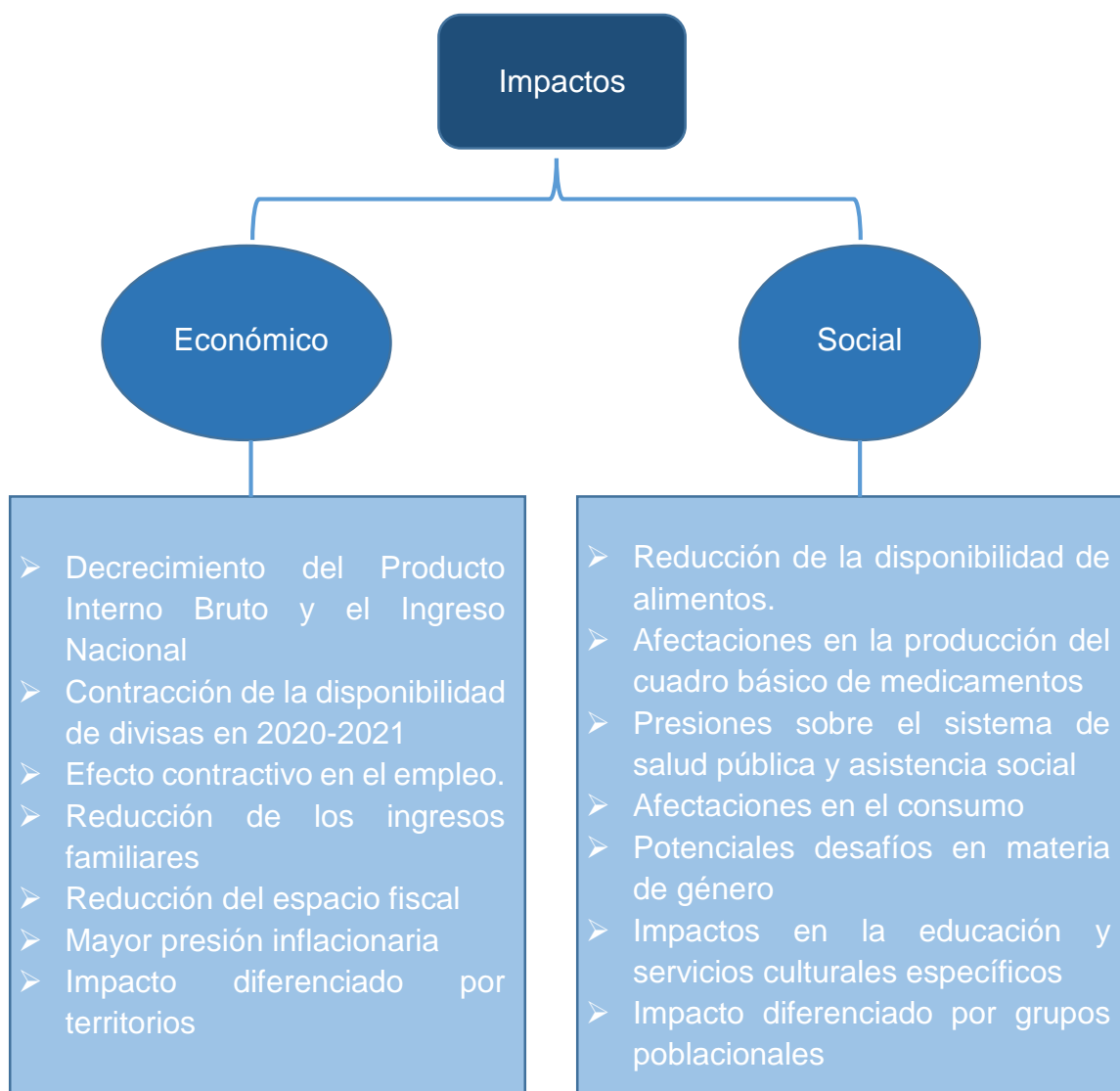


Figura 2. Impactos de la pandemia en la economía y la sociedad cubana.

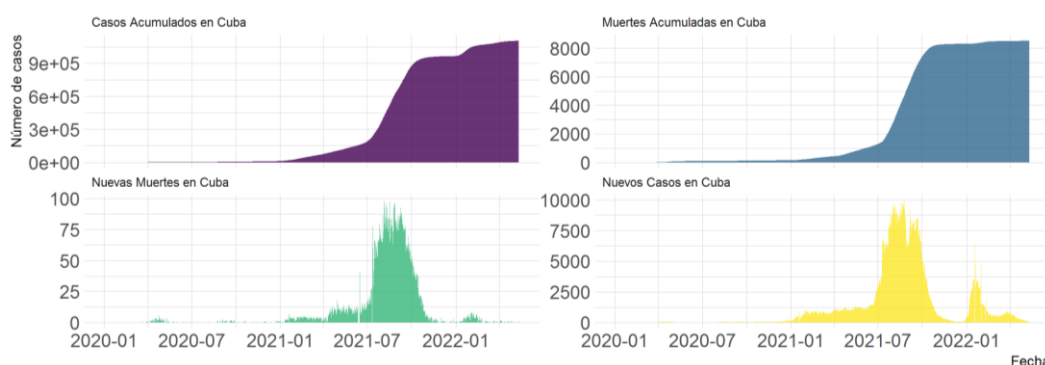


Figura 3. Casos Acumulados, Muertes Acumuladas, Nuevos Casos, Nuevas muertes en Cuba desde marzo del 2020 hasta la fecha. Datos tomados de: <https://github.com/owid/covid-19-data/tree/master/public/data>

Otro parámetro epidemiológico de gran importancia para describir el avance de la epidemia es el índice de mortalidad. Este caracteriza la efectividad de los programas de mitigación y contribuye a la vigilancia del alcance del sistema de salud sobre los casos graves y críticos, dado que, un incremento en la letalidad puede ser, entre otras cosas, un indicio de que no se está llegando a tiempo a los casos positivos o que el sistema de salud está arribando al límite del colapso.

Cuba es un país con altos niveles de envejecimiento y la mortalidad por COVID-19 se correlaciona positivamente con la edad [7], por esto se debe prestar atención sobre los índices de mortalidad en el país.

Los virus como el SARS-CoV-2 evolucionan constantemente a medida que se producen cambios en el código genético (provocados por las mutaciones genéticas o la recombinación viral) durante la replicación del genoma. Su capacidad para mutar está inducida por disímiles factores como son [8]:

- las características específicas de la partícula viral
- la forma de transmisión
- el alto índice de transmisibilidad
- la propagación global

En Cuba se han detectado muchas de las variantes reconocidas por la Organización Mundial de la Salud (OMS) como variantes de interés [8]. Estas variantes se descubrieron en diferentes países y se propagaron rápidamente por el mundo.

Las mutaciones que se han producido en la proteína de la espícula del SARS-CoV-2 provocaron que las variantes Alfa, Beta, Gamma y Delta podrían ser más infecciosas en células humanas en comparación con el SARS-CoV-2 aislado en Wuhan-2019; mientras que las variantes Gamma y Delta serían las más infecciosas entre ellas [9, 10]. La introducción de las nuevas variantes en Cuba ocasionó que los parámetros y las características de la epidemia variaran, provocando en algunos casos enormes olas de nuevos contagios.

Dada esta situación y lo difícil y costoso de detener el avance de la epidemia de SARS-CoV-2 en Cuba, y sumado el amplio avance científico y tecnológico con que se cuenta, el país y sus instituciones, al igual que el mundo, diferentes grupos científicos se enrolaron en la búsqueda de vacunas que fuesen capaces de prevenir los contagios. Estas metas también se lograron en Cuba y siendo creadas diferentes vacunas que contribuyeron a mitigar el avance de la epidemia y sus efectos sobre la sociedad.

Por todo lo antes expuesto es de vital importancia para Cuba dar solución al siguiente problema.

Problema

Crear un modelo matemático que describa los parámetros epidemiológicos de la epidemia de SARS-CoV-2 en Cuba y que incluya el efecto que tiene sobre estos parámetros la introducción de nuevas variantes, así como la vacunación masiva de la población.

Para resolver el problema se plantearon los siguientes objetivos:

Objetivo General

Modelar los parámetros de la epidemia de SARS-CoV-2 en Cuba, teniendo en cuenta el efecto de las nuevas variantes y de la vacunación.

Objetivos específicos

1. Identificar los modelos matemáticos que permitan describir los parámetros de la epidemia de SARS-CoV-2.
2. Construir un modelo Bayesiano para identificar el efecto que tuvo la campaña de vacunación y la introducción de nuevas variantes en Cuba.
3. Estimar los parámetros asociados al modelo construido utilizando los lenguajes de programación R y Stan.
4. Evaluar el ajuste y convergencia del modelo.

Estructura de la tesis

El trabajo está estructurado en cuatro capítulos además de la Introducción, conclusiones, recomendaciones y referencias bibliográficas.

El primer capítulo se dedica al estudio de los modelos empleados para describir los parámetros de las epidemias. Para su análisis, los modelos se dividen en dos grandes grupos de clasificación, modelos deterministas y modelos estocásticos. Con la finalidad de seleccionar el tipo de modelo que mejor se ajuste a las características específicas de la epidemia de SARS-CoV-2 en Cuba se establece una comparación entre los modelos explicados en el capítulo.

El segundo capítulo describe el enfoque Bayesiano, así como el método empleado para resolver el modelo diseñado a partir de la simulación computacional. Además, se describen los modelos jerárquicos y se explican los fundamentos del paradigma Bayesiano introduciéndose los métodos de Monte Carlo Markov Chain (MCMC). También se presentan las bondades de los software R y Stan, los cuales son utilizados en la simulación computacional del modelo diseñado.

En el tercer capítulo se presenta el complejo modelo jerárquico Bayesiano diseñado para describir los parámetros de la epidemia de SARS-CoV-2 en Cuba. Además, se evalúa el ajuste de los resultados obtenidos mediante la simulación a los datos, así como la convergencia de las transiciones iniciadas durante el proceso de simulación. Por último, en este capítulo, se describen los parámetros epidemiológicos obtenidos y se realizan comparaciones con resultados descritos por otros autores para determinar en qué medida el modelo diseñado se ajusta a la realidad.

Finalmente, en las conclusiones se exponen los resultados fundamentales a los que se llegó a partir de los objetivos planteados y en las recomendaciones se indican pautas a seguir con la finalidad de ampliar la investigación presentada.

Capítulo I

Modelos empleados para describir la epidemia de SARS-Cov-2

I Modelos empleados para describir la epidemia de Covid 19

I.1 Aspectos generales

La Epidemiología es una disciplina científica que se concentra en el estudio de la frecuencia y distribución de los eventos de salud y de sus determinantes en las poblaciones humanas, y la aplicación de este estudio en la prevención y control de los problemas de salud [11, 12].

Esta es una rama importante de la ciencia y contribuye a [13-15]:

- Definir los problemas e inconvenientes de salud importantes de una comunidad;
- Describir la historia natural de una enfermedad;
- Descubrir los factores que aumentan el riesgo de contraer una enfermedad (su etiología);
- Predecir las tendencias de una enfermedad;
- Determinar si la enfermedad o problema de salud es prevenible o controlable;
- Determinar la estrategia de intervención (prevención o control) más adecuada;
- Probar la eficacia de las estrategias de intervención;
- Cuantificar el beneficio conseguido al aplicar las estrategias de intervención sobre la población;
- Evaluar los programas de intervención;

Así mismo, se puede dividir en cuatro ramas fundamentales:

1. Epidemiología descriptiva
2. Epidemiología analítica
3. Epidemiología experimental
4. Ecoepidemiología

La epidemiología analítica, es la que nos ocupa en este estudio pues esta se basa en la observación o experimentación, a través de las relaciones causales entre “factores de riesgo” a los que se exponen las personas, las poblaciones y las enfermedades, de acuerdo con la probabilidad de riesgo absoluto (probabilidad de ocurrencia de una enfermedad en un período de tiempo dado) o riesgo relativo (probabilidad relativa de ocurrencia de dos o más riesgos). Además, busca las causas y los efectos, el por qué y el cómo de una enfermedad. Se utiliza la epidemiología analítica para cuantificar, a través de modelos matemáticos, la asociación entre exposiciones y resultados y para probar las hipótesis sobre las relaciones causales [12, 16, 17]. Así mismo, los modelos matemáticos se pueden clasificar atendiendo a una serie de características dicotómicas como [12, 16-18]:

- Empíricos o teóricos. Un modelo teórico se basa en las leyes físicas que rigen los procesos, mientras que un modelo empírico se basa en relaciones estadísticamente significativas entre variables [12, 16-18].
- Estocásticos o deterministas. Los primeros incluyen generadores de procesos aleatorios dentro del modelo que modifican ligeramente algunas de las variables. Un modelo determinista es aquel en el que dado un conjunto de parámetros y variables de entrada va a producir siempre el mismo conjunto de variables de salida [12, 16-18].
- Estáticos o dinámicos. Se refiere a la forma en que se trata el tiempo. Los modelos estáticos dan un resultado agregado para todo el período de tiempo considerado. Los modelos dinámicos devuelven las series temporales de las variables consideradas a lo largo del período de estudio [12, 16-18].
- Agregados o distribuidos. En el primer caso toda el área de estudio se considera de forma conjunta. Se tiene un único valor para todos los parámetros del modelo. El modelo predice unas salidas para las entradas aportadas sin informar de lo que ocurre dentro del sistema. En un modelo distribuido, se tiene que el área de estudio se divide en porciones cada

una de ellas con su propio conjunto de parámetros y sus propias variables de estado [12, 16-18].

En la bibliografía consultada existe gran variabilidad de modelos matemáticos utilizados para describir las características epidemiológicas del SARS-Cov-2 por lo que se pueden encontrar modelos y clasificarlos en cada uno de los grupos antes expuestos. Así mismo se hizo una selección de los modelos que mayor impacto tuvieron en la sociedad y que fueron más utilizados por la comunidad científica para describir el avance del SARS-Cov-2 y sus características.

Los modelos seleccionados se agrupan según la clasificación de modelos estocásticos o modelos deterministas.

I.2 Modelos deterministas

I.2.1 Modelo SIRM

El modelo SIRM es uno de los modelos epidemiológicos más simples capaces de capturar muchas de las características típicas de los brotes epidémicos. El nombre del modelo proviene de las iniciales S (población susceptible), I (población infectada), R (población recuperada) y M (población fallecida). El modelo relaciona las variaciones de las cuatro poblaciones (Susceptible, Infectada, Recuperada y Muerta) a través del índice de infección y el período infeccioso promedio.

La población total es $N = S + I + R + M$. Hecha esta compartimentación se hace necesario especificar ecuaciones que describan la variación temporal del número de individuos en cada compartimento, las variables S, I, R, M pueden ser tratadas como variables continuas, y el modelo SIRM viene dado por las siguientes ecuaciones diferenciales [19-22]:

$$\frac{dS}{dt} = -\beta \frac{S(t)I(t)}{S(t)+I(t)} \quad (1)$$

$$\frac{dI}{dt} = \beta \frac{S(t)I(t)}{S(t)+I(t)} - \gamma I(t) - \nu I(t) \quad (2)$$

$$\frac{dR}{dt} = \gamma I(t) \quad (3)$$

$$\frac{dD}{dt} = \nu I(t) \quad (4)$$

Donde β, γ, ν son los respectivos índices de transmisión, de recuperación y de mortalidad.

I.2.2 Modelo Logístico de Crecimiento de Richards

Otro modelo altamente utilizado por la comunidad científica para determinar los parámetros de una epidemia es el modelo logístico de crecimiento de Richards. Este es un caso particular del modelo generalizado de Richards el cual está definido como [23, 24]:

$$\frac{dC(t)}{dt} = R_0 [C(t)]^p \left[1 - \left(\frac{C(t)}{K} \right)^\alpha \right] \quad (5)$$

Donde $C(t)$ representa el número acumulado de infecciones en el tiempo t , R_0 número reproductivo básico, K es el tamaño final de la epidemia, $p \in [0, 1]$ es un parámetro que permite al modelo capturar diferentes perfiles de crecimiento y α mide la desviación de la simetría de la curva logística simple. Para $p=1$ y $\alpha=1$, se tiene el Modelo Logístico de Crecimiento de Richards:

$$\frac{dC(t)}{dt} = R_0 [C(t)] \left[1 - \left(\frac{C(t)}{K} \right) \right] \quad (6)$$

Esta ecuación tiene solución analítica:

$$C(t) = \frac{K}{1 + Ae^{-rt}} \quad (7)$$

Donde $A = \frac{K - C_0}{C_0}$ y C_0 es el número inicial de infectados.

I.2.3 Consideraciones sobre los modelos deterministas

El modelo SIRM y el modelo logístico de crecimiento de Richards son modelos teóricos pues se describen mediante ecuaciones físicas que permiten transformar totalmente las variables de entrada en variables de salida. Además, ambos entran dentro del grupo de los modelos deterministas pues las variables de entrada y los parámetros determinan de forma única el conjunto de variables de salida.

Asimismo, ambos se pueden clasificar como modelos dinámicos y distribuidos. La primera se debe a que tanto el modelo SIRM como el logístico de crecimiento de Richards devuelven las series temporales de las variables consideradas a lo largo del período de estudio. En otras palabras, para cada momento t se obtienen los parámetros de la epidemia que se están midiendo. Además, son modelos distribuidos ya que el área de estudio está dividida en porciones, cada una de ellas tiene sus propias variables y parámetros.

I.3 Modelos estocásticos

I.3.1 Estimación del índice de mortalidad por infección

En el mundo las estadísticas reportadas sobre las muertes por Covid son más confiables que la cantidad de persona infectadas [25]. Específicamente en el caso cubano, los protocolos de actuación adoptados durante la pandemia favorecen la detección de las personas que mueren enfermos con Covid 19. Este factor contribuye a la confiabilidad y a disminuir el sesgo en las estadísticas de las muertes reportadas en Cuba a causa del SAR-Cov-2.

Uno de los aspectos del SAR-Cov-2 es que muchas de las personas infectadas son asintomáticas, o sea, no manifiestan síntomas clínicos de la infección. Por

lo tanto, es evidente que en las estadísticas oficiales de infectados se subestima severamente el número real de individuos. Asimismo, el índice de mortalidad por infección (*Infection Fatality Rate* (IFR), en inglés) se define como el número de muertes por infección. Estos datos son más fiables que el índice de letalidad (*Case Fatality Rate* (CFR), en inglés) el cual se calcula como el número de muertes por caso reportado. Por ello, es fundamental estimar el IFR para poder describir las características de la epidemia [26]. Para ello se presentarán dos modelos.

Debido a las características propias de la infección, existe un proceso latente asociado al número de contagios diarios reportados. Entonces sea X_i la variable aleatoria asociada a este proceso, se tiene [26]:

$$X_i = \begin{cases} 1 & \text{con probabilidad } p \\ 0 & \text{con probabilidad } 1-p \end{cases}$$

donde p corresponde a la probabilidad de estar infectado por SARS-CoV-2, es decir, la prevalencia de infección por SARS-Cov-2. Además, sea:

$$Z_i = \begin{cases} 1 & \text{si el individuo } i - \text{ésimo fallece} \\ 0 & \text{en otro caso} \end{cases}$$

Entonces:

$$\begin{aligned} P(Z_i = 1 | X_i = 1) &= w \\ P(Z_i = 1 | X_i = 0) &= 0 \end{aligned} \tag{8}$$

donde w corresponde a la probabilidad de fallecer por COVID-19 si se está infectado por SARS-Cov-2, es decir, el IFR.

Entonces por el Teorema de la Probabilidad Total, se obtiene:

$$P(Z_i=1) = \underbrace{P(Z_i=1|X_i=1)}_w \underbrace{P(X_i=1)}_p + \underbrace{P(Z_i=1|X_i=0)}_0 \underbrace{P(X_i=0)}_0 = pw$$

Dada una población de tamaño N el par $\left\{ \sum_i Z_i = n, N \right\}$ describe la cantidad de personas que fallecen a causa del SARS-Cov-2 en la población. Nótese que la probabilidad de fallecer por causa de este virus en esta población corresponde a la proporción muestral $\frac{n}{N}$. Note que esta proporción no es el IFR.

En la prevalencia y en el IFR existen diferencias según sexos y grupos de edad, de acuerdo a la literatura consultada. Dadas estas evidencias, se ha considerado oportuno incorporar las covariables Edad y Sexo en la estimación de los parámetros de interés. Para ello, se propone utilizar el siguiente modelo jerárquico bayesiano [26]:

$$\begin{aligned} n_{sj} &\sim \text{Bin}(N_{sj}, p_j w_{sj}), \\ \log\left(\frac{p_j}{1-p_j}\right) &= \alpha_0 + \alpha_1 (Edad_j - \xi)_-, \\ \log\left(\frac{w_{sj}}{1-w_{sj}}\right) &= \beta_0 + \beta_1 \text{Sexo}_s + \beta_2 Edad_j + \beta_3 \text{Sexo}_s \text{Edad}_j, \end{aligned} \quad (9)$$

$$\left. \begin{aligned} \alpha_k &\sim \text{N}(\mu_{\alpha_k}, \sigma_{\alpha_k}) \\ \beta_l &\sim \text{N}(\mu_{\alpha_l}, \sigma_{\alpha_l}) \\ \xi &\sim \text{Unif}(\xi_1, \xi_2) \end{aligned} \right\} \begin{aligned} &s = 0, 1, \quad j = 0, 1, \dots, 9 \\ &k = 0, 1, \quad l = 0, 1, 2, 3 \\ &1 \leq \xi_1, \xi_2 \leq 9 \end{aligned}$$

I.3.2 Estimación del número reproductivo básico (R0) y el número variable en el tiempo (Rt)

La propagación de la enfermedad en la población depende del R0 y del tiempo de generación (usualmente establecido según el intervalo serial).

