

COVID-19 y Modelos No Lineales

Sarai Elisabet Gómez Ibarra
Universidad Autónoma de Nuevo León
Facultad de Ciencias Físico - Matemáticas

Este estudio se realizó con el apoyo del proyecto de CB-15-252996 bajo
la supervisión de la Dra. Graciela González Farías

Septiembre 2020

Resumen

Presentamos una comparación del desempeño de varios modelos no lineales para ajustar datos de crecimiento de la epidemia COVID-19 en varios países.

Índice

1. Agradecimientos	2
2. Introducción	2
3. Teoría	2
3.1. Modelos considerados	3
3.1.1. Modelo Gompertz	3
3.1.2. Modelo Logístico	4
3.1.3. Modelo Normal	5
3.2. Estimación de los parámetros	6
4. Resultados	6
4.1. Modelos para México	6
4.1.1. Modelo Gompertz	6
4.1.2. Modelo Logístico	8
4.1.3. Modelo Normal	10
4.2. Modelos para Egipto	11
4.3. Modelos para Rusia	13
4.4. Modelos para Brasil	15
4.5. Modelos para Alemania	17
5. Comparación de los modelos	19
5.1. Comparación para México	19
5.2. Comparación para Egipto	20
5.3. Comparación para Rusia	20
5.4. Comparación para Brasil	21
5.5. Comparación para Alemania	21
6. Conclusiones	22

1. Agradecimientos

Quiero agradecer primeramente a todos mis asesores; al Dr. Eduardo Cordero por instruirme en la estadística, al Dr. Fernando Camacho por creer en mí, al Dr. Rogelio Ramos, Dra. Gaciela González y los demás instructores del CIMAT y al CONACYT por darme la oportunidad de formarme como investigador.

2. Introducción

La COVID-19 es la enfermedad infecciosa causada por el coronavirus; dicha enfermedad tuvo origen en Wuhan (China) en diciembre de 2019. Actualmente la COVID-19 es una pandemia que afecta a muchos países de todo el mundo.

Este estudio tiene como objetivo el comparar los modelos Gompertz, Logístico y Normal, estimados con los datos arrojados de diferentes países, para validar cual es el desempeño que tiene cada uno de estos y así finalmente descubrir cual es el modelo que mejor se acople al comportamiento de la pandemia en los países afectados.

Dentro del estudio se mantiene como base la utilización de tres modelos no lineales, para esto se toma en cuenta los datos de la pandemia causada por la COVID-19 arrojados en diferentes países.

3. Teoría

DEF Una función sigmoidea es una función real diferenciable y acotada que se define para todos los valores de entrada reales y tiene una derivada no negativa en cada punto.

DEF La función de distribución acumulativa de una variable aleatoria de valor real es la función dada por X .

$$F_X(x) = P(X \leq x)$$

DEF El Error Cuadrático Medio (MSE) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre lo estimado y el valor real; Sea \hat{Y} el vector de n predicciones y Y el vector de los valores reales, entonces el MSE del predictor es:

$$MSE = \frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2$$

DEF La Raíz del Error Cuadrático medio (RMSE) es una medida mayormente usada para valores de muestra o población; nos ayuda a calcular la diferencia entre los valores predichos y los valores observados; se define como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{Y}_i - Y_i)^2}$$

DEF El Error Absoluto Medio de igual manera es una medida de diferencia entre dos variables continuas; normalmente se usa cuando los datos observados (Y) contienen valores atípicos, esta definido como:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{Y}_i - Y_i|$$

Teorema fundamental del calculo. Sea f una función continua en el intervalo cerrado $[a, b]$ y sea $g(x)$ cualquier función definida en $[a, b]$. Si F es la función definida por:

$$F(x) = \int_0^{g(x)} f(t)dt$$

entonces $F'(x) = f(g'(x)) \iff \frac{d}{dx} \int_0^{g(x)} f(t)dt = F(g(x))g'(x)$

3.1. Modelos considerados

Los modelos a elegir son funciones sigmoidales o bien funciones de distribución acumulada. Se escogen este tipo de funciones ya que las sigmoides tienen su dominio en todos los números reales, el valor de retorno aumenta monótonamente y normalmente la primer derivada tiene forma de campana. Por esto mismo se escogieron los siguientes modelos:

1. Modelo Gompertz
2. Modelo Logístico
3. Modelo Normal

3.1.1. Modelo Gompertz

La curva de Gompertz tiene nombre de su diseñador Benjamin Gompertz, el cual diseño el modelo para la Royal Society para detallar su ley de mortalidad humana. El modelo Gompertz esta definido como:

$$Y_{ij} = G(t_{ij}, \theta_i) + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, n_{ij} \quad (1)$$

Donde ε_{ij} es una variable aleatoria independiente y se distribuye $N(0, \sigma^2)$, θ_i tambien una variable aleatoria independiente y se distribuye $N(0, \Omega)$, $F(t_{ij}, \theta_i)$ es el modelo Gompertz; para este caso de estudio $i = 1, 2, 3$, esto es, el modelo $G(t_{ij}, \theta_i)$ tiene 3 parámetros a estimar, entonces $G(t_{ij}, \theta_i)$ es de la forma:

$$G(t) = \alpha e^{-\beta e^{-\kappa t}} \quad (2)$$

Donde:

- $G(t)$ es el numero acumulado de casos confirmados en el tiempo t .
- α corresponde a la asíntota, en este caso α estima el numero de casos al final de la epidemia.

Ahora, si derivamos la ec. (2) nos arroja la curva que modela el número casos diarios confirmados en el tiempo t , donde la derivada queda de la siguiente manera:

$$g(t) = \beta \kappa G(t) e^{-\kappa t} = \beta \kappa \alpha e^{-\beta e^{-\kappa t}} e^{-\kappa t} \quad (3)$$

De las ecuaciones (2) y (3) podemos estimar datos de interés para la pandemia, los cuales son:

- El tiempo t que corresponde al pico de la pandemia, o bien, el tiempo en el cual se espera la máxima incidencia diaria, lo obtenemos buscando el máximo global de la ec. (3), el cual es:

$$t_{max} = \frac{\log(\beta)}{\kappa}$$

- Cuantos casos se estiman para el tiempo t_{max} , es decir, cuantos casos se esperan para el pico; esto se estimaría de la siguiente manera:

$$g(t_{max}) = \beta \kappa \alpha e^{-\beta e^{-\kappa t_{max}}} e^{-\kappa t_{max}}$$

- Estimar el numero de casos acumulados al tiempo t_{max} , o bien, cuantos casos acumulados se esperan desde el tiempo 0 hasta el pico, esto se estima como $G(t_{max})$

Para estimar en que tiempo t se tiene el 95 % de avance en la pandemia, se realiza el siguiente análisis:

$$\text{Sea } G(t) = \alpha e^{-\beta e^{-\kappa t}} \quad ; \quad 0 \leq t < \infty$$

$$\text{Buscamos que } G(t) = 0,95\alpha$$

$$\text{esto es,} \quad \alpha e^{-\beta e^{-\kappa t}} = 0,95\alpha$$

$$\implies e^{-\beta e^{-\kappa t}} = 0,95$$

$$\implies e^{-\kappa t} = \frac{-\ln(,95)}{\beta}$$

$$\therefore t = \frac{-\ln(\frac{-\ln(,95)}{\beta})}{\kappa}$$

3.1.2. Modelo Logístico

El Modelo de Regresión Logística fue publicado en 1844 por Pierre François, con el fin de estudiar el crecimiento de la población. el modelo se ha usado en estudios para relacionar el tiempo con la población, por ejemplo, el numero de especies (Y) con el tiempo (X). El modelo logístico es de la siguiente manera:

$$Y_{ij} = H(t_{ij}, \theta_{1j}) + \varepsilon_{ij} \quad i = 1, \dots, n \quad j = 1, \dots, n_{ij} \quad (4)$$

Donde:

- ε_{ij} es una variable aleatoria independiente y $\varepsilon_{ij} \sim N(0, \sigma_1^2)$
- θ_i tambien es una variable aleatoria independiente, la cual se distribuye $N(0, \Omega_1)$

Para este estudio θ_{1j} es de dimensión 3 y la función $H(t_{ij}, \theta_{1j})$ esta definida como:

$$H(t) = \frac{\gamma_0}{1 + \gamma_1 e^{-\gamma_2 t}} \quad (5)$$

Donde:

- $H(t)$ es el número acumulado de casos confirmados en el tiempo t .
- γ_0 es el valor de crecimiento máximo, para este estudio γ_0 representa los casos predichos para el final de la pandemia.
- γ_2 es la tasa de crecimiento logístico o la pendiente de la curva.

Al derivar la ec. (5) nos arroja la curva que modela los casos diarios en el tiempo t , la curva queda de la siguiente manera:

$$\frac{\gamma_0 \gamma_1 \gamma_2 e^{(-\gamma_2 t)}}{(1 + \gamma_1 e^{-\gamma_2 t})^2} \quad (6)$$

De las ecuaciones (5) y (6) podemos estimar datos interesantes para la pandemia, los cuales son:

- En que tiempo t se encuentra la máxima incidencia diaria, esto se obtiene al buscar el punto máximo global de la ec. (6), el cual es:

$$t_{maxL} = \frac{\ln(\gamma_1)}{\gamma_2}$$

- Cuantos casos se estiman para el pico, esto se obtiene con:

$$h(t_{maxL}) = \frac{\gamma_0 \gamma_1 \gamma_2 e^{(-\gamma_2 t_{maxL})}}{(1 + \gamma_1 e^{-\gamma_2 t_{maxL}})^2}$$

- Cuantos casos acumulados se esperan para la máxima incidencia diaria, esto se estima como $H(t_{maxL})$.

