

# Modelos Gompertz

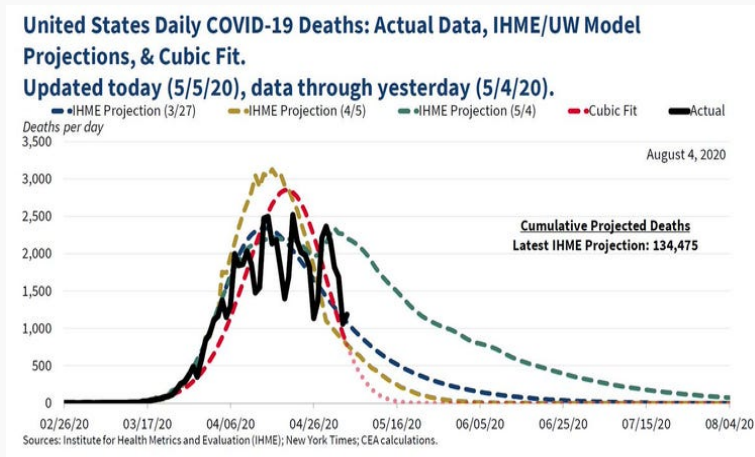
**Rogelio Ramos Quiroga**  
**Graciela González Farías**

`rramosq@cimat.mx`

8 de Mayo, 2020

# Usos y abusos de modelos empíricos

Modelo cúbico usado por la Casa Blanca para justificar que para mediados de mayo ya no habría muertes por COVID-19.



## Ideas iniciales de modelación

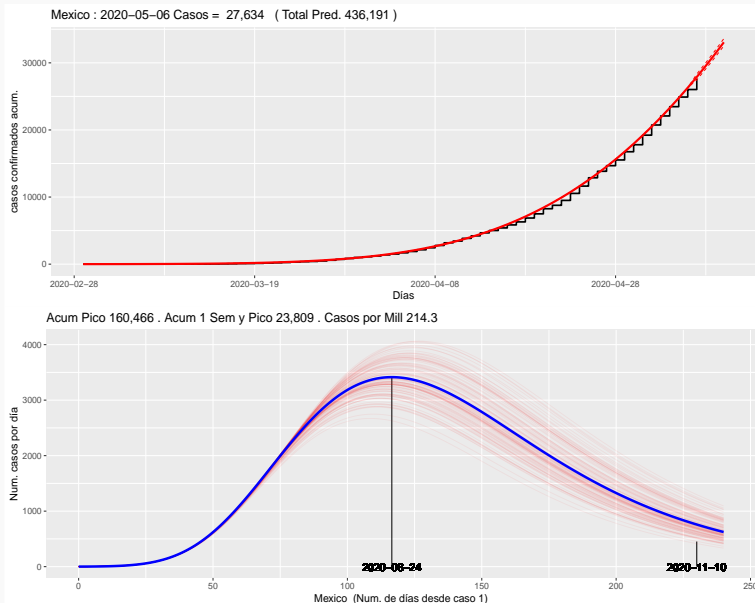
- Modelo Poisson autorregresivo.
  - **Agosto, A. and Giudici, P.** (2020) A Poisson autoregressive model to understand COVID-19 contagion dynamics. ISSN: 2281-1346. University of Pavia

$$y_t \sim \text{Poisson}(\lambda_t)$$

$$\log(\lambda_t) = \omega + \alpha \log(1 + y_{t-1}) + \beta \log(\lambda_{t-1})$$

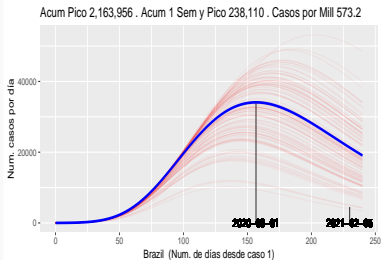
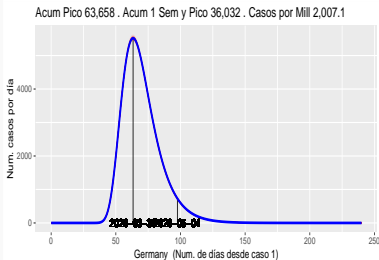
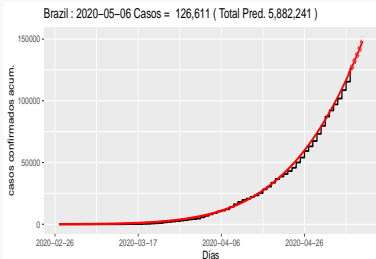
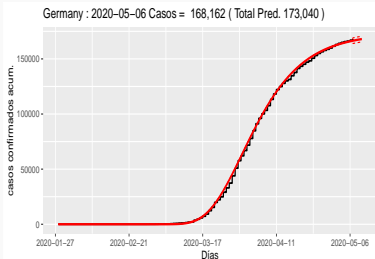
- Modelo Gompertz
  - **Prats, C. et al.** (2020) Analysis and prediction of COVID-19 for different regions and countries. Daily report 27-03-2020. UPC, BioComSC, CMCiB, IGTP.
  - **IHME COVID-19 forecasting team** (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. Report (Modelo Normal).