Cada enfermedad tiene un R_0 , el cual se utiliza para calcular la velocidad con que se propaga en una población, así como para comprender qué tan infecciosa es una enfermedad y, por lo tanto, el nivel de riesgo asociado con ella.

El R_0 es el número de personas a las que una persona infectada puede transmitir la enfermedad o el número de casos secundarios que cada caso primario genera en promedio (durante el tiempo que es contagioso) [11].

La epidemia se propaga si y solo si R_0 es mayor de 1 (cada caso contagia a más de una persona); si R_0 es igual a 1 entonces se dice que hay equilibrio epidémico (o endémico) y si el R_0 es menor de 1 la situación, invariablemente, se dirige al agotamiento epidémico. Para que R_0 alcance valores menores que 1, se debe actuar sobre sus tres componentes [11]:

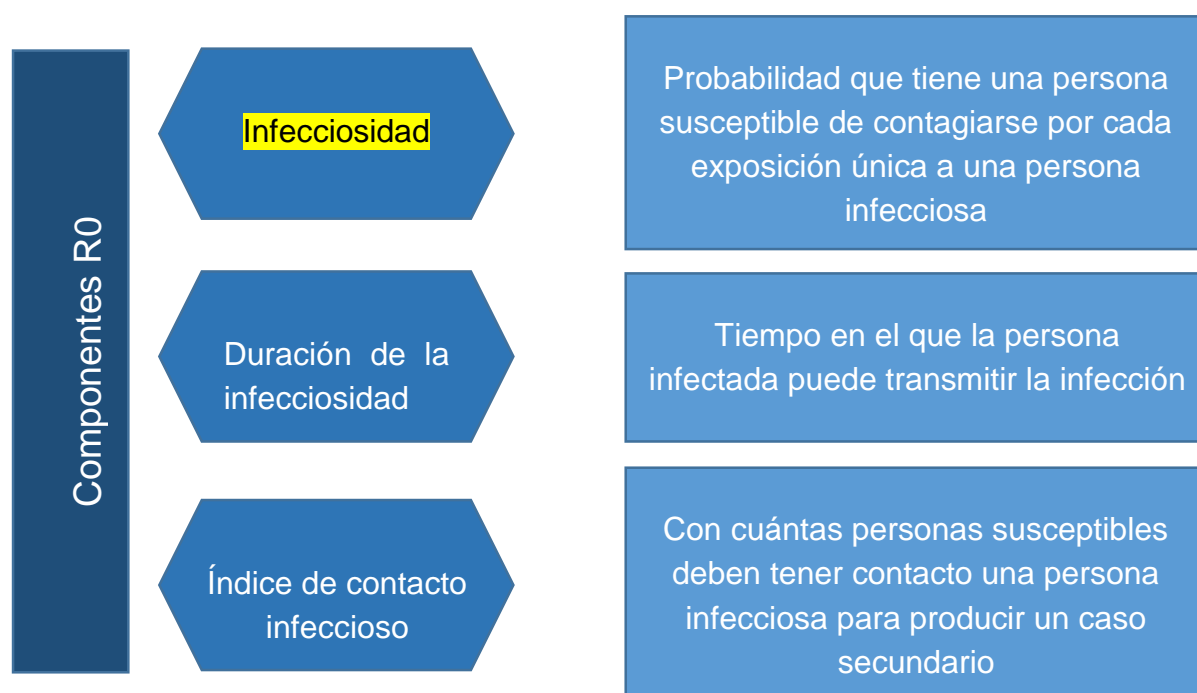


Figura 4. Factores que influyen sobre R_0 .

Con el actuar de los gobiernos, las instituciones y las personas para detener el avance de la epidemia, se adoptan determinadas medidas que influyen directamente sobre este número y provoca que tenga variaciones en el tiempo (R_t). Debido a la importancia que presenta R_t para describir las características

de la epidemia además para verificar si las iniciativas realizadas surgen efecto es fundamental estimarlo. A continuación, se presentan algunos de los modelos que utilizados para estimar el R_0 y R_t :

Bibliografía	Modelo	Variables Dependientes
Reporte 13. Colegio Imperial de Londres (CIL) [25] (10)	$R_0 \sim N(2.4, \kappa), \kappa \sim N(0, 0.5)$ $R_t \sim R_0 \exp\left(-\sum_{k=1}^6 \alpha_k I_{k,t}\right)$ $\alpha_k \sim \text{Gamma}(0.5, 1)$	$I_{k,t}$ son las intervenciones adoptadas por los países, (distanciamiento social, cierre de escuelas, aislamiento de los casos sospechosos, cuarentena, suspensión de eventos públicos), fecha en que se tomó la primera intervención
Reporte 20 y 21. Colegio Imperial de Londres (CIL) [27, 28] (11)	$R_0 \sim N(3.28, \kappa), \kappa \sim N(0, 0.5)$ $R_{t,m} \sim R_0 f\left(-\sum_{k=1} (\alpha_k + \beta_{m,k}) I_{k,t,m}\right)$ $\alpha_k \sim N(0, 0.5), \beta_{m,k} \sim N(0, \gamma)$ $\gamma \sim N(0, 0.5), f(x) = \frac{2e^x}{1+e^x}$	<p>$I_{k,t,m}$ son variables que caracterizan la movilidad de la población.</p> <p>α_k es un parámetro asociado a la variable movilidad</p> <p>$\beta_{m,k}$ es un parámetro que además de estar asociado a la movilidad se asocia además a una región determinada.</p>

<p>Reporte 23. Colegio Imperial de Londres (CIL) [29] (12)</p>	$R_0 \sim N(3.28, \kappa), \kappa \sim N(0, 0.5)$ $R_{t,m} \sim R_0 f \begin{bmatrix} -\left(\sum_{k=1} \alpha_k X_{k,t,m}\right) - \\ -Y_{t,m} \alpha_{p(m)}^{provincia} \\ -Z_{t,m} \alpha_m^{region} - \varepsilon_{m,w_m}(t) \end{bmatrix}$ $\alpha_k \sim N(0, 0.5), f(x) = \frac{2e^x}{1+e^x}$ $\alpha_{p(m)}^{provincia} \sim N(0, \gamma_p), \gamma_p \sim N^+(0, 0.5)$ $\alpha_m^{region} \sim N(0, \gamma_r), \gamma_r \sim N^+(0, 0.5)$ $\varepsilon_{w,m} \sim N(\rho_1 \varepsilon_{w-1,m} + \rho_2 \varepsilon_{w-2,m}, \sigma_w^*)$ $\varepsilon_{1,m} = 0, \varepsilon_{2,m} \sim N(0, \sigma_w^*),$ $\rho_1 \sim N(0.8, 0.05), \rho_2 \sim N(0.1, 0.05)$ $\sigma_w^* = \sigma_w \sqrt{1 - \rho_1^2 - \rho_2^2 - \frac{2\rho_1^2 \rho_2}{1 - \rho_2}}$ $\sigma_w \sim N^+(0, 0.2)$	<p>X, Y, Z hacen referencia a variables que caracterizan la movilidad en el municipio, provincia y región respectivamente.</p> <p>$\varepsilon_{m,w_m}(t)$ es un proceso semanal, centrado alrededor del cero, que registra variaciones entre las regiones que no son explicadas por las variables X, Y, Z</p> <p>α_k parámetro asociado a la movilidad</p> <p>$\alpha_{p(m)}^{provincia}$ parámetro asociado a la movilidad en la provincia</p> <p>α_m^{region} parámetro asociado a la movilidad en la región</p>
<p>Reporte 25. Colegio Imperial de Londres (CIL) [30] (13)</p>	$I_t \sim \text{Poisson} \left(R_t \sum_{s=1}^t I_{t-s} w_s \right)$ $w_s \sim \text{Gamma}(6.46, 3.83)$ $R_t \sim \text{Gamma}(1, 10)$	<p>I_t incidencia de casos positivos en el día t</p> <p>w_s intervalo de serie</p> <p>$\text{Gamma}(1, 10)$ a priori distribución no informativa</p>

Reporte 25. Colegio Imperial de Londres (CIL) [31] (14)	$\log(R_t) = \log(R_0) - \beta(1 - m_t)$	m_t movilidad en el día t R_0 y β parámetros
Modelo SIRM	Modelo SIRM ecuación (2)	S (población susceptible), I (población infectada), R (población recuperada) y M (población fallecida)
Logístico de crecimiento	Modelo Logístico de Crecimiento de Richards ecuación (6)	$C(t)$ número acumulado de infecciones

Tabla 4. Diferentes modelos que estiman R_t y R_0 .

I.3.3 Modelo del Colegio Imperial de Londres (CIL) Reporte 26

El modelo descrito a continuación utiliza las muertes reportadas como parámetro de calibración. Los autores sustentan la hipótesis de que las muertes reportadas por los países son los datos que menor sesgo estadístico presentan, entonces se relacionan las muertes notificadas observadas en el día t con el número reproductivo en el tiempo R_t y el parámetro β descritos en la fórmula (14) utilizando la ecuación [31]:

$$D_t \sim \text{BinomialNegativa} \left(R_t^D \sum_{s=0}^t D_s w_{t-s}, \delta \right) \quad (15)$$

Donde $R_t^D = \sum_{s=0}^t R_s h(t-s)$, R_s se describe en (14), h es el tiempo desde que se infecta una persona hasta que muere, el cual sigue una distribución Gamma con media de 18.8 días y desviación estándar de 8.46 días y w es el intervalo serial, el cual sigue una distribución Gamma con media de 6.48 días y desviación estándar de 3.83 días.

I.3.4 Modelo del Colegio Imperial de Londres Reportes 13, 20, 21 y 23

El Colegio Imperial de Londres a partir del modelo descrito en el Reporte 13 [25] estimó los parámetros de la epidemia en diferentes países del mundo. A medida que realizaba actualizaciones y mejoras al modelo entonces publicaba los resultados de las estimaciones correspondientes (Reporte 20 y 21 [27, 28]), hasta llegar al modelo definitivo el cual está explicado en el Reporte 23 [29]. Este último modelo por ser el más completo y el que mejor se ajusta a los parámetros de la epidemia es el que se describe a continuación:

Se introduce un marco bayesiano para estimar la intensidad de transmisión y el índice de ataque (porcentaje de la población que ha sido infectada). Este marco utiliza el número de reproducción variable en el tiempo R_t para informar una función latente para las infecciones, y luego estas infecciones, junto con los parámetros, se calibran frente a las muertes observadas. Las muertes observadas, aunque aún susceptibles según los informes y demoras en los reportes, son más confiables que el número informado de casos confirmados.

El modelo bayesiano se puede expresar [29]:

$$D_{t,m} \sim \text{BinomialNegativa} \left(d_{t,m}, d_{t,m} + \frac{d_{t,m}^2}{\psi} \right) \quad (16)$$

$$d_{t,m} = ifr_m^* \sum_{\tau=0}^{t-1} c_{\tau,m} \pi_{t-\tau}, \quad (17)$$

$$c_{t,m} = S_{t,m} R_{t,m} \sum_{\tau=0}^{t-1} c_{\tau,m} g_{t-\tau}, \quad (18)$$

$$S_{t,m} = 1 - \frac{\sum_{i=0}^{t-1} c_{i,m}}{N_m} \quad (19)$$

$$\begin{aligned} D_{t,m} &: \text{Muertes reportadas} & c_{\tau,m} &: \text{nuevas infecciones} \\ d_{t,m} &= E[D_{t,m}] & \pi_s &= \begin{cases} \int_{s-0.5}^{s+0.5} \pi(\tau) d\tau, s = 2, 3, \dots \\ \int_{s-0.5}^{s+0.5} \pi(\tau) d\tau, s = 1 \end{cases} \\ \psi &\sim N^+(0, 0.5) & \pi(\tau) &: \{\text{función de densidad de } \pi\} \\ ifr_m^* &\sim N(ifr_m, 0.1) \\ ifr_m &: \begin{cases} \text{IFR previamente} \\ \text{estimado} \end{cases} \\ \pi &\sim \underbrace{Gamma(5.1, 0.86)}_{\text{T tiempo de incubación (Infección - Síntomas)}} + \underbrace{Gamma(17.8, 0.45)}_{\text{Periodo de aparición de los síntomas hasta la muerte (Síntomas - Muerte)}} \end{aligned}$$

$$\begin{aligned} g_s &= \begin{cases} \int_{s-0.5}^{s+0.5} g(\tau) d\tau, s = 2, 3, \dots \\ \int_{s-0.5}^{s+0.5} g(\tau) d\tau, s = 1 \end{cases} & g &\sim \underbrace{Gamma(6.5, 0.62)}_{\text{Inicio de la enfermedad en el caso primario y en el caso secundario (Intervalo Serial)}} \\ & & S_{t,m} &: \text{susceptible} \\ & & N_m &: \text{población total} \\ g(\tau) &: \{\text{función de densidad de } g\} \end{aligned}$$

$R_{t,m}$ se estima en (12)

I.3.5 Consideraciones sobre los modelos estocásticos

Los modelos descritos en esta sección se clasifican como modelos estocásticos pues incluyen generadores de procesos aleatorios que modifican ligeramente algunas de las variables. De esta manera, para un mismo conjunto de datos de entrada, las salidas no serían siempre las mismas. En estos casos las salidas son distribuciones de probabilidad. Las razones por las que las componentes de los modelos resultan aleatorias son [12, 32]:

- Existencia de procesos realmente estocásticos (las distribuciones de los tiempos)
- Existencia de procesos aparentemente estocásticos debido a la falta de conocimiento (no se puede determinar exactamente la cantidad de personas contagiadas o el número de muertes reportadas)
- Errores aleatorios en la medición de las variables (fluctuaciones de la población debido a los nacimientos, muertes y migraciones)
- Procesos muy complejos que es preferible modelizar como estocásticos (IFR, R_0 o R_t)

Así mismo y para todos los casos de esta sección, los modelos se clasifican como empíricos y distribuidos. La primera de las clasificaciones se debe a que estos están basados en relaciones estadísticamente significativas entre variables y la segunda se debe a que el área de estudio se divide en procesos y cada uno de ellos tiene su propio conjunto de parámetros y sus propias variables [12, 32].

La característica de ser un modelo dinámico la cumplen todos menos el modelo que se utiliza para estimar el IFR, ya que este último, a diferencia del resto de los modelos, ofrece un resultado que explica todo el periodo de tiempo. Esta es una característica poco deseable para este parámetro epidemiológico pues el índice de fatalidad depende de diferentes factores que provocan su variabilidad en el tiempo [12, 32].

I.4 Comparación entre modelos

El último de los modelos presentados, es un modelo general, que integra y estima muchos parámetros que permiten describir y caracterizar la epidemia de SARS-Cov-2. Entre los parámetros que se estiman con este modelo están:

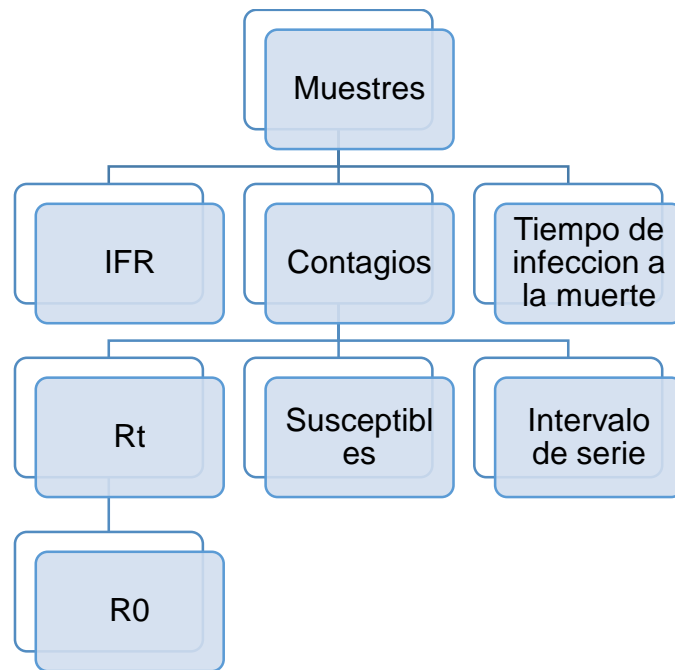


Figura 5. Parámetros epidemiológicos estimados en el modelo del Colegio Imperial de Londres Reporte 23

Así mismo, el modelo determinado en el Reporte 23 del CIL guarda relación con todos los modelos que se referencian en este trabajo pues los procesos que los integran están estrechamente relacionados entre sí. En la Tabla 5. se comparan los procesos del reporte 23 de CIL con los procesos del resto de los modelos.

Proceso Reporte 23 CIL	Modelo	Ecuación	Observación
(17)	SIRM	(4)	Fallecidos=IFR*Contagiados
IFR, Rt	SIRM	β, ν	$Rt = \beta, IFR = \nu$

(18)	Logístico de Crecimiento de Richards $p=1$, $q=1$	(6)	Contagiados=Susceptibles $*R_t*Contagiados$
IFR	Modelo Bayesiano Para estimar el IFR	(9)	Las estimaciones se realizan de diferentes formas $IFR = w_{sj}$

Tabla 5. Comparación entre los procesos del modelo del Reporte 23 del CIL con los procesos del resto de los modelos.

El ajuste de los parámetros, para todos los casos, se hace utilizando los datos de la epidemia reportados por las instancias pertinentes. Tanto el modelo SIRM como el logístico de crecimiento de Richards utilizan además de las muertes reportadas datos como el número de nuevos contagiados y/o el número de recuperados. Estos últimos datos presentan un sesgo incorporado que depende directamente de las características inherentes de la enfermedad. Por citar un ejemplo, la infección por SARS-Cov-2 presenta un alto número de pacientes asintomáticos. Si estas personas no son testadas no se determinan como contagiadas. Este es uno de los mayores sesgos que se le introduce al número de nuevos contagios, el cual está presente además en el número de recuperados.

Otro factor que influye en la calidad de los datos es el retraso en los reportes. Por su parte el modelo del Reporte 13 del CIL solo utiliza para el ajuste las muertes reportadas, siendo esta una de las fuentes que presenta menores errores.

Entre las ventajas de incorporar una componente estocástica al modelo se encuentran [12, 16-18]:

1. La utilización de generadores aleatorios de series de las variables de entrada (cuando no se cuenta con un valor absoluto de la variable de entrada)
2. La utilización de generadores aleatorios para dar valores a los parámetros del modelo y a su distribución espacial, de esta manera se evita el problema que supone utilizar parámetros estimados con cierto grado de incertidumbre.

La introducción del componente estocástico permite además comprobar cómo se comportaría el modelo para diferentes conjuntos de parámetros o valores de las variables de entrada. De este modo en lugar de obtener un resultado a partir de un conjunto verosímil de datos, obtendremos un conjunto de resultados a partir de varios conjuntos verosímiles de parámetros y variables. De este modo no tenemos por qué conformarnos con un valor esperable, sino que tendremos un rango de variación dentro del cual estarán los resultados esperables [12, 17, 18, 32].

Además, en los modelos estocásticos, aparece un efecto que (casi) nunca aparecen en modelos deterministas: la población es capaz de extinguirse [33] (se puede estimar el momento en que todas las personas dejarán de ser susceptibles). En modelos deterministas, sólo se encuentra una solución que se aproxima a cero si el tiempo tiende a infinito [33].

En los modelos explicados en esta sección ocurre que las ecuaciones diferenciales tienen variables de estado real, mientras que el sistema epidemiológico tiene entidades discretas, es decir, solo asume valores en \mathbb{N} [33].

Cuando se trabaja con poblaciones pequeñas (la cantidad de personas que fallecen cuando se está en la base de las oleadas) los modelos deterministas pierden información, por lo que se espera que los efectos estocásticos jueguen un papel importante y describan mejor el proceso [33]. Por ejemplo:

Si el número esperado de contagios por persona (R_t) es inferior a uno, la propagación de la epidemia se extinguirá con probabilidad uno. En promedio, un contagiado será reemplazado por menos de otro individuo. Finalmente, la cantidad de enfermos están destinados a disminuir.

Un modelo determinista falla completamente en mostrar este efecto, lo cual constituye una diferencia entre el enfoque estocástico y el determinista, ya que muchos gobiernos emplean todo su potencial y esfuerzo en entrar en ese punto o pasar por debajo de él, para detener el avance de la epidemia [33].

