**ANÁLISIS DE REGRESIÓN MEDIANTE LA PARÁBOLA DE LOS MÍNIMOS CUADRADOS**

**INTRODUCCIÓN**

Los primeros y más importantes estudios al respecto se deben a los científicos Francis Galton (1822-1911) y Karl Pearson (1857-1936). Fue Galton quien utilizó por primera vez el término regresión para indicar que, aunque influida por la estatura de sus padres, la estatura de los hijos “regresaba” a la media general.

La regresión examina la relación entre dos variables, pero restringiendo una de ellas con el objeto de estudiar las variaciones de una variable cuando la otra permanece constante. En otras palabras, la regresión es un método que se emplea para predecir el valor de una variable en función de valores dados a la otra variable. En estadística la palabra predecir no se utiliza en el sentido empleado por los astrólogos, futurólogos y mentalistas, sino mas bien en un sentido lógico como es el de utilizar el conocimiento del comportamiento de una variable para obtener información sobre otra variable. Por ejemplo, puede predecirse el resultado que obtendrá un estudiante en su examen final, basados en el conocimiento de las calificaciones promedio de sus exámenes parciales, o predecir la preferencia de los estudiantes por profesiones científicas, conociendolos promedios de sus calificaciones en los estudios escolares.

En todos los casos de regresión existe una dependencia funcional entre las variables. En el caso de dos variables, siendo una de ellas (X) variable independiente y la otra (Y) la dependiente, se habla de regresión de Y sobre X; Por ejemplo, los ingenieros forestales utilizan la regresión de la altura de los árboles sobre su diámetro, lo cual significa que midiendo el diámetro (variable independiente) y reemplazando su valor en una relación definida según la clase de árbol se obtiene la altura, y aun sin necesidad de cálculos aprecian la altura utilizando gráficas de la función de dependencia, altura = función del diámetro.

**LA PARÁBOLA DE LOS MÍNIMOS CUADRADOS**

La parábola de mínimos cuadrados que aproxima el conjunto de puntos (X1,Y1) , (X2,Y2), (X3,Y3),…..(XN,YN) tiene ecuación dada por$ Y=a\_{0}+a\_{1}X+a\_{2}X^{2}$, donde las constantes a0, a1 y a2 se determinan al resolver simultáneamente el sistema de ecuaciones que se forma al multiplicar la ecuación $Y=a\_{0}+a\_{1}X+a\_{2}X^{2}$por 1, X, Y sucesivamente, y sumando después.

$$\left\{\begin{matrix}ΣY=a\_{0}N+a\_{1}ΣX+a\_{2}ΣX^{2}\\ΣXY=a\_{0}ΣX+a\_{1}ΣX^{2}+a\_{2}ΣX^{3}\\ΣX^{2}Y=a\_{0}ΣX^{2}+a\_{1}ΣX^{3}+a\_{2}ΣX^{4}\end{matrix}\right.$$

**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 $Y=a\_{0}+a\_{1}X+a\_{2}X^{2}$

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, eligiendo X de modo que el año central, 1985, corresponda a X= 0, para que se hagan más fáciles los cálculos.

**Solución:**

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

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Año | X | Y | X2 | X3 | X4 | XY | X2Y |
| 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 reemplaza valores en el sistema y se obtiene:

$$\left\{\begin{matrix}ΣY=a\_{0}N+a\_{1}ΣX+a\_{2}ΣX^{2}\\ΣXY=a\_{0}ΣX+a\_{1}ΣX^{2}+a\_{2}ΣX^{3}\\ΣX^{2}Y=a\_{0}ΣX^{2}+a\_{1}ΣX^{3}+a\_{2}ΣX^{4}\end{matrix}\right.$$

$$\left\{\begin{array}{c}102,96=a\_{0}∙11+a\_{1}∙0+a\_{2}∙110\\109,46=a\_{0}∙0+a\_{1}∙110+a\_{2}∙0\\1020,66=a\_{0}∙110+a\_{1}∙0+a\_{2}∙1958\end{array}⇒\left\{\begin{array}{c}11a\_{0}+0a\_{1}+110a\_{2}=102,96\\0a\_{0}+110a\_{1}+0a\_{2}=109,46\\110a\_{0}+0a\_{1}+1958a\_{2}=1020,66\end{array}\right.\right.$$

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

****

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

****

$$a\_{1}=\frac{23577549,48+0+0-1324466-0-0}{1038180}=\frac{2357549,48}{1038180}=0,995$$

****

$$a\_{2}=\frac{1234998,6+0+0-1245816-0-0}{1038180}=\frac{-10817,4}{1038180}=-0,01$$

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

****

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

$Y=a\_{0}+a\_{1}X+a\_{2}X^{2}$ $⇒ $Y = 9,464 + 0,995X - 0,01X2

2) Los valores de tendencia se obtienen al reemplazar 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 tendenciaY = 9,464 + 0,995X - 0,01X2 |
| 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 - 0,01X2 =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 - 0,01X2 =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 | X2 | XY | Y2 |
| 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 |

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

$$r=0,996$$

Elevando al cuadrado coeficiente de Pearson queda calculado el coeficiente de determinación.

Coeficiente de determinación = $r^{2}=\left(0,996\right)^{2}=0,992$

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



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



**REFERENCIAS BIBLIOGRÁFICAS**

SPIEGEL, Murray, (2000), Estadística,Serie de Compendios Schaum, Ed. McGraw-Hill, México.

SUÁREZ, Mario, (2011), Interaprendizaje de Estadística Básica,

TAPIA , Fausto Ibarra, Ecuador.

SUÁREZ, Mario, (2004), Interaprendizaje Holístico de Matemática, Ed. Gráficas Planeta, Ibarra,

 Ecuador.

SUAREZ IBUJÉS MARIO ORLANDO

mgsmariosuarez@gmail.com

mosuarez@utn.edu.ec

Telf: 06 2632 166

 085619601