

Modelo de análisis predictivo del crecimiento poblacional: Estimación de la capacidad de carga

Guillermo Cepero Garcia

Grupo D111

THEWILLYJAKE53@GMAIL.COM

Luis Ernesto Serras Rimada

Grupo D111

LUISERNESTOSERRAS@GMAIL.COM

Miguel Vadim Vilarino Pedraza

Grupo D111

A.UNO@LAB.MATCOM.UH.CU

Tutor(es):

Dr. Tutor Uno, *Centro*

Lic. Tutor Dos, *Centro*

Resumen

El análisis del crecimiento poblacional forma parte fundamental de este proyecto, abordando la problemática desde cuestiones básicas como "por qué", "qué resultado se desea obtener" y "con qué objetivos". Este tema reviste gran importancia debido a su relevancia en diversos campos como la demografía, la epidemiología y la ecología. El estudio del crecimiento poblacional resulta útil en múltiples ámbitos, incluyendo estudios demográficos, planificación urbana y análisis de recursos naturales. Conocer un resultado aproximado que se acerque lo máximo posible a una solución real del crecimiento poblacional permitiría beneficiar significativamente la dirección y gestión de recursos en nuestro país. Los resultados obtenidos en el análisis demuestran la importancia de adaptar modelos matemáticos a patrones de crecimiento reales complejos, considerando factores como la heterogeneidad temporal en los datos y la posibilidad de cambios abruptos en las tasas de crecimiento. La comparación entre diferentes modelos, como el modelo logístico tradicional versus uno ajustado, proporciona valiosas lecciones sobre cómo mejorar la precisión de las predicciones demográficas. Se observó que el segundo modelo, con intervalos de 8 años, se ajustaba mejor, reflejando valores cercanos a la densidad poblacional promedio. Además, se identificaron problemas en las respuestas de los parámetros estimados dentro de ciertos intervalos, atribuidos a una tendencia casi vertical en los valores durante ciertos periodos. Esto llevó a proponer un ajuste diferente al utilizado con todos los datos, basado en un crecimiento más acelerado. En los años 1980 en adelante, Cuba experimentó un crecimiento poblacional significativo, seguido de un período de estabilización y eventual disminución de la tasa de crecimiento. Este análisis demuestra la necesidad de considerar factores específicos del contexto cubano, como el impacto de eventos históricos en el patrón de crecimiento poblacional.

Abstract

The analysis of population growth is a fundamental part of this project, addressing the problem from basic questions such as "why", "what result is desired" and "with what objectives". This topic is of great importance due to its relevance in various fields such as demography, epidemiology and ecology. The study of population growth is useful in multiple areas, including demographic studies, urban planning and natural resource analysis. Knowing an approximate result that is as close as possible to a real solution to population growth would significantly benefit the direction and management of resources in our country. The results obtained in the analysis demonstrate the importance of adapting mathematical models to complex real growth patterns, considering factors such as temporal heterogeneity in the data and the possibility of abrupt changes in growth rates. Comparison between different models, such as the traditional logistic model versus an adjusted one, provides valuable lessons on how to improve the accuracy of demographic predictions. It was observed that the second model, with intervals of 8 years, fit better, reflecting values close to the average population density. In addition, problems were identified in the responses of the estimated parameters within certain intervals, attributed to an almost vertical trend in the values during certain periods. This led to proposing a different adjustment than that used with all the data, based on more accelerated growth. In the 1980s onwards, Cuba experienced significant population growth, followed by a period of stabilization and eventual decline in the growth rate. This analysis demonstrates the need to consider factors specific to the Cuban context, such as the impact of historical events on the population growth pattern.

Palabras Clave: Predicción, carga, poblacion, datos, modelo, entrenamiento, capacidad.

Tema: Predicción de crecimiento poblacional, estimación del valor de la capacidad de carga.