# Optamos por el Gompertz



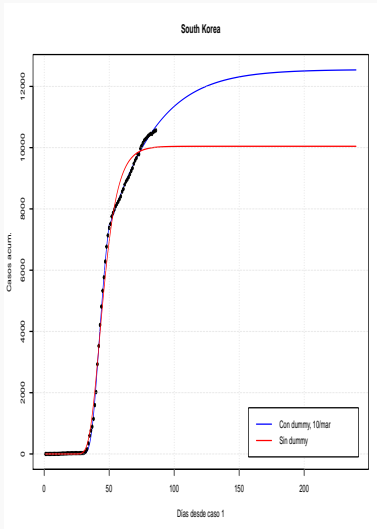
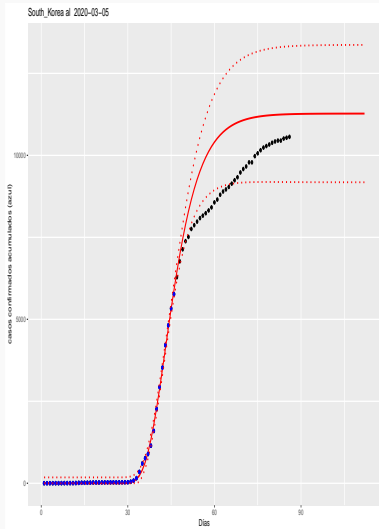
# Es flexible...

Regiones en diferentes etapas de la pandemia.



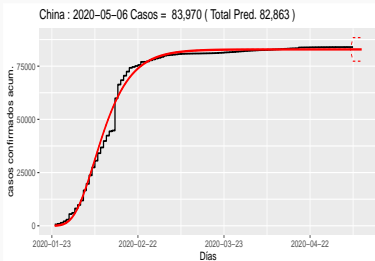
# Es adaptable...

Permite incorporar intervenciones

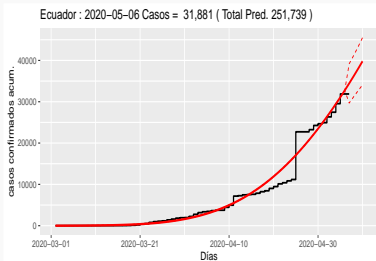
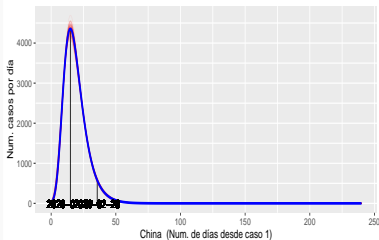


# No es perfecto...

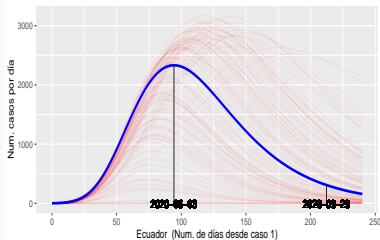
Como todo modelo



Acum Pico 30,483 . Acum 1 Sem y Pico 25,032 . Casos por Mill 58.3



Acum Pico 92,610 . Acum 1 Sem y Pico 16,236 . Casos por Mill 1,807



## Modelo Gompertz

- Forma paramétrica del modelo

$$N(t) = \alpha \exp(-\beta e^{-\kappa t})$$

donde  $N(t)$  es el número acumulado de casos confirmados.

