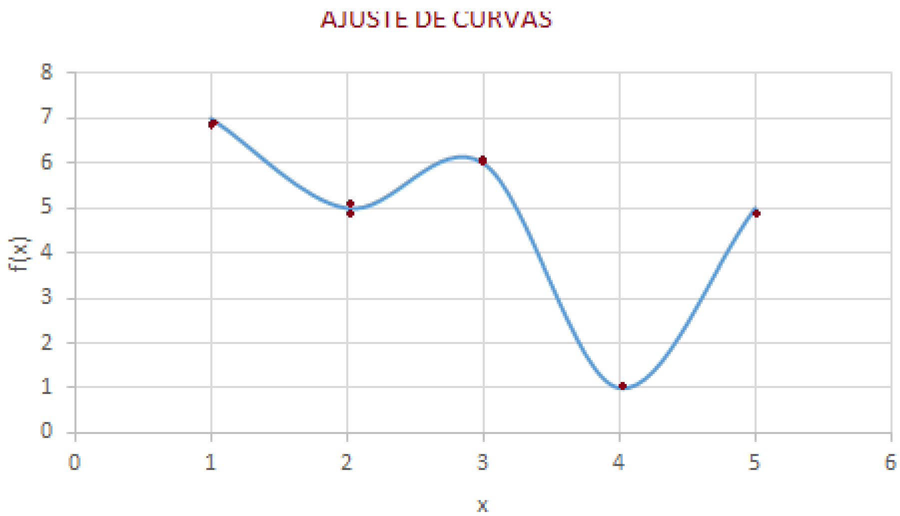
4. AJUSTE DE CURVAS

Resumen.

El presente capitulo tiene como objetivo ajustar una curva de acuerdo con una serie de datos presentados en una tabla, con el objeto de poder realizar estimaciones internas a la tabla, lo que denominamos interpolación, así mismo buscaremos hacer estimaciones externas a las tablas de datos la cual denominamos extrapolaciones. Como primera instancia presentaremos los llamados polinomios de colocación, los cuales son utilizados para interpolar, entre ellos destacamos los polinomios de Newton y Lagrange. Posteriormente explicaremos los llamados polinomios de mínimos cuadrados, los cuales además de interpolantes también son extrapolantes. Por último, haremos ajustes a modelos no polinomiales, para ello haremos uso del proceso llamado linealización.

Palabras claves: polinomios, interpolación, extrapolación

4.1 POLINOMIOS DE COLOCACIÓN



En este grafico hemos ajustado los datos (1,7),(2,5),(3,6),(4,1) y (5,5) a lo que se llama un polinomio de colocación, cuya característica principal es que el modelo pasa exactamente por los puntos tabulares. Estos polinomios de colocación se utilizan para realizar INTERPOLACIONES, es decir estimar imágenes a valores internos de la tabla.

Dentro de estos polinomios de colocación hablaremos de los de Newton y Lagrange.

4.1.1 POLINOMIOS DE NEWTON

Los polinomios de Newton los daremos para tablas igualmente espaciadas y para todo tipo de tabla.

Diremos que una tabla esta igualmente espaciada si la diferencia entre dos entradas consecutivas siempre es igual, es decir: $(\forall x_i, x_{i+1})(x_{i+1} - x_i = h)$, donde h se denomina longitud de paso.

Los polinomios de Newton para tablas igualmente espaciadas requieren de unas tablas llamadas tablas de diferencia, las cuales utilizan un operador llamado operador de paso hacia adelante, el cual lo definimos de la siguiente manera:

$$\Delta f(x) = f(x+h) - f(x)$$

$$\Delta f_k = f_{k+1} - f_k$$

La primera definición es a nivel de entradas de la tabla y la segunda en función de la posición k-ésima en la tabla.

