LA PARÁBOLA DE LOS MÍNIMOS CUADRADOS

Autor: Mario Suárez

La parábola de mínimos cuadrados que aproxima el conjunto de puntos $(X_1, Y_1), (X_2, Y_2), (X_3, Y_3), \ldots (Y_N, Y_N)$ tiene ecuación dada por $Y = a_0 + a_1X + a_2X^2$, donde las constantes a_0, a_1 y a_2 se determinan al resolver simultáneamente el sistema de ecuaciones que se forma al multiplicar la ecuación $Y = a_0 + a_1X + a_2X^2$ por $1, X, Y$ sucesivamente, y sumando después.

$$\begin{cases} \Sigma Y = a_0N + a_1\Sigma X + a_2\Sigma X^2 \\ \Sigma XY = a_0\Sigma X + a_1\Sigma X^2 + a_2\Sigma X^3 \\ \Sigma X^2Y = a_0\Sigma X^2 + a_1\Sigma X^3 + a_2\Sigma X^4 \end{cases}$$

Ejemplo ilustrativo

La siguiente tabla muestra la población de un país en los años 1960-2010 en intervalos de 5 años.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Población (millones)	4,52	5,18	6,25	7,42	8,16	9,12	10,92	11,62	12,68	13,12	13,97

1) Ajustar una parábola de mínimos cuadrados de la forma $Y = a_0 + a_1X + a_2X^2$

2) Calcular los valores de tendencia para los años dados.

3) Estimar la población para los años 2015 y 2020.

4) Calcular el coeficiente de determinación.

5) Elaborar un diagrama de dispersión, y en el mismo diagrama graficar la parábola de los mínimos cuadrados.

Nota: Se recomienda codificar o cambiar la numeración de los años, tratando que X = 0 esté ubicado en lo posible en el centro.

Solución:

1) Para ajustar una parábola de mínimos cuadrados se llena la siguiente tabla:

Año	X	Y	X^2	X^3	X^4	XY	X^2Y
1960	-5	4,52	25	-125	625	-22,6	113
1965	-4	5,18	16	-64	256	-20,72	82,88
1970	-3	6,25	9	-27	81	-18,75	56,25
1975	-2	7,42	4	-8	16	-14,84	29,68
1980	-1	8,16	1	-1	1	-8,16	8,16
1985	0	9,12	0	0	0	0	0
1990	1	10,92	1	1	1	10,92	10,92
1995	2	11,62	4	8	16	23,24	46,48
2000	3	12,68	9	27	81	38,04	114,12
2005	4	13,12	16	64	256	52,48	209,92
2010	5	13,97	25	125	625	69,85	349,25
Σ	0	102,96	110	0	1958	109,46	1020,66

Se remplaza valores en el sistema y se obtiene:

$$\begin{cases} \Sigma Y = a_0 N + a_1 \Sigma X + a_2 \Sigma X^2 \\ \Sigma XY = a_0 \Sigma X + a_1 \Sigma X^2 + a_2 \Sigma X^3 \\ \Sigma X^2 Y = a_0 \Sigma X^2 + a_1 \Sigma X^3 + a_2 \Sigma X^4 \end{cases}$$

$$\begin{cases} 102{,}96 = \mathrm{a}_0 \cdot 11 + \mathrm{a}_1 \cdot 0 + \mathrm{a}_2 \cdot 110 \\ 109{,}46 = \mathrm{a}_0 \cdot 0 + \mathrm{a}_1 \cdot 110 + \mathrm{a}_2 \cdot 0 \\ 1020{,}66 = \mathrm{a}_0 \cdot 110 + \mathrm{a}_1 \cdot 0 + \mathrm{a}_2 \cdot 1958 \end{cases} \Rightarrow \begin{cases} 11\mathrm{a}_0 + 0\mathrm{a}_1 + 110\mathrm{a}_2 = 102{,}96 \\ 0\mathrm{a}_0 + 110\mathrm{a}_1 + 0\mathrm{a}_2 = 109{,}46 \\ 110\mathrm{a}_0 + 0\mathrm{a}_1 + 1958\mathrm{a}_2 = 1020{,}66 \end{cases}$$