Los modelos deterministas son más simples, pero contiene menos información. Además, las varianzas o la probabilidad de extinción pueden ser sólo calculadas en el marco estocástico [33].

Asimismo, los modelos estocásticos no necesitan encontrar una solución óptima debido a que los estados futuros del sistema nunca serán los mismos como consecuencia de las causas de variación incontables a las que está sometido dicho sistema. Los modelos probabilistas producen entonces información llamada empírica o experimental, por lo que se hace necesario hacer uso del proceso de inferencia.

Cuando no es posible del todo o es muy complejo encontrar una solución óptima, el procedimiento de cálculo es ineficiente, entonces se recurre a procedimientos sistemáticos que reducen el esfuerzo necesario para encontrar una solución satisfactoria, aunque no necesariamente esta solución es óptima. Este tipo de procedimientos se conocen con el nombre de modelos heurísticos, y corresponden tanto a modelos determinísticos como a probabilísticos [34, 35].

Por todos los puntos antes expuestos se decide escoger para estimar los parámetros de la epidemia de SARS-Cov-2 en Cuba una modelación estocástica a través del método Bayesiano.

Capítulo II

Enfoque bayesiano y simulación del modelo

II Enfoque bayesiano y simulación del modelo

II.1 Enfoque Bayesiano

La inferencia bayesiana constituye un enfoque alternativo para el análisis estadístico de datos que contrasta con los métodos convencionales de inferencia, en parte por hacer un manejo subjetivo, no frecuentista, del concepto de probabilidad [36-38].

El florecimiento de la metodología bayesiana en los últimos años se expresa en un número creciente de artículos y, más recientemente, en no pocos libros especializados. La necesidad de un enfoque alternativo puede ser fácilmente comprendida; resulta natural que se aspire a contar con un procedimiento inferencial libre de las serias impugnaciones que se hacen a las pruebas de significación [36-38].

De modo que va resultando crecientemente obvia la necesidad de un nuevo paradigma, que esté en lo posible exento de las objeciones que cuestionan al procedimiento inferencial viejo. En este sentido, el enfoque bayesiano constituye una alternativa atractiva puesto que cumple las siguientes condiciones [36-38]:

1. No está lastrado por el tamaño muestral en el sentido de que, si éste es pequeño, el impacto informativo también lo será, pero, en cualquier caso, cuanto mayor sea, más adecuadamente se podrá valorar la realidad que esa muestra representa.
2. Lejos de operar en un vacío total de información, el modelo de análisis bayesiano exige contemplar formal y explícitamente el conocimiento previo.
3. Valora la credibilidad o verosimilitud de las hipótesis en lugar de obligarnos a adoptar decisiones dicotómicas sobre ellas

II.1.1 Modelos Jerárquicos

Un tipo de modelo empleado para explicar las características de determinados procesos son los modelos jerárquicos. Estos son modelos probabilísticos para colecciones de variables formulados como combinaciones de diversas

componentes denominadas niveles, capas o etapas. Esta estrategia es útil especialmente en la construcción de modelos complejos [39, 40].

Este tipo de modelización permite enlazar diferentes fuentes de información, empleando relaciones entre las variables estudiadas. Esta capacidad de adaptarse a situaciones complejas y gracias al desarrollo de técnicas inferenciales asequibles mediante simulación han permitido que se conviertan en una herramienta principal en la modelización estadística de problemas epidemiológicos [39-41].

La construcción de un modelo jerárquico se hace a través de las distribuciones condicionales, con las cuales, se construye un encadenamiento de dependencias lo que ayuda a potenciar la conexión entre modelos complejos. Este tipo de metodología permite introducir no sólo factores de confusión indeterminados sino combinar fuentes de variabilidad y unir modelos parciales.

Así mismo, desde una perspectiva Bayesiana las capas de un modelo son vistas cada una, como un proceso estocástico compuesto de observaciones, factores ocultos y parámetros a estimar. A través del Teorema de Bayes es posible que la información de un dato se transfiera a factores asociados a otro dato, para esto se requiere incorporar incertidumbre tanto en las observaciones como en los parámetros de interés [39, 42, 43].

El proceso de aprendizaje a través de la distribución posterior es enorme y constante, es así, que los modelos jerárquicos Bayesianos se convierten en una herramienta potencial para el análisis de problemas complejos.

II.1.2 Paradigma Bayesiano

La estadística/inferencia bayesiana supone un cambio de paradigma en el proceso de inferencia en auge en los últimos años. Esta es un enfoque de la inferencia basado en gran medida en el teorema de Bayes y que consta de los siguientes pasos principales [40, 44]:

1. Definir un modelo de probabilidad total, que recoja la distribución de probabilidad conjunta para todos los parámetros cuantificables en un

problema. El modelo debe ser coherente con el conocimiento disponible sobre la materia y el proceso de recolección de datos.

2. Actualizar nuestro conocimiento con los datos observados; calcular e interpretar la distribución a *posteriori*, esto es, la distribución de probabilidad condicional de los parámetros no observados de interés, dados los datos observados. La distribución a *posteriori* de los parámetros es entonces proporcional al producto de la información que teníamos sobre su distribución a *priori*, que recoge el conocimiento inicial acerca de los parámetros, por su función de verosimilitud dados los datos observados.
3. Evaluar la idoneidad del modelo y las implicaciones de la distribución a *posteriori* resultante. El investigador debe analizar si el modelo se ajusta adecuadamente a los datos, si se extraen conclusiones razonables, y cómo se integran los resultados en las premisas del paso 1.

La idea básica, es resolver un problema dividiéndolo en sub-problemas más simples. Cada uno de estos estados pueden a su vez dividirse en muchos sub-estados. En los últimos 40 años, se han producido grandes avances en la inferencia bayesiana. La motivación principal del pensamiento bayesiano es que facilita interpretaciones razonadas ante conclusiones estadísticas. De esta forma, un intervalo de probabilidad bayesiano para un valor desconocido de interés puede calificarse como un intervalo de alta probabilidad de contener dicho valor, en contraste con el intervalo de confianza, que sólo debe interpretarse de acuerdo a una proporción de resultados en una colección de inferencias con muestras equivalentes [44, 45].

Asimismo, bajo la perspectiva Bayesiana, la incertidumbre o falta de información sobre el parámetro θ puede ser incorporada a través de distribuciones previas, considerando este parámetro como una variable aleatoria. Sea $\pi(\theta|\lambda)$ la distribución previa, donde λ es un vector de hiperparámetros. Si λ es conocida la inferencia sobre θ se hace a partir de la distribución posterior de $p(\theta|y, \lambda)$ que se obtiene mediante el teorema de Bayes, el cuál combina la distribución previa y la verosimilitud [39, 43, 45, 46]:

$$p(\theta | y, \lambda) = \frac{p(y, \theta | \lambda)}{p(y | \lambda)} = \frac{p(y, \theta | \lambda)}{\int p(y, \theta | \lambda) d\theta} = \frac{f(y, \theta) \pi(\theta | \lambda)}{\int f(y, \theta) \pi(\theta | \lambda) d\theta} \quad (20)$$

En la práctica, λ no es conocido y por tanto, es necesario definir un segundo estado para los hiperparámetros (distribución para $p(\lambda)$), quedando como [39, 43, 45, 46]:

$$p(\theta | y) = \frac{p(y, \theta)}{p(y)} = \frac{\int f(y, \theta) \pi(\theta | \lambda) h(\lambda) d\lambda}{\int \int f(y, \theta) \pi(\theta | \lambda) h(\lambda) d\theta d\lambda} \quad (21)$$

Al condicionar en el valor conocido de los datos y , usando la regla de Bayes se obtiene la distribución posterior:

$$p(\theta, y) = p(y, \theta) = p(y | \theta) p(\theta) \quad (22)$$

Entonces sustituyendo (22) en (21) se tiene:

$$p(\theta | y) = \frac{p(y | \theta) p(\theta)}{p(y)} \quad (23)$$

Pero como la distribución de y es conocida y no depende de θ se puede considerar $p(y)$ como una constante.

Entonces específicamente, en el enfoque Bayesiano, hay tres componentes asociados con la estimación de parámetros: la distribución *a priori*, la función de verosimilitud y la distribución *a posteriori*. Estos tres componentes son formalmente combinados por las reglas de Bayes de la siguiente manera [46]:

Distribución a posteriori \propto *distribución a priori* \times *la función de verosimilitud*

Esto es:

$$p(\theta | y) \propto p(y | \theta) p(\theta) \quad (24)$$

En términos simples, la ecuación anterior indica que la información contenida en la muestra (reflejada en la función de verosimilitud) es combinada con la información de otras fuentes (resumida por la distribución a *priori*) para obtener la distribución a *posteriori*. La distribución a *posteriori* contiene todo el conocimiento disponible acerca de los parámetros en el modelo [46].

La solución de las ecuaciones integrales antes descritas presenta un alto nivel de complejidad por lo que se han desarrollado diferentes metodologías para resolver este tipo de ecuaciones. Una de las más utilizadas es a través de las cadenas de Markov mediante la simulación de Monte Carlos (MCMC). En este trabajo se utiliza además el algoritmo de Metrópolis-Hastings.

II.1.3 Métodos Markov Chain Monte Carlo (MCMC)

El método de Monte Carlo fue inventado por Stanislaw Ulam en 1940 mientras trabajaba en el Proyecto Manhattan en Los Álamos, EE.UU. Lo llamaron así porque la base para que funcione bien es un buen generador de números aleatorios. Con los números aleatorios obtenidos se generan múltiples casos con los cuales es posible hacer varias pruebas que permiten realizar cálculos. El poder del método Monte Carlo reside en que repitiendo algo muchas veces se puede obtener aproximaciones sin hacerlo analíticamente con papel y bolígrafo [47].

Entonces, la unión de estos dos términos en los métodos de Markov Monte Carlo MCMC (generación de cadenas de Markov para usarlas en estudios Monte Carlo) están diseñadas con la finalidad de poder estudiar empíricamente las características de distribuciones complejas. Básicamente consisten en lo siguiente: Dada una distribución $p(\theta)$ completamente conocida, se trata de generar una o varias realizaciones de una cadena de Markov cuya distribución estacionaria sea $p(\theta)$. Esas realizaciones se utilizarán posteriormente para obtener, por Monte Carlo, aproximaciones a todas las características de $p(\theta)$ que se desean conocer [39, 43, 45, 46].

Desde la perspectiva de las técnicas MCMC, el estudio de las cadenas de Markov es completamente distinto. Se parte suponiendo que la distribución que deseamos es invariante, $\pi(\theta)$, y se pretende construir un núcleo de transición que produzca una cadena de Markov con distribución estacionaria $\pi(\theta)$ para la que las medias sean estimadores consistentes y podamos aplicar alguna versión del Teorema del Límite [39, 43, 45, 46]. La mayoría de estos algoritmos tienen los siguientes pasos [47]:

1. Se empieza en la posición actual.
2. Se propone moverse a una nueva posición (investigar los puntos de alrededor).
3. Aceptar o rechazar la nueva posición basándose en la cohesión de la posición a los datos y las distribuciones *a priori* (preguntar si esos puntos están cerca o no de nuestro objetivo).
4. Si acepta: Desplazar a la nueva posición. Volver al paso 1.
5. Si no: No se traslada a la nueva posición. Volver al paso 1.
6. Después de un gran número de iteraciones, devolver todas las posiciones aceptadas.

De esta manera, el algoritmo se mueve en la dirección hacia las regiones donde existen las distribuciones *a posteriori* $\pi(\theta)$, y las muestras realizadas se recogen en este proceso una vez que se consigue llegar a la posterior. Si la posición actual del algoritmo MCMC se encuentra en zonas de probabilidad extremadamente bajas, lo que suele ocurrir cuando el algoritmo comienza, el algoritmo se moverá en posiciones que no sean de la posterior pero que son mejores que todo lo que hay cerca. Por lo tanto, los primeros movimientos no reflejan las posterioridades. A continuación, se detalla concisamente algunos de los algoritmos más populares de MCMC:

- Algoritmo Metrópolis-Hastings Es el más simple y flexible de los algoritmos MCMC. El muestreo por rechazo es el precursor de este algoritmo, el cual proporciona de una manera muy eficiente la simulación de valores a partir de una variable aleatoria. En este caso, los valores

aleatorios se toman de distribuciones escogidas de tal manera que se comporten asintóticamente como valores de la distribución objetivo [46, 47].

- Algoritmo Hamiltoniano de Monte Carlo (HMC) El HMC surgió de la ineficiencia del Algoritmo de Metrópolis-Hastings para muchos modelos. En este contexto, y cuando el muestreo se realiza a partir de variables continuas, el HMC puede resultar muy potente. Evita el comportamiento de paso aleatorio entre los estados, al simular un sistema físico gobernado por la dinámica hamiltoniana, evitando distribuciones excesivamente complejas en el proceso. Por tanto, las muestras en el algoritmo HMC se obtienen mediante la simulación de un sistema físico, en el que las partículas se mueven por un marco de altas dimensiones, sujetas a energías potenciales y cinéticas. Utiliza también la eficiencia demostrada del muestreo basado en el gradiente, usado en disciplinas de redes neuronales y *machine learning* [45, 48-50].
- No U-turn Sampler (NUTS) La mayor desventaja del algoritmo HMC es la cantidad de ajustes que se requieren para que el muestreo sea efectivo. Entre los parámetros que deben ser especificados por el usuario está la escala de la distribución del momento, el tamaño del paso de salto para el algoritmo y el número de pasos a dar. Cuando estos parámetros están mal elegidos, el algoritmo HMC sufre de ineficiencia. Un algoritmo MCMC eficiente busca la optimización manteniendo un equilibrio, y aunque el HMC puede ajustarse sobre la marcha, requiere costosas ejecuciones continuadas para hacerlo. Por ello, se creó el algoritmo de muestreo ‘sin giro’ (NUTS) que ajusta automáticamente los parámetros de tamaño y número de pasos, sin intervención del usuario. Como resultado, NUTS es el algoritmo de muestreo por defecto para variables continuas en muchos de los programas de simulación [45, 48-50].

La complejidad de las estructuras estocásticas que se derivan a partir de la formulación de un modelo jerárquico Bayesiano dificulta su inferencia. Esta dificultad se debe a la variedad de posibilidades para la especificación de la distribución previa y a la dificultad de resumir la distribución posterior resultante.

Sin embargo, en la actualidad están disponibles algoritmos y algunas herramientas informáticas que permiten realizar tal inferencia a pesar de su complejidad.

Esta inferencia supone un desafío computacional ya que, en problemas reales, las integrales requeridas para hacer las estimaciones son generalmente intratables al no tener una forma analítica cerrada. Este obstáculo numérico se resuelve usando métodos de integración MCMC. Por lo que se utiliza el software R para simular el problema mediante la librería Rstan.

II.2 Simulación del Modelo

II.2.1 R

R es un lenguaje de programación con software libre, que, a su vez, es un tipo de lenguaje interpretado. Es decir, los comandos o instrucciones son ejecutados directamente sin ningún tipo de software externo o compilación de datos de lenguaje máquina. Este tipo de lenguaje en R consiste en la planificación o creación de programas, aplicaciones o herramientas de medición estadística y gráfica [51-53].

Del mismo modo, su capacidad permite que se puedan desarrollar distintas técnicas de medición, como lineales o no lineales, análisis de series de tiempo, agrupaciones, estadísticas clásicas o clasificaciones.

El código R es desarrollado actualmente por R Development Core Team, y su proyecto es colaborativo y abierto, por lo que otros usuarios programadores pueden ser partícipes del proyecto. Pudiendo publicar paquete de códigos que extienden su configuración básica. Este tipo de lenguaje permite [51-53]:

- El manejo efectivo de almacenamiento de datos.
- Desarrollo de herramientas para cálculo de matrices y estadística.
- Utilidades gráficas.
- Un lenguaje de programación diverso, con bucles, saltos condicionales, utilidades recursivas y más.

- Su documentación posee un formato LaTeX, que proporciona documentación completa en formato físico o digital.

Al ser un tipo de lenguaje que está dedicado a análisis de *big data*, se aplica en situaciones que requieran análisis profundo. Así, se puede manipular, procesar y visualizar gráficas con dichos datos.

El análisis de datos conlleva un proceso bastante riguroso, y en cada uno de las partes que componen ese proceso, está presente el lenguaje de programación en R [51-53].

- Adquisición de datos en las fuentes disponibles,
- Preparación de los datos,
- Análisis de los datos,
- Aplicación de resultados

Entre sus ventajas se tiene [51-53]:

- Acumula, en un único programa, todo lo necesario para analizar los datos con eficacia. Las diversas herramientas de medición de datos y de realización de estadísticas se convierten en imprescindibles para ahorrar tiempo y esfuerzo.
- Incluye modelos lineales y no lineales usados en el sector de la estadística. Es idóneo para la clasificación y el agrupamiento de los datos.
- Incluir gráficos permite estudiar y sacar conclusiones más precisas en menor tiempo
- El lenguaje posibilita el desarrollo de un lenguaje de programación limpio que facilita la gestión de datos.
- Funciona sin problemas en sistemas operativos UNIX, Linux, macOS y Windows.
- Es gratuito.
- Admite la colaboración altruista de cualquier usuario que quiera ir mejorando el lenguaje con nuevos aportes.

El amplio espectro en términos de colaboración que posee R, permite la utilización y creación de un inmenso número de librerías, entre las que se encuentra Rstan la librería especializada en MCMC que se utiliza en este trabajo.

II.2.2 RStan

Stan es una plataforma de última generación para el modelado estadístico y el cálculo estadístico de alto rendimiento. Miles de usuarios confían en Stan para el modelado estadístico, el análisis de datos y la predicción en las ciencias sociales, biológicas y físicas, la ingeniería y los negocios. Los usuarios pueden realizar [54-56]:

- inferencia bayesiana completa en MCMC (NUTS, HMC)
- inferencia bayesiana aproximada con inferencia variacional (ADVI)
- estimación de máxima verosimilitud penalizada con optimización (L-BFGS)

El objetivo del proyecto Stan es proporcionar un lenguaje de programación probabilístico flexible para modelado estadístico junto con un conjunto de herramientas de inferencia para ajustar modelos que son robustos, escalable y eficiente.

Stan difiere de BUGS y JAGS de dos maneras principales. En primer lugar, Stan se basa en un nuevo lenguaje de programación probabilístico imperativo que es más flexible y expresivo que los lenguajes de modelado gráfico declarativo que subyacen a BUGS o JAGS, en formas tales como la declaración de variables con tipos y variables locales de apoyo y sentencias condicionales. Segundo, el método MCMC de Stan se basa en el HMC, un muestreador más eficiente y robusto que el muestreo de Gibbs o Metrópolis-Hastings para modelos con posteriores complejos [54-56].

A través de la librería Rstan se llama desde R al motor de Stan y esta permite realizar las simulaciones correspondientes mediante la función *sampling* y el *método no tur sampling*.

Capítulo III

Modelación para Cuba

III Modelación para Cuba

Existen muchos tipos de modelos matemáticos que pueden ser utilizados para describir los parámetros de la epidemia de Covid-19. Seleccionar y construir un modelo de este nivel de complejidad, implica un profundo análisis pues es necesario obtener estimaciones con elevados niveles de exactitud. Así mismo, es deseable que el modelo [57-59]:

- sea simple (parsimonia),
- alcance sólo objetivos asequibles. No debe aspirarse a imitar la realidad sino sólo a resaltar aquellos aspectos de interés para su aplicación (modesta),
- reproduzca en la medida de lo posible el funcionamiento del sistema y genere valores para las variables de salida similares a los observados en la realidad (exactitud).
- los resultados se deben poder comparar con datos reales y determinar de este modo el grado de exactitud del modelo (verificabilidad).
- funcione bien por las razones correctas.

Según lo planteado hasta el momento, se escoge para describir los parámetros epidemiológicos del Covid-19 en Cuba un modelo empírico debido a los procesos que intervienen en las características de la epidemia, no se conocen exactamente las leyes físicas que los rigen por lo tanto algunos de los parámetros se describen a partir de relaciones estadísticamente significativas con las variables, así mismo, será un modelo estocástico ya que incluye generadores de procesos aleatorios dentro del mismo que modifiquen ligeramente algunas de las variables, de esa manera, para un mismo conjunto de datos de entrada, las salidas serán distribuciones probabilísticas, además será distribuido y dinámico, el primero se debe a que durante el proceso de estimación se contará con facetas intermedias para estimar parámetros que a su vez estos darán lugar a otras estimaciones y la segunda clasificación es porque los parámetros se estiman en función del tiempo [12, 16-18].

III.1 Descripción del modelo

Una descripción a grandes rasgos del modelo sería: a partir de las variables independientes se estiman las personas susceptibles en la población, pues para que haya transmisión de un determinado patógeno en una comunidad es necesario que existan personas que puedan contraer la infección, posteriormente se estima el número reproductivo en el tiempo (R_t) a partir del número reproductivo básico (R_0) que posee el patógeno en cuestión, así mismo y utilizando el mismo procedimiento se estima el IFR, seguido se determina la cantidad de personas contagiadas a partir de la cantidad de contagios anteriores, luego, a partir de los contagios y de la proporción de personas que se recuperan de la infección se estiman los casos activos, después haciendo uso del IFR y de la cantidad de contagios se obtienen los fallecidos, y por último se ajusta la distribución de los fallecidos a partir de los fallecidos reportados por las instancias de salud. Este modelo sigue la premisa de que el dato de los fallecidos es el dato que menor sesgo incorporado posee, Figura 6.