Este operador lo utilizaremos para aproximar las derivadas de la siguiente manera $f^{(n)} \cong \frac{\Delta^n}{h^n}$, donde la potencia del operador se obtiene al aplicar en forma recurrente $\Delta^n = \Delta(\Delta^{n-1})$, y para esto elaboramos precisamente la tabla de diferencias:

x_k	f_k	Δ	Δ^2	Δ^3	$\dots$	Δ^n
x_0	f_0	Δf_0	$\Delta^2 f_0$	$\Delta^3 f_0$	$\dots$	$\Delta^n f_0$
x_1	f_1	Δf_1	$\Delta^2 f_1$			
x_2	f_2	Δf	$\Delta^2 f_2$		$\dots$	
. . .	. . .	Δf_{n-1}	$\Delta^2 f_{n-2}$	. .	$\dots$	
x_n	f_n					

Una de las características de la tabla es que de acuerdo al grado del polinomio de colocación veremos que la tabla estandariza.

La forma del polinomio de Newton para esta tabla será:

$$P_n(x) = f_0 + \frac{\Delta f_0}{h}(x-x_0)^1 + \frac{\Delta^2 f_0}{h^2}\frac{(x-x_0)^1(x-x_1)}{2!} + \frac{\Delta^3 f_0}{h^3}\frac{(x-x_0)^1(x-x_1)(x-x_2)}{3!} + \cdots + \frac{\Delta^n f_0}{h^n}\frac{(x-x_0)^1(x-x_1)(x-x_2)\dots(x-x_{n-1})}{n!}$$

Y su parte residual se aproxima con:

$$R_n(x) \cong \frac{\triangle^{n+1}}{h^{n+1}}\frac{(x-x_0)^1(x-x_1)(x-x_2)\dots(x-x_n)}{(n+1)!}$$

Veamos un ejemplo para aterrizar lo anteriormente dicho:

Ajustar un polinomio de colocación a la tabla de datos:

x	-1	0	1	2	3
f(x)	-4	-3	0	5	12

Como vemos en la tabla el valor de h es uno, ahora realizamos la tabla de diferencias.

x	f	Δ	Δ2	Δ3	Δ4
-1	-4	1	2	0	0
0	-3	3	2	0	
1	0	5	2		
2	5	7			
3	12				

Démonos cuenta que la tabla estandarizo en $\triangle^2$, lo cual quiere decir que el polinomio de colocación es cuadrático. Ahora remplazando en la fórmula de Newton tendremos:

$$P_2(x) = -4 + \frac{1}{1}\frac{(x+1)}{1!} + \frac{2}{1^2}\frac{(x+1)x}{2!}$$

$$P_2(x) = -4 + (x+1) + (x^2+x)$$
$$P_2(x) = x^2 + 2x - 3$$

La restricción del método anterior es que la tabla debe ser igualmente espaciada, para el caso en que no lo sea e inclusive si lo es viene la otra versión del polinomio de Newton para todo tipo de tabla, en este caso definiremos las tablas de diferencias divididas..

Tablas de diferencias divididas:

Xk	f(Xk)	primer orden	segundo orden	tercer orden	Columna1	orden n
X0	f(X0)	f(X0,X1)	f(X0,X1,X2)	.	.	f(X0,X1...,Xn)
X1	f(X1)	f(X1,X2)	f(X1,X2,X3)			
X2	f(X2)	f(X2,X3)	f(X2,X3,X4)			
.	.	.	.			
.	.	.	f(Xn-2,Xn-1,Xn)			
.	.	f(Xn-1,Xn)				
Xn	f(Xn)					

Donde $f(x_0, x_1, \ldots . x_k) = \frac{f(x_0,\ldots,x_{k-1}) - f(x_1,\ldots,x_k)}{x_k - x_0}$, y el polinomio de colocación será:
$P_n(x) = f(x_0) + f(x_0, x_1) * (x - x_0) + f(x_0, x_1, x_2) * (x - x_0)(x - x_1) + \cdots +$
$f(x_0, \ldots x_n) * (x - x_0) \ldots (x - x_{n-1})$.
Y su parte residual seria:
$R_n(x) = f(x_0, \ldots . x_{n+1}) * (x - x_0) \ldots (x - x_n)$.

Ejemplo: La siguiente tabla relaciona el tiempo t y la temperatura T

t	0	2	3	5
T	20	40	70	100

Estimar la temperatura en t=1 usando polinomios de colocación de grados 1 y 2 e indicar mejor estimación.