1. Introducción

Como objetivo de análisis para este proyecto se seleccionó el asunto del crecimiento poblacional, donde se desea abordar la problemática partiendo de unas cuestiones fundamentales: ¿Porqué?, ¿qué resultado se desea obtener y con qué objetivos?. Y pues, se tiene en cuenta que constituye un problema de alta significación en muchos campos, como la demografía, la epidemiología y la ecología, que puede resultar útil en estudios demográficos, planificación urbana y análisis de recursos naturales, entre otros. En otras palabras, por ejemplo, conociendo un resultado aproximado que converja lo máximo posible a una solución real del crecimiento poblacional se podría beneficiar enormemente a la dirección y gestión de recursos de nuestro país.

2. Desarrollo

Se plantea trabajar el asunto con un modelo matemático que nos pueda conducir a dicha predicción, y se considera el **modelo logístico**, que es ecuación diferencial que describe cómo la tasa de crecimiento de la población (dP/dt) cambia con el tamaño de la población ($P(t)$). Cuando la población es pequeña, la tasa de crecimiento es alta, ya que hay muchos recursos disponibles para cada individuo. A medida que la población crece, la tasa de crecimiento disminuye porque hay menos recursos disponibles por individuo. Cuando la población alcanza su capacidad de carga (K), la tasa de crecimiento se vuelve cero, indicando que la población ha alcanzado un equilibrio sostenible. *Entonces 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.

Ecuación Diferencial de Modelo Logístico:

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

Para la tasa de crecimiento intrínseca tenemos que:

$$r = \left(\frac{N(t) - F(t)}{P(t)}\right),$$

donde $N(t)$ es la natalidad en función del tiempo y $F(t)$ es la mortalidad en función del tiempo.

Y para obtener la capacidad de carga de la población cubana siguiendo el contexto del modelo logístico y utilizando datos reales, se debe seguir un proceso que involucre la recopilación de datos sobre la población cubana y los factores que podrían influir en su capacidad de carga. Estos pueden incluir la disponibilidad de recursos naturales, la infraestructura, la salud pública, y la economía.

Y posteriormente el análisis de esos datos y la interpretación de los resultados (Análisis Exploratorio).

Por ejemplo, supongamos que después de nuestro análisis, encontramos que la tasa de crecimiento intrínseca (r) de la población cubana es de 0.02 (un 2% anual) y que la capacidad de carga (K) estimada es de 15 millones de habitantes. Esto significaría que, teóricamente, nuestro país podría soportar hasta 15 millones de personas sin agotar sus recursos vitales, siempre y cuando se mantengan las condiciones actuales y se gestionen adecuadamente los recursos.

2.1 Modelo matemático

Como idea de solución se plantea la función logística, que describe cómo la población crece hacia su capacidad de carga y luego se estabiliza.

La función logística tiene la forma: $P(t) = \frac{K}{1 + Ae^{-rt}}$

Donde (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.

Para encontrar (A), se requiere conocer el valor inicial de la población, ($P(0)$), y usarlo junto con (K) y (r) para resolver para (A). La condición inicial ($P(0)$) te da el valor de la población en el momento inicial, antes de que comience el crecimiento logístico.

La ecuación para encontrar (A) es: $P(0) = \frac{K}{1 + Ae^0}$

Resolviendo para (A), obtenemos: $A = \frac{K}{P(0)} - 1$

Una vez que se tenga el valor de (A), se puede sustituir en la solución general de la ecuación diferencial logística para obtener la función que describe el crecimiento de la población hacia su capacidad de carga. Por ejemplo, si se sabe que la población inicial es de 1000 individuos ($P(0) = 1000$), la capacidad de carga es de 10,000 individuos ($(K = 10,000)$), y la tasa de crecimiento intrínseca es de 0.01 ($(r = 0.01)$), se puede calcular (A) de la siguiente manera:

$$A = \frac{10,000}{1000} - 1 = 9$$

Luego, se puede sustituir ($A = 9$) en la solución general para obtener la función específica que describe cómo la población crecerá hacia su capacidad de carga máxima.

2.2 Modelo numérico

Se toma como punto de partida el hecho de que se necesita realizar un proceso de optimización no lineal para encontrar los parámetros que mejor ajusten una función dada a un conjunto de datos determinados.