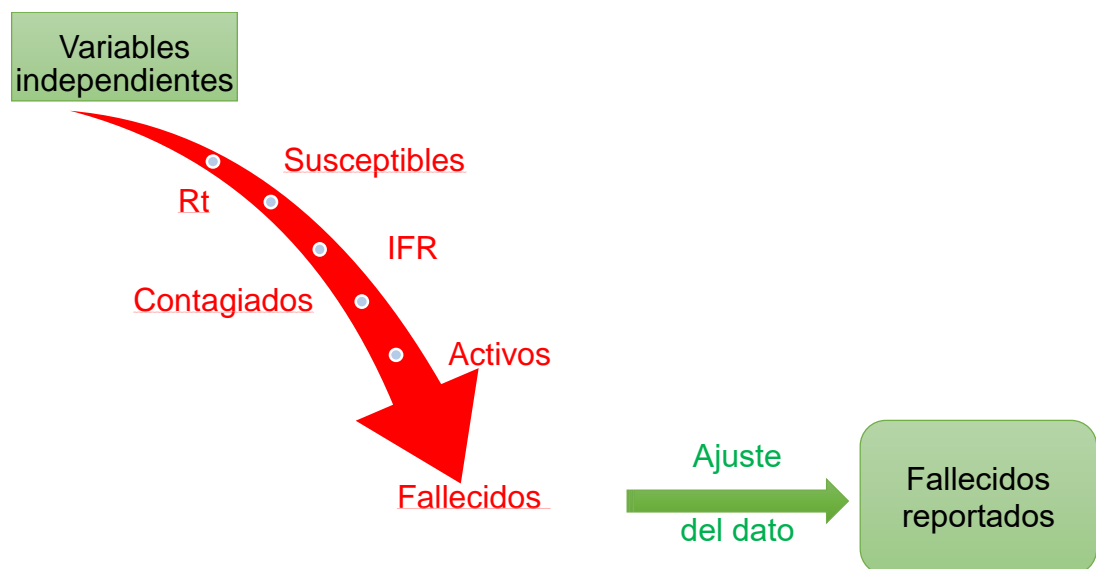


Figura 6. Modelo para Cuba

Cada variable estimada depende de otras variables o parámetros. Para una mejor comprensión, a continuación, se explican las dependencias de cada una de las variables que intervienen en el modelo.

III.1.1 Variables independientes:



Figura 7. Variables independientes del modelo

Dado que la epidemia de Covid-19 en Cuba se ha prolongado más de dos años, las estimaciones en este trabajo se realizarán semanalmente, lo que quiere decir que cada unidad de tiempo representa siete días.

III.1.2 Susceptibles dependencias:

- tamaño de la población, en el primer momento de la epidemia las personas susceptibles son toda la población,
- tiempo, a medida que pasa el tiempo se van contagiando, vacunando y muriendo las personas, lo que provoca una disminución de la población susceptible,
- contagiados estimados y vacunados,
- tiempo a ser susceptible después de estar vacunado o de haber superado la enfermedad, los anticuerpos que se generan durante la infección o la vacunación no protegen para siempre, así que las personas susceptibles también dependen de estos tiempos [60].

III.1.3 Número reproductivo en el tiempo

- susceptibles, a medida que la población susceptible disminuye se le dificulta al patógeno encontrar un hospedero,
- número reproductivo básico, el número reproductivo en el tiempo se estima como el producto del número reproductivo básico por una proporción que representa el incremento o detrimento de este.
- variantes mutaciones, cada variante posee características que las diferencian en cuanto a su transmisibilidad.
- tiempo, el parámetro se estima en función al tiempo,
- momento en que se encuentra la epidemia, el número reproductivo en el tiempo depende del momento epidémico pues las medidas adoptadas por el gobierno y la forma de actuar de la población son diferentes.

III.1.4 IFR

- variantes mutaciones, cada variante posee características que las diferencian en cuanto a su letalidad.
- vacunados, las personas vacunadas que contraen el patógeno tienen menos probabilidad de fallecer.
- contagios reportados, se buscan relaciones estadísticamente significativas con la cantidad de contagios reportados, cuando el número de contagios es muy grande los sistemas de salud pueden colapsar o ser de menor accesibilidad para las personas enfermas.
- tiempo, el parámetro se estima en función al tiempo,
- momento en que se encuentra la epidemia,

III.1.5 Contagiados

- contagios y número reproductivo en el tiempo, se relacionan todas estas variables similares al modelo Logístico de Crecimiento de Richards.
- intervalo serial, el tiempo desde que una persona se infecta hasta que presenta síntomas, los nuevos contagios están relacionados con los contagios anteriores mediante este parámetro [61],

- tiempo, el parámetro se estima en función al tiempo,

III.1.6 Activos

- contagios los casos activos se determinan como la cantidad de casos menos los recuperados.
- tiempo desde que una persona se infecta hasta que se recupera, se relacionan los recuperados con los contagiados mediante este parámetro [62],
- tiempo, el parámetro se estima en función al tiempo,

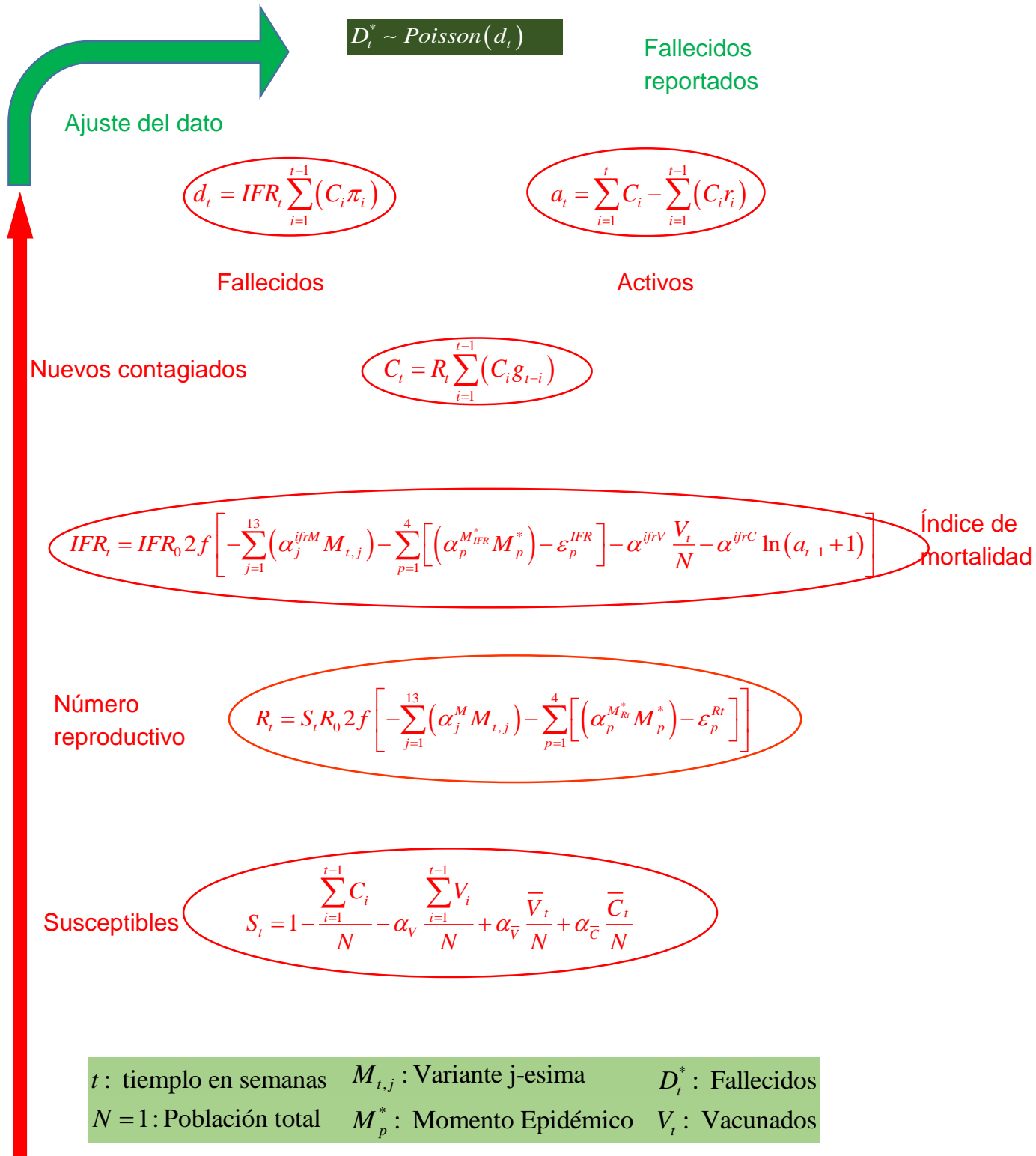
III.1.7 Fallecidos

- Contagios, IFR, dada la proporción de personas que fallecen se estiman la cantidad de muertes
- tiempo desde que una persona se infecta hasta que se muere, se relacionan los fallecidos con los contagiados mediante este parámetro [62],
- tiempo, el parámetro se estima en función al tiempo,

Las relaciones entre las variables y parámetros descritos se pueden representar mediante una red Bayesiana. Estas son grafos dirigidos acíclicos cuyos nodos representan variables aleatorias en el sentido de Bayes, las mismas pueden ser cantidades observables, variables latentes, parámetros desconocidos o hipótesis [63-65].

Para especificar completamente la red bayesiana y por lo tanto representar plenamente a la distribución de probabilidad conjunta, es necesario detallar para cada nodo la distribución de probabilidad condicionada, esta puede tener cualquier forma. A veces sólo restricciones sobre una distribución son conocidas. En la figura 7 se presentan los nodos centrales de la red bayesiana asociada al complejo modelo construido para describir el comportamiento de los parámetros de la epidemia de SARS-Cov-2 en Cuba.

III.2 Red Bayesiana



Variables
independientes

Figura 7. Nodos centrales de la Red Bayesiana del modelo para Cuba.

A continuación, se describen las relaciones que tiene cada variable y parámetros con los nodos centrales de la Red Bayesiana del modelo matemático cubano.

Susceptibles:

Ecuación general:

$$S_t = 1 - \frac{\sum_{i=1}^{t-1} C_i}{N} - \alpha_V \frac{\sum_{i=1}^{t-1} V_i}{N} + \alpha_{\bar{V}} \frac{\bar{V}_t}{N} + \alpha_{\bar{C}} \frac{\bar{C}_t}{N}$$

$$\begin{aligned} S_1 &= 1 \\ \bar{V}_t &= \sum_{i=1}^{t-1} \alpha_V V_i c_{t-i} & c_{t-i} &= \int_0^{t-i} c(x) dx \\ & & c(x) & \text{f.d. de v.a.} \sim \text{Gamma}(\text{shape}=2.061856, \text{rate}=0.003436426) \\ \bar{C}_t &= \sum_{i=1}^{t-1} C_i c_{t-i} & \alpha_V &\sim N(0.9, \sigma_{\alpha_V}), \sigma_{\alpha_V} \sim N(0, 0.09) \end{aligned}$$

Donde:

$\alpha_{\bar{V}} \bar{V}_t, \alpha_{\bar{C}} \bar{C}_t$ cantidad de vacunados y contagiados, respectivamente, que se hacen susceptibles,

$1 - \alpha_{\bar{V}}, 1 - \alpha_{\bar{C}}$ proporción de vacunados y contagiados, respectivamente, que no se convierten en personas susceptibles,

c distribución del tiempo desde que una persona se vacuna o se infecta hasta que vuelve a ser susceptible,[60].

Número reproductivo

Ecuación general,

$$R_t = S_t R_0 2f \left[- \sum_{j=1}^{13} (\alpha_j^M M_{t,j}) - \sum_{p=1}^4 \left[(\alpha_p^{M^*} M_p^*) - \varepsilon_p^{Rt} \right] \right]$$

$$f(x) = \frac{1}{1 + e^{-x}}, R_1 = R_0,$$

$R_0 \sim N(1.40, \kappa)$, $\kappa \sim N^+(0, 0.1)$ El número reproductivo básico tiene media 1.40 según determina *Guinovart* en [19].

$$\alpha_j^M \sim N(0, 0.5), \quad \alpha_p^{M_{Rt}^*} \sim N(0, \sigma_p^{M_{Rt}^*}), \quad \sigma_p^{M_{Rt}^*} \sim N(0, 0.5)$$

El elemento auto recursivo se describe:

$$\begin{aligned} \varepsilon_p^{Rt} &\sim N(\rho_1^{Rt} \varepsilon_{p-1}^{Rt} + \rho_2^{Rt} \varepsilon_{p-2}^{Rt}, \sigma_p^{Rt*}) \\ \varepsilon_1^{Rt} &= 0, \varepsilon_2^{Rt} \sim N(0, \sigma_p^{Rt*}), \quad \sigma_p^{Rt*} = \sigma_p^{Rt} \sqrt{1 - (\rho_1^{Rt})^2 - (\rho_2^{Rt})^2 - \frac{2(\rho_1^{Rt})^2 \rho_2^{Rt}}{1 - \rho_2^{Rt}}} \\ \rho_1^{Rt} &\sim N(0.8, 0.05), \rho_2^{Rt} \sim N(0.1, 0.05) \quad \sigma_p^{Rt} \sim N^+(0, 0.2) \end{aligned}$$

En estadística un modelo autorregresivo (AR) es una representación de un proceso aleatorio, en el que la variable de interés depende de sus observaciones pasadas. Específicamente, la variable de interés o de salida, depende linealmente de sus valores anteriores. Por esto decimos que existe dependencia lineal entre las distintas observaciones de la variable.

Una forma de entender un modelo autorregresivo es un proceso que luego de enfrentado a una perturbación tarda tiempo en regresar a su equilibrio de largo plazo. [29, 66-68]

Índice de mortalidad

Ecuación general,

$$IFR_t = IFR_0 2f \left[-\sum_{j=1}^{13} (\alpha_j^{ifrM} M_{t,j}) - \sum_{p=1}^4 \left[(\alpha_p^{M_{IFR}^*} M_p^*) - \varepsilon_p^{IFR} \right] - \alpha^{ifrV} \frac{V_t}{N} - \alpha^{ifrC} \ln(a_{t-1} + 1) \right]$$

Donde:

$$f(x) = \frac{1}{1 + e^{-x}}, \quad IFR_1 = IFR_0,$$

$$IFR_0 \sim N(0.01, \kappa^{IFR}), \quad \kappa^{IFR} \sim N^+(0, 0.0001)$$

$$\alpha_j^{ifrM} \sim N(0, 0.5), \quad \alpha_p^{M_{IFR}^*} \sim N(0, \sigma_p^{M_{IFR}^*}), \quad \sigma_p^{M_{IFR}^*} \sim N(0, 0.5)$$

$$\alpha^{ifrV} \sim N(0.01, \kappa^{ifrV}), \kappa^{ifrV} \sim N^+(0, 0.1), \alpha^{ifrC} \sim N(0.01, \kappa^{ifrC}), \kappa^{ifrC} \sim N^+(0, 0.1)$$

El elemento auto recursivo se describe:

$$\begin{aligned} \varepsilon_p^{IFR} &\sim N(\rho_1^{IFR} \varepsilon_{p-1}^{IFR} + \rho_2^{IFR} \varepsilon_{p-2}^{IFR}, \sigma_p^{IFR*}) \\ \varepsilon_1^{IFR} &= 0, \varepsilon_2^{IFR} \sim N(0, \sigma_p^{IFR*}), \quad \sigma_p^{IFR*} = \sigma_p^{IFR} \sqrt{1 - (\rho_1^{IFR})^2 - (\rho_2^{IFR})^2 - \frac{2(\rho_1^{IFR})^2 \rho_2^{IFR}}{1 - \rho_2^{IFR}}} \\ \rho_1^{IFR} &\sim N(0.8, 0.05), \rho_2^{IFR} \sim N(0.1, 0.05) \quad \sigma_p^{IFR} \sim N^+(0, 0.2) \end{aligned}$$

$$\alpha_j^{ifrM} \sim N(0, 0.5), \alpha^{ifrC} \sim N(0, \gamma^{ifrC}), \alpha^{ifrV} \sim N(0, \gamma^{ifrV}), \{\gamma^{ifrV}, \gamma^{ifrC}\} \sim N^+(0, 0.5),$$

Nuevos contagiados

Ecuación general,

$$C_t = R_t \sum_{i=1}^{t-1} (C_i g_{t-i})$$

Donde:

$$g_{t-i} = \int_{7(t-i)-3.5}^{7(t-i)+3.5} g(x) dx \quad [61]$$

$g(x) f.d. \text{ de v.a. } \sim \text{Gamma}(\text{shape}=415.5325, \text{rate}=78.40237)$

Activos

Ecuación general,

$$a_t = \sum_{i=1}^t C_i - \sum_{i=1}^{t-1} (C_i r_{t-1-i})$$

Donde:

$$a_1 = C_1$$

$$r_{t-i} = \int_0^{t-i} r(x) dx \quad [62]$$

$r(x)$ f.d. de v.a. $\sim \text{Gamma}(\text{shape}=18.62067, \text{rate}=1/1.18892)$

r representa la distribución del intervalo de tiempo desde la infección hasta la recuperación.

Fallecidos

Ecuación general,

$$d_t = IFR_t \sum_{i=1}^{t-1} (C_i \pi_{t-i})$$

Donde:

$$\pi_{t-i} = \int_{7(t-i)-3.5}^{7(t-i)+3.5} \pi(x) dx \quad [62]$$

$\pi(x)$ f.d. de v.a. $\sim \text{Gamma}(\text{shape}=21.33660, \text{rate}=1/1.03174)$

π representa la distribución del intervalo de tiempo desde la infección hasta el deceso.

Distribución de las muertes reportadas

$$D_t^* \sim \text{Poisson}(d_t),$$

Los datos siguen una distribución de Poisson pues la frecuencia con que se dan los máximos en la distribución de las nuevas muertes es extremadamente baja, un factor que contribuye a amplificar este fenómeno es que el estudio se realiza por semanas, por lo que al sumar las muertes de las semanas donde se está en un pico epidémico, se generan resultados extremadamente altos y muy poco probables, siendo la distribución de Poisson la que mejor explica estos fenómenos [69].

En este modelo se estiman 1268 parámetros de la epidemia por lo que se requiere de un gasto computacional importante.

III.3 Evaluación del modelo

El uso de métodos MCMC presenta dos dificultades que no enfrentan los métodos de Monte Carlo de muestras independientes [70, 71].

1. El primer problema es determinar cuándo una cadena de Markov inicializada aleatoriamente ha convergido a su distribución de equilibrio.
2. El segundo problema es que los muestreos de una cadena de Markov pueden estar correlacionados y, por lo tanto, el teorema central del límite en el error de estimación ya no se aplica.

Las herramientas de análisis posterior de Stan calculan una serie de estadísticas resumidas, estimaciones y diagnósticos para muestras de cadena de Markov Monte Carlo (MCMC). Los estimadores y diagnósticos de Stan son más sólidos frente a la no convergencia y las correlaciones de la cadena de Markov [70, 71].

El algoritmo tiene como hipótesis que una cadena de Markov genera muestras de la distribución objetivo solo después de que ha convergido al equilibrio. En teoría, la convergencia solo se garantiza asintóticamente a medida que el número de sorteos crece sin límite. En la práctica, se deben aplicar diagnósticos para monitorear la convergencia cuando el número de sorteos disponibles es realmente finito [70, 71].

El modelo se ejecutó desde el software R en el motor de compilación de Stan a través de la función *'sampling'* del paquete *'rstan'* de R. El cómputo utilizado fue una laptop de trabajo la cual cuenta con un procesador *'Intel(R) Core(TM) i7-6700HQ CPU @ 2.60 GHz'* y *16 GB de RAM*.

El modelo fue ejecutado en cuatro cadenas paralelas siendo en cada cadena 200 000 iteraciones y de ellas 100 000 fueron de calentamiento, lo que implica que en total se realizaron 800 000 iteraciones. El tiempo de procesamiento total fue de 104239 segundos lo que equivale aproximadamente a 30 horas generándose aproximadamente 15.8 GB de datos.

Es importante realizar un análisis de convergencia de los parámetros del modelo y además del ajuste de los datos.

III.3.1 R-hat

Una forma de monitorear si una cadena ha convergido a la distribución de equilibrio es comparar su comportamiento con otras cadenas inicializadas aleatoriamente, le llamaremos R-hat (\hat{R}). Este indicador mide la relación entre la varianza promedio de las muestras dentro de cada cadena y la varianza de las muestras agrupadas en las cadenas; si todas las cadenas están en equilibrio, estas serán iguales y R-hat será uno. Si las cadenas no han convergido a una distribución común, el R-hat será mayor que uno. Así mismo, en el algoritmo las cadenas de Markov se inicialicen con valores iniciales difusos para los parámetros y se muestrean hasta que todos los valores para R-hat estén por debajo de 1.1, por esto es que se acepta un R-hat hasta 1.1 [70-75].