Ahora, para estimar en que valor t se encuentra el 95 % de avance en la pandemia, se realiza el siguiente análisis:

Consideremos la función $H(t) = \frac{\gamma_0}{1+\gamma_1 e^{-\gamma_2 t}}$; $0 \leq t < \infty$

Buscamos el valor t^* que nos arroje $H(t^*) = 0,95\gamma_0$

esto es, $\frac{\gamma_0}{1+\gamma_1 e^{-\gamma_2 t^*}} = 0,95\gamma_0$

$$\implies \frac{1}{1+\gamma_1 e^{-\gamma_2 t^*}} = 0,95$$

$$\implies \frac{20}{19} - 1 = \gamma_1 e^{-\gamma_2 t^*}$$

$$\implies \ln\left(\frac{1}{\gamma_1}\right) = -\gamma_2 t^*$$

$$\therefore t^* = \frac{-\ln\left(\frac{1}{\gamma_1}\right)}{\gamma_2}$$

3.1.3. Modelo Normal

El modelo normal es la distribución normal acumulada con modificaciones, este modelo asume que los casos acumulados siguen una función de error gaussiana parametrizada, la cual esta definida como:

$$N(t) = \frac{p}{2}(\Psi(\alpha(t - \beta))) = \frac{p}{2} \left(1 + \frac{2}{\sqrt{\pi}} \int_0^{\alpha(t-\beta)} \exp(-\tau^2) d\tau \right) \quad (7)$$

Donde:

- la función Ψ es la función de error gaussiana (escrita explícitamente arriba).
- $\frac{p}{2}$ es el valor de crecimiento máximo, o bien, la estimación de cuantos casos se esperan para el final de la pandemia.
- α es un parámetro de crecimiento.
- β es el tiempo en el que la tasa de aumento es máxima.

Para conseguir la curva que modele los casos diarios, derivamos la ec. (7) con el primer teorema fundamental del calculo, el cual nos da:

$$n(t) = \frac{p\alpha}{2\sqrt{\pi}} \left(e^{\frac{-1}{2}\alpha^2(t-\beta)^2} \right) \quad (8)$$

De las ecuaciones (7) y (8), se estiman información de interés, los cuales son:

- Para estimar el tiempo t en el cual se encuentre la incidencia máxima de casos diarios, basta con estimar el parámetro β
- Cuantos casos se esperan para el pico, esto se calcula evaluando β en la ec. (8).
- Cuantos casos acumulados se esperan en el tiempo $t = \beta$, esto se obtiene evaluando β en la ec. (7).

Para estimar en que tiempo t se encuentra el 95 % de avance en la pandemia, se realiza lo siguiente:

Consideremos la función $N(t) = \frac{p}{2}(\Psi(\alpha(t - \beta)))$; $0 \leq t < \infty$

Buscamos el valor t' que nos arroje $N(t') = \frac{0,95p}{2}$

esto es, $\frac{p}{2}(\Psi(\alpha(t' - \beta))) = \frac{0,95p}{2}$

$$\implies \Psi(\alpha(t' - \beta)) = 0,95$$

$$\implies \alpha(t' - \beta) = qnorm(0,95) \quad (\star)$$

$$\implies t' = \beta + \frac{1}{\alpha}qnorm(0,95)$$

(\star) $qnorm$ es una función que determina el cuantil de una probabilidad dada.

3.2. Estimación de los parámetros

Para realizar la estimación de los parámetros se utilizaron las funciones `nls` y `nlsLM`, las cuales son funciones que realizan las estimaciones por el método de mínimos cuadrados no lineales.

La función `nlsLM` en especial, realiza la estimación por mínimos cuadrados no lineales con el método de Levenberg-Marquardt, el cual es un algoritmo interactivo semejante a los métodos de máximo descenso y el método de Gauss-Newton; esto es porque, cuando la solución se encuentra lejos del mínimo local el algoritmo se comporta como el método de máximo descenso y cuando la solución esta cercana al mínimo local, el método se asemeja al método de Gauss-Newton.

Levenberg-Marquardt es una buena alternativa para evitar los problemas que le método Gauss-Newton presenta cuando la matriz Jacobiana es no singular.

4. Resultados

En esta sección veremos los modelos resultantes para cada uno de los países seleccionados, los cuales son:

1. México
2. Egipto
3. Rusia
4. Brasil
5. Alemania

Los cálculos y gráficos se realizaron en RStudio.

4.1. Modelos para México

4.1.1. Modelo Gompertz

La estimación de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

$$\begin{aligned}\alpha &= 1,046,175 \\ \beta &= 11,36096 \\ \kappa &= 0,01627938\end{aligned}$$

Datos importantes de la pandemia:

1. Se estima que al final de la pandemia se tendrían $\alpha = 1,046,175$ casos acumulados.
2. Se predice que el pico de la pandemia fue el día 26 de Julio del 2020.

$$t_{max} = \frac{\ln(\beta)}{\kappa} = 149 = 2020 - 07 - 26$$

3. El numero de casos acumulados que se esperaban en el pico (2020-07-26) es

$$G(149) = 383,113$$

4. Se estima que el 95 % de avance de la pandemia sea el 2021-01-25

$$t = \frac{-\ln\left(\frac{-\ln(.95)}{\beta}\right)}{\kappa} \implies t = 322 = 2021 - 01 - 25$$