Resolviendo el sistema empleando determinantes (regla de Cramer) se obtiene:

$$a_0 = \frac{\Delta a_0}{\Delta} = \frac{\begin{vmatrix} 102{,}96 & 0 & 110 \\ 109{,}46 & 110 & 0 \\ 1020{,}66 & 0 & 1958 \end{vmatrix}}{\begin{vmatrix} 11 & 0 & 110 \\ 0 & 110 & 0 \\ 110 & 0 & 1958 \end{vmatrix}} = \frac{\begin{array}{c|ccc|c} & 102{,}96 & 0 & 110 & \\ & 109{,}46 & 110 & 0 & \\ - & 1020{,}66 & 0 & 1958 & + \\ - & 102{,}96 & 0 & 110 & + \\ - & 109{,}46 & 110 & 0 & + \end{array}}{\begin{array}{c|ccc|c} & 11 & 0 & 110 & \\ & 0 & 110 & 0 & \\ - & 110 & 0 & 1958 & + \\ - & 11 & 0 & 110 & + \\ - & 0 & 110 & 0 & + \end{array}}$$

$$\mathrm{a}_0 = \frac{22175524{,}8 + 0 + 0 - 12349986 - 0 - 0}{2369180 + 0 + 0 - 1331000 - 0 - 0} = \frac{9825538{,}8}{1038180} = 9{,}464$$

$$a_1 = \frac{\Delta a_1}{\Delta} = \frac{\begin{vmatrix} 11 & 102{,}96 & 110 \\ 0 & 109{,}46 & 0 \\ 110 & 1020{,}66 & 1958 \end{vmatrix}}{1038180} = \frac{\begin{array}{c|ccc|c} & 11 & 102{,}96 & 110 & \\ & 0 & 109{,}46 & 0 & \\ - & 110 & 1020{,}66 & 1958 & + \\ - & 11 & 102{,}96 & 110 & + \\ - & 0 & 109{,}46 & 0 & + \end{array}}{1038180}$$

$$a_1 = \frac{23577549{,}48 + 0 + 0 - 1324466 - 0 - 0}{1038180} = \frac{2357549{,}48}{1038180} = 0{,}995$$

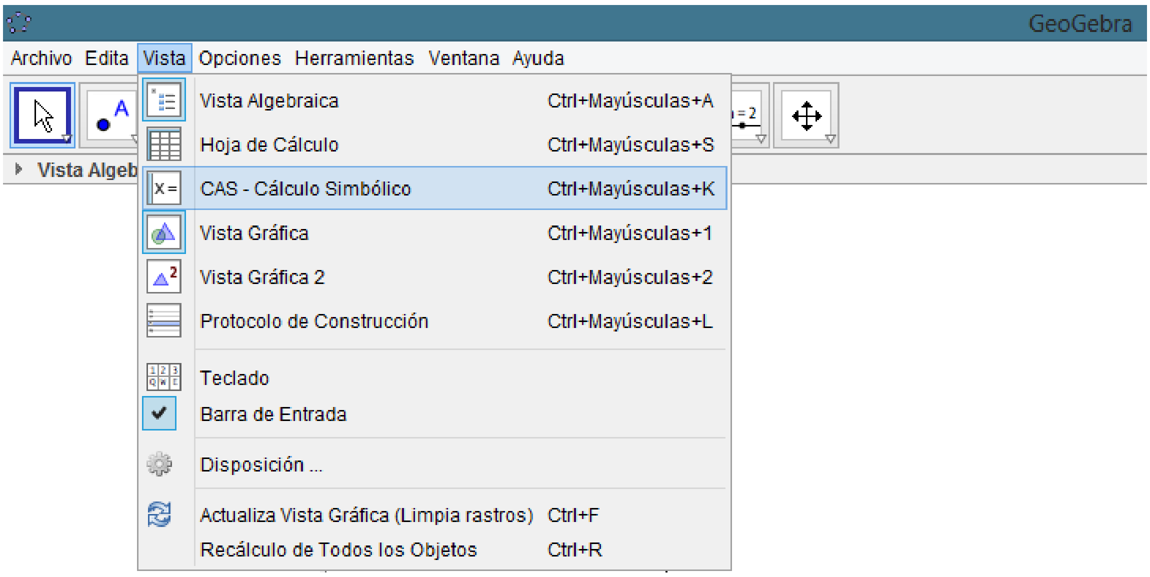
$$a_2 = \frac{\Delta a_2}{\Delta} = \frac{\begin{vmatrix} 11 & 0 & 102{,}96 \\ 0 & 110 & 109{,}46 \\ 110 & 0 & 1020{,}66 \end{vmatrix}}{1038180} = \frac{\begin{array}{c|ccc|c} & 11 & 0 & 102{,}96 & \\ & 0 & 110 & 109{,}46 & \\ - & 110 & 0 & 1020{,}66 & + \\ - & 11 & 0 & 102{,}96 & + \\ - & 0 & 110 & 109{,}46 & + \end{array}}{1038180}$$

