

# Modelo predictivo del crecimiento poblacional

## Equipo #2

*Estimación de la capacidad de carga*

September 15, 2024

### **Integrantes:**

Guillermo Cepero García

Luis Ernesto Serras Rimada

Miguel Vadim Vilariño Pedraza

# El crecimiento poblacional



El análisis del crecimiento poblacional reviste gran importancia debido a su relevancia en diversos campos como la economía, demografía, epidemiología, la ecología, entre muchos otros. Resulta útil en múltiples ámbitos, incluyendo estudios demográficos, planificación urbana y análisis de recursos naturales.

Conocer un resultado aproximado que se acerque a una solución real del crecimiento poblacional permitiría beneficiar significativamente la dirección y gestión de recursos en nuestro país.

# Crecimiento poblacional en Cuba

Según datos de la Oficina Nacional de Estadísticas e Información de Cuba (ONEI), la población del país aumentó de aproximadamente 6.5 millones en 1953 hasta alcanzar un máximo de 11.2 millones en 2016. Desde entonces, ha experimentado una ligera disminución, llegando a los 11.0 millones en 2022.

Las causas principales de este patrón de crecimiento han incluido factores como:

- La política socialista de Cuba (que priorizó la educación y la salud pública, lo que llevó a un aumento en la esperanza de vida y la reducción de la mortalidad infantil)
- Las migraciones internacionales (especialmente desde la década de 1990)
- Los cambios en las políticas familiares, como la legalización del aborto y la anticoncepción, que afectaron la fertilidad de la población.
- Y el impacto de la crisis económica de 1991, que llevó a un cambio en las actitudes hacia la planificación familiar; entre otros.

# Objetivos

Por ende, los objetivos principales de este análisis son:

- Estimar el número máximo de personas que pueden ser sostenidamente alojadas en un área geográfica determinada y evaluar el equilibrio entre la población existente y la capacidad de carga ambiental
- Contribuir al diseño de estrategias de planificación urbana y rural que equilibren el crecimiento económico con la protección del medio ambiente y los servicios básicos.
- Permitir evaluar el impacto potencial del cambio climático y otros factores externos en la capacidad de carga poblacional a largo plazo.
- Contribuir al diseño de programas de educación ambiental y concientización sobre las implicaciones del crecimiento poblacional.
- Ayudar a establecer límites razonables para el crecimiento demográfico, evitando excederse en la explotación de recursos naturales y servicios públicos.

# Modelo logístico

Se plantea trabajar el asunto con un modelo matemático que nos pueda conducir a dicha predicción, y se considera el **modelo del crecimiento logístico**, que es solución de la ecuación diferencial que describe cómo la tasa de crecimiento de la población ( $dP/dt$ ) cambia con el tamaño de la población ( $P(t)$ )

Ecuación Diferencial:

$$\frac{dP}{dt} = r \cdot P(t) \left(1 - \frac{P(t)}{K}\right)$$

# Parámetros a estimar

## Donde se tiene que:

- $(P(t))$  es la población en función del tiempo  $(t)$ .
- $(r)$  es la tasa de crecimiento intrínseca de la población.
- $(K)$  es la capacidad de carga o tamaño máximo sostenible de la población.

## Parámetros a Estimar:

- $(r)$ : Tasa de crecimiento intrínseca de la población.
- $(K)$ : Capacidad de carga de la población.

# Modelo del crecimiento logístico

La ecuación tiene la forma:

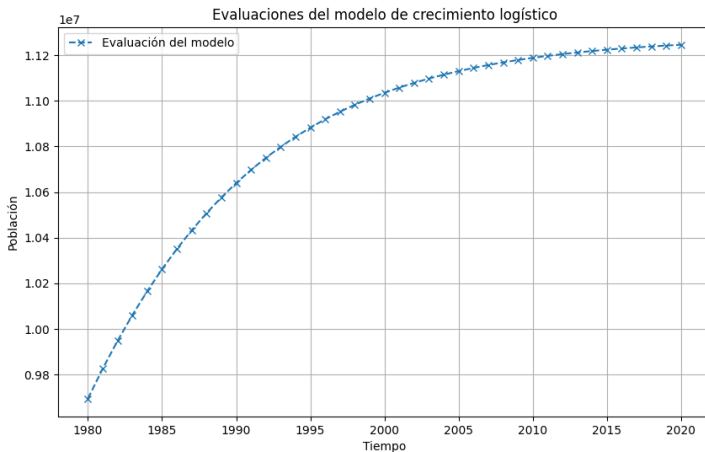
$$P(t) = \frac{K}{1 + Ae^{-rt}}$$

Y para encontrar (A) es:

$$P(0) = \frac{K}{1 + Ae^0}$$

Nota: **(A)** es una constante que depende de las condiciones iniciales de la población.

Esta función muestra claramente el punto de inflexión, donde la tasa de crecimiento cambia de positiva a negativa, indicando que la población ha alcanzado su capacidad de carga y está comenzando a estabilizarse.





Se utilizaron los datos históricos de densidad poblacional desde 1980 hasta 2020 publicados en las series estadísticas del sitio web de la ONEI. Y se desean ajustar los parámetros de la capacidad de carga ( $K$ ) y la tasa ( $r$ ). Para ello se decide utilizar la aproximación por mínimos cuadrados por medio de la función `curvefit` del módulo `scipy.optimize` en Python, cuya función matemáticamente se puede representar como:

Minimizar:  $\sum_{i=1}^N (y_i - f(x_i, \theta))^2$

Y toma los parámetros:

- $f$ : es la función modelo para la optimización.
- $xdata$ : son los valores independientes (el tiempo ( $t$  como  $x_i$ )).
- $ydata$ : son los valores dependientes (los valores de densidad poblacional en función del tiempo ( $P$  como  $y_i$ )).
- $p0$ : Estimación inicial.

La función `curvefit` intentará minimizar la suma de los cuadrados de los residuales para encontrar los valores óptimos de  $K$  y  $r$  que minimizan esta expresión.

### **Paso 1: Definición de la Función de Pérdida:**

El objetivo es minimizar la suma de los cuadrados residuales, que es la suma de las diferencias cuadradas entre los valores de la población observada ( $Pdata$ ) y los valores de la población ajustados ( $f(tdata, K, r)$ ).

#### Suma de cuadrados residuales:

$$\sum_{i=0}^N [(Pdata(t_i) - f(tdata_i, K, r))^2]$$

### **Paso 2: Inicialización de los Parámetros:**

Se proporcionan valores iniciales para los parámetros  $K$  y  $r$  (initial-guess).

### Paso 3: Iteración hacia los Parámetros Óptimos:

La función curve-fit utiliza el algoritmo de **Levenberg-Marquardt** para encontrar los valores de  $K$  y  $r$  que minimizan la suma de cuadrados residuales. Este algoritmo iterativo implica los siguientes pasos:

- 1 Cálculo de la matriz  $\omega$ : Se calcula de la función de pérdida, la matriz  $\omega$  que contiene las derivadas parciales de la función de pérdida con respecto a los parámetros  $K$  y  $r$ .
- 2 Actualización de Parámetros: Utilizando la matriz  $\omega$ , se calculan nuevas estimaciones para  $K$  y  $r$  que minimizan localmente la función de pérdida.
- 3 Repetición: Los pasos 1 y 2 se repiten hasta que se alcanzan los criterios de convergencia (por ejemplo, un cambio insignificante en los parámetros o un número máximo de iteraciones).

# Calculo de la matriz $\omega$

$\omega$  es una matriz que contiene las derivadas parciales de la función de pérdida con respecto a los parámetros  $K$  y  $r$ .

Se calcula como:

$$\omega = \left[ \frac{\delta(r_{min})}{\delta(K)}, \frac{\delta(r_{min})}{\delta(r)} \right]$$

siendo  $r_{min}$  la suma de cuadrados residuales

Donde  $f(t_{data}, K, r)$  es la función logística:

Se tiene la función de pérdida:

$$f(t, K, r) = \frac{K}{(1 + ((\frac{K}{P_0}) - 1)e^{-rt})}$$

donde  $P_0$  es el valor inicial de la población en el período de tiempo.