En la figura 8 se muestra el valor que toma el R-hat de cada parámetro estimado en el modelo.

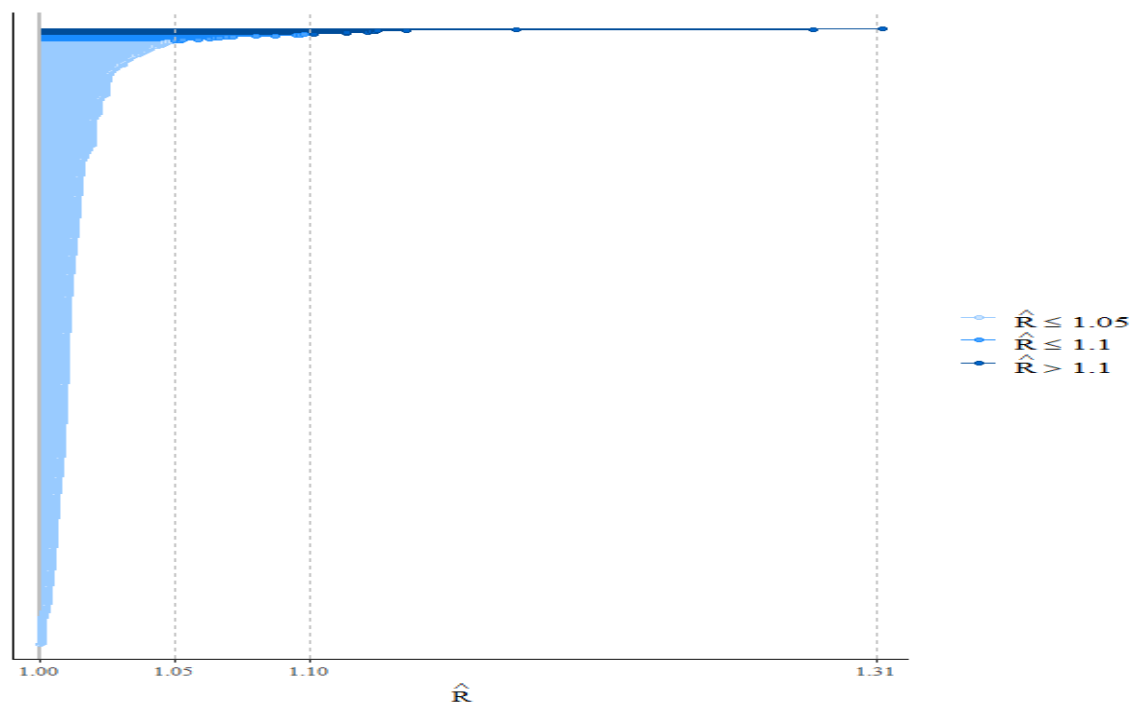


Figura 8. Valor del R-hat para cada parámetro estimado en el modelo.

De un total de 1268 parámetros estimados en el modelo existen 11 que presentan un R-hat mayor que 1.10. Este es un problema común en este tipo de modelos y se resuelve aumentando la cantidad de iteraciones.

Los parámetros que presentaron un R-hat mayor que 1.10 son:

$$\alpha_j^M, j = 5, 6, 12, 14, \alpha_p^{M_{Rt}^*}, p = 1, 3, \alpha_p^{M_{IFR}^*}, p = 2, 3, \varepsilon_p^{IFR}, p = 2, 3, \varepsilon_p^{Rt}, p = 1$$

Estos parámetros presentan un papel muy importante en las estimaciones, por lo que los resultados del modelo se deben tomar con reserva.

III.3.2 Logaritmo de la densidad posterior $\ln p_{\text{post}}$

No es aceptable monitorear la convergencia de solo un subconjunto de los parámetros o cantidades generadas. La convergencia de la cadena de Markov es una propiedad global en el sentido de que no depende de la elección de la función de los parámetros que se monitorean. Por lo que puede ocurrir que de manera local no exista convergencia para algunos de los parámetros, pero si se garantiza el equilibrio posterior. Lo que sucede es que a medida que el número de estados en la cadena se acerca a infinito, la distribución de posibles estados en la cadena se acerca a la distribución objetivo y en ese límite el valor esperado del estimador de Monte Carlo de cualquier función integrable converge a la verdadera esperanza, en el límite, los efectos del estado inicial se eliminan por completo [70-75].

Por esto se considera el valor $\ln p_{\text{post}}$, que es el logaritmo de la densidad posterior. Por lo tanto, es un error declarar convergencia en cualquier sentido práctico si $\ln p_{\text{post}}$ no ha convergido, porque las diferentes cadenas están realmente en diferentes partes del espacio [70-75].

El valor del R-hat de $\ln p_{\text{post}}$ es de 1.09765 aunque está por debajo de 1.1, se encuentra muy cerca de este, indicando una vez más que se deben tomar con reserva los resultados obtenidos. La figura 9 muestra la distribución estimada del parámetro $\ln p_{\text{post}}$

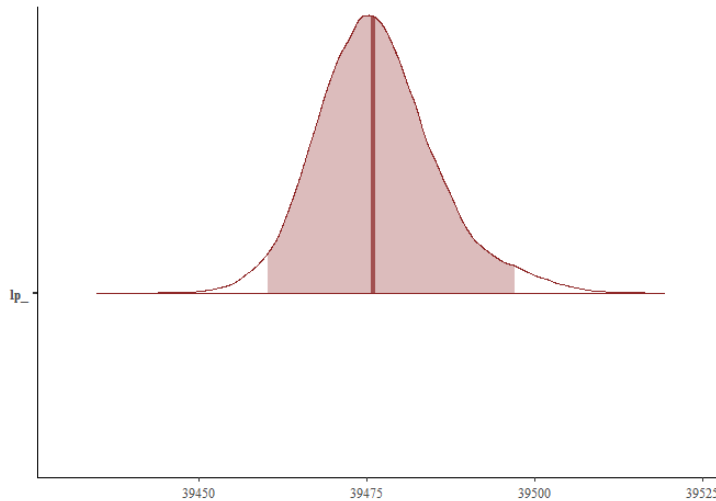


Figura 9. Distribución de $lp_{__}$

Una manera para mejorar esta estimación es incrementar el número de iteraciones.

III.3.3 Tamaño efectivo de la muestra

La segunda dificultad técnica que plantean los métodos MCMC es que las muestras normalmente estarán autocorrelacionadas dentro de una cadena. Esto aumenta la incertidumbre de la estimación de cantidades posteriores de interés, como medias, varianzas o cuantiles [76]. Stan estima un tamaño de muestra efectivo para cada parámetro.

La cantidad por la cual la autocorrelación dentro de las cadenas aumenta la incertidumbre en las estimaciones, se puede medir mediante el tamaño de muestra efectivo (ESS). Dadas muestras independientes, el teorema central del límite restringe la incertidumbre en las estimaciones basadas en el número de muestras N . Dadas las muestras dependientes, se puede calcular el número de muestras independientes o efectivas N_{eff} . Estos valores particulares son arbitrarios en el sentido de que no tienen un significado teórico particular, pero una heurística útil es preocuparse por cualquier valor de N_{eff} / N menor que 0.1 [70-75].

En la figura 10 se observa que el tamaño de muestra efectivo para cada parámetro es inferior a 0.1 en la gran mayoría de parámetros, indicando, una vez más, la necesidad de incrementar el número de iteraciones en el momento de ejecutar el modelo.

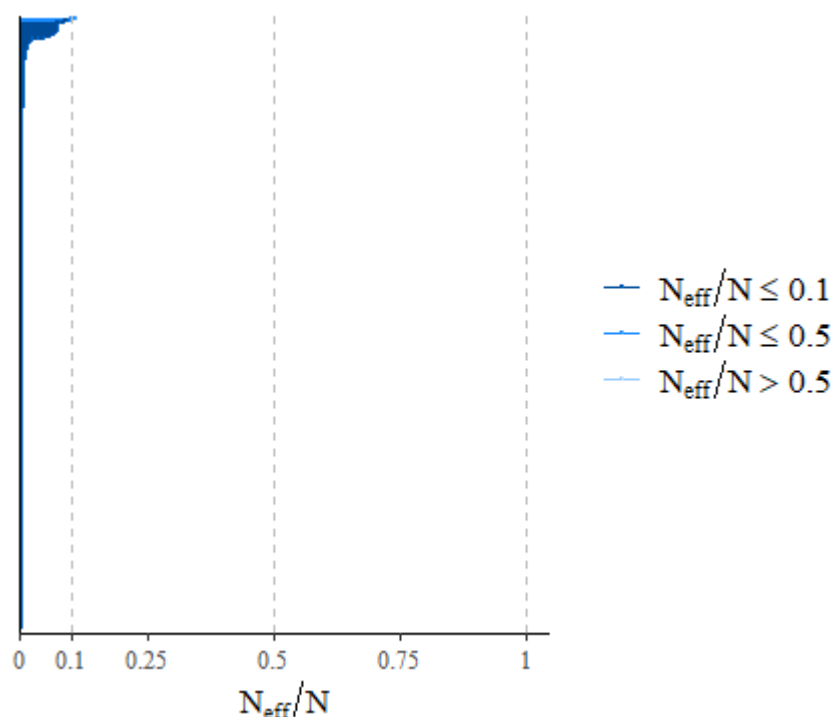
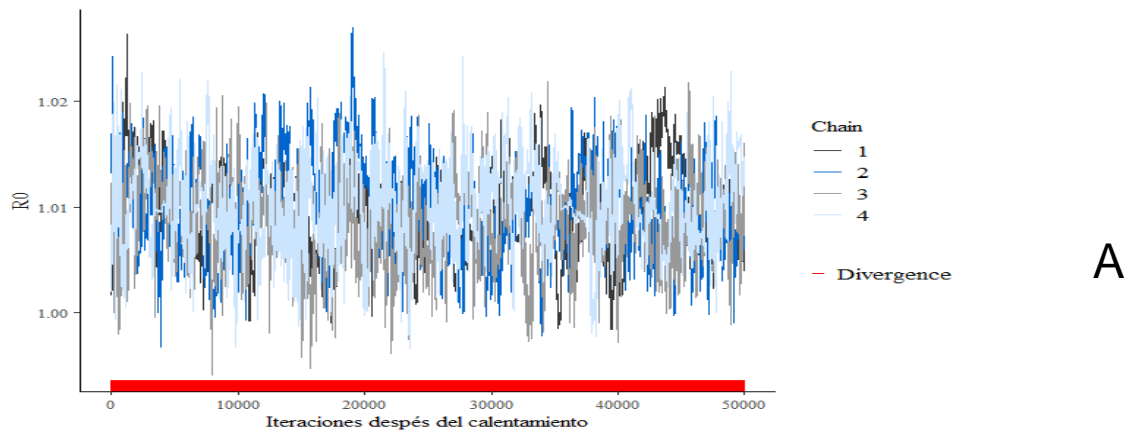


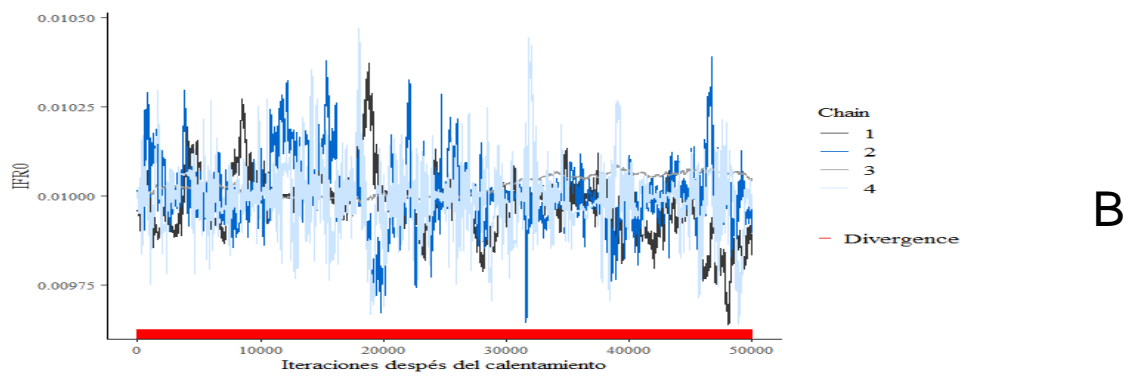
Figura 10. Tamaño de muestra efectivo para cada parámetro.

III.3.4 Autocorrelación y transiciones divergentes en R0 e IFR0

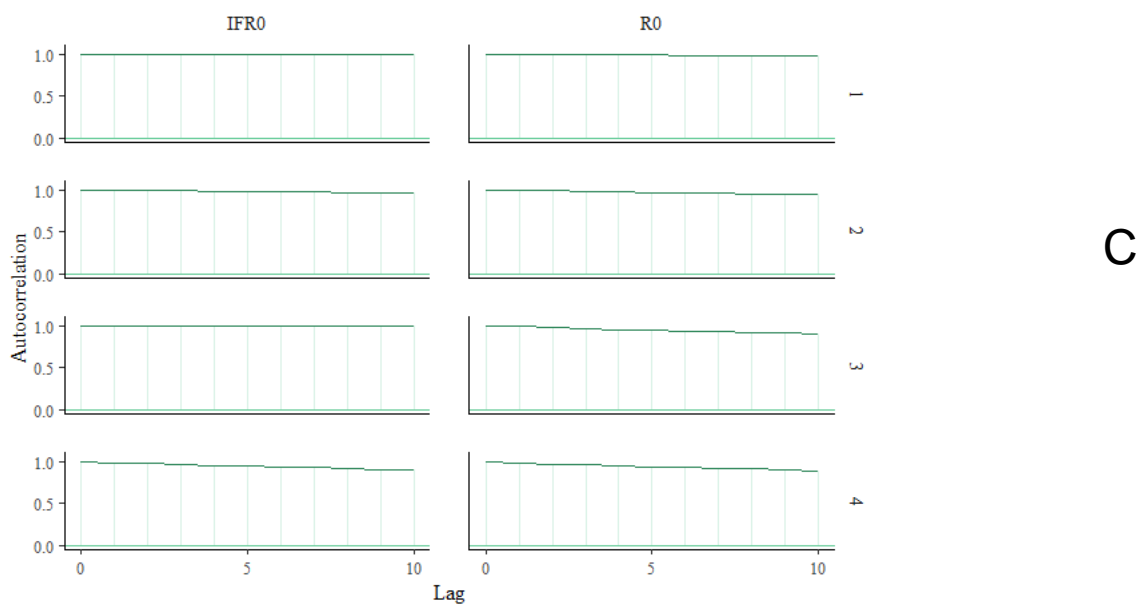
Además, se puede visualizar la autocorrelación para cada cadena de Markov por separado hasta un número de retrasos especificado por el usuario. La autocorrelación positiva es mala (significa que la cadena tiende a permanecer en la misma área entre iteraciones) y se desea que caiga rápidamente a cero con un retraso creciente. La autocorrelación negativa es posible y es útil porque indica una convergencia rápida de la media de la muestra hacia la media verdadera [70-75]. En la figura 11 (C) se puede observar que las cadenas presentan altos niveles de autocorrelación. Asimismo, se aprecia el elevado número (197 270) de transiciones divergentes después del calentamiento, figura 11 (A), (B). además, en (B) se aprecia que las cadenas no se mezclan correctamente.



A



B



C

Figura 11. (A) traza del parámetro R_0 en las cuatro cadenas, (B) traza del parámetro IFR_0 en las cuatro cadenas, (C) autocorrelación de los parámetros R_0 e IFR_0 respectivamente.

III.3.5 El coeficiente de determinación

El coeficiente de determinación es una medida estadística de la bondad del ajuste o fiabilidad del modelo estimado a los datos. Se representa por R^2 y este indica cuál es la proporción de la variación total en la variable dependiente, que es explicada por el modelo estimado, es decir, mide la capacidad explicativa del modelo estimado [77-79]. Para este modelo el coeficiente de determinación es de $R^2 = 0.98$, infiriéndose que de manera general el modelo se ajusta correctamente a los datos, resultado que coincide con lo expuesto en el apartado IV.3.2.

III.3.6 Pruebas de normalidad

Es deseable que los residuos de todo modelo de regresión tengan una distribución normal, para esto se realiza la prueba de hipótesis propuesta por Shapiro [80, 81] [82, 83], donde:

H_0 : los datos presentan una distribución normal.

H_A : los datos no presentan una distribución normal

El resultado de la prueba es:

Shapiro-Wilk normality test

Data: Residuals

W = 0.82331, p-value = 1.572e-10

Por lo que se rechaza la hipótesis nula concluyendo que los residuos no siguen una distribución normal con un nivel de significación del 95%. La figura 12 muestra la dispersión que presentan los valores y la existencia de resultados extremos, los cuales, contribuyen negativamente a que la distribución sea normal.

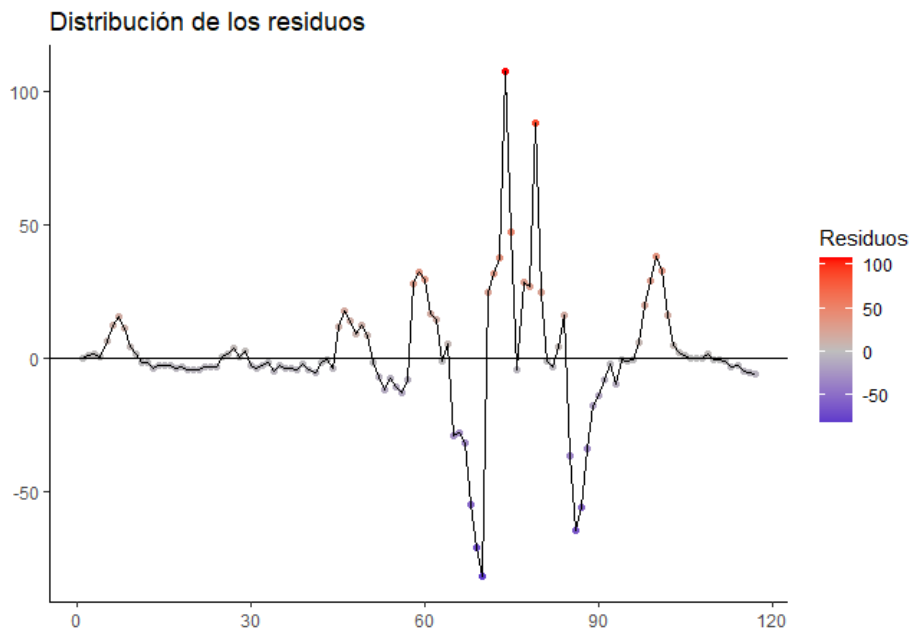


Figura 12. Residuos del modelo

Según lo que se ha presentado, se hace necesario ejecutar el modelo con un número mayor de iteraciones, también incrementar el número de cadenas, así como variar algunos parámetros con el objetivo de mejorar la convergencia de las transiciones. Para ello, es necesario de un sistema de cómputo que presente determinadas características como los llamados '*cluster*'.

Dado el correcto ajuste global del modelo a los datos, así como la convergencia de la mayoría de las distribuciones de los parámetros, se pueden hacer inferencias sobre los resultados, pero es importante alertar que dichas inferencias se deben tomar con precaución.

III.4 Análisis de los parámetros epidemiológicos

III.4.1 Análisis descriptivo

Hasta la semana epidemiológica 117 en Cuba se habían reportado 1 105 443 casos confirmados de SARS-Cov-2, 9 966 280 vacunados y 8 530 fallecidos [84]. Los momentos epidemiológicos (M) tenidos en cuenta en este estudio se basaron en las diferentes olas de contagios observadas en Cuba, las cuales estuvieron determinadas por las medidas de mitigación tomadas por el país ante

el incremento de los casos de SAR-Cov-2, la percepción de riesgo de las personas, las variantes circulantes y el comportamiento inherente de la epidemia

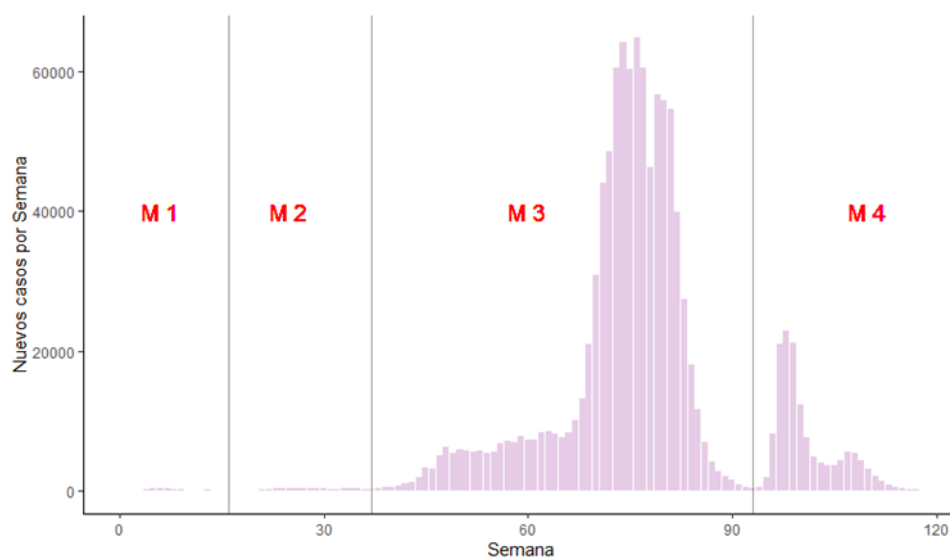


Figura 13. Nuevos casos de SARS-Cov-2 detectados en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

Los momentos epidemiológicos 1 y 2 se caracterizaron por una baja transmisibilidad del virus en la población, así como un bajo número de muertes mientras que los momentos 3 y 4 evidenciaron grandes números de contagios y consecuentemente un incremento del número de personas fallecidas (Fig. 13 y 14).