$$a_2 = \frac{1234998{,}6 + 0 + 0 - 1245816 - 0 - 0}{1038180} = \frac{-10817{,}4}{1038180} = -0{,}01$$

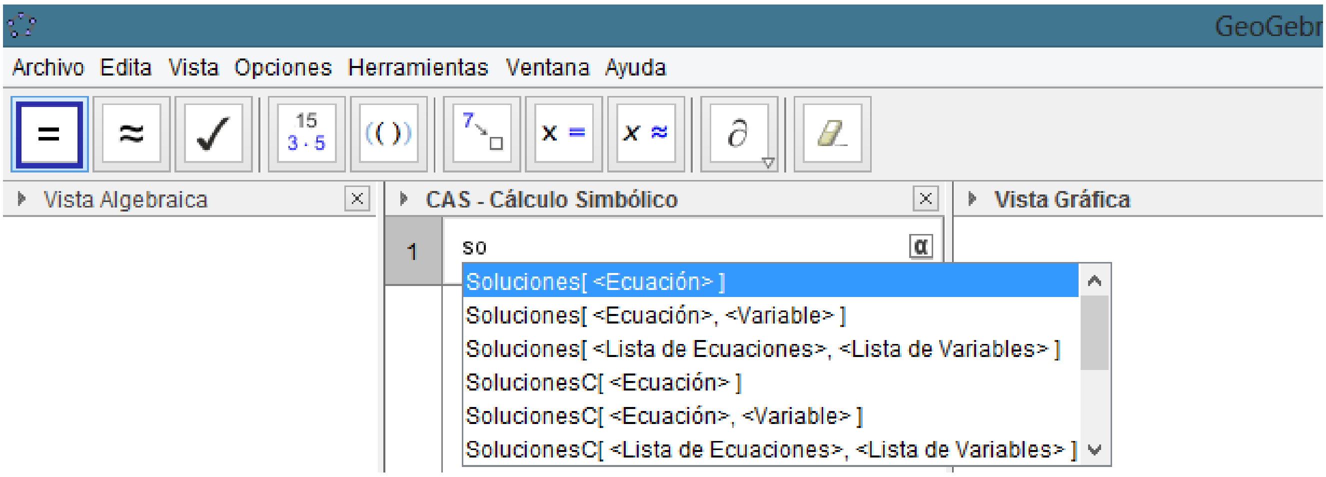
El sistema resuelto en Excel se muestra en la siguiente figura:

	A	B	C	D	E	F	G	H
16						a_0	a_1	a_2
17	$11a_0 + 0a_1 + 110a_2 = 102{,}96$				102,96	11	0	110
18	$0a_0 + 110a_1 + 0a_2 = 109{,}46$				109,46	0	110	0
19	$110a_0 + 0a_1 + 1958a_2 = 1020{,}66$				1020,66	110	0	1958
20								
21	102,96	0	110					
22	109,46	110	0		$a_0 = \frac{\Delta a_0}{\Delta}$	9,464	=B24/B40	
23	1020,66	0	1958					
24	Δa_0	9825538,8	=MDETERM(A21:C23)					
25								
26								
27	11	102,96	110					
28	0	109,46	0		$a_1 = \frac{\Delta a_1}{\Delta}$	0,995	=B30/B40	
29	110	1020,66	1958					
30	Δa_1	1033083,5	=MDETERM(A27:C29)					
31								
32	11	0	102,96					
33	0	110	109,46					
34	110	0	1020,66					
35	Δa_2	-10817,4	=MDETERM(A32:C34)					
36								
37	11	0	110					
38	0	110	0		$a_2 = \frac{\Delta a_2}{\Delta}$	-0,010	=B35/B40	
39	110	0	1958					
40	Δ	1038180	=MDETERM(A37:C39)					

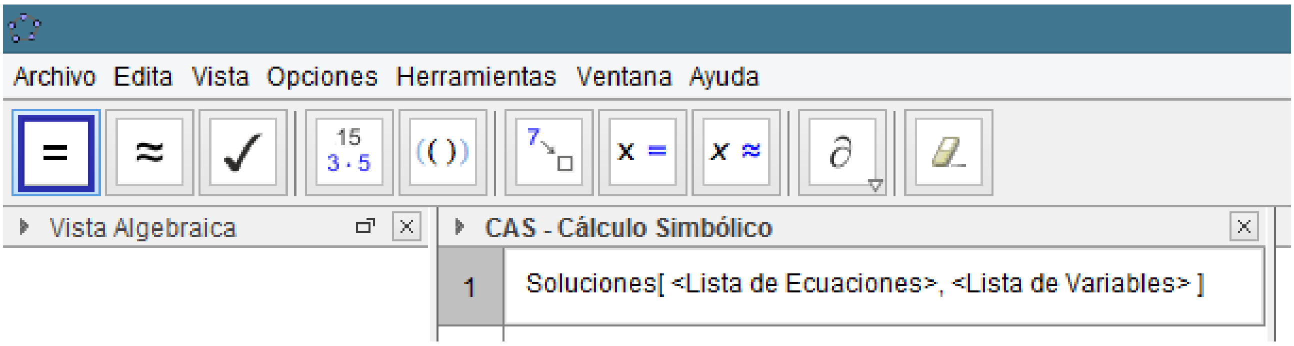
Para resolver el sistema en GeoGebra se sigue los siguientes pasos:
a) Clic en Vista



b) Clic en CAS-Cálculo Simbólico. Escribir soluciones en la casilla 1

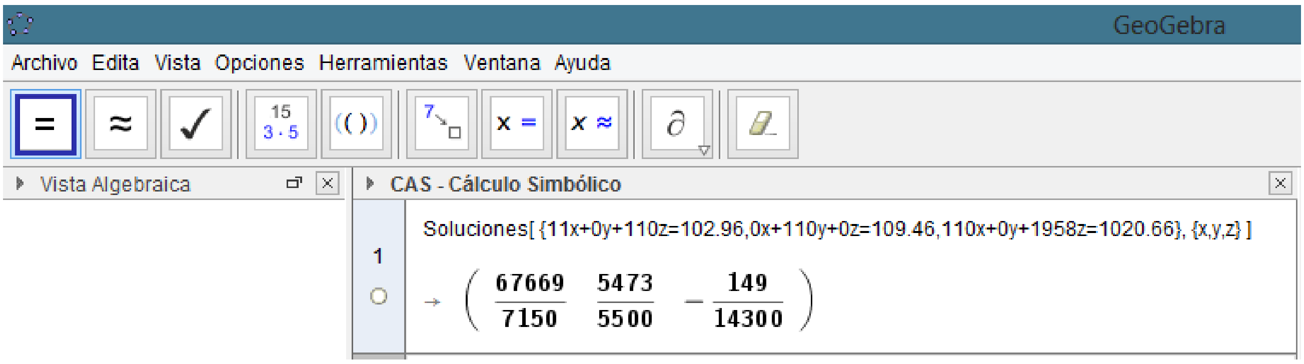


c) Escoger la opción Soluciones[<Lista de Ecuaciones>, <Lista de Variables>]