- La asíntota  $\alpha$  corresponde al número total de casos al final de la epidemia.
- La derivada del modelo es una proxy para el número diario de casos nuevos confirmados

$$C(t) = N'(t) = \beta \kappa N(t) e^{-\kappa t}$$

- El tiempo en el cual se tendría la máxima incidencia diaria, lo obtenemos derivando la función de número de casos,  $C(t)$

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

- La máxima incidencia de casos, en  $t_{\max}$ , se estimaría mediante

$$C(t_{\max}) = \kappa \beta N(t_{\max}) e^{-\kappa t_{\max}}$$



## Modelo Gompertz

- La cantidad acumulada de casos al momento de máxima incidencia es  $N(t_{\max})$ .
- El fin de la epidemia la podemos aproximar usando el tiempo en el cual se alcanza una cierta fracción del total  $\alpha$ .
- La estimación del modelo puede hacerse mediante mínimos cuadrados no lineales (equivalente a máxima verosimilitud bajo normalidad). Intervalos de confianza para predicción de niveles de incidencia acumulada al tiempo  $t^*$ , pueden calcularse como:

$$y(t^*; \hat{\theta}) \pm z_{1-\nu/2} \left( \hat{\sigma}^2 + \hat{S}^2 \right)^{\frac{1}{2}}$$

$$\hat{S}^2 = \hat{\sigma}^2 \left( \frac{\partial}{\partial \theta} y(t^*; \hat{\theta}) \right)^T \left[ \left( \frac{\partial}{\partial \theta} y(\mathbf{t}; \hat{\theta}) \right)^T \left( \frac{\partial}{\partial \theta} y(\mathbf{t}; \hat{\theta}) \right) \right]^{-1} \left( \frac{\partial}{\partial \theta} y(t^*; \hat{\theta}) \right)$$

(Huet, S. *et al.* (2004). Statistical Tools for Nonlinear Regression. Springer).

## México: Modelo específico

- Estimación con datos hasta el 6 de Mayo:

$$N(t) = \alpha \exp(-\beta e^{-\kappa t})$$

$$\alpha = 436,191$$

$$\beta = 11.94$$

$$\kappa = 0.0212775$$

- El número máximo para el total de casos

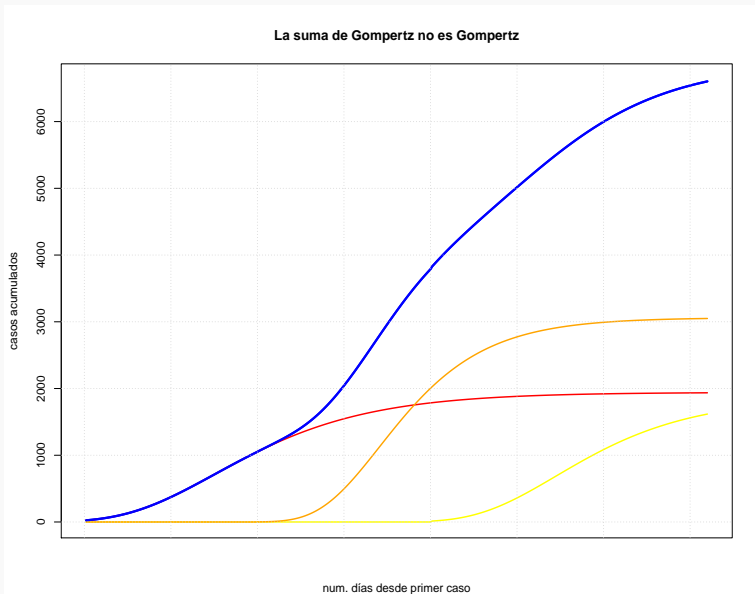
$$N_{\max} = \alpha = 436,191$$

- El horizonte de tiempo en el cual se alcanzaría este máximo sería alrededor del 10 de Noviembre, 2020.
- El pico de casos diarios se alcanzaría en

$$t_{\max} = \frac{1}{\kappa} \log(\beta) = 117 \equiv 24 \text{ de Junio, 2020}$$

- El número máximo de casos en el día  $t_{\max}$  se estima en  $N(t_{\max}) = 3,414$ .