La activa campaña de vacunación llevada a cabo en Cuba a partir de la semana epidemiológica 65 conllevó a la consecuente disminución de los casos y el número de personas fallecidas en el momento epidemiológico 4 (Fig. 13, 14 y 15).

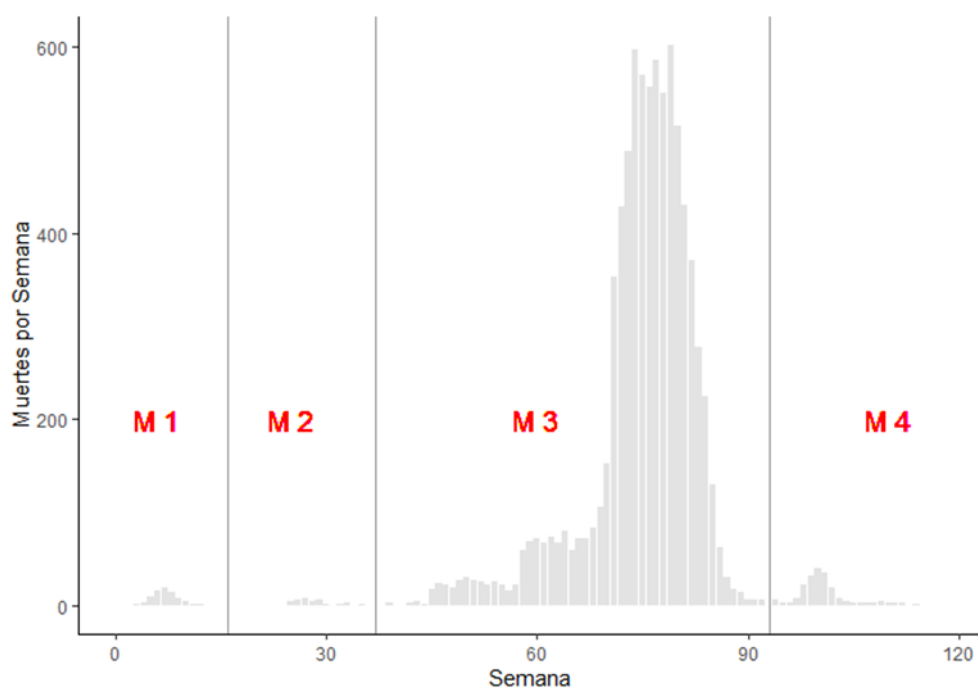


Figura 14. Muertes reportadas en Cuba a causa de la infección por SARS-Cov-2 hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

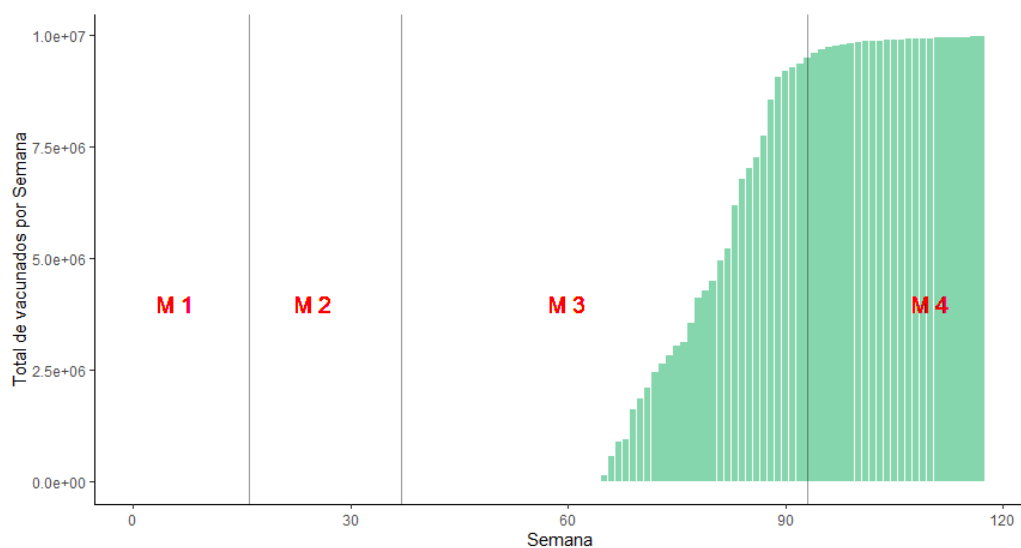


Figura 15. Número de vacunados ante el virus SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

Desde inicios del año 2020 se ha monitoreado a nivel global la aparición de nuevas variantes genómicas de SARS-Cov-2 impulsado por los avances en la vigilancia genómica y la secuenciación nucleotídica. Según la base de datos de GISAID (del inglés, Global Initiative on Sharing All Influenza Data) se han reportado más de 15 variantes del virus SARS-CoV-2 (<https://www.gisaid.org/>). Los primeros casos de SARS-Cov-2 en Cuba fueron reportados el 11 de marzo de 2020 y se han reportado hasta la fecha más de un millón de casos confirmados a la infección viral [85]. Como parte de la vigilancia genómica molecular, llevada a cabo en el Instituto de Medicina Tropical Pedro Kouri, han sido detectadas en Cuba algunas de las variantes genómicas de SARS-Cov-2 reportadas por la OMS, así como patrones mutaciones nunca antes reportados (Fig. 16).

Dentro de las variantes que mayor impacto han tenido en la pandemia de SARS-Cov-2 se encuentra alfa, beta, delta y recientemente omicron (<https://www.gisaid.org/>). Todas ellas fueron reportadas por primera vez en Cuba durante las semanas 42, 43, 59 y 91 respectivamente (Fig. 16).

Hasta la semana epidemiológica número 42 la única variante detectada en Cuba había sido la variante original de Wuhan. Sin embargo, a partir de esta fecha comenzaron a detectarse en el país otras variantes genómicas y varios patrones mutacionales. En la semana epidemiológica 59 fue detectada por primera vez en Cuba la variante delta, la cual desplazó [86] a las variantes que se habían detectado hasta esa fecha y se mantuvo de forma persistente hasta la semana epidemiológica 91, en la cual la aparición de la variante omicron comenzó a desplazar a la variante delta circulante (Fig. 16). Omicron, es una variante de preocupación, según la OMS, y es reconocida como una de las variantes más contagiosas de SARS-Cov-2 [87].

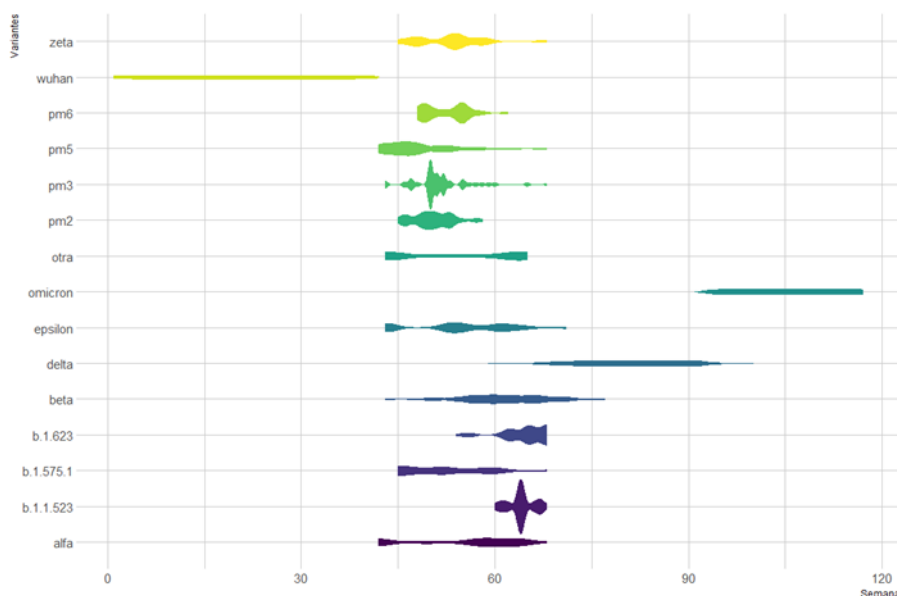


Figura 16. Detección de las variantes genómicas introducidas en Cuba hasta la SE 117.

III.4.2 Susceptibles y parámetros asociados

Los individuos susceptibles son aquellas personas que se encuentran en riesgo de sufrir una infección determinada [88]. Durante el periodo comprendido entre la semana 1 y la semana 65 la cantidad de personas susceptibles en el país disminuyó solamente en un 1% (Fig. 17), esto se debe fundamentalmente al bajo índice de contagios y a que las personas que se infectaban y lograban la inmunidad, regresaban a la condición de susceptibles nuevamente al transcurrir las semanas [60]. Estos resultados evidencian lo expresado por *Franco Escobar* [89] donde plantea que es necesaria una campaña de vacunación para lograr que un porcentaje de la población alcance los niveles de anticuerpos necesarios para lograr la inmunidad de rebaño y de esa manera detener la rápida propagación del virus.

El número de personas susceptibles al SARS-Cov-2 en Cuba, comenzó a disminuir notablemente en la semana epidemiológica número 65 desde 0.99 [0.992-0.993] hasta 0.49 [0.490-0.505], lo cual estuvo mediado por la intensa campaña de vacunación llevada a cabo en el país y el alto índice de contagios (Fig. 17). Hacia el momento epidemiológico 4, se observó un notable incremento

en el número de susceptibles, este fenómeno se debe a que las personas que alcanzaron los niveles de anticuerpos necesarios, ya sea por la vacunación o por la infección, regresaron a la condición de susceptibles [60].

La vacunación constituye una fuente de protección a las infecciones virales, siendo en algunos casos comparables a los obtenidos mediante la infección natural [90]. En el presente estudio se pudo determinar que el 50% [50.0%-5.03%] de las personas vacunadas resultaron protegidas ante el virus SARS-Cov-2 (Fig. 18 A), este resultado se asemeja al valor propuesto en un estudio realizado por el Ministerio de Salud del Gobierno de Chile y OPS donde plantean que la efectividad de la vacunación en Chile para un esquema de vacunación completo y sin refuerzo fue de 62,3% (IC95%: 52,9% - 69,8%) [91].

El 96% [81,9%- 99,9%] de las personas infectadas, así como el 96% [84,6%-99,9%] de los vacunados que logran alcanzar los niveles de anticuerpos necesarios, se pueden re-infectar al transcurrir las semanas (Fig. 18 B), esto se debe a que los anticuerpos obtenidos no son duraderos [60], solo el 4% de este grupo poblacional crea una inmunidad de memoria duradera que los protege ante futuras infecciones.

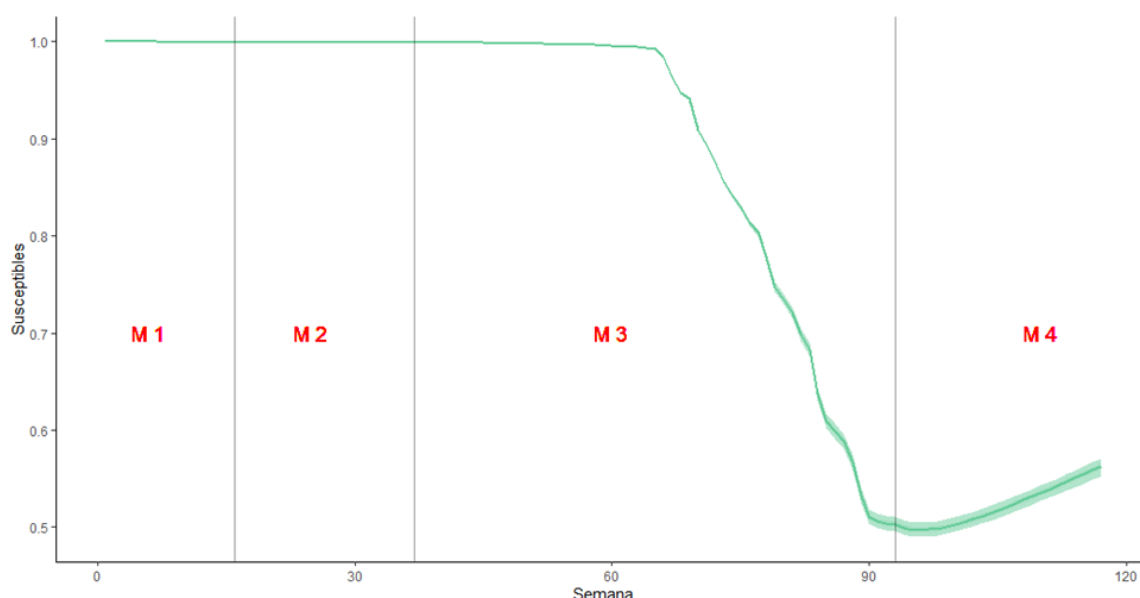


Figura 17. Número de susceptibles ante el virus SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

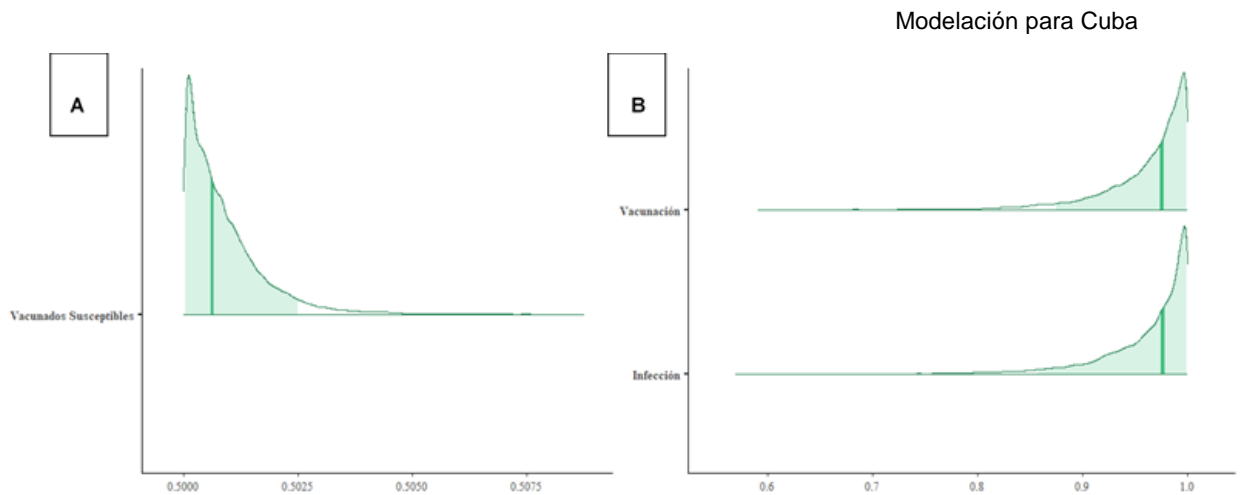


Figura 18. Vacunados susceptibles sin protección conferida por la vacuna (A) y proporción de individuos protegidos por vacunación y por infección natural que se pueden re infectar al transcurrir las semanas (B). M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

III.4.3 Número reproductivo en el tiempo y parámetros asociados

El número reproductivo en el tiempo se determina a partir de variaciones del número reproductivo básico R_0 . El valor estimado por el modelo para este índice fue de 1.01 [1.002- 1.017] (fig. 19). Este resultado difiere significativamente del valor de 1.40 reportado por *Guinovart en [19]*, esto puede estar ocasionado por la forma en que se estima ya que ambos modelos lo estiman de manera diferente, además otro factor que puede estar influyendo es que en este estudio la unidad de tiempo es siete días y las variaciones a medida que se avanza en una unidad de tiempo es menor con respecto a las variaciones que se pueden dar cuando la unidad de tiempo es equivalente a un día.

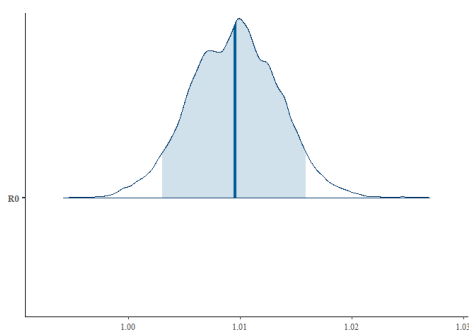


Figura 19. Número reproductivo básico de la infección por SARS-Cov-2 en Cuba.

Asimismo, en la figura 19 se aprecia la distribución de R_0 , se nota a simple vista que la distribución posterior de R_0 convergió, resultado que se puede corroborar con el valor 1.02 del $R\text{-hat}$ del R_0 , valor que al ser menor de 1.1 indica la correcta convergencia de la distribución posterior.

Para estimar cómo afectaron las diferentes variantes genómicas el valor de R_0 fue determinado el parámetro asociado a cada una de ellas. Las variantes para las cuales el cero se encontraba dentro del 95% de su distribución, como por ejemplo la variante de Wuhan, no influyeron significativamente sobre el parámetro R_0 . Las variantes para las cuales las distribuciones de sus parámetros con valores superiores a cero favorecieron positivamente en el decrecimiento de valor R_0 y por tanto en la disminución de la transmisión del virus y la consecuente disminución del número de casos (Fig. 20).

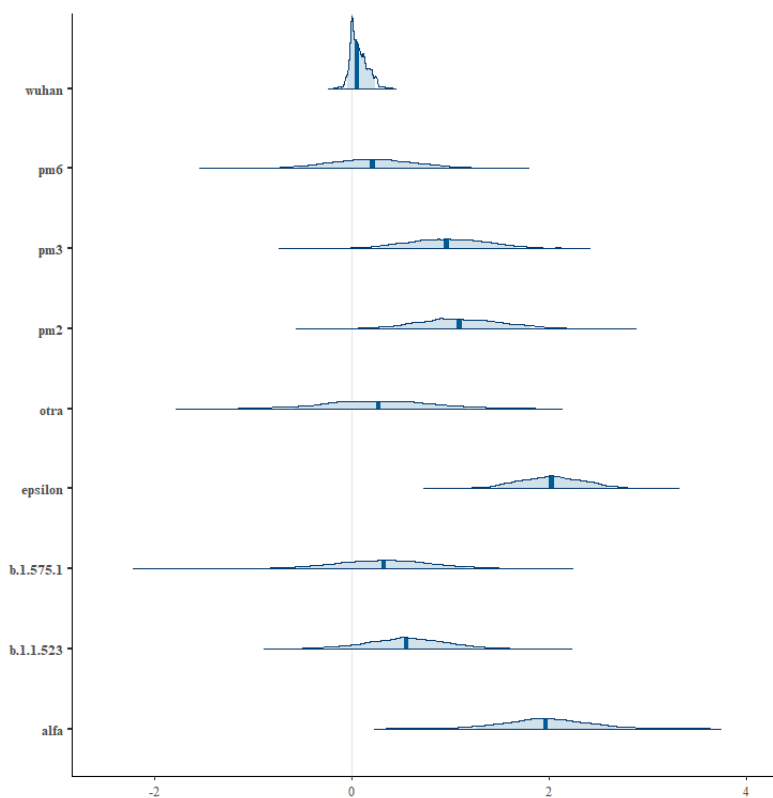


Figura 20. Variantes genómicas que contribuyen al detrimento o a la estabilidad del número reproductivo básico de la infección por SARS-Cov-2 en Cuba.

Asimismo, en la figura 20 se puede apreciar que el parámetro asociado a la variante de Wuhan tiene una distribución posterior bien definida sobre el cero,

esto se explica ya que la variante de Wuhan fue la variante que dominaba durante el periodo donde se estimó el R_0 .

Las variantes como delta, beta y omicron mostraron distribuciones con valores inferiores a cero evidenciando que influyeron en el incremento del valor estimado de R_0 y, como consecuencia, en el incremento del número de casos (Fig. 21). Según explica *Liu* en [92] la alta transmisibilidad relativa de la variante omicron se debe fundamentalmente a su capacidad de evasión de la respuesta inmune.