d) Escribir la lista de ecuaciones y la lista de variables. Enter

Soluciones[{11x+0y+110z=102.96,0x+110y+0z=109.46,110x+0y+1958z=1020.66}, {x,y,z}]



$$\frac{67669}{7150} = 9{,}464\,; \frac{5473}{5500} = 0{,}995\,; -\frac{149}{14300} = -0{,}01$$

Remplazando los valores encontrados se obtiene la ecuación de la parábola de mínimos cuadrados:

$$Y = a_0 + a_1X + a_2X^2 \Rightarrow Y = 9{,}464 + 0{,}995X - 0{,}01X^2$$

2) Los valores de tendencia se obtienen al remplazar los valores de X en la ecuación de la parábola de mínimos cuadrados, los cuales se presenta en la siguiente tabla:

Año	X	Y	Valores de tendencia $Y = 9{,}464 + 0{,}995X - 0{,}01X^2$
1960	-5	4,52	4,24
1965	-4	5,18	5,32
1970	-3	6,25	6,39
1975	-2	7,42	7,43
1980	-1	8,16	8,46
1985	0	9,12	9,46
1990	1	10,92	10,45
1995	2	11,62	11,41
2000	3	12,68	12,36
2005	4	13,12	13,28
2010	5	13,97	14,19

3) Para estimar la población de los años 2015 y 2020 se transforma estos años a X siguiendo la secuencia de la tabla anterior, siendo X = 6 para el año 2015 y X= 7 para el 2020

Entonces para el 2015 se tiene:
$Y = 9{,}464 + 0{,}995X - 0{,}01X^2 = 9{,}464 + 0{,}995(6) - 0{,}01(6)^2 = 9{,}464 + 5{,}97 - 0{,}36 = 15{,}074$

Para el 2020 se tiene:
$Y = 9{,}464 + 0{,}995X - 0{,}01X^2 = 9{,}464 + 0{,}995(7) - 0{,}01(7)^2 = 9{,}464 + 6{,}965 - 0{,}49 = 15{,}939$

4) Se llena la siguiente tabla y se aplica la ecuación para calcular el coeficiente de Pearson

Año	X	Y	X^2	XY	Y^2
1960	-5	4,52	25	-22,6	20,430
1965	-4	5,18	16	-20,72	26,832
1970	-3	6,25	9	-18,75	39,063
1975	-2	7,42	4	-14,84	55,056
1980	-1	8,16	1	-8,16	66,586
1985	0	9,12	0	0	83,174
1990	1	10,92	1	10,92	119,246
1995	2	11,62	4	23,24	135,024
2000	3	12,68	9	38,04	160,782
2005	4	13,12	16	52,48	172,134
2010	5	13,97	25	69,85	195,161
Σ	0	102,96	110	109,46	1073,490

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{[N\sum X^2 - (\sum X)^2][N\sum Y^2 - (\sum Y)^2]}} = \frac{11 \cdot 109{,}46 - 0 \cdot 102{,}96}{\sqrt{[11 \cdot 110 - (0)^2][11 \cdot 1073{,}490 - (102{,}96)^2]}}$$
$$r = 0{,}996$$