# Derivadas Parciales

Para calcular las derivadas parciales, diferenciamos la función de pérdida con respecto a  $K$  y  $r$ .

Derivada parcial con respecto a  $K$ :

$$\omega = \frac{\delta(r_{min})}{\delta(K)} = \sum_{i=0}^N [2(P(t_i) - f(tdata, K, r)) \cdot \left(\frac{-f(tdata, K, r)}{K^2}\right)]$$

Y substituyendo  $f(tdata, K, r)$  en la derivada, obtenemos:

$$\omega = \frac{\delta(r_{min})}{\delta(K)} = -2 \sum_{i=0}^N \left[ \left( P(t_i) - \frac{K}{(1 + ((\frac{K}{P_0}) - 1) \cdot e^{-r \cdot t})} \right) \cdot \left( \frac{K}{(1 + ((\frac{K}{P_0}) - 1) \cdot e^{-r \cdot t})^2} \right) \right)]$$

Derivada parcial con respecto a r:

$$\omega = \frac{\delta(r_{min})}{\delta(r)} =$$
$$= \sum_{i=0}^N [2(P(t_i) - f(tdata, K, r)) \cdot (f(tdata, K, r) \cdot \frac{(\frac{-K}{P_0}) \cdot e^{-rt}}{(1 + ((\frac{K}{P_0}) - 1) \cdot e^{-rt})^2})]$$

Y sustituyendo  $f(tdata, K, r)$  en la derivada, obtenemos:

$$\omega = \frac{\delta(r_{min})}{\delta(r)} =$$
$$= -2 \sum_{i=0}^N [(P(t_i) - \frac{K}{(1 + ((\frac{K}{P_0}) - 1) \cdot e^{-rt})}) \cdot \frac{K}{(1 + ((\frac{K}{P_0}) - 1) \cdot e^{-rt})^2} \cdot (\frac{-K}{P_0}) \cdot e^{-rt}]$$

Evaluando estas derivadas en los valores iniciales de K y r, obtenemos  $\omega$

# Ecuación normal

Luego se forma una ecuación normal utilizando  $\omega$  y la matriz de Hessiana ( $H$ ), que es la matriz de segundas derivadas parciales de la función de pérdida. La ecuación normal es:

$$(\omega^T \omega + \lambda I) \Delta p = -\omega^T r$$

donde:

- $\Delta p$  es un vector de correcciones para los parámetros  $K$  y  $r$
- $\lambda$  es un parámetro de regularización (que se ajusta para controlar el paso de la iteración)
- $I$  es la matriz identidad
- $r$  es el vector residual, que contiene las diferencias entre los valores observados y ajustados

# Actualización de parámetros

Se calculan las correcciones para los parámetros resolviendo la ecuación normal.

Los parámetros actualizados son:

$$K_{\text{nuevo}} = K_{\text{anterior}} + \Delta K$$

$$r_{\text{nuevo}} = r_{\text{anterior}} + \Delta r$$

Nota:  $\Delta K$  y  $\Delta r$  representan el cambio o corrección que se aplica a los valores anteriores de  $K$  y  $r$  para obtener los nuevos valores. Este cambio se calcula a partir de la solución de la ecuación normal, que busca minimizar la función de pérdida.