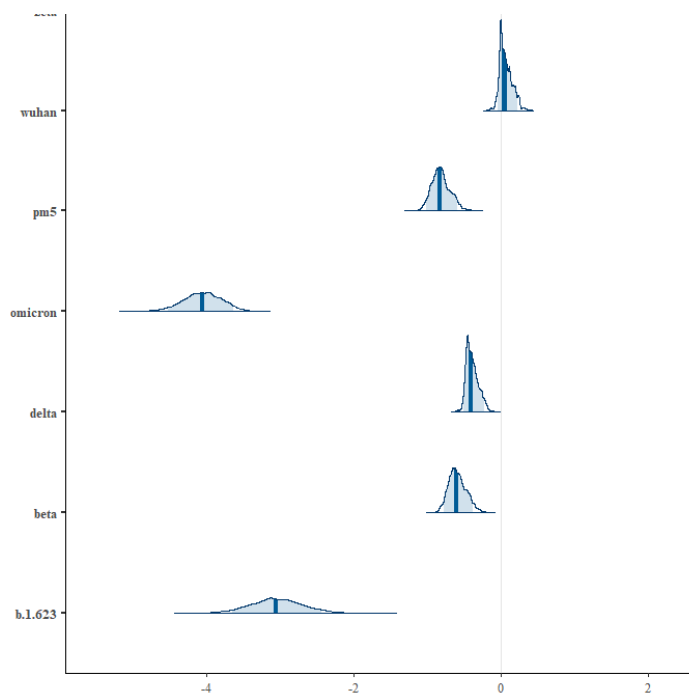


Figura 21. Wuhan y variantes genómicas que contribuyen al incremento del número reproductivo básico de la infección por SARS-Cov-2 en Cuba.

Fueron estimadas además las distribuciones de los diferentes momentos epidemiológicos y se pudo observar que para todos los momentos el cero se encontraba dentro del 95% de su distribución y por tanto no influyeron significativamente sobre el parámetro R_0 (Fig.22). este resultado se debe tomar con precaución pues las distribuciones a posterior de los momentos 1 y 3 presentaron valores de R -hat superiores a 1.1 indicando la no convergencia.

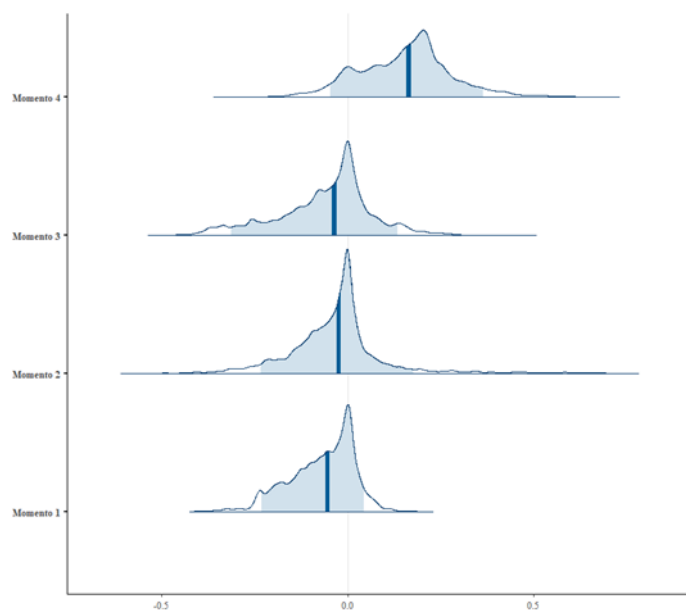


Figura 22. Distribución posterior de los momentos epidemiológicos de la infección por SARS-Cov-2 en Cuba, respecto al R_0 .

El número reproductivo (R_t) permite medir la velocidad de propagación de las enfermedades en un intervalo de tiempo [4]. Cuando R_t es mayor que uno indica que el número de nuevos casos se encuentra en crecimiento y por tanto no existe un control sobre la epidemia. Cuando R_t es menor que uno entonces el número de casos decrece indica que la epidemia se encuentra bajo control [33]. En Cuba, se estimó un R_t que oscilaba alrededor de uno hasta la semana epidemiológica 45, lo cual sugiere que si bien la epidemia se mantuvo controlada hasta esa fecha las medidas de mitigación no fueron suficientes para alcanzar valores de R_t menores de uno. Esto pudo estar influido por el alto número de susceptibles y la ausencia de vacunas disponibles para este virus durante ese periodo (Fig. 23).

Asimismo, el incremento de R_t a partir de la SE 45, el cual alcanzó un valor máximo de 1.41 [1.348-1.459] en la SE 68, estuvo propiciado entre otras cosas por la detección de la variante delta, la cual como ya se explicó, contribuye al incremento del R_t . En la semana epidemiológica 68 (M3), el valor estimado de R_t comienza a disminuir alcanzando valores inferiores a uno, lo cual evidencia la efectividad de la vacunación masiva llevada a cabo en el país. Sin embargo, tras ser detectada la circulación en la población de la variante genómica omicron, el R_t aumentó dramáticamente, alcanzando valores superiores a uno a partir de la

SE 102. Hacia la SE 117 el valor de R_t estimado en Cuba fue de 1.11 [1.096-1.128] (Fig.23). Otros factores como la falta de percepción de riesgo de las personas y el agotamiento por los prolongados periodos en los que se han sostenido las medidas restrictivas relacionadas a la epidemia de SARS-Cov-2 en Cuba pudieran influir sobre estos resultados.

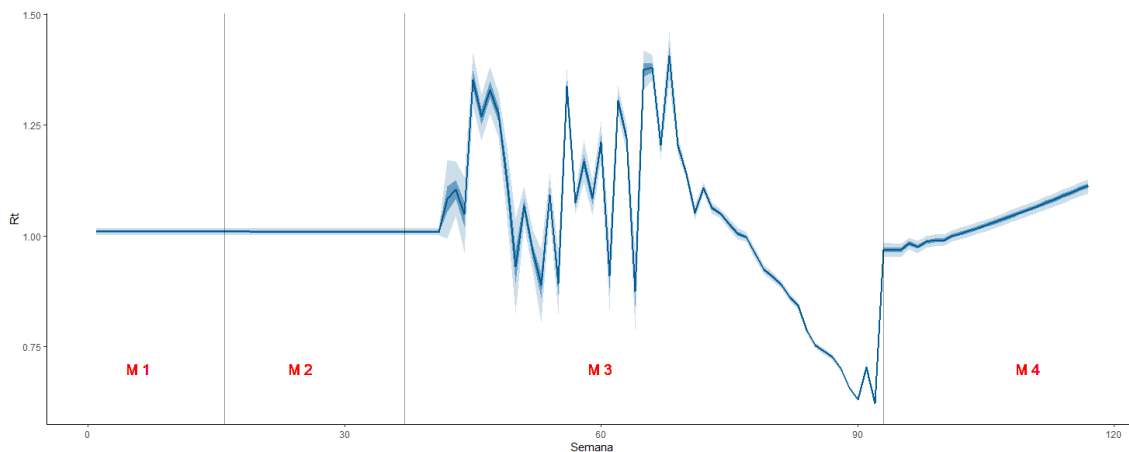


Figura 23. Número reproductivo en el tiempo de la infección por SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

III.4.4 Índice de mortalidad por infección y parámetros asociados

El índice de mortalidad por infección (IFR del inglés *Infection fatality ratio*) permite estimar la probabilidad de fallecer estando infectado [26]. El valor inicial de IFR estimado para este estudio fue de 1.01% [1.002% -1.019%] (Fig. 24).

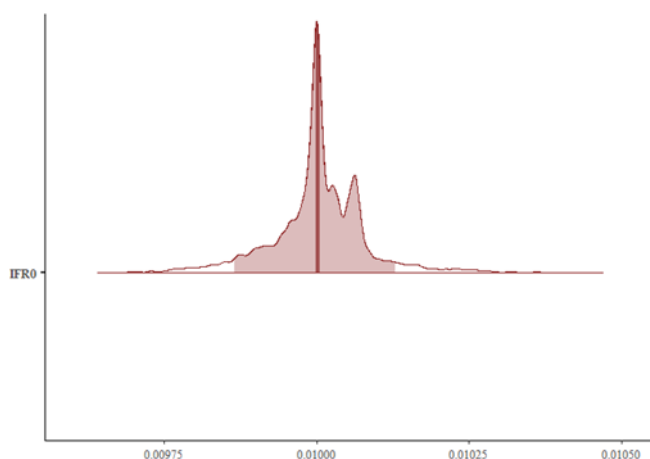


Figura 24. El índice de mortalidad por infección (IFR0) por SARS-Cov-2 en Cuba

La distribución a priori que se definió para el IFR₀, es una distribución informativa, otorgándole una varianza muy pequeña, el objetivo es obtener un valor muy cercano a 1.01%, dado a que *Brazeau en* [93] estimó en su estudio un valor de IFR de 0,23 % (0,14%-0,42%) para los países de bajos ingresos mientras que para países de altos ingresos el IFR estimado sería de 1,15 % (0,78%-1,79%). Además, en su investigación concluyen que las diferencias presentadas entre ambos grupos de países están dadas por el alto nivel de envejecimiento que muestran los países de altos ingresos frente a un bajo nivel de envejecimiento en el caso de los países de bajos ingresos. Como Cuba es un país de bajos ingresos, pero que presenta un índice elevado de envejecimiento, se pretendía que el IFR₀ para Cuba estuviese en la parte baja de la distribución de los países de altos ingresos.

Dado que la distribución descrita para el IFR₀ inicial permite una variabilidad baja alrededor del 1.01%, se logró obtener ese valor para el IFR₀ (Fig. 24). Pero inmediatamente el IFR_t comienza a crecer hasta llegar a 1,41% [1,285%- 1,488 %] (Fig. 28), siendo este último resultado el valor real del índice IFR al comienzo de la epidemia de SARS-Cov-2 en Cuba.

Para determinar cómo afectan las variantes, la cantidad de vacunados, los momentos de la epidemia y la cantidad de casos activos se estimaron parámetros asociados a cada una de estas variables

Las variantes para las cuales el cero se encontraba dentro del 95% de su distribución no influyeron significativamente sobre el parámetro IFR₀. La variante omicron para la cual la distribución posterior de su parámetro asociado, mostró valores inferiores a cero favoreció positivamente en el incremento de valor IFR₀ y por tanto en el aumento de las muertes asociadas al SARS-Cov-2. Las variantes Wuhan, delta y beta mostraron distribuciones con valores superiores a cero e influyeron en el detrimento del valor estimado de IFR₀ y, como consecuencia, en el decrecimiento de las muertes asociadas al SARS-Cov-2 (Fig. 25).

Fueron estimadas además las distribuciones de los diferentes momentos epidemiológicos y se pudo observar que para M 2-4 el cero se encontraba dentro del 95% de su distribución y por tanto no influyeron significativamente sobre el parámetro IFR₀ (Fig. 26). Sin embargo, para M-1 la distribución posterior de su parámetro mostró valores inferiores a cero favoreciendo positivamente en el incremento de valor IFR₀ y por tanto el incremento de las muertes asociadas al SARS-Cov-2.

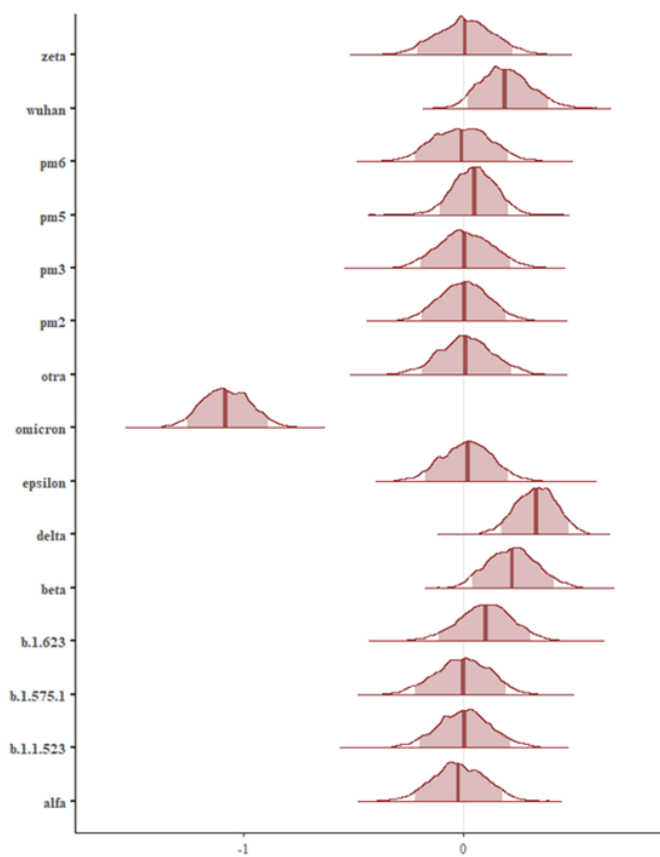


Figura 25. Distribución posterior de los parámetros asociados a las variantes genómicas de SARS-Cov-2 en Cuba, respecto al IFR0.

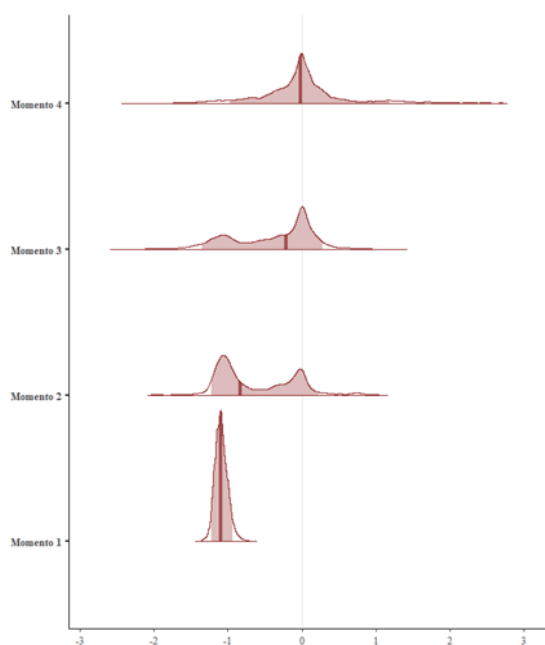


Figura 26. Distribución posterior de los parámetros asociados a los momentos epidemiológicos de la infección por SARS-Cov-2 en Cuba respecto al IFR0.

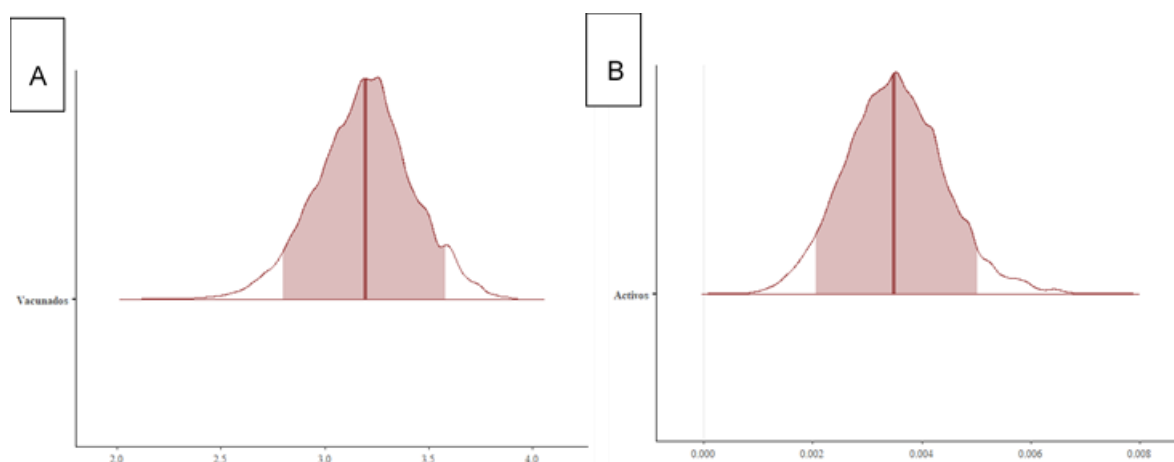


Figura 27. Distribución posterior de los individuos vacunados (A) y los casos activos (B) de la infección por SARS-Cov-2 en Cuba según IFR0.

Otro de los parámetros que influyen en la estimación de IFR0 es la distribución de individuos vacunados en la población. En este caso los valores de esta distribución fueron superior a cero, lo cual indica el éxito de la campaña de vacunación masiva llevada a cabo en el país en la reducción del número de fallecidos (Fig. 27A). Estos resultados avalan el uso de las vacunas cubanas en poblaciones susceptibles como medida para mitigar las muertes provocadas por la infección de SARS-Cov-2. Además, la distribución de casos activos en la población mostro valores superiores a cero (Fig. 27B). Estos resultados sugieren que al incrementarse el número de casos activos disminuye la probabilidad de fallecer. Algunos autores plantean que la vacunación es muy eficaz para reducir las enfermedades graves y la muerte por covid-19 [90]. En un estudio reciente, los investigadores estimaron que, según las muertes por COVID-19 notificadas oficialmente a nivel mundial, las vacunas previnieron 14,4 millones (13,7%–15,9%) muertes por COVID-19 en 185 países y territorios entre el 8 de diciembre de 2020 y el 8 de diciembre de 2021 [94].

Durante el inicio de M3, en la SE 44, el valor de IFRt se incrementó ligeramente alcanzando un valor máximo de 1,47% [1,421%-1,498%] lo cual pudo estar influenciado por la circulación de nuevas variantes virales en el país. Sin embargo, debido a la disminución de los individuos susceptibles y el éxito de la campaña de vacunación, los valores de IFRt comenzaron a disminuir a partir de la SE 62 alcanzando valores mínimos de 0,24% [0,030%- 0. 308%] en la SE 92 (Fig. 28). Un estudio realizado por Watson en 2022 [94] sugiere que la vacunación ha alterado positivamente el curso de la pandemia, contribuyendo a salvar millones de vidas en todo el mundo.

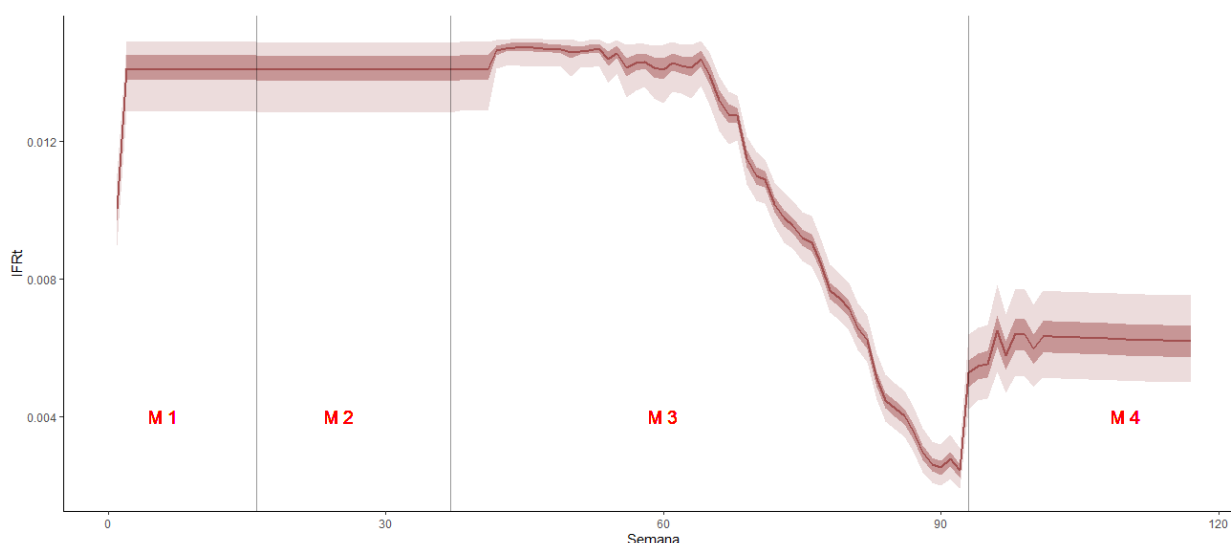


Figura 28. IFR en el tiempo de la infección por SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

Luego de ser detectada la circulación en la población de la variante genómica omicron, el IFR aumentó dramáticamente a partir de la SE 93, alcanzando valores sostenidos de aproximadamente 0,62%. Estos resultados, si bien evidencian un incremento de la letalidad ante la presencia de nuevas variantes virales, reafirman la importancia de la vacunación, pues si bien ocurrió un aumento de los susceptibles (Fig. 17) y consecuentemente un aumento de la letalidad hacia M4, los valores de IFRt obtenidos fueron inferiores a los estimados en etapas previas (Fig.28).

III.4.5 Nuevos contagios y nuevas muertes

A partir de los parámetros estimados se estimaron los contagios semanales de SARS-Cov-2, así como el número de muertes (Fig. 29 y 30). El mayor número de nuevos contagios en una semana se observó durante las semanas epidemiológicas 76 con 69 627 [67 318-76 550] casos, correspondientes al momento M3. Asimismo, se observa que el mayor número de muertes se dio en la SE 76 con 560 [543- 577] decesos.