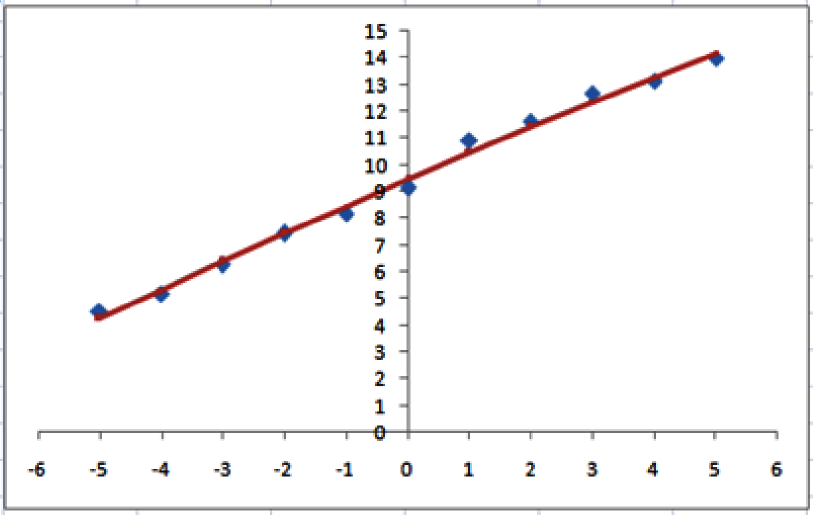
Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.
Coeficiente de determinación = $r^2 = (0{,}996)^2 = 0{,}992$

El coeficiente de determinación calculado en Excel se muestra en la siguiente figura:

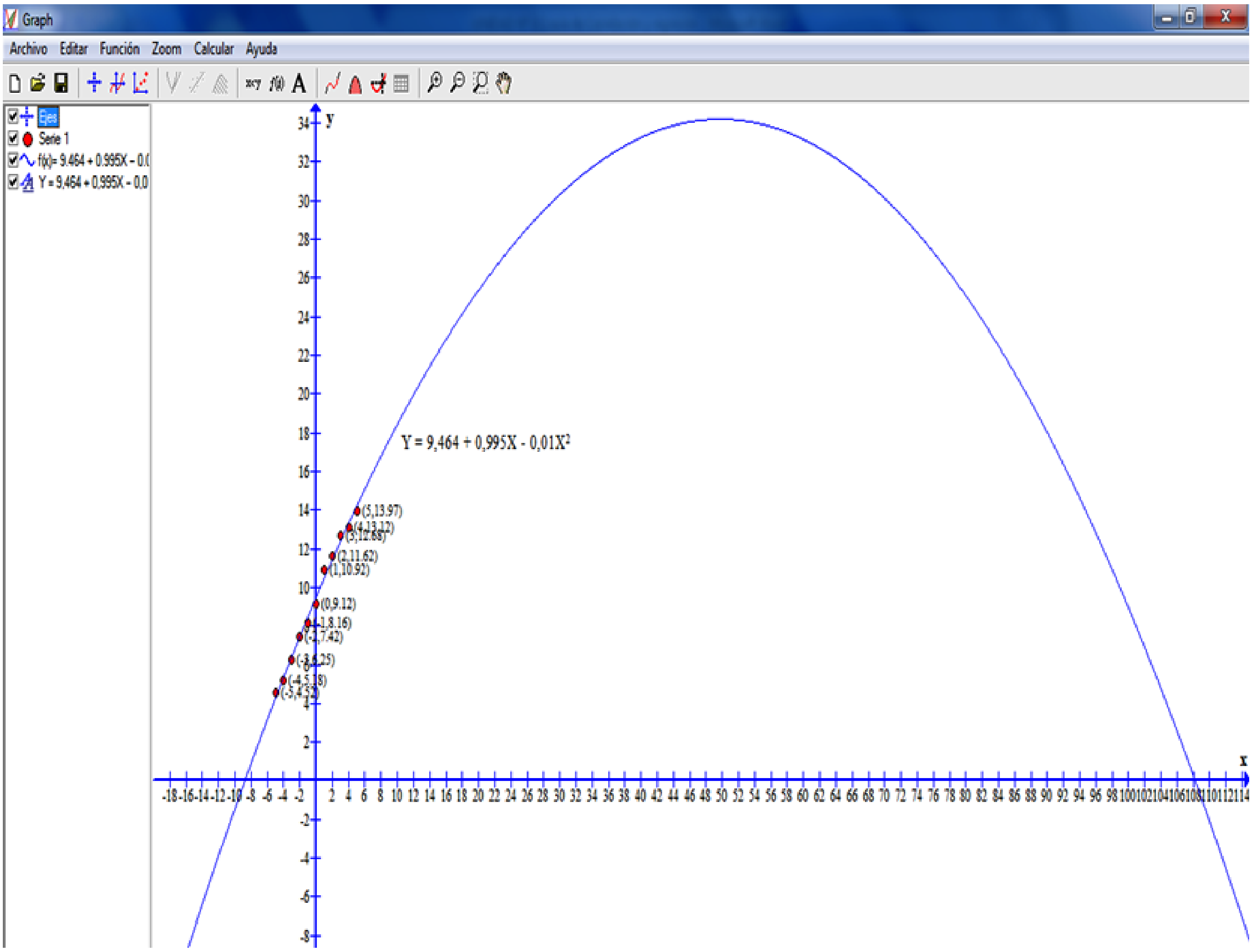
	A	B	C	D	E
1	Año	X	Y		
2	1960	-5	4,52		
3	1965	-4	5,18		
4	1970	-3	6,25		
5	1975	-2	7,42		
6	1980	-1	8,16		
7	1985	0	9,12		
8	1990	1	10,92		
9	1995	2	11,62		
10	2000	3	12,68		
11	2005	4	13,12		
12	2010	5	13,97		
13					
14	r	0,9960666	=COEF.DE.CORREL(B2:B12;C2:C12)		
15		0,9960666	=COEF.DE.CORREL(A2:A12;C2:C12)		
16	r^2	0,9921487	=E15^2		
17		0,9921487	=COEFICIENTE.R2(C2:C12;B2:B12)		