A continuación, se muestran los gráficos de la modelación Gompertz

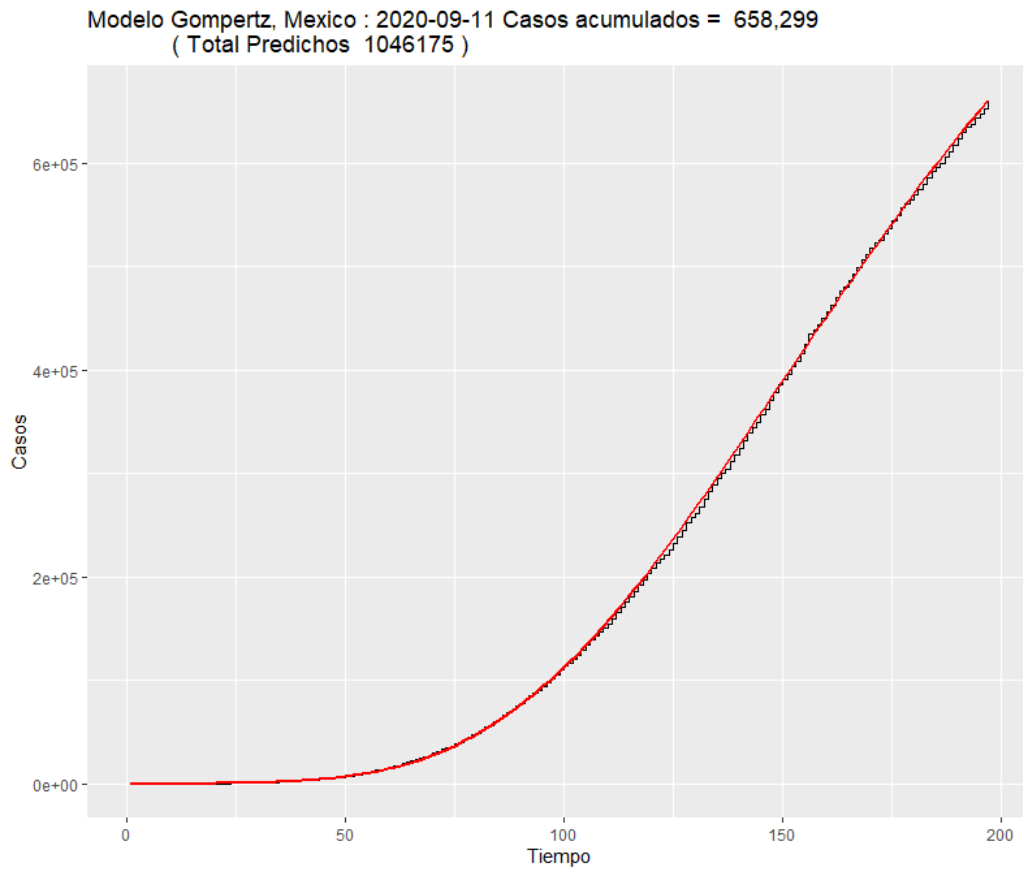


Figura 1: Modelo Gompertz para los casos acumulados en México

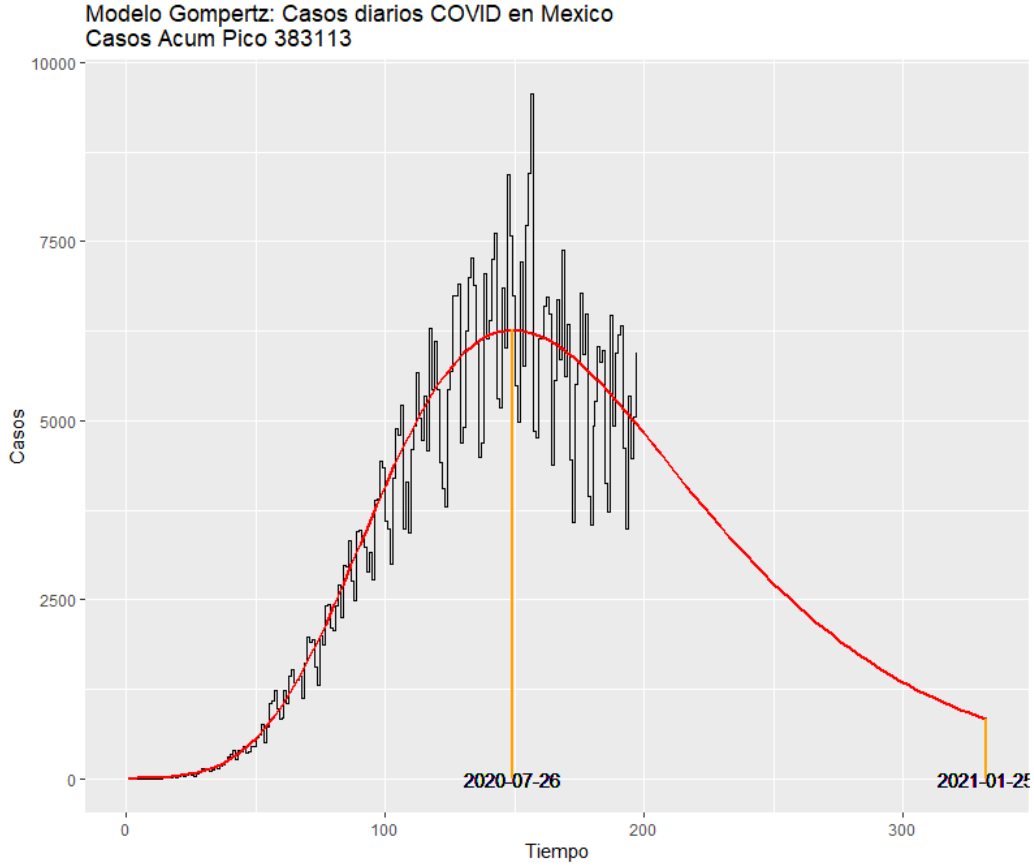


Figura 2: Modelo Gompertz para los casos diarios en México

4.1.2. Modelo Logístico

La estimación de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

$$\begin{aligned}\gamma_0 &= 731,650 \\ \gamma_1 &= 261,2094 \\ \gamma_2 &= 0,03812863\end{aligned}$$

Datos importantes de la pandemia:

1. Se estima que al final de la pandemia se tendrían $\gamma_0 = 731,650$ casos acumulados.
2. Se predice que el pico de la pandemia fue el día 23 de Julio del 2020.

$$t_{maxL} = \frac{\ln(\gamma_1)}{\gamma_2} = 146 = 2020 - 07 - 23$$

3. El numero de casos acumulados que se esperaban en el pico (2020-07-23) es

$$H(146) = 366,092$$

4. Se estima que el 95 % de avance de la pandemia sea el 2020-10-08

$$t^* = \frac{-\ln(\frac{1}{\gamma_1})}{\gamma_2} \implies t^* = 223 = 2020 - 10 - 08$$

A continuación, se muestran los gráficos de la modelación Logística

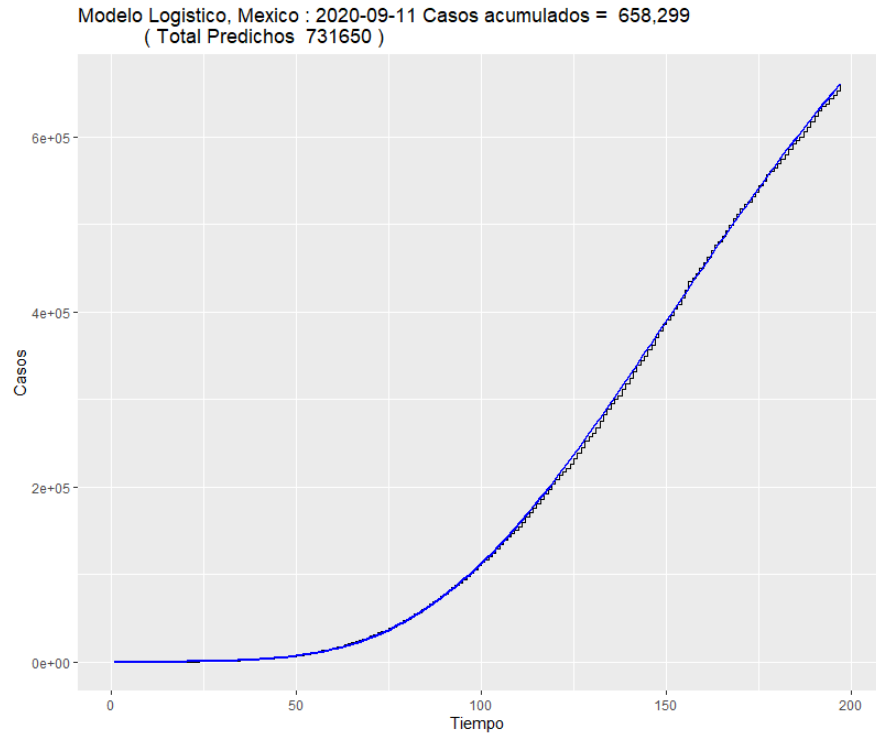


Figura 3: Modelo Logístico para los casos acumulados en México

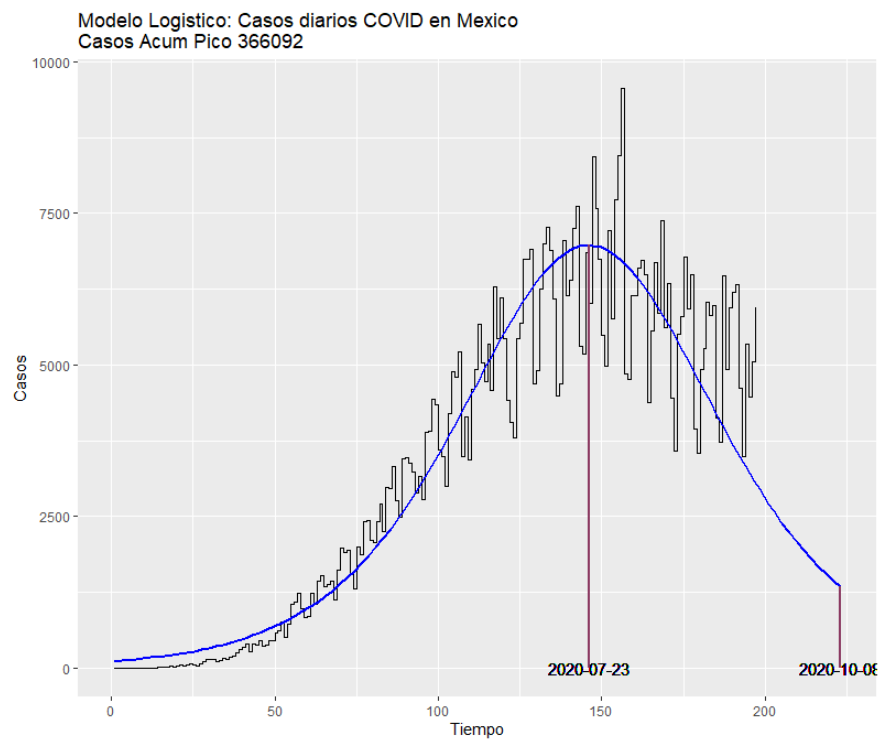


Figura 4: Modelo Logístico para los casos diarios en México

4.1.3. Modelo Normal

La estimación de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

$$\begin{aligned}p &= 1,510,456 \\ \alpha &= 0,02194987 \\ \beta &= 147,9341\end{aligned}$$

Datos importantes de la pandemia:

1. Se estima que al final de la pandemia se tendrían $\frac{p}{2} = 755,228$ casos acumulados.
2. Se predice que el pico de la pandemia fue el día 23 de Julio del 2020.

$$t_{maxL} = \frac{\ln(\gamma_1)}{\gamma_2} = 148 = 2020 - 07 - 25$$

3. El numero de casos acumulados que se esperaban en el pico (2020-07-25) es

$$H(148) = 378,050$$

4. Se estima que el 95 % de avance de la pandemia sea el 2020-10-08

$$t' = \beta + \frac{1}{\alpha} qnorm(0,95) \quad \implies t' = 223 = 2020 - 10 - 08$$