Para ello se decide utilizar la aproximación por mínimos cuadrados de forma iterativa por medio de la función `curvefit` del módulo `scipy.optimize`, cuya función matemáticamente se puede representar como:

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

- y_i son los valores medidos

- x_i son los valores independientes
- $f(x_i, \theta)$ es la función modelo con parámetros
- \sum denota la suma sobre todos los puntos de datos

Definición inicial: Se proporciona una función modelo $f(x, \theta)$ y los datos experimentales ($xdata$, $ydata$).

Inicialización: Se establecen valores iniciales para los parámetros θ .

Cálculo de residuales: Se calcula la diferencia entre los datos mediantes y la función modelo:
 $r = ydata - f(xdata, \theta)$

Derivación: Se calcula la derivada de los residuales respecto a cada parámetro:

$$\frac{\partial r}{\partial \theta_j} = -\frac{\partial f}{\partial \theta_j}$$

Iteración: Se actualiza cada parámetro θ_j según la fórmula de Levenberg-Marquardt:

$$\theta_j^{(n+1)} = \theta_j^n - [J^T J]^{-1} J^T r$$

Donde J es la matriz Jacobiana:

$$\frac{\partial r}{\partial \theta_j}$$

Convergencia: El proceso se repite hasta que se alcance un criterio de convergencia.

Adaptación al modelo logístico:

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.

2.3 Implementación en Python

Se propone el desarrollo por medio de dos modelos diferentes, uno sin intervalos y otro con iteraciones por intervalos

Primero se define la función logística sin intervalos.

```
1 def gen_log(p0):
2     def logistic_function(t, K, r):
3         return K / (1 + (K/p0 - 1) * np.exp(-r * t))
4     return logistic_function
5
6 f = gen_log(P_data[0])
```

Listing 1: Código función logística sin intervalos

La siguiente celda de código ajustará la función logística a los datos observados, buscando los valores óptimos de (K) y (r) que minimicen la diferencia entre los datos observados y los predichos por el modelo.

Observaciones:

- La función ‘`curvefit`’ del modulo ‘`scipy.optimize`’ devuelve dos arrays: ‘`popt`’, que contiene los *parámetros estimados*, y ‘`pcov`’, que contiene la *covarianza de los parámetros estimados*.

- La función ‘`curvefit`’ es parte de la biblioteca ‘`SciPy`’ y se utiliza para ajustar una función a un conjunto de datos mediante el método de *mínimos cuadrados*. Su objetivo principal es encontrar los parámetros de la función que minimizan la diferencia entre los valores observados y los predichos por la función

```
1 # Par metros iniciales:
2 initial_guess = [10000000, 0.01] # Asume
   un valor inicial para K y r
3
4 # Ajustar el modelo a los datos observados
5 popt, pcov = curve_fit(f, t_data, P_data,
   p0=initial_guess, sigma=None,
   absolute_sigma=False)
6
7 # Imprimir los par metros estimados
8 print("Par metros estimados:", popt)
9 #Se formatean los datos:
10 params_formateados = [format(param, ".2f")
   for param in popt]
11 print("Parametros formateados: \n",
   params_formateados)
12
13 #Asignacion:
14 K = float(params_formateados[0])
15 K = int(K) #Parametro K
16 r_intr = float(params_formateados[1]) #
   Tasa intrinseca de crecimiento r
17 df_sin_intervalos = DataFrame({"K":K,
   "r":r_intr
18 }, index=[0])
```

Listing 2: Estimación de parámetros

De este modelo inicial lo que se obtuvo como resultado al evaluar y comparar con valores reales fueron valores muy cercanos a la linea real de los datos que se tiene, pero logicamente no conviene para intetar realizar una predicción exacta.