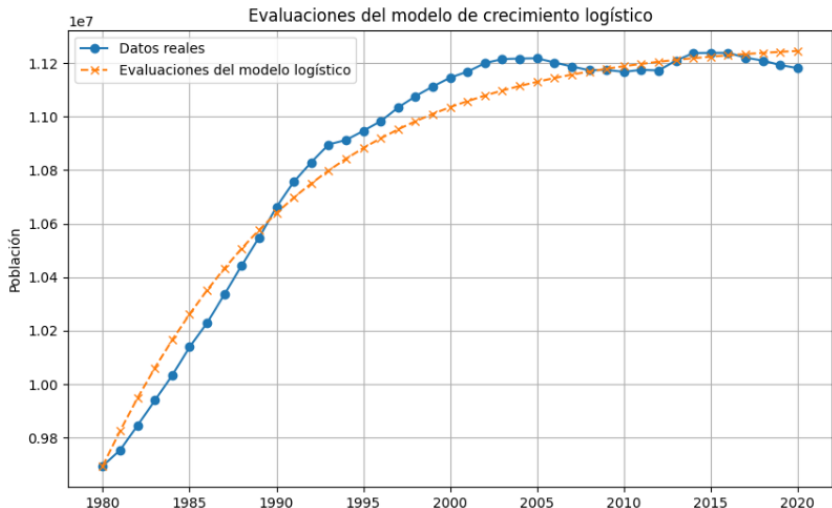


## Estimación de parámetros:

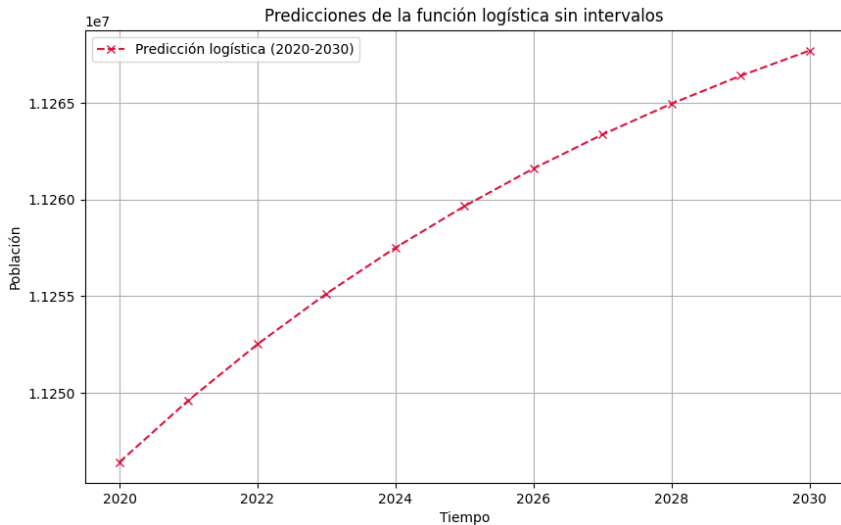
Finalmente, se repiten los pasos anteriores hasta que se alcanzan los criterios de convergencia, como un cambio insignificante en los parámetros o un número máximo de iteraciones. Y una vez que el algoritmo converge, los valores estimados de  $K$  y  $r$  se obtienen como los valores finales de  **$K_{\text{nuevo}}$**  y  **$r_{\text{nuevo}}$** .

# Resultados

Primeramente se visualiza una comparación de valores reales vs estimaciones del modelo sin intervalos para las estimaciones de  $[11280122.85, 0.10]$  devueltas por la función utilizada.



# Predicción del modelo sin intervalos



# Predicciones hasta el 2030

Años	Predicción
2020	11246416
2021	11249615
2022	11252511
2023	11255133
2024	11257506
2025	11259654
2026	11261599
2027	11263359
2028	11264952
2029	11266394
2030	11267699

Figure: Resultados del modelo sin intervalos

# DataFrame para intervalos de 5 años

	K	r
0	20442337054	0.008
1	340143106946	0.003
2	10660179	-0.352
3	5105238279	0.000
4	2575174752	0.000
5	1223441	0.000
6	999003076	0.000
7	1798999	0.000