Solución: como vemos la tabla no es igualmente espaciada por lo tanto tomaremos la segunda versión de polinomios de Newton, por tanto realizamos la tabla de diferencias divididas:

t	T2	primero	segundo	tercero
0	20	10	6,66666667	-2,33333333
2	40	30	-5	
3	70	15		
5	100			

Luego la aproximación lineal seria.$T_1(t) = 20 + 10(t - 0)$, $\quad R_1(t) = 6.66667(t - 0)(t - 2)$

Reemplazando $T_1(1) = 30$, $R_1(1) = -6.66667$

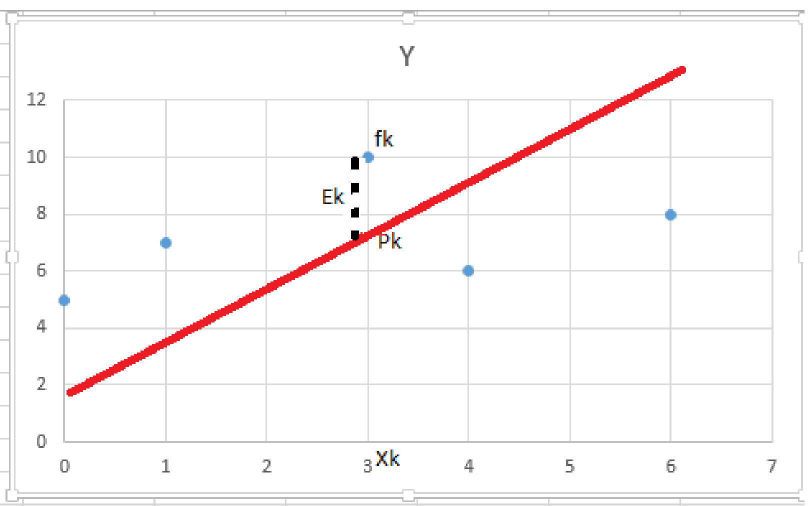
La aproximación cuadrática seria: $T_2(t) = T_1(t) + R_1(t)$, $\quad R_2(t) = -2.333(t)(t - 2)(t - 3)$

Reemplazando tendremos $T_2(1) = 23.3333$, $\quad R_2(1) = -4.6666$

Comparando los dos residuos observamos que en valor absoluto el más pequeño es el cuadrático, por tanto, la mejor estimación seria 23.3333.

4.2 POLINOMIOS DE MÍNIMOS CUADRADOS

X	Y
0	5
1	7
3	10
4	6
6	8



A diferencia de los polinomios de colocación que, si pasan exactamente por los puntos tabulares, los de mínimos cuadrados a veces ni siquiera toca alguno de los puntos tabulares. En la gráfica identificamos algunos elementos como son:

$$f_k = valor\ exacto\ de\ tabla$$

$$p_k = valor\ estimado\ mediante\ polinomio\ de\ minimos\ cuadrados\ p(x)$$
$$\in_k = error$$
$$\in_k = f_k - p_k$$

Estos errores pueden ser positivos o negativos, para efectos de que siempre nos dé positivo los elevaremos al cuadrado y lo denominaremos error cuadrático. La suma de todos estos errores cuadráticos la denominaremos bondad de ajuste $I = \sum_{k=1}^{n}(p_k - f_k)^2$, la cual pretendemos optimizar para que sea lo más pequeña posible para la cual aplicamos calculo vectorial la cual indica que las derivadas parciales se deben igualar a cero para hallar el punto crítico.