Por tanto, acto seguido: se procede a desarrollar un segundo modelo para grupos de intervalos de 5, 8 y 10 años

```
1 initial_guess = [10000000, 0.01]
2 def model(t, P, times):
3     results = {}
4     def estimate(times):
5         x = 0
6         y = times
7         estimate_k = []
8         estimate_r=[]
9         estimate_lenght = int((n-3)/times)
10        for _ in range(estimate_lenght):
11            par_t = t[x:y]
12            par_P = P[x:y]
13            f = gen_log(par_P[0])
14            # Ajustar el modelo a los
   datos observados
15            popt, pcov = curve_fit(f,
   par_t, par_P, p0=initial_guess, sigma=
   None, absolute_sigma=False)
16            params_formateados = [format(
   param, ".2f") for param in popt]
17            popt_formateado = list(map(
   lambda x: float(x), popt))
```

```

18         estimate_k.append(
19             popt.formateado[0])
20         estimate_r.append(
21             popt.formateado[1])
22         x += times
23         y += times
24         return [int(x) for x in estimate_k
25                 ], [np.round(x,3) for x in estimate_r],
26                 pcov
27     for i in times:
28         K, r, cov = estimate(i)
29         results[str(i) + " a os"] = {"K":
30                                     K, "r" : r, "Covarianza" : cov}
31     return results
32 #Llamada a la funci n del modelo por
33 intervalos
34 results = model(t_data[:-3], P_data[:-3],
35                 [5,8,10])

```

Listing 3: Estimación de parámetros

De los resultados se pueden identificar varios problemas en las respuestas de los parámetros estimados dentro de ciertos intervalos para cada rango. Esta situación se debe a que los valores en esos periodos presentan una tendencia casi vertical, caracterizada por una inclinación pronunciada hacia arriba. Como resultado, el valor de K podría verse afectado, ya que teóricamente, cuando el tiempo tiende a infinito, la función poblacional tiende a K. En consecuencia, en la primera aproximación global, K proporciona un valor cercano a la población total de Cuba (11 millones).

Es importante tener en cuenta que los datos reales no siguen exactamente la curva logística ideal. La función real presenta segmentos más concavos y otros más convexos que la función teórica. Por esta razón, se realizará un ajuste diferente al utilizado con todos los datos, basándose en un crecimiento más acelerado. Este ajuste se justifica porque los puntos en cuestión muestran una tendencia más vertical, lo que implica aumentos significativos en la población. Dado que estos aumentos son tan pronunciados, es razonable esperar que K también experimente una subida considerable.

Con respecto a los resultados de la evaluación de predicción se tuvo que el segundo modelo (intervalos de 8 años) fue el que mejor se ajustó y al representarlo se reflejaron valores bastante cercanos a la densidad poblacional promedio lo que indica un correcto ajuste.

Por otro lado cabe resaltar que en los años 1980 en adelante la isla experimentó un crecimiento poblacional significativo, pero luego se comenzó a estabilizar y eventualmente disminuir.

3. Conclusiones

Análisis de los Resultados El análisis de los resultados reveló varios problemas en las respuestas de los parámetros estimados dentro de ciertos intervalos para cada rango. Estas dificultades se deben principalmente

a la presencia de una tendencia casi vertical en los valores durante ciertos periodos, caracterizada por una inclinación pronunciada hacia arriba.

Impacto en K:

La tendencia vertical observada en los datos puede afectar significativamente el valor de K. Teóricamente, cuando el tiempo tiende a infinito, la función poblacional tiende a K. Sin embargo, en la primera aproximación global, K proporcionó un valor cercano a la población total de Cuba (11 millones), lo cual sugiere que el modelo logístico tradicional no captura adecuadamente el comportamiento real de la población.

Ajuste del Modelo:

Dado que los datos reales no siguen exactamente la curva logística ideal, se propone un ajuste diferente al utilizado con todos los datos. Este nuevo ajuste se basaría en un crecimiento más acelerado, especialmente en los periodos donde se observa una tendencia más vertical. Esta modificación se justifica por los aumentos significativos en la población que se han registrado en esos momentos.

Evaluación de Predicción:

Los resultados de la evaluación de predicción mostraron que el segundo modelo (intervalos de 8 años) fue el que mejor se ajustó. Al representarlo, se reflejaron valores bastante cercanos a la densidad poblacional promedio, lo que indica un correcto ajuste del modelo.