A continuación, se muestran los gráficos del modelo Normal

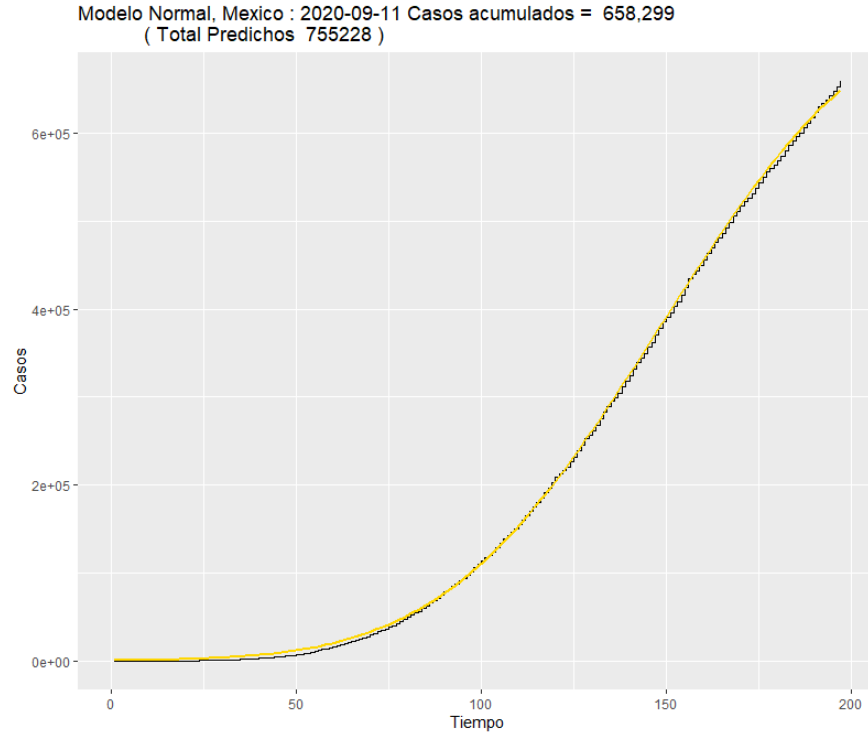


Figura 5: Modelo Normal para los casos acumulados en México

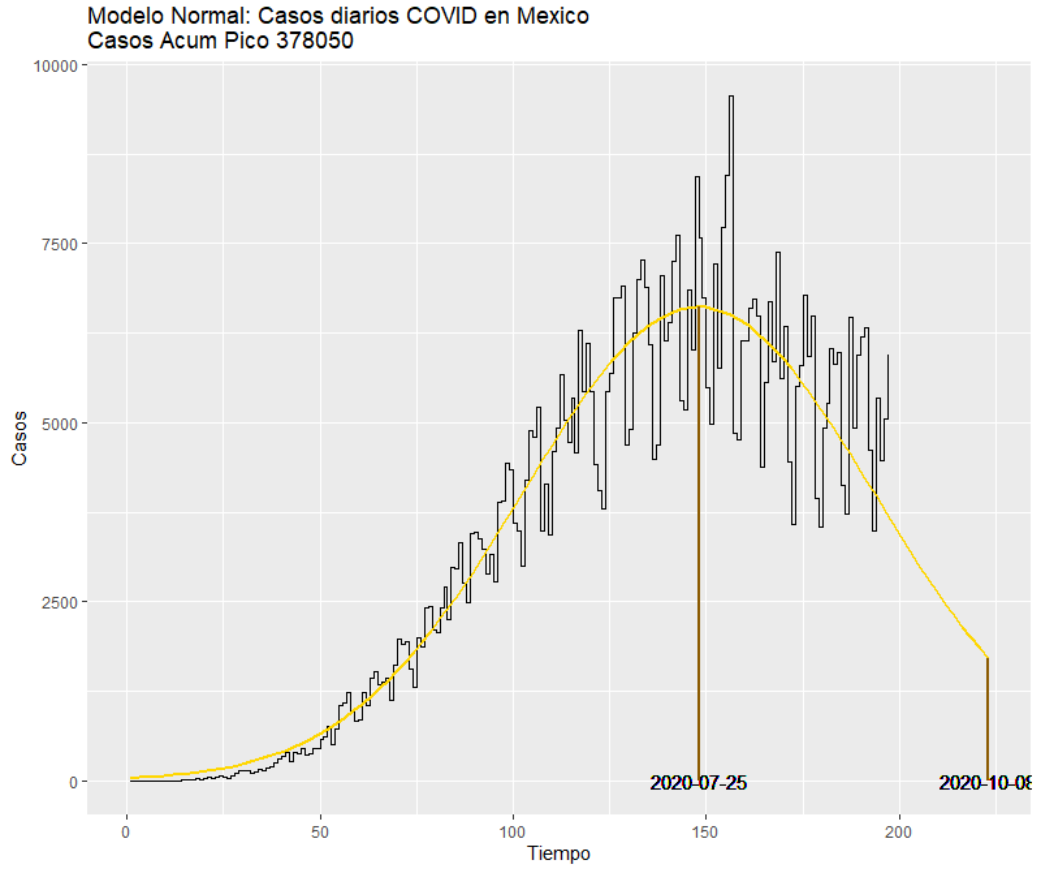


Figura 6: Modelo Normal para los casos diarios en México

4.2. Modelos para Egipto

Las estimaciones de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

	Modelo		
	<i>Gompertz</i>	<i>Logístico</i>	<i>Normal</i>
Parámetros	$\alpha = 104630.3$	$\gamma_0 = 99820.28$	$p = 197865.6$
	$\beta = 87.88444$	$\gamma_1 = 2407.838$	$\alpha = 0.03722599$
	$\kappa = 0.03853526$	$\gamma_2 = 0.06205153$	$\beta = 125.0506$
Casos acumulados al final de la pandemia	104630	99820	98933
Fecha del pico	2020-06-09	2020-06-18	2020-06-18
Casos acumulados estimados para el pico	38263	49160	49392
Fecha del 95% de avance en la pandemia	2020-08-25	2020-08-05	2020-08-01

Figura 7: Datos relevantes de la pandemia en Egipto

Modelo Gompertz:

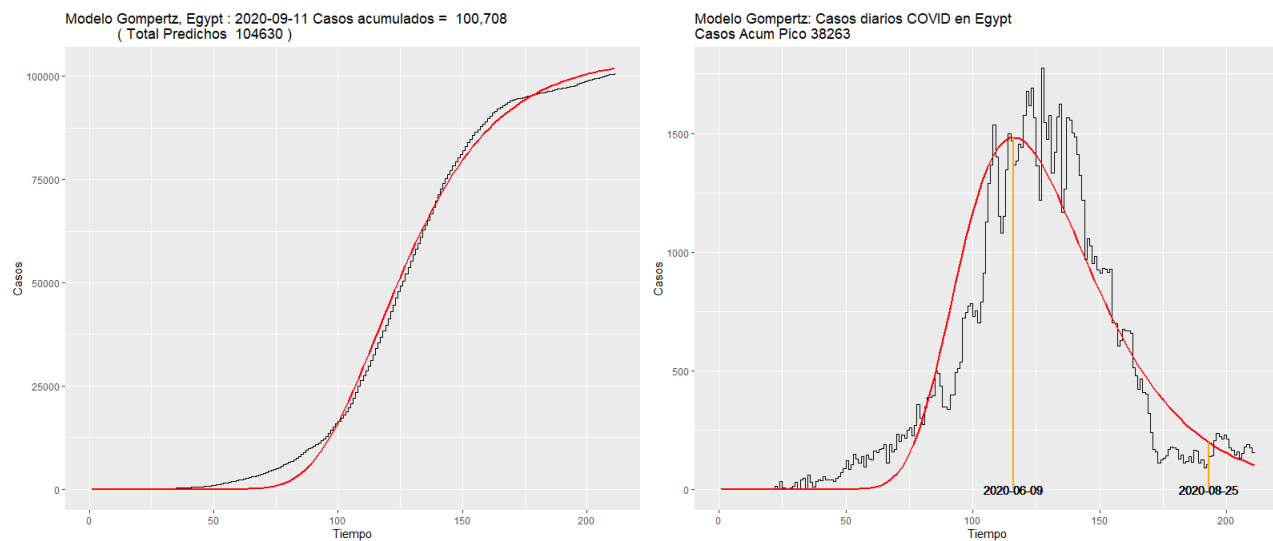


Figura 8: Modelo Gompertz para casos acumulado y modelo Gompertz para casos diarios en Egipto

Modelo Logístico:

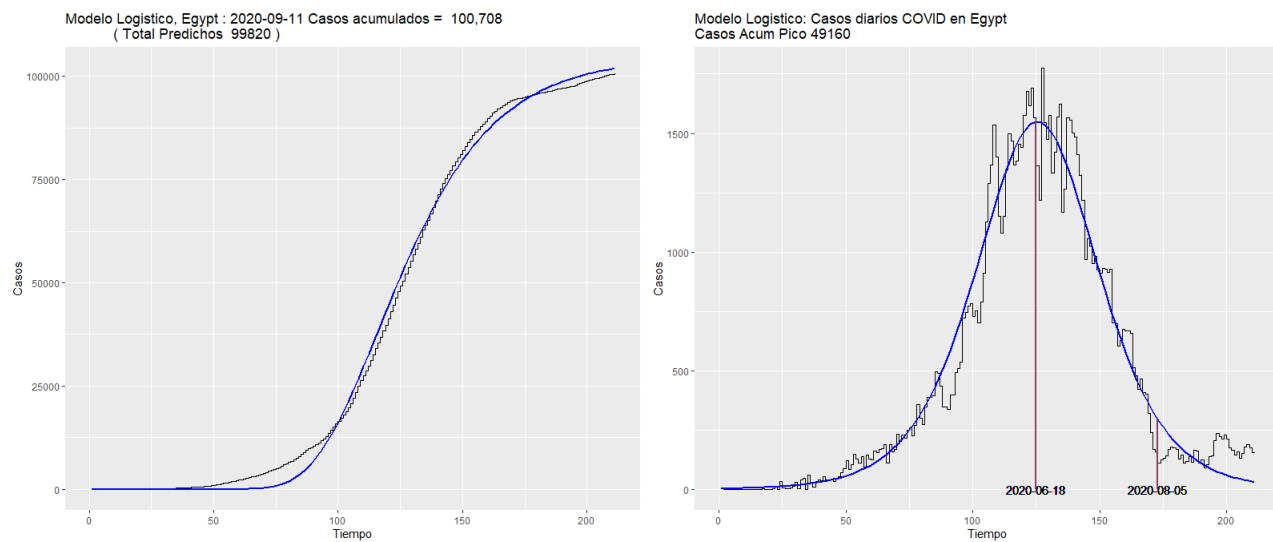


Figura 9: Modelo Logístico para casos acumulado y modelo Logístico para casos diarios en Egipto