# Regiones con diferente punto de arranque



## Modelos mixtos no lineales

- Para datos de incidencia a nivel estatal o municipal optamos por modelación usando efectos aleatorios, suponiendo que los parámetros específicos de cada ente pueden conceptualizarse como realizaciones de variables aleatorias bajo cierta distribución.
- Esto es razonable para, por ejemplo, los estados o municipios de México, para los cuales se tienen los mismos criterios de reporte de casos.
- La estructura del modelo es

$$y_{ij} = g(t_{ij}; \theta_i) + e_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_{ij}$$

donde  $g$  es el modelo Gompertz y

$$e_{ij} \sim N(0, \sigma^2)$$

$$\theta_i \sim N(\theta, \Omega)$$

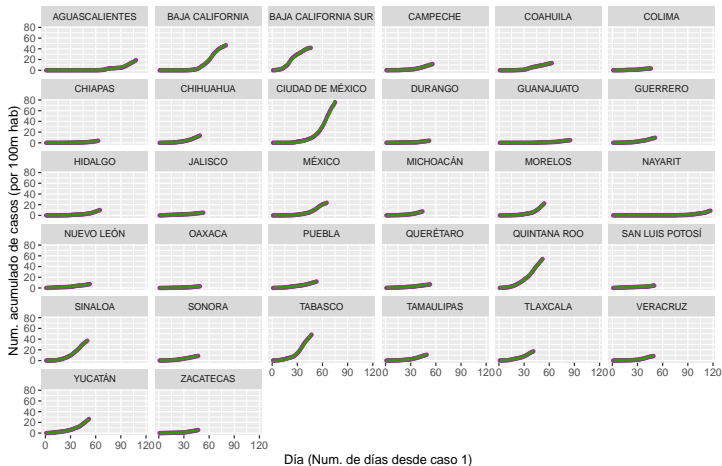
son variables aleatorias independientes.

## Modelos mixtos no lineales

- El supuesto de normalidad sobre las  $\theta_i$ 's puede flexibilizarse, sobre todo por la naturaleza no negativa de los parámetros  $\theta_i = (\alpha_1, \beta_i, \kappa_i)$ .
- La inferencia sobre las unidades específicas se hace considerando las distribuciones condicionales  $p(\theta_i|y_i, \hat{\theta}, \hat{\Omega})$ , lo cual es equivalente a un método empírico Bayesiano.
- En diferentes etapas del proceso hemos usado las bases de datos facilitadas por el Centro Geo. Las estimaciones han sido efectuadas usando el software `saemix` (Stochastic approximation EM) de Marc Lavielle.
- Se producen los reportes con tablas y resúmenes gráficos bajo los dos modelos (estatal y áreas metropolitanas). Las gráficas siguientes son sólo ejemplos de la visualización generada.

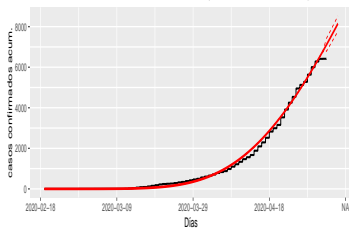


## Resultados para los Estados de México

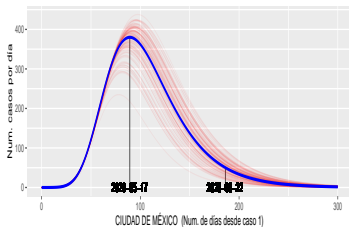


# CDMX y Edo. de México

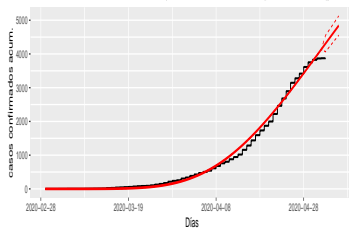
CIUDAD DE MÉXICO : 2020-05-02 Casos = 6,417 ( Total Pred. 33,620 IC 99% : [ 33,344 - 33,896 ]



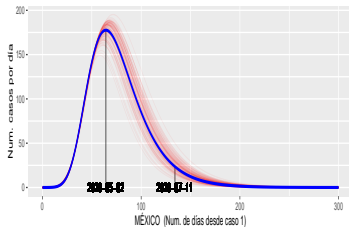
Acum Pico 12,368 . Acum 1 Sem y Pico 2,638 . Casos por Mil 758.3



MÉXICO : 2020-05-02 Casos = 3,875 ( Total Pred. 11,368 IC 99% : [ 11,154 - 11,582 ]