Crecimiento Poblacional en Cuba:

En los años 1980 en adelante, Cuba experimentó un crecimiento poblacional significativo. Sin embargo, posteriormente se produjo un proceso de estabilización y eventual disminución de la tasa de crecimiento.

4. Recomendaciones

Mejorar el Modelo Matemático:

- Desarrollar un modelo matemático más sofisticado que pueda capturar mejor la complejidad del crecimiento poblacional cubano. Esto podría incluir:
- Incorporar factores económicos, sociales y políticos que influyen en el crecimiento poblacional.
- Utilizar técnicas de modelado avanzadas, como modelos dinámicos o agentes, que puedan simular mejor la heterogeneidad temporal observada.
- Realizar un ajuste más preciso del modelo basado en el crecimiento acelerado observado en ciertos periodos. Esto podría implicar:
- Analizar en detalle los factores que causan los aumentos significativos en la población.
- Implementar un sistema de pesos diferentes para distintos periodos, dando mayor importancia a los segmentos de crecimiento más rápido.
- Ampliar la Base de Datos:

- Recopilar y analizar nuevos datos históricos que permitan una mejor comprensión del patrón de crecimiento poblacional en Cuba. Esto podría incluir:
- Información sobre migraciones internacionales y su

impacto en la población.

-Datos detallados sobre la distribución geográfica de la población.

-Indicadores socio-económicos relacionados con el crecimiento poblacional.

-Integrar fuentes de información adicionales, como encuestas de salud pública o estudios demográficos nacionales, para obtener una visión más completa del fenómeno.

Validar y Refinar el Modelo:

-Realizar pruebas de validación cruzada para evaluar la robustez del modelo frente a diferentes escenarios y condiciones.

-Comparar el rendimiento del modelo ajustado con el tradicional logístico, utilizando indicadores cuantitativos como la raíz cuadrática media (RMSE) o la correlación de Pearson.

-Realizar simulaciones prospectivas utilizando el modelo refinado para predecir el futuro crecimiento poblacional y compararlos con proyecciones oficiales.

Investigar Factores Externos

-Analizar el impacto de eventos macroeconómicos y geopolíticos en el crecimiento poblacional, como cambios en las relaciones internacionales de Cuba o reformas económicas.

-Estudiar cómo factores ambientales y climáticos podrían influir en el patrón de crecimiento poblacional, especialmente en un país insular como Cuba.

Divulgar los Resultados

-Presentar los hallazgos y recomendaciones en foros académicos y profesionales, buscando colaboraciones interdisciplinarias con economistas, sociólogos y científicos políticos.

-Proporcionar herramientas de visualización intuitiva de los resultados para facilitar la comprensión del público general y las autoridades locales.

-Sugirir aplicaciones prácticas de los resultados, como orientación para planificación urbana, desarrollo económico sostenible o formulación de políticas públicas.

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.

Información(ONEI) en la sección de estadísticas demográficas. URL: <https://www.onei.gob.cu/>
Consultado en September 7, 2024.

[4] Documentación de SciPy. URL: https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.curve_fit.html.

[5] StackOverflow. URL: <https://stackoverflow.com/questions/49790974/what-is-curve-fits-algorithm-when-fitting-data-to-a>

[6] IRIC's Bioinformatics Plataform. URL: <https://bioinfo.irc.ca/scipy-and-logistic-regressions/>.

[7] Herramientas para proyectar la población. URL: https://ccp.ucr.ac.cr/cursos/icamacho/public_html/planificacion/contenido/tema6.htm.

[8] Documentación Matplotlib. URL: <https://matplotlib.org/stable/index.html>.

References

[1] DataFrame utilizado. URL: <https://github.com/LFrench03/Modelo-de-Crecimiento-Poblacional/blob/main/data/csv/poblacion-residente.csv>.

[2] Blog del Instituto de Matemáticas de la Universidad de Sevilla. URL: <https://institucional.us.es/blogimus/2018/01/un-ejemplo-sencillo-de-modelizacion-matematica-el-crecimiento-de-poblaciones/>

[3] Sitio web de la Oficina Nacional de Estadísticas e