5) *El diagrama de dispersión y la parábola de los mínimos cuadrados mediante Excel se muestra en la siguiente figura:*

	A	B	C	D
1	Año	X	Y	$Y=9{,}464+0{,}995X-0{,}01X^2$
2	1960	-5	4,52	4,24
3	1965	-4	5,18	5,32
4	1970	-3	6,25	6,39
5	1975	-2	7,42	7,43
6	1980	-1	8,16	8,46
7	1985	0	9,12	9,46
8	1990	1	10,92	10,45
9	1995	2	11,62	11,41
10	2000	3	12,68	12,36
11	2005	4	13,12	13,28
12	2010	5	13,97	14,19
13				
14		a_0	9,464	
15		a_1	0,995	
16		a_2	-0,01	



Empleando el programa Graph se obtiene la siguiente figura:



TAREA DE INTERAPRENDIZAJE

1) La siguiente tabla muestra la población aproximada de la Provincia de Imbabura en los años 1960-2010 en intervalos de 5 años.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Población (miles)	123	140	170	201	221	247	296	315	344	356	379

1.1) Ajuste una parábola de mínimos cuadrados de la forma $Y = a_0 + a_1X + a_2X^2$ manera manual, empleando Excel y GeoGebra.

$Y = 256,464 + 26,991X - 0,265X^2$

1.2) Calcule los valores de tendencia para los años dados de manera manual y empleando Excel.

Año	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010
Valor de tendencia	114,88	144,26	173,11	201,42	229,21	256,46	283,19	309,39	335,05	360,19	384,79

1.3) Estime la población para los años 2015 y 2020

Año 2015 = 408,87 miles de habitantes
Año 2020 = 432,42 miles de habitantes

1.4) Calcule el coeficiente de determinación de manera manual y empleando Excel.

0,992

1.5) Elabore un diagrama de dispersión, y en el mismo diagrama graficar la parábola de los mínimos cuadrados de manera manual, empleando Excel y empleando Graph.

2) Cree y resuelva un ejercicio de aplicación de la parábola de los mínimos cuadrados con datos de la población del Ecuador o de cualquier otro país de manera manual, empleando Excel y Graph.

3) Consulte en la biblioteca o en el internet un ejercicio de aplicación de la Parábola de los mínimos cuadrados. Presente el ejercicio resuelto con GeoGebra y Graph.