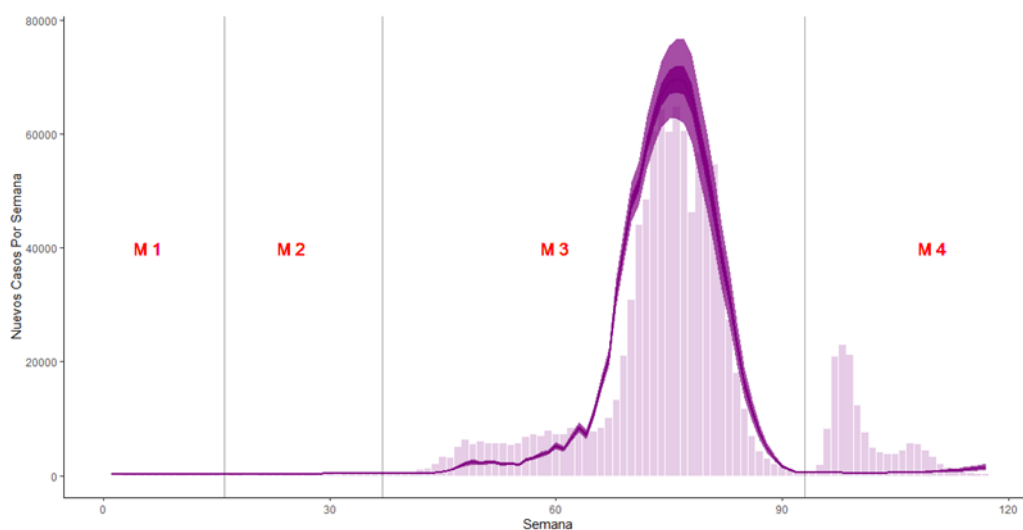


Figura 24. Nuevos casos por semana de la infección por SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

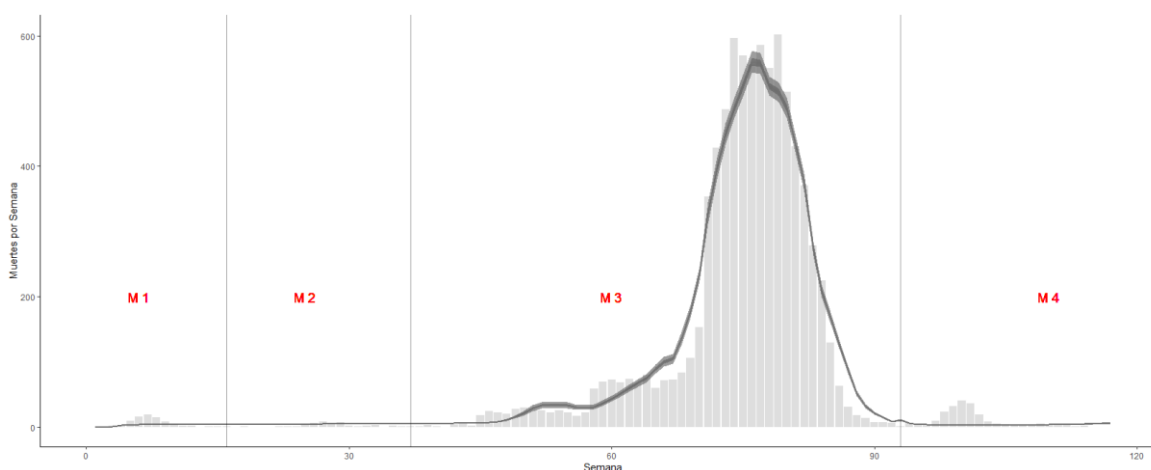


Figura 25. Nuevas muertes por semana de la infección por SARS-Cov-2 en Cuba hasta la SE 117. M1, M2, M3 y M4 se refieren a los cuatro momentos epidémicos observados en el país.

Conclusiones

Conclusiones

1. Se pudo identificar como construir un modelo de tipo estocástico que permita encontrar las relaciones significativas entre los parámetros de los cuales se quiere obtener información.
2. Se construyó un modelo Bayesiano jerárquico, que posibilita determinar la influencia que tuvo el proceso de vacunación y las diferentes cepas de la epidemia sobre el resto de parámetros epidemiológicos.
3. Los parámetros R_t e IFR_t fueron estimados en función de los individuos vacunados y las variantes genómicas virales, mediante la simulación computacional en los software R y Stan, determinándose que el período de mayor circulación y decesos fue el M3, mientras que hacia M4 la vacunación masiva y la disminución de los susceptibles contribuyó notablemente en la disminución de estos parámetros.
4. Las pruebas realizadas permitieron demostrar el correcto ajuste del modelo y la convergencia de los parámetros correspondientes.

Recomendaciones

Recomendaciones

1. Ejecutar el modelo en un sistema de cómputo con mayores capacidades, para lograr la convergencia de las cadenas.
2. Introducir la estimación de otros parámetros epidemiológicos de interés como por ejemplo el efecto que tuvo la campaña del refuerzo a la población.
3. Hacer estimaciones desagregadas, por provincias, por grupos de edades, según el género de las personas o según las comorbilidades que presenta la población cubana.
4. Modelar futuros escenarios.

Referencias Bibliográficas

Referencias bibliográficas

1. Pulcha-Ugarte, R., et al., *¿ Qué lecciones nos dejará el covid-19?: Historia de los nuevos coronavirus*. Revista de la Sociedad Peruana de Medicina Interna, 2020. **33**(2): p. 68-76.
2. Webster, P., *COVID-19 timeline of events*. Nature medicine, 2021. **27**(12): p. 2054-2055.
3. Nazario, D.B., *¿Cómo se propaga el coronavirus?* WebMD, July 27, 2021.
4. Sy, K.T.L., L.F. White, and B.E. Nichols, *Population density and basic reproductive number of COVID-19 across United States counties*. PloS one, 2021. **16**(4): p. e0249271.
5. Yesudhas, D., A. Srivastava, and M.M. Gromiha, *COVID-19 outbreak: history, mechanism, transmission, structural studies and therapeutics*. Infection, 2021. **49**(2): p. 199-213.
6. Rodríguez, D.J.L., D.S. Odriozola, and E.c.c.e.e.d.P.y.l.O.e. Cuba, *Impactos Económicos y Sociales de la COVID 19 en Cuba: Opciones de políticas*. PENNUD, Mayo 30 / 2020.
7. Wimalawansa, S., *COVID-19: Evolution and prevention*. Trends Telemed E-Health, 2020. **2**(3): p. 1-5.
8. OMS. *Seguimiento de las variantes del SARS-CoV-2*. 2022 [cited 2022; Available from: <https://www.who.int/es/activities/tracking-SARS-CoV-2-variants>].
9. Sanches, P.R., et al., *Recent advances in SARS-CoV-2 Spike protein and RBD mutations comparison between new variants Alpha (B. 1.1. 7, United Kingdom), Beta (B. 1.351, South Africa), Gamma (P. 1, Brazil) and Delta (B. 1.617. 2, India)*. Journal of Virus Eradication, 2021. **7**(3): p. 100054.
10. Shahab, M.S., S.S. Imam, and M.A. Jahangir, *A review on the contemporary status of mutating coronavirus and comparative literature*

- study of current COVID-19 vaccines. International Journal of Pharmaceutics & Pharmacology*, 2021. **5**(1): p. 1-19.
11. OPS, *COVID-19 Glosario sobre brotes y epidemias. Un recurso para periodistas y comunicadores*. 2020.
 12. Mesa Mazo, M., *Modelos epidemiológicos. Marco Teórico. Blog Modelos Matemáticos en Epidemiología*. 2011 [citado 8 Jun 2020].
 13. Krieger, N., *Glosario de epidemiología social*. Revista Panamericana de Salud Pública, 2002. **11**: p. 480-490.
 14. Altamirano, L.M., F.C. Valle, and H.G. Romero, *Epidemiología clínica*. 1994: Interamericana. McGraw-Hill.
 15. Celentano, D.D., S. Mhs, and M. Szklo, *Gordis. Epidemiología*. 2019: Elsevier.
 16. Sánchez Valdés, L., et al., *Modelos de transmisión, pronóstico y gravedad del dengue*. Dengue [Internet]. La Habana: Editorial Ciencias Médicas, 2014.
 17. Vidal Ledo, M., et al., *Modelos matemáticos para el control epidemiológico*. Educación Médica Superior, 2020. **34**(2).
 18. Montesinos-López, O.A. and C.M. Hernández-Suárez, *Modelos matemáticos para enfermedades infecciosas*. Salud pública de México, 2007. **49**(3): p. 218-226.
 19. Díaz, R.G., et al., *Modelos SIR para el seguimiento de la COVID-19 en Cuba SIR models for monitoring COVID-19 in Cuba*. 2020.
 20. Fernández-Villaverde, J. and C.I. Jones, *Estimating and simulating a SIRD model of COVID-19 for many countries, states, and cities*. Journal of Economic Dynamics and Control, 2022: p. 104318.

21. Calafiore, G.C., C. Novara, and C. Possieri, *A time-varying SIRD model for the COVID-19 contagion in Italy*. Annual reviews in control, 2020. **50**: p. 361-372.
22. Weiss, H.H., *The SIR model and the foundations of public health*. Materials mathematics, 2013: p. 0001-17.
23. Bliznashki, S., *A Bayesian logistic growth model for the spread of COVID-19 in New York*. medRxiv, 2020.
24. Wu, K., et al., *Generalized logistic growth modeling of the COVID-19 outbreak: comparing the dynamics in the 29 provinces in China and in the rest of the world*. Nonlinear dynamics, 2020. **101**(3): p. 1561-1581.
25. Flaxman, S., et al., *Report 13: Estimating the number of infections and the impact of non-pharmaceutical interventions on COVID-19 in 11 European countries*. 2020.
26. Blasco-Moreno, A., et al., *Estimación bayesiana de la tasa de mortalidad por COVID-19 dada la infección por SARS-Cov-2: el caso español*. 2020.
27. Vollmer, M.A., et al., *Report 20: Using mobility to estimate the transmission intensity of COVID-19 in Italy: A subnational analysis with future scenarios*. MedRxiv, 2020.
28. Mellan, T.A., et al., *Report 21: Estimating COVID-19 cases and reproduction number in Brazil*. 2020.
29. Unwin, H., et al., *Report 23: State-level tracking of COVID-19 in the United States*. Imperial College London, 2020.
30. Dighe, A., et al., *Response to COVID-19 in South Korea and implications for lifting stringent interventions*. BMC medicine, 2020. **18**(1): p. 1-12.
31. Nouvellet, P., S. Bhatia, and A. Cori, *Report 26: Reduction in mobility and COVID-19 transmission*. 2020.

32. Alzate, Ó.E.T., *Modelos y modelización en la enseñanza y el aprendizaje de las ciencias*. Enseñanza de las ciencias: revista de investigación y experiencias didácticas, 2013(Extra): p. 3484-3487.
33. Müller, J. and C. Kuttler, *Methods and models in mathematical biology*. Lecture Notes on Mathematical Modelling in Life Sciences, Springer, Berlin, 2015.
34. Gómez Pavón, M.S., *Modelo de inventario aplicando modelos heurísticos para la microempresa Santiagoscreen*. 2022.
35. Martínez Carranza, C.A. and D.E. Medina Turizo, *Modelacion de una heuristica para el análisis del desempeño de un modelo deterministico de ruteo de vehículos multiples depósitos bajo un ambiente estocástico*. 2014, Universidad de la Costa.
36. Silva, L.C. and P. Suárez, ¿ *Qué es la inferencia bayesiana*. Jano, 2000. **58**: p. 65-66.
37. Silva, L., *La alternativa bayesiana*. Brotes, 2000. **1**: p. 1-4.
38. Silva, L. and A. Benavides, *El enfoque bayesiano: otra manera de inferir*. Gaceta Sanitaria, 2001. **15**(4): p. 341-346.
39. Monsalve Graterol, N.C., *Modelos jerárquicos bayesianos espaciales en epidemiología agrícola*. 2013, Universitat Politècnica de València.
40. Everitt, B.S. and A. Skrondal, *The Cambridge dictionary of statistics*. 2010.
41. Garcia Mondeja, Y., A. Rodríguez García, and Y. Díaz Villasuso, *Herramienta computacional para hacer inferencias bayesianas, aplicaciones a bioinformática*. 2006, Universidad Central "Marta Abreu" de Las Villas.
42. Gelman, A., et al., *Bayesian data analysis*. 1995: Chapman and Hall/CRC.

43. Lesaffre, E. and A.B. Lawson, *Bayesian biostatistics*. 2012: John Wiley & Sons.
44. Mateos Marcos, C., J. Amador Pacheco, and R. Susi García, *Definición y estudios de redes bayesianas aplicadas a ciencias de la salud y de la vida*. 2021.
45. Gelman, A., et al., *Bayesian data analysis*. 2013: Chapman and Hall/CRC.
46. González, L.D., *Selección de modelos en regresión logística binaria bajo el paradigma bayesiano*. 2015: D-Universidad de La Habana.
47. Cabria Fuente, E., *IA explicable: programación probabilística con PyMC para prevención de blanqueo de capitales*. 2021.
48. Nishio, M. and A. Arakawa, *Performance of Hamiltonian Monte Carlo and No-U-Turn Sampler for estimating genetic parameters and breeding values*. Genetics Selection Evolution, 2019. **51**(1): p. 1-12.
49. Durán, U., et al., *Inferencia Variacional para la estimación de mortalidad COVID-19 en Portoviejo*. Revista Ibérica de Sistemas e Tecnologias de Informação, 2021(E41): p. 244-257.
50. Hoffman, M.D. and A. Gelman, *The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo*. J. Mach. Learn. Res., 2014. **15**(1): p. 1593-1623.
51. Smite, D., et al. *From collaboration to solitude and back: remote pair programming during Covid-19*. in *International Conference on Agile Software Development*. 2021. Springer, Cham.
52. Kennedy, R. and P. Waggoner, *Introduction to R for Social Scientists: A Tidy Programming Approach*. 2021: Chapman and Hall/CRC.
53. Braun, W.J. and D.J. Murdoch, *A first course in statistical programming with R*. 2021: Cambridge University Press.

54. Carpenter, B., et al., *Stan: A probabilistic programming language*. Journal of statistical software, 2017. **76**(1).
55. Baudart, G., et al. *Compiling Stan to generative probabilistic languages and extension to deep probabilistic programming*. in *Proceedings of the 42nd ACM SIGPLAN International Conference on Programming Language Design and Implementation*. 2021.
56. REN, Y., *Bayesian inference for female labour force participation in South Europe*. 2021.
57. Villa-Ochoa, J., et al., *El proceso de modelación matemática. Una mirada a la práctica del docente*. 2009.
58. Domínguez, E., *Modelación matemática, Una introducción al método*. Recuperado de: <http://www.mathmodelling.org>, 2010.
59. Arcos, J.H., R. Borromeo-Ferri, and J.J.F. Mena-Lorca, *El conocimiento de la modelación matemática desde la reflexión en la formación inicial de profesores de matemática*. Enseñanza de las Ciencias. Revista de investigación y experiencias didácticas, 2018. **36**(1): p. 99-115.
60. Townsend, J.P., et al., *The durability of immunity against reinfection by SARS-CoV-2: a comparative evolutionary study*. The Lancet Microbe, 2021. **2**(12): p. e666-e675.
61. Alene, M., et al., *Serial interval and incubation period of COVID-19: a systematic review and meta-analysis*. BMC Infectious Diseases, 2021. **21**(1): p. 1-9.
62. Paul, S. and E. Lorin, *Estimation of COVID-19 recovery and decease periods in Canada using delay model*. Scientific Reports, 2021. **11**(1): p. 1-15.
63. Kareem, S.W., et al., *Structure Learning of Bayesian Network: A Review*. QALAAI ZANIST JOURNAL, 2022. **7**(1): p. 956-975.

64. Rojas, J.C.S., D.U. Pérez, and C.E.H. Reyes, *Definición de Redes Bayesianas y sus aplicaciones*. Revista Vinculando, 2012.
65. Sucar, L.E. and M. Tonantzintla, *Redes bayesianas*. Aprendizaje Automático: conceptos básicos y avanzados, 2006. **77**: p. 100.
66. Tipán Charro, L.E., *Estimación de la demanda para micro-zonas basado en el modelo autorregresivo integrado de promedio móvil*. 2022.
67. Scheidereiter, G.D. and O.R. Faure, *Aplicación de la Teoría de los Residuos de las Funciones Complejas a los Modelos Autorregresivos*. 2022.
68. Alvarez-De-Toledo, P., et al., *Introducción de elementos autorregresivos en modelos de dinámica de sistemas*. Revista de dinámica de sistemas, 2006. **2**(1): p. 37-66.
69. Martínez Gómez, M. and M.D. Marí Benlloch, *La distribución de Poisson*. 2010.
70. Gabry, J. and M. Modrák. *Visual MCMC diagnostics using the bayesplot package*. 2022 2022-03-10 [cited 2022 2022-06-04]; Available from: <https://mc-stan.org/bayesplot/articles/visual-mcmc-diagnostics.html>.
71. Stan Development Team. 2022. Stan Modeling Language Users Guide and Reference Manual, V., *Stan Reference Manual Version 2.29*. 2022.
72. Vats, D., et al., *Analyzing MCMC output*. arXiv preprint arXiv:1907.11680, 2019.
73. Roy, V., *Convergence diagnostics for markov chain monte carlo*. arXiv preprint arXiv:1909.11827, 2019.
74. Lambert, B. and A. Vehtari, *R*: A robust MCMC convergence diagnostic with uncertainty using decision tree classifiers*. Bayesian Analysis, 2022. **17**(2): p. 353-379.

75. Vehtari, A., et al., *Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of MCMC (with Discussion)*. Bayesian analysis, 2021. **16**(2): p. 667-718.
76. Geyer, C.J., *Introduction to markov chain monte carlo*. Handbook of markov chain monte carlo, 2011. **20116022**: p. 45.
77. Escobar López, L.A., *Coeficiente de determinación y análisis de varianza, para la regresión polinomial local, en muestreo de poblaciones finitas*. 2011.
78. López Jáuregui, A. and P. Elosua Oriden, *Estimaciones bootstrap para el coeficiente de determinación: un estudio de simulación*. Revista Electrónica de Metodología Aplicada, 2004. **9**(2): p. 1-14.
79. Rodríguez, E.M., *Errores frecuentes en la interpretación del coeficiente de determinación lineal*. Anuario jurídico y económico escorialense, 2005(38): p. 315-331.
80. Torres, V., et al., *Criterios de bondad de ajuste en la selección de modelos no lineales en la descripción de comportamientos biológicos*. Revista Cubana de Ciencia Agrícola, 2012. **46**(4): p. 345-350.
81. Macías Cabrera, S.V. and C.A. Pincay Chiquito, *Construcción de software para regresión el caso de selección del modelo y pruebas de homocedasticidad*. 2012, Espol.
82. González-Estrada, E., J.A. Villaseñor, and R. Acosta-Pech, *Shapiro-Wilk test for multivariate skew-normality*. Computational Statistics, 2022: p. 1-17.
83. Royston, P., *Approximating the Shapiro-Wilk W-test for non-normality*. Statistics and computing, 1992. **2**(3): p. 117-119.
84. *Data on COVID-19 (coronavirus) by Our World in Data*. 2022; Available from: <https://github.com/owid/covid-19-data/tree/master/public/data>.

85. Beldarraín-Chaple, E., I.R. Alfonso-Sánchez, and M. Vidal-Ledo, *La segunda ola de la epidemia de COVID-19 en Cuba: los rebrotes del sexto y del séptimo mes/The second wave of the COVID-19 epidemic in Cuba: the sixth-and seventh-month flare-ups*. Medicina Social, 2021. **14**(2): p. 47-58.
86. *Variantes de SARSCoV-2 en España: linaje BA.2 de Ómicron. Evaluación Rápida de Riesgo*. 2022, Centro de Coordinación de Alertas y Emergencias Sanitarias.
87. Tian, D., et al., *The emergence and epidemic characteristics of the highly mutated SARS-CoV-2 Omicron variant*. Journal of Medical Virology, 2022. **94**(6): p. 2376-2383.
88. Española, R.A. *Susceptible*. 2005 [cited 2022 Julio]; Available from: <https://www.rae.es/dpd/susceptible>.
89. Franco Escobar, M.F., Y.C. Gil Ocampo, and J.E. Moreno Cruz, *Vacuna contra el Covid 19 mitos y realidades*. 2021.
90. Mohapatra, P.R., B. Mishra, and B. Behera, *BCG vaccination induced protection from COVID-19*. indian journal of tuberculosis, 2021. **68**(1): p. 119-124.
91. OPS, M.d.S.d.G.d.C.y., *Evaluación de la efectividad de la vacuna contra la COVID-19 en Chile*. 2022.
92. Liu, Y., et al., *Reduction in the infection fatality rate of Omicron (B. 1.1. 529) variant compared to previous variants in South Africa*. Available at SSRN 4010080, 2022.
93. Brazeau, N.F., et al., *Report 34 - COVID-19 Infection Fatality Ratio Estimates from Seroprevalence*. 2020, Imperial College London.

94. Watson, O.J., et al., *Global impact of the first year of COVID-19 vaccination: a mathematical modelling study*. The Lancet Infectious Diseases, 2022.