Figure: Resultados del segundo modelo para el grupo de intervalos de 5 años

# DataFrame para intervalos de 10 años

	K	r
0	2059434657188	0.009
1	10623571	-0.135
2	726227	-0.000
3	5837330070	0.000

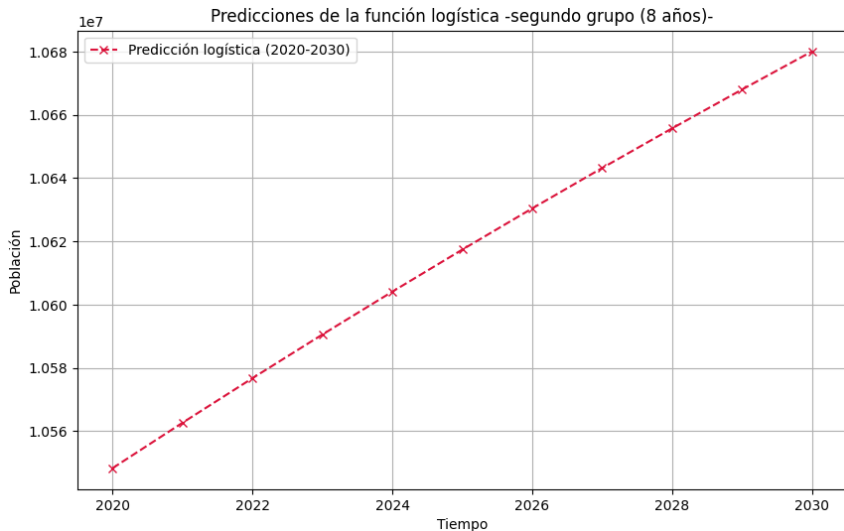
Figure: Resultados del segundo modelo para el grupo de intervalos de 10 años

# DataFrame para intervalos de 8 años

	K	r
0	2055969070415	0.009
1	10396071	-0.167
2	10981262	-0.209
3	1000853	0.000
4	11251162	0.022

Figure: Resultados del segundo modelo para el grupo de intervalos de 8 años

# Predicción del modelo en intervalos de 8 años



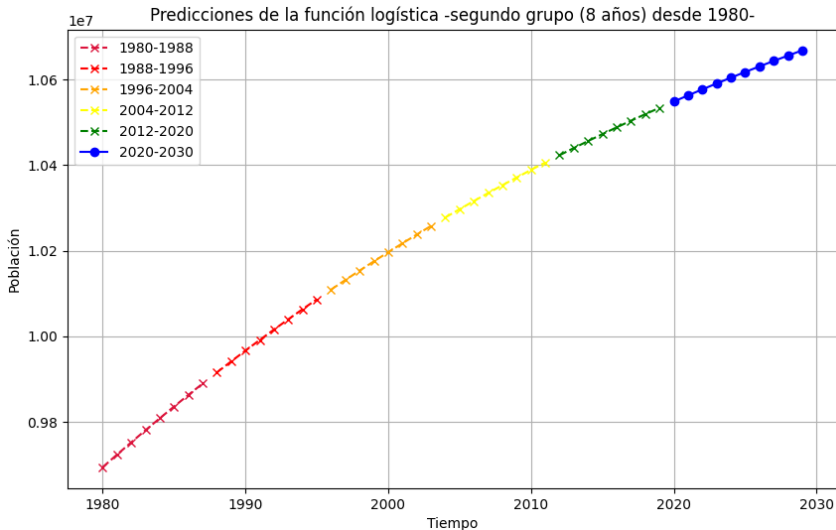


# Predicciones hasta el 2030

Años	Predicción
2020	10548308
2021	10562666
2022	10576750
2023	10590563
2024	10604111
2025	10617398
2026	10630427
2027	10643205
2028	10655733
2029	10668018
2030	10680063

Figure: Resultados del modelo en intervalos de 8 años

# Predicción del modelo en intervalos de 8 años desde 1980 hasta 2030



- Mejorar el Modelo Matemático
- Ampliar la Base de Datos
- Validar y Refinar el Modelo
- Investigar Factores Externos
- Divulgar los Resultados

Estas recomendaciones buscan fortalecer el modelo actual, ampliar su alcance y utilidad, y contribuir al conocimiento demográfico de Cuba, teniendo en cuenta las particularidades observadas en el análisis inicial.

Muchas gracias :)

