Análisis del Método de Mínimos Cuadrados en el Crecimiento Poblacional

Ramón Cherta González
John Mauris López Ramos

1 Introducción

En el marco de nuestro proyecto, llevamos a cabo tres investigaciones aplicando el modelo logístico para analizar datos de población. En estas investigaciones, nos propusimos determinar los valores de los parámetros a y b, que son fundamentales para describir el comportamiento del crecimiento poblacional. Para ello, encontramos la recta que mejor se ajustaba a los puntos $(P_i, P_i P_i')$, utilizando inicialmente herramientas como la calculadora gráfica de GeoGebra. Sin embargo, en este trabajo profundizaremos en un enfoque alternativo: el método numérico de mínimos cuadrados.No obstante, surge la pregunta: **¿por qué elegir el método de mínimos cuadrados para este tipo de análisis?**

La respuesta radica en que este método es especialmente adecuado para modelos lineales debido a su simplicidad, eficiencia y robustez frente a pequeñas desviaciones en los datos. Además, al penalizar los errores grandes mediante el uso de diferencias cuadráticas, minimiza significativamente su impacto en el ajuste final. A continuación, explicaremos detalladamente cómo aplicamos este método para obtener los valores óptimos de a y b, destacando su relevancia en el análisis de datos poblacionales.

2 Método de Mínimos Cuadrados

El método de mínimos cuadrados busca encontrar los parámetros de un modelo que minimicen la suma de los errores cuadráticos entre los valores observados (y_i) y los valores predichos $(\hat{y}_i)$ por el modelo. En términos simples, trata de "ajustar" una curva (o recta) a los datos de tal manera que las diferencias entre los datos reales y los predichos sean lo más pequeñas posible.

Matemáticamente, si tenemos un conjunto de n puntos (x_i, y_i), el error cuadrático total se define como:

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

donde:

- y_i: valor observado,
- $\hat{y}_i$: valor predicho por el modelo.

El objetivo es minimizar S, lo que implica encontrar los parámetros del modelo que mejor se ajustan a los datos.

Uno de los casos más comunes del método de mínimos cuadrados es el ajuste de una recta a los datos. La ecuación de una recta es:

$$y = a + bx \tag{2}$$

donde:

- a: intercepto (valor de y cuando $x = 0$),
- b: pendiente (tasa de cambio de y con respecto a x).

Para encontrar a y b, minimizamos la suma de los errores cuadráticos:

$$S = \sum_{i=1}^{n} (y_i - (a + bx_i))^2 \tag{3}$$

Derivando S con respecto a a y b, e igualando a cero, obtenemos el siguiente sistema de ecuaciones normales:

$$na + \sum x_i b = \sum y_i \tag{4}$$

$$\sum x_i a + \sum x_i^2 b = \sum x_i y_i \tag{5}$$

3 Aplicación del Método

P_i (millones)	Tasa $y_i = \frac{P_i'}{P_i}$
5.308	0.0313
7.240	0.0300
9.638	0.0291
12.861	0.0289
17.064	0.0303
23.192	0.0310
31.443	0.0244
38.558	0.0243
50.189	0.0243
62.980	0.0207
76.212	0.0192

Table 1: Datos de población y tasa relativa de crecimiento

Aquí:

- $x_i = P_i$: población en millones,
- $y_i = \frac{P_i'}{P_i}$: tasa relativa de crecimiento.

Queremos ajustar una recta de la forma:

$$y = a + bP \tag{6}$$

El sistema de ecuaciones normales es:

$$na + (\sum x_i)b = \sum y_i \tag{7}$$
$$(\sum x_i)a + (\sum x_i^2)b = \sum x_i y_i \tag{8}$$

Donde:

- n: número de puntos de datos,
- $\sum x_i$: suma de las poblaciones,
- $\sum y_i$: suma de las tasas relativas de crecimiento,
- $\sum x_i^2$: suma de los cuadrados de las poblaciones,
- $\sum x_i y_i$: suma de los productos de población y tasa relativa.

Número de puntos de datos (n):

$$n = 11 \tag{9}$$

Suma de x_i:

$$\sum x_i = 5.308 + 7.240 + 9.638 + 12.861 + ... + 76.212 \tag{10}$$

$$\sum x_i = 334.685 \tag{11}$$

Suma de y_i:

$$\sum y_i = 0.0313 + 0.0300 + 0.0291 + 0.0289 + ... + 0.0192 \tag{12}$$

$$\sum y_i = 0.2935 \tag{13}$$

Suma de x_i^2::

$$\sum x_i^2 = 5.308^2 + 7.240^2 + 9.638^2 + 12.861^2 + ... + 76.212^2 \tag{14}$$

$$\sum x_i^2 = 16192.72 \tag{15}$$

Suma de $x_i y_i$:

$$\sum x_i y_i = (5.308 \times 0.0313) + (7.240 \times 0.0300) + ... + (76.212 \times 0.0192) \tag{16}$$

$$\sum x_i y_i = 7.962 \tag{17}$$

Substituyendo los valores calculados:

$$11a + 334.685b = 0.2935 \tag{18}$$

$$334.685a + 16192.72b = 7.962 \tag{19}$$

Resolvemos el sistema:

$$11a + 334.685b = 0.2935$$
$$334.685a + 16192.72b = 7.962$$

Obtenemos:

$$a \approx 0.0316, \quad b \approx -0.0001616$$

Por lo que la ecuación ajustada es:

$$y = 0.0316 - 0.0001616x \tag{20}$$

Utilizando GeoGebra, se obtuvo la siguiente ecuación:

$$y = -0.0001714045365P + 0.0318603444775. \tag{21}$$

Para evaluar qué tan cercanos son ambos modelos, comparamos los parámetros a y b:

Parámetro	**Cálculo Manual**	**GeoGebra**	**Error Absoluto**	**Error Relativo (%)**
a	0.0316	0.0318603	0.0002603	0.817%
b	−0.0001616	−0.0001714	0.0000098	5.72%

Table 2: Comparación de parámetros entre el cálculo manual y GeoGebra.

El error relativo en a es menor al 1%, lo que indica que el punto de intersección con el eje vertical es prácticamente el mismo en ambos modelos. En cuanto a b, el error relativo del 5.72% sugiere una ligera diferencia en la pendiente, pero sigue siendo aceptable dentro del contexto del ajuste.

4 Cálculo del Error Cuadrático Total (ECT)

Para un análisis más detallado, calculamos el **Error Cuadrático Total (ECT)** para ambos modelos. Recordemos que el ECT se define como:

$$S = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{22}$$

donde:

- y_i: valor observado,
- $\hat{y}_i$: valor predicho por el modelo.

1. Modelo Manual

Los valores predichos por el modelo manual son:

Tiempo (T)	$P(T)_{\textbf{real}}$ **(millones)**	$P(T)_{\textbf{predicho}}$ **(millones)**	**Diferencia ($y_i - \hat{y}_i$)**
10	7.24	7.21	$7.24 - 7.21 = 0.03$
20	9.638	9.75	$9.638 - 9.75 = -0.112$
30	12.681	13.13	$12.681 - 13.13 = -0.449$
40	17.064	17.58	$17.064 - 17.58 = -0.516$
50	23.192	23.33	$23.192 - 23.33 = -0.138$
60	31.443	30.56	$31.443 - 30.56 = 0.883$
70	39.98	39.77	$39.98 - 39.77 = 0.21$
80	50.189	50.56	$50.189 - 50.56 = -0.371$
90	62.89	63.36	$62.89 - 63.36 = -0.47$
100	77.212	77.75	$77.212 - 77.75 = -0.538$

Table 3: Valores predichos y diferencias para el modelo manual.

Sumando todos los errores cuadráticos:

$$\text{ECT}_{\text{manual}} = 0.0009 + 0.0125 + 0.2016 + 0.2662 + ... + 0.2894 = 1.9719$$

2. Modelo GeoGebra

Los valores predichos por el modelo de GeoGebra son:
Sumando todos los errores cuadráticos:

$$\text{ECT}_{\text{GeoGebra}} = 0.0004 + 0.0231 + 0.2798 + 0.4173 + ... + 0.000004 = 1.2016$$

Comparación de los Errores Cuadráticos Totales

El modelo obtenido mediante GeoGebra presenta un error cuadrático total significativamente menor (ECT = 1.2016) en comparación con el modelo calculado manualmente (ECT = 1.9719). Esto indica que el ajuste realizado con GeoGebra es más preciso y se adapta mejor a los datos históricos de población.

Tiempo (T)	$P(T)_{\mathbf{real}}$ (millones)	$P(T)_{\mathbf{predicho}}$ (millones)	Diferencia ($y_i - \hat{y}_i$)
10	7.24	7.22	$7.24 - 7.22 = 0.02$
20	9.638	9.79	$9.638 - 9.79 = -0.152$
30	12.681	13.21	$12.681 - 13.21 = -0.529$
40	17.064	17.71	$17.064 - 17.71 = -0.646$
50	23.192	23.45	$23.192 - 23.45 = -0.258$
60	31.443	30.80	$31.443 - 30.80 = 0.643$
70	39.98	39.98	$39.98 - 39.98 = 0.00$
80	50.189	50.18	$50.189 - 50.18 = 0.009$
90	62.89	62.86	$62.89 - 62.86 = 0.03$
100	77.212	77.21	$77.212 - 77.21 = 0.002$

Table 4: Valores predichos y diferencias para el modelo de GeoGebra.

Método	Error Cuadrático Total (ECT)
Modelo Manual	1.9719
Modelo GeoGebra	1.2016

Table 5: Comparación de los errores cuadráticos totales.

5 Conclusión

El modelo obtenido mediante GeoGebra presenta un error cuadrático total significativamente menor en comparación con el modelo calculado manualmente. No obstante, el método de mínimos cuadrados demostró ser una herramienta poderosa y confiable para ajustar el modelo logístico a los datos históricos de población. A pesar de ser un enfoque manual, los resultados obtenidos fueron consistentes y cercanos a los generados por herramientas computacionales avanzadas.