Modelo Normal:

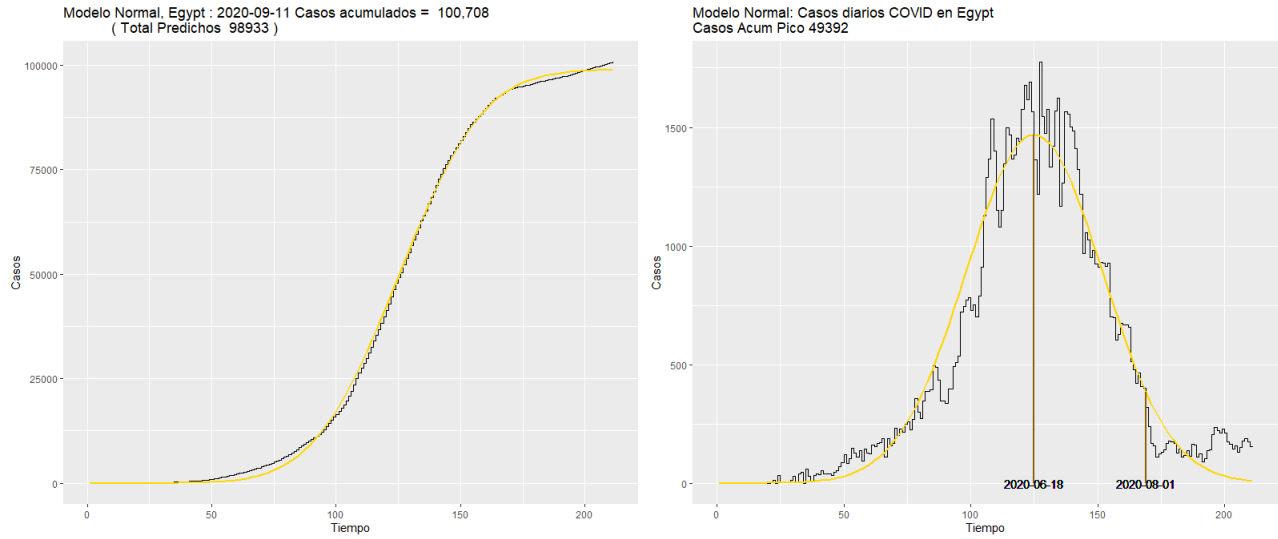


Figura 10: Modelo Normal para casos acumulado y modelo Normal para casos diarios en Egipto

4.3. Modelos para Rusia

Las estimaciones de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

	Modelo		
	<i>Gompertz</i>	<i>Logístico</i>	<i>Normal</i>
Parámetros	$\alpha = 1110127$	$\gamma_0 = 1000001$	$p = 1993877$
	$\beta = 17.48475$	$\gamma_1 = 249.8297$	$\alpha = 0.0241385$
	$\kappa = 0.02300265$	$\gamma_2 = 0.04064987$	$\beta = 135.7528$
Casos acumulados al final de la pandemia	1110127	1000001	996939
Fecha del pico	2020-06-03	2020-06-15	2020-06-15
Casos acumulados estimados para el pico	404717	501901	500842
Fecha del 95% de avance en la pandemia	2020-10-11	2020-08-26	2020-08-22

Figura 11: Datos relevantes de la pandemia en Rusia

Modelo Gompertz:

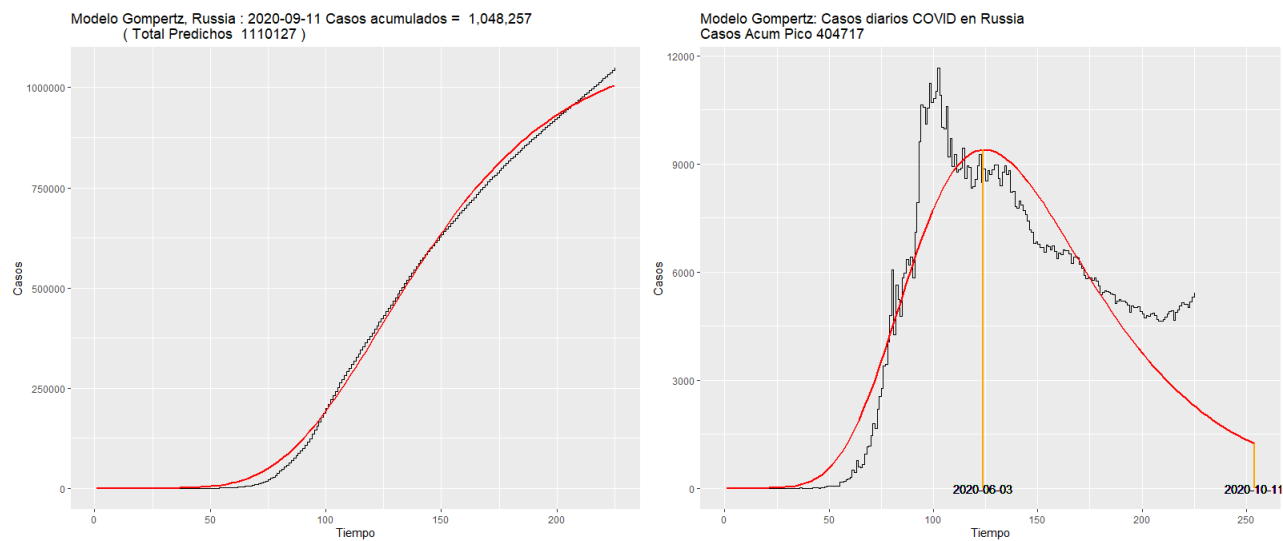


Figura 12: Modelo Gompertz para casos acumulado y modelo Gompertz para casos diarios en Rusia

Modelo Logístico:

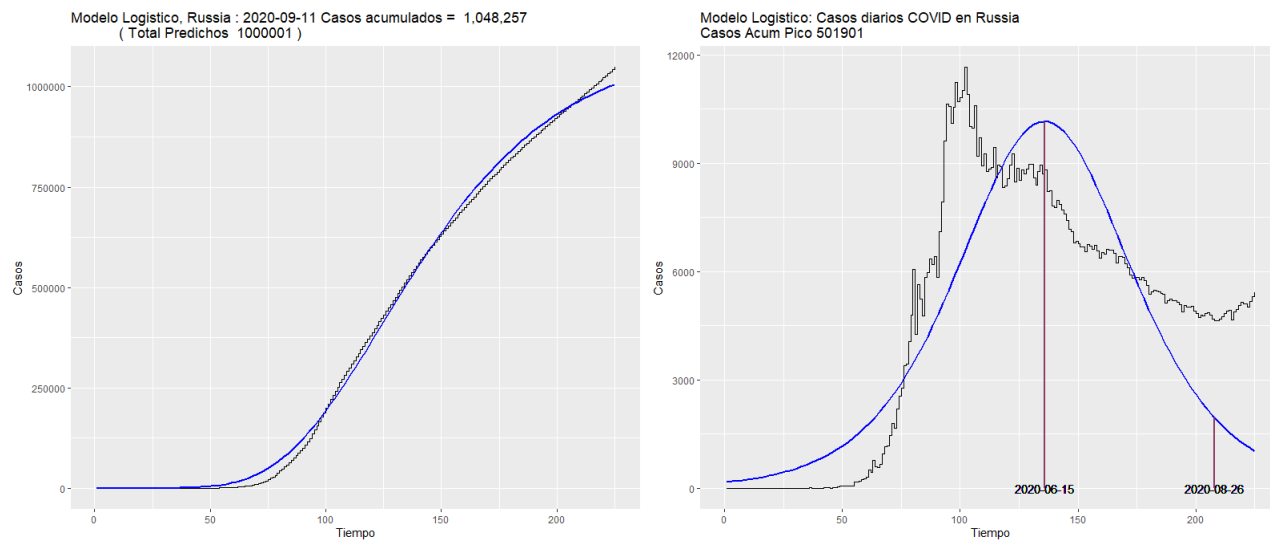


Figura 13: Modelo Logístico para casos acumulado y modelo Logístico para casos diarios en Rusia

Modelo Normal:

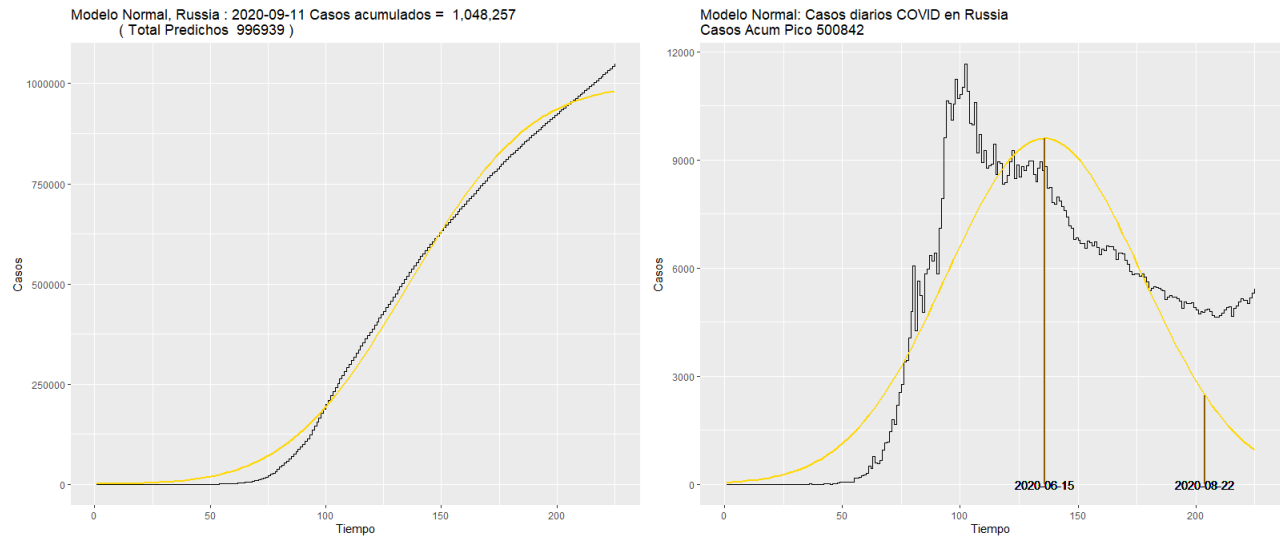


Figura 14: Modelo Normal para casos acumulado y modelo Normal para casos diarios en Rusia

4.4. Modelos para Brasil

Las estimaciones de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

	Modelo		
	<i>Gompertz</i>	<i>Logístico</i>	<i>Normal</i>
Parámetros	$\alpha = 7088709$	$\gamma_0 = 4852805$	$p = 10085787$
	$\beta = 12.47444$	$\gamma_1 = 335.9201$	$\alpha = 0.0220306$
	$\kappa = 0.01618119$	$\gamma_2 = 0.03843174$	$\beta = 153.7158$
Casos acumulados al final de la pandemia	7088709	4852805	5042894
Fecha del pico	2020-07-31	2020-07-26	2020-07-29
Casos acumulados estimados para el pico	2609312	2409805	2534044
Fecha del 95% de avance en la pandemia	2021-01-31	2020-10-11	2020-10-11

Figura 15: Datos relevantes de la pandemia en Brasil

Modelo Gompertz:

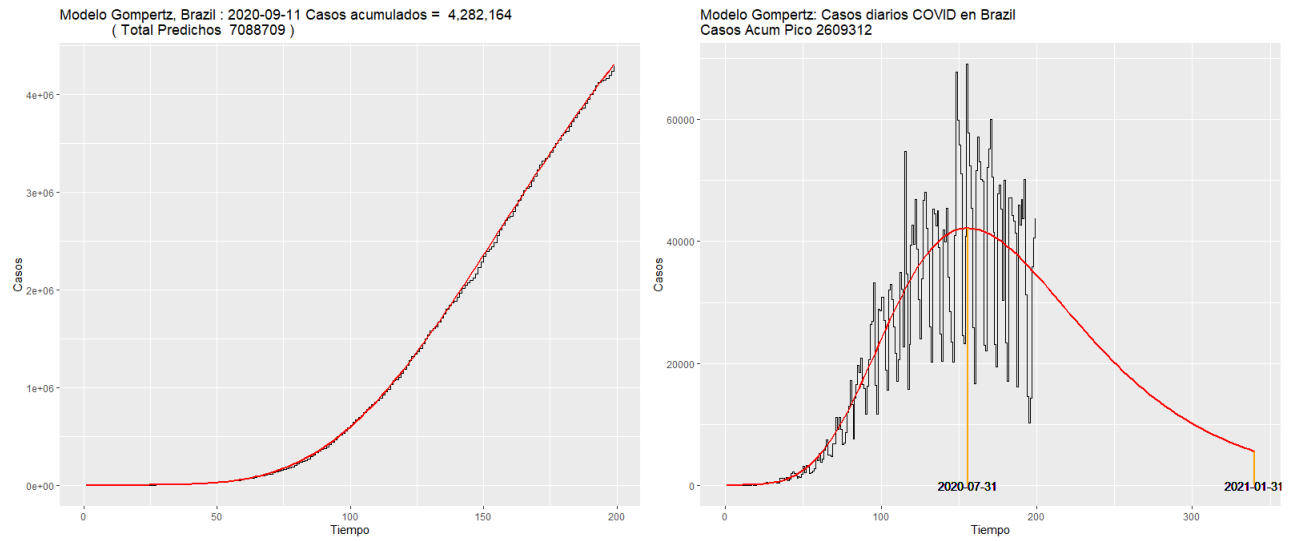


Figura 16: Modelo Gompertz para casos acumulado y modelo Gompertz para casos diarios en Brasil

Modelo Logístico:

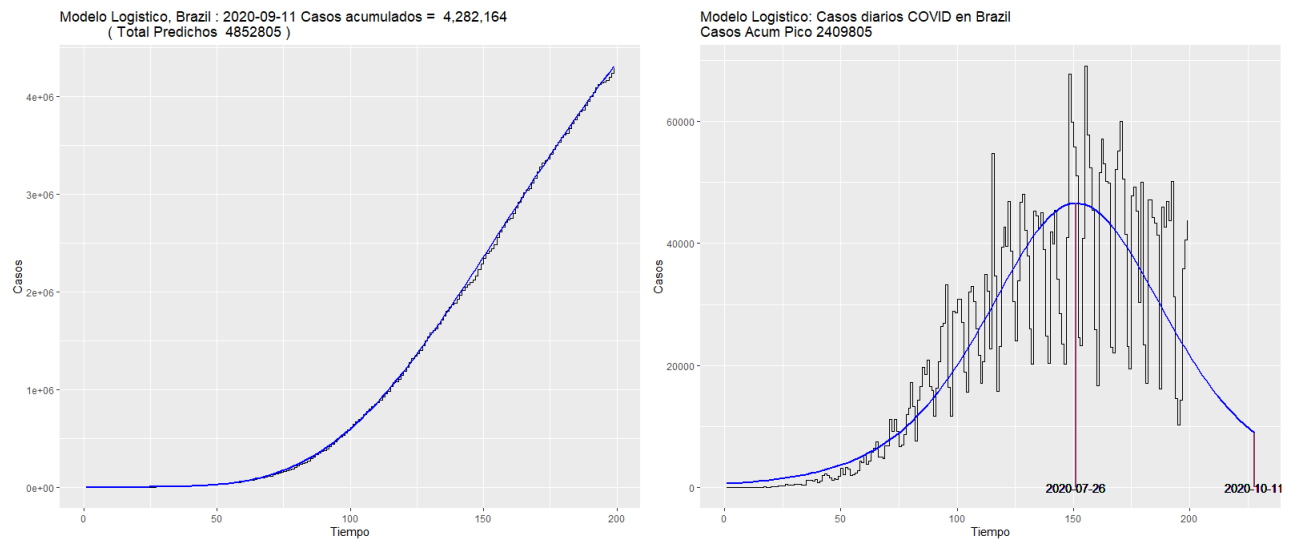


Figura 17: Modelo Logístico para casos acumulado y modelo Logístico para casos diarios en Brasil

Modelo Normal:

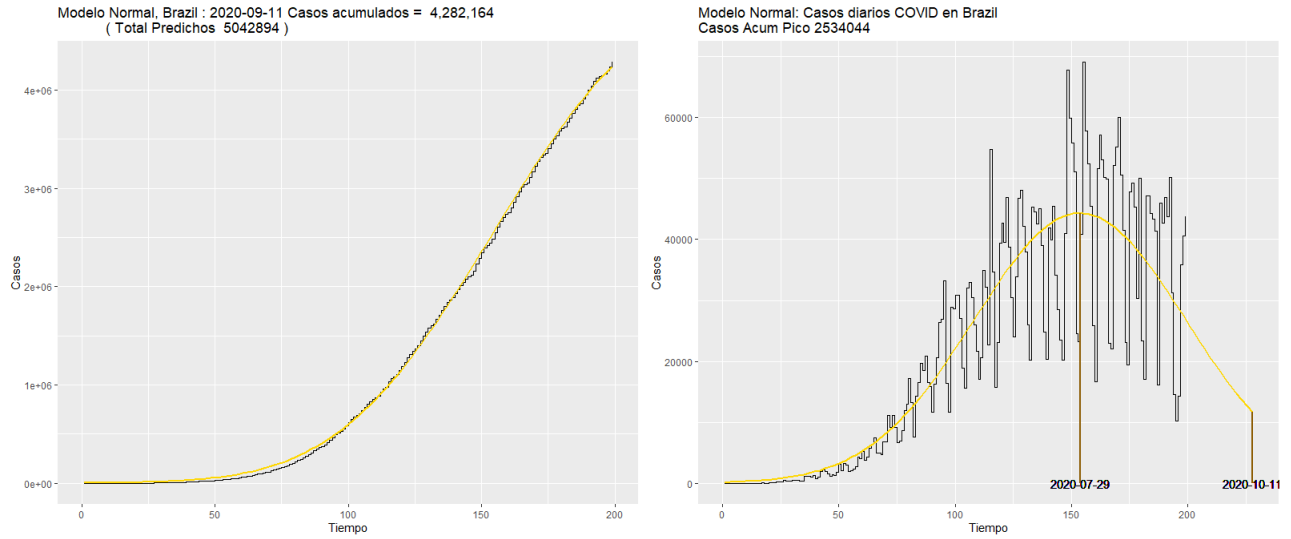


Figura 18: Modelo Normal para casos acumulado y modelo Normal para casos diarios en Brasil