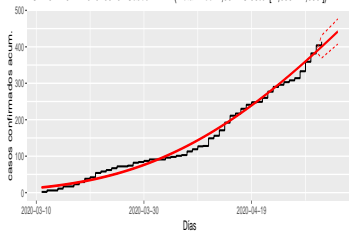


Acum Pico 4,182 . Acum 1 Sem y Pico 1,224 . Casos por Mil 236.3

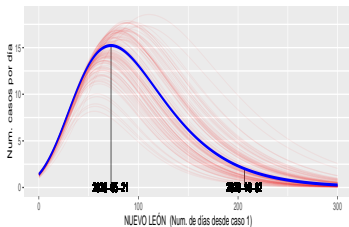


# Nuevo León, Yucatán

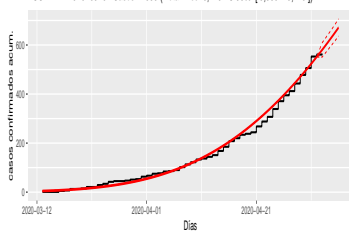
NUEVO LEÓN : 2020-05-01 Casos = 414 ( Total Pred. 1,867 IC 99% : [ 1,839 - 1,895 ])



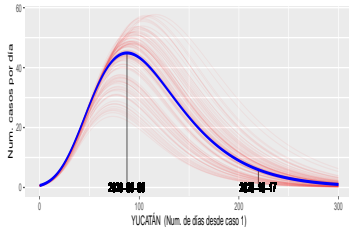
Acum Pico 687 . Acum 1 Sem y Pico 106 . Casos por Mil 73.5



YUCATÁN : 2020-05-02 Casos = 565 ( Total Pred. 5,415 IC 99% : [ 5,390 - 5,440 ])



Acum Pico 1,992 . Acum 1 Sem y Pico 313 . Casos por Mil 263.3

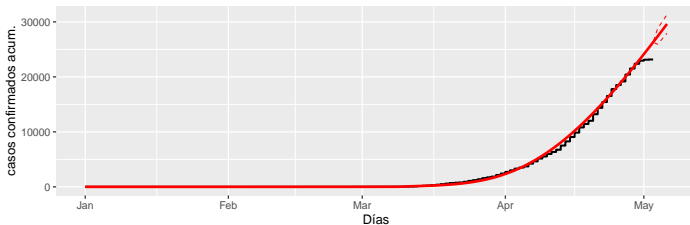




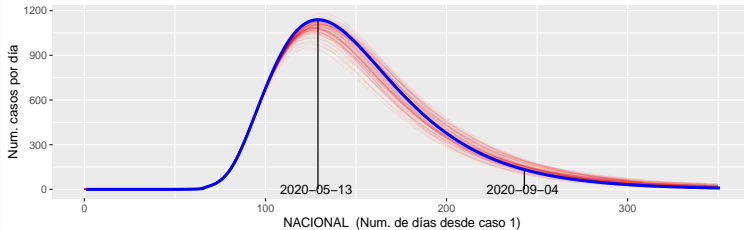
# Nacional



NACIONAL : 2020-05-03 Casos = 23,161 ( Total Pred. 108,153 )



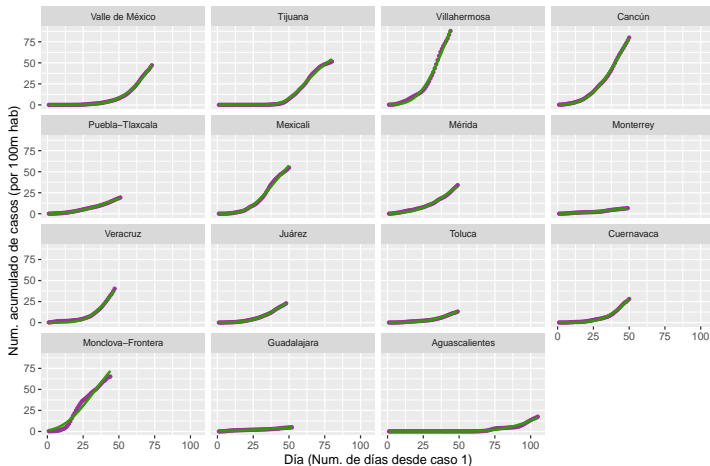
Acum Pico 37,449 . Acum 1 Sem y Pico 7,924 . Casos por Mill 181.2394