4.5. Modelos para Alemania

Las estimaciones de los parámetros se realizaron con los datos hasta el 11 de septiembre del 2020:

	Modelo		
	<i>Gompertz</i>	<i>Logístico</i>	<i>Normal</i>
Parámetros	$\alpha = 215464.3$	$\gamma_0 = 209292.1$	$p = 417016.1$
	$\beta = 17.20218$	$\gamma_1 = 195.1534$	$\alpha = 0.04191388$
	$\kappa = 0.04194959$	$\gamma_2 = 0.06923798$	$\beta = 76.16083$
Casos acumulados al final de la pandemia	215464	209292	208508
Fecha del pico	2020-04-04	2020-04-12	2020-04-12
Casos acumulados estimados para el pico	79862	104034	103693
Fecha del 95% de avance en la pandemia	2020-06-14	2020-05-25	2020-05-21

Figura 19: Datos relevantes de la pandemia en Alemania

Modelo Gompertz:

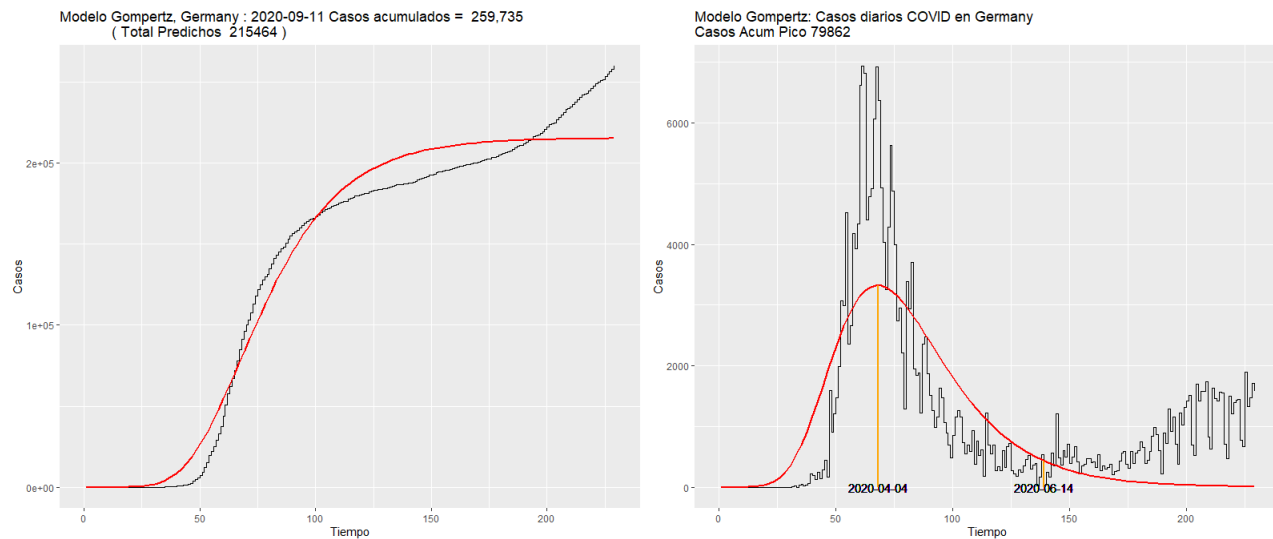


Figura 20: Modelo Gompertz para casos acumulado y modelo Gompertz para casos diarios en Alemania

Modelo Logístico:

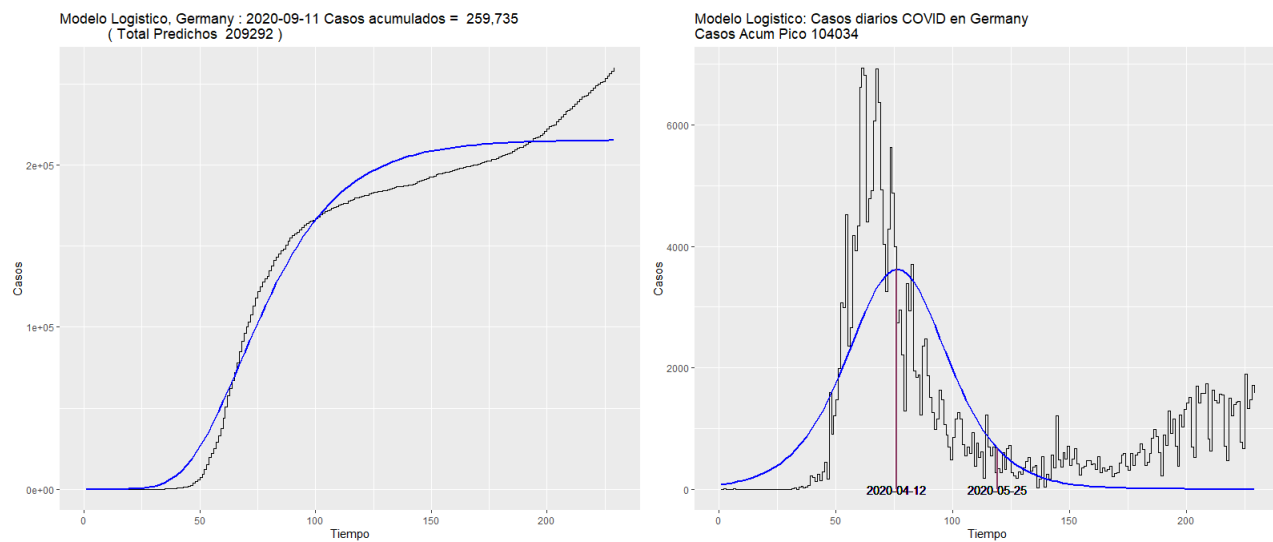


Figura 21: Modelo Logístico para casos acumulado y modelo Logístico para casos diarios en Alemania

Modelo Normal:

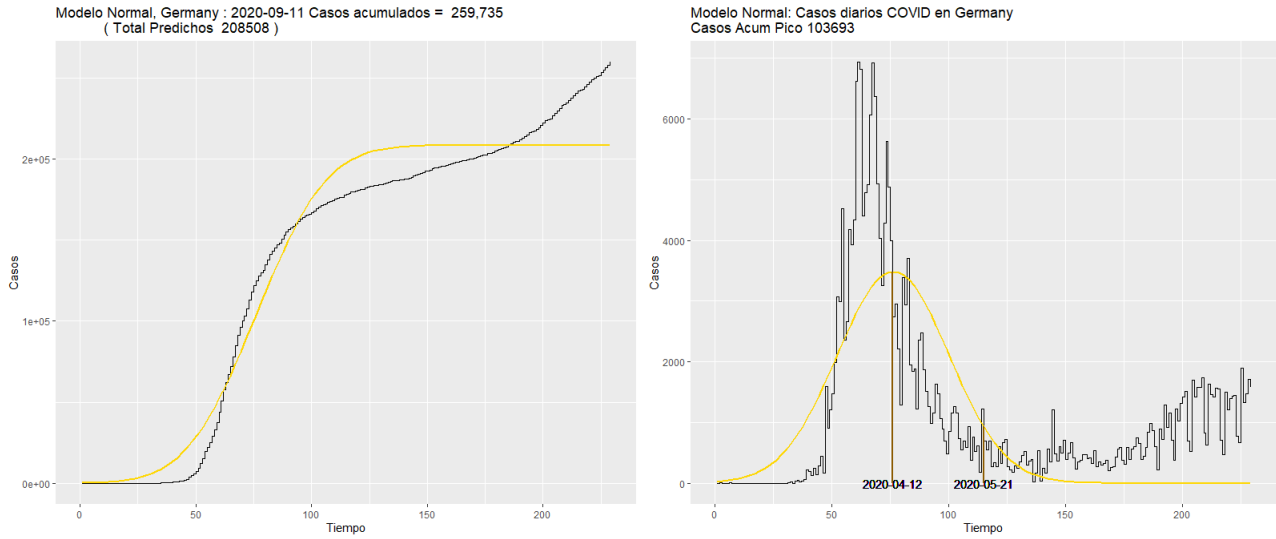


Figura 22: Modelo Normal para casos acumulado y modelo Normal para casos diarios en Alemania

5. Comparación de los modelos

La comparación de modelos a través de procesos estocásticos es una tarea muy complicada, aun mas si los modelos tienen diferente tipo de parámetros; Por tal motivo se optó por realizar la comparación de los modelos midiendo el MSE de los casos acumulados con los casos predichos por el modelo. Aunque esto no sea lo mejor para realizar las comparaciones, nos da buena información para saber cuál modelo se ajusta mejor, esto es, de los tres modelos considerados, el que tenga menor MSE, RMSE o MSA nos indica que ajusta mejor los datos a comparación de los demás modelos considerados.

5.1. Comparación para México

Ah simple vista podemos ver que los tres modelos se ajustan bien para México, pero al ver las gráficas de los casos diarios, es notable que el comportamiento después del pico tiene una caída pesada, por tal motivo es coherente pensar que el modelo Gompertz es el que se ajusta mejor a los datos.

Se realizan los cálculos de cada uno de las medidas de los errores y nos arroja lo siguiente:

		Modelo		
		Gompertz	Logístico	Normal
Métricas del error	MSE	4929030	49934247	8537168
	RMSE	2220.142	7066.417	2921.843
	MAE	1625.118	6250.701	2443.945

Esto comprueba la hipótesis mencionada anteriormente, ya que tanto el MSE, RMSE y MAE del modelo Gompertz son menores a la de los otros dos modelos. Entonces decimos que el Modelo Gompertz genera menor error y por tanto es el modelo que genera mayor desempeño en México a comparación de los Modelos Logístico y Normal.

5.2. Comparación para Egipto

Egipto es un país ubicado en el extremo noreste de África; Egipto tiene un clima desértico, su territorio abarca $1,001,450 \text{ km}^2$ de superficie, con una población de 96.980.000 personas (Calculo realizado el 2018).

Al ver los datos diarios y acumulados, es notorio que Egipto tubo un rápido crecimiento en los casos diarios y al llegar al pico, el descenso es semejante al crecimiento, por tal motivo es factible pensar que el modelo logístico o normal serian buenos para ajustar el modelo, ya que ambos modelos son simétricos con respecto al pico.

Al realizar los callos de las medidas de los errores nos arroja los siguientes datos:

		Modelo		
		Gompertz	Logístico	Normal
Métricas del error	MSE	4512989	277487	931420.8
	RMSE	2124.38	526.7704	965.1014
	MAE	1680.521	414.5745	752.739

El Modelo Logístico es el que tiene menor MSE, RMSE y MAE, esto es, el modelo que genera menor índice de error es el Logístico, por tanto, el modelo Logístico es el que tiene un mejor desempeño a comparación de los modelos Gompertz y Normal.

5.3. Comparación para Rusia

Rusia es un país que se extiende en el norte de Eurasia, con una superficie de $17,098,242 \text{ km}^2$ de territorio, con una población de 146,800,000 personas (calculado al 2018), por su extenso territorio, Rusia tiene una gran variedad de climas, que por lo general predomina el clima continental.

En Rusia el COVID-19 tuvo un crecimiento acelerado, al llegar al pico el descenso es pesado, por tal motivo se espera que el modelo Gompertz sea el que mejor lo estime.

Las métricas del error para cada modelo son las siguientes:

		Modelo		
		Gompertz	Logístico	Normal
Métricas del error	MSE	244663284	1031991531	752250183
	RMSE	15641.72	32124.62	27427.18
	MAE	12384.46	27771.95	22894.85

En efecto el Modelo Gompertz es el modelo que genero menor MSE, RMSE y MAE, por tanto decimos que el Modelo Gompertz es el que tiene mejor desempeño a comparación del modelo logístico y normal, sin embargo, notamos que el modelo Gompertz no es la mejor opción, ya que la cola derecha es mucho más pesada que las colas generadas por el Modelo Gompertz, había que buscar otro modelo o quizá mezclas de modelos.

5.4. Comparación para Brasil

Brasil es un país ubicado en América del sur, con una superficie de 8,5 millones de km^2 con una población de 207,999,646 personas (calculado el 2016), el clima es mayormente tropical.

Las métricas del error para cada modelo son las siguientes:

		Modelo		
		Gompertz	Logístico	Normal
Métricas del error	MSE	272160709	3143640103	907970244
	RMSE	16497.29	56068.17	30132.54
	MAE	10991.82	50541.51	25270

Por el tipo de datos en Brasil no es fácil ver cual modelo podría ajustarse mejor, ya que tiene muchos cambios significativos en los casos diarios; al realizar los cálculos de las métricas del error vemos que el que tiene menor MSE, RMSE y MAE, es el modelo Gompertz, por tanto, el modelo Gompertz es el que tiene mejor desempeño para Brasil.

5.5. Comparación para Alemania

Alemania está ubicado en Europa con un territorio de 357,022 km^2 , Alemania tiene una población de 83,166,711 habitantes (conteo realizado el 2019), con un clima mayormente templado.

Al ver los datos en Alemania es notorio ver el crecimiento es exponencial, llega al pico y el descenso es pesado, también es importante recalcar que cuando está por concluir el descenso, hay un aumento en los casos diarios.

Las métricas del error para cada modelo son las siguientes:

		Modelo		
		Gompertz	Logístico	Normal
Métricas del error	MSE	210275104	292430804	303170363
	RMSE	14500.87	17100.61	17411.79
	MAE	11382.71	13903.18	13864.07

El modelo Gompertz es el que arroja los valores mínimos en las métricas del error, por tanto decimos que el modelo Gompertz es el que tiene mejor desempeño para Alemania, sin embargo, aunque el modelo Gompertz nos arroje los valores mínimos, es importante recalcar el aumento en los nuevos casos, por tal motivo es recomendable realizar mezclas de modelos para tener un mejor panorama de la pandemia en Alemania.

6. Conclusiones

Después de contrastar los modelos con los respectivos datos de cada uno de los países, podemos concluir particularmente para cada uno de ellos.

En el caso de México el modelo Gompertz es el que tiene un mayor desempeño con un $MAE = 1625,118$ a comparación del modelo Logístico que arrojó un $MAE = 6250,701$ y el modelo Normal con un $MAE = 2443,945$.

En cambio, en Egipto es notorio que del inicio de la pandemia al pico tuvo un crecimiento acelerado, llega al pico y el descenso es semejante al ascenso; por tanto es razonable el resultado, el modelo que tiene el mejor desempeño es el modelo Logístico con un $MAE = 414,5745$ a comparación del Gompertz con un $MAE = 1680,521$ y el Normal con $MAE = 752,739$, pero es importante recalcar que el modelo Normal también es aceptable para los datos, esto es por la naturaleza de ambos modelos que son simétricos.

En Rusia el modelo que tuvo el mejor desempeño es el Gompertz con un $MAE = 12384,46$ a comparación del logístico con un $MAE = 27771,95$ y del normal con $MAE = 22894,85$; es importante observar que, aunque el modelo Gompertz tiene un buen desempeño habría que buscar algún otro modelo que tenga un descenso más lento que el Gompertz.

En el caso de Brasil también el modelo Gompertz fue el que arrojó un mejor desempeño con un $MAE = 10991,81$ a comparación del logístico con $MAE = 50541,51$ y del normal con $MAE = 25270$.

El modelo que tiene un mejor desempeño para Alemania es el Gompertz con un $MAE = 11382,71$ a comparación del Logístico con un $MAE = 13903,18$ y del normal con $MAE = 13864,07$. A comparación de los demás países Alemania ya se encontraba a punto de concluir la pandemia según las predicciones de los modelos, en cambio se presenta un aumento en los casos, por tal motivo es recomendable hacer uso de algún otro modelo o mezclas de modelos.

Nótese que el modelo Gompertz es el que tuvo un mejor desempeño en la mayoría de los países, por tanto, es recomendable usar el modelo Gompertz como primera opción para la modelación de la pandemia, en cambio también es importante hacer uso de los demás modelos mencionados ya que cada país tiene un comportamiento distinto ante la situación, estas variaciones pueden ser efecto de muchas cosas, como las medidas preventivas, el tiempo que duro la cuarentena, las bases económicas, entre otras cosas.

Bibliografía

- [1] Prats, C. et al. (2020) Analysis and prediction of COVID-19 for different regions and countries. Daily report 27-03-2020. UPC, BioComSC, CMCiB, IGTP.
- [2] IHME COVID-19 forecasting team (2020). Forecasting COVID-19 impact on hospital beddays, ICU-days, ventilator days and deaths by US state in the next 4 months. Report.
- [3] Huet, S. et al. (2004). Statistical Tools for Nonlinear Regression. Springer.
- [4] Michael H. Kutner, Christopher J. Nachtsheim. (2005). Applied Linear Statistical Models. New York: McGraw-Hill Irwin.
- [5] Cornejo-Zúñiga, Ó.; Rebolledo-Vega, R. (2016). Estimación de parámetros en modelos no lineales: algoritmos y aplicaciones. Revista EIA, 13(25), enero-junio, pp. 81-98. [Online]. Disponible en: DOI: <http://dx.doi.org/10.14508/reia.2016.13.25.81-98>
- [6] Organización Mundial de la Salud (2020). Preguntas y respuestas sobre la enfermedad por coronavirus (COVID-19), de Organización Mundial de la Salud Sitio web: <https://www.who.int/es/emergencies/diseases/novel-coronavirus-2019/advice-for-public/q-a-coronaviruses>