# Áreas metropolitanas

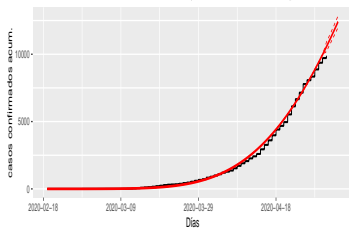


## Resultados para los Zonas Metropolitanas de México

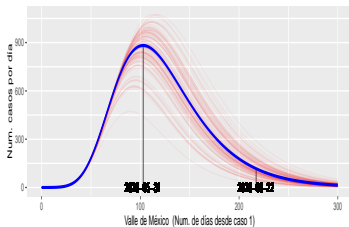


# CDMX, Monterrey

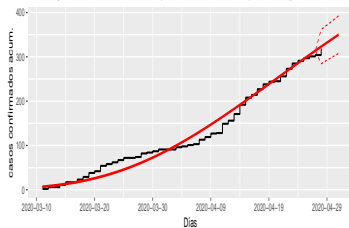
Valle de México : 2020-04-30 Casos = 9,888 ( Total Pred. 92,377 IC 99%: [ 92,082 - 92,673 ] )



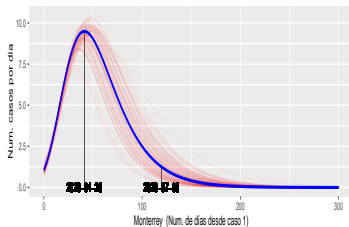
Acum Pico 33,984 . Acum 1 Sem y Pico 6,140 . Casos por Mill 470.5



Monterrey : 2020-04-27 Casos = 319 ( Total Pred. 682 IC 99%: [ 647 - 716 ] )



Acum Pico 251 . Acum 1 Sem y Pico 66 . Casos por Mill 66.7



## Estado de avance del coronavirus en México

- Para el modelo Gompertz

$$N(t) = \alpha \exp(-\beta e^{-\kappa t})$$

- se tiene que

$$\frac{N(t)}{\alpha} = \left[ \frac{N(t+1)}{N(t)} \right]^{1/(e^{-\kappa} - 1)}$$

- La cantidad del lado izquierdo es la proporción de casos confirmados al tiempo  $t$ , con respecto al total  $\alpha$ . La variabilidad con que se estima  $\alpha$ , puede ser grande, consecuentemente, una estimación del avance de la epidemia puede ser inestable, si la estimamos directamente con la expresión del lado izquierdo.
- Alternativamente, podemos usar el lado derecho, con la información de los casos acumulados durante los últimos días.

## Estado de avance (al 3 de mayo 2020)

Estado	Avance		Estado	Avance
Aguascalientes	9		Morelos	7
Baja California	78		Nayarit	4
Baja California Sur	83		Nuevo León	16
Campeche	10		Oaxaca	10
Coahuila	66		Puebla	19
Colima	19		Querétaro	22
Chiapas	8		Quintana Roo	30
Chihuahua	17		San Luis Potosí	8
Distrito Federal	20		Sinaloa	39
Durango	4		Sonora	24
Guanajuato	14		Tabasco	37
Guerrero	14		Tamaulipas	15
Hidalgo	7		Tlaxcala	9
Jalisco	20		Veracruz	19
México	36		Yucatán	8
Michoacán	9		Zacatecas	7

## Comentarios

- El modelo Gompertz sigue la dinámica de su ecuación diferencial correspondiente. En nuestro caso, la modelación básica parte de su forma funcional. Esto es, lo estamos tomando como un modelo empírico con propiedades afines al comportamiento de una epidemia.
- El modelo de Gompertz usado por el equipo de la UPC en España depende de dos parámetros. En nuestro caso, tenemos un modelo más flexible, con tres parámetros.
- El equipo del IHME de la Universidad de Washington consideraron un modelo basado en la función acumulada de la distribución Normal. Este modelo es más rígido que el Gompertz, pues asume, al igual que la curva logística, comportamiento simétrico antes y después del máximo, lo cual no es razonable.

## Comentarios finales

- Con los modelos Gompertz, no pretendemos explicar factores de impacto ni evaluar escenarios alternativos. La utilidad de estos resultados estriba en que son herramientas de apoyo para la toma de decisiones y pueden ser usados en conjunción con fuentes externas, por ejemplo, niveles de uso de camas de hospitales, niveles de uso de respiradores, etc.
- Es importante señalar la importancia de tener bases de datos actualizadas en los sistemas de salud de un país. La relevancia de un modelo depende en gran parte de la calidad de la información de la cual se alimenta.

# Referencias



## Referencias

- **IHME COVID-19 forecasting team** (2020). Forecasting COVID-19 impact on hospital bed-days, ICU-days, ventilator days and deaths by US state in the next 4 months. Report
- **UPC, BioComSC, CMCiB, IGTP** (2020). Analysis and prediction of COVID-19 for different regions and countries. Daily report 27-03-2020