Deduciremos los modelos para la recta y parábola de mínimos cuadrados y al final daremos una generalización

RECTA DE MINIMOS CUADRADOS:

$$p(x) = Ax + B$$

Bondad de ajuste. $I = \sum_{K=1}^{n}(Ax_k + B - f_k)^2$

Optimizando:

$$\frac{\partial I}{\partial A} = \sum_{k=1}^{n} 2(Ax_k + B - f_k)(x_k) = 0$$
$$\frac{\partial I}{\partial B} = \sum_{k=1}^{n} 2(Ax_k + B - f_k)(1) = 0$$

Aplicando propiedades de sumatorias generamos el sistema normal:

$$\begin{cases} A\sum_{k=1}^{n} x_k^2 + B\sum_{k=1}^{n} = \sum_{K=1}^{n} f_k x_k \\ A\sum_{k=1}^{n} x_k + B * n = \sum_{K=1}^{n} f_k \end{cases}$$

Al resolverlos obtenemos los valores de A y B para obtener nuestra recta de mínimos cuadrados o también llamada regresión lineal

Ejemplo: dada la tabla que relaciona tiempo t y temperatura T

t	0	2	3	5
T	20	40	70	100

Estimar la temperatura en t=1 por regresión lineal, y además estimar la temperatura en t=6 (extrapolación).

t	T	t^2	t*T
0	20	0	0
2	40	4	80
3	70	9	210
5	100	25	500
10	230	38	790

En la tabla adjunta en rojo están las sumatorias que son las que vamos a reemplazar en el sistema normal:

$$\begin{cases} 38A + 10B = 790 \\ 10A + 4B = 230 \end{cases}$$

Al resolver obtenemos: A=16.5384615 y B=16.1538462, quedando el modelo lineal

$$T(t) = 16.5384615t + 16.1538462$$

De donde $T(1) = 32.6923077, \qquad T(6) = 115.3846152.$

El estadístico más común de medida de tendencia central se tiene que es la media aritmética es decir $\bar{y} = \frac{\sum y_i}{n}$ donde n es el tamaño de la muestra, asimismo la medida de dispersión más común es la desviación estándar dada por $s_y = \sqrt{\frac{s_t}{n-1}}$ donde $s_t = \sum(y_i - \bar{y})^2$.
El estadístico que tiene utilidad para cuantificar la dispersión es el llamado coeficiente de variación (c.v.) dado por $cv = \frac{s_y}{\bar{y}} * 100\%$.
Adaptando estos conceptos a nuestra recta de mínimos cuadrados tenemos que s_t es lo que hemos llamado la bondad de ajuste I y de esta manera definimos el error estándar del estimado como $s_{y/x} = \sqrt{\frac{I}{n-2}}$, el cual cuantifica la dispersión alrededor de la recta de mínimos cuadrados a diferencia de s_y que lo hace alrededor de la media.
El coeficiente de determinación se define como $r^2 = \frac{s_t - I}{s_t}$ y $r = \sqrt{r^2}$ se denomina el coeficiente de correlación. El ajuste se dice perfecto si I=0, es decir cuando $r = r^2 = 1$, lo cual implica que la recta explica el 100% de la variabilidad de los datos. Si $s_t = I$ el ajuste no es el mejor donde r=0.
En el ejemplo que dimos al principio de esta sección tenemos:

$$T(t) = 16.5384615\text{t} + 16.1538462$$

De donde al observar la tabla que realizamos para este caso obtenemos que $\bar{y} = \frac{230}{4} = 57{,}5$

$$s_t = (20 - 57{,}5)^2 + (40 - 57{,}5)^2 + (70 - 57{,}5)^2 + (100 - 57{,}5)^2 = 3675$$

De donde $s_y = \sqrt{\frac{3675}{4-1}} = 35$, el error estándar estimado es $s_{y/x} = \sqrt{\frac{119.230769}{4-2}} = 7.721099954$,
Donde tenemos que $r^2 = \frac{3675 - 119.230769}{3675} = 0.9675$, lo cual indica que el modelo lineal explica el 96.75% de la incertidumbre original lo cual es bueno por cuanto es alta.

PARABOLA DE MINIMOS CUDRADOS:

$$p(x) = Ax^2 + Bx + C$$

Bondad de ajuste: $I = \sum_{k=1}^{n}(Ax_k^2 + Bx_k + C - f_K)^2$
Optimizando:

$$\frac{\partial I}{\partial A} = \sum_{k=1}^{n} 2(Ax_k^2 + Bx_k + C - f_K)(x_k^2) = 0$$

$$\frac{\partial I}{\partial B} = \sum_{k=1}^{n} 2(Ax_k^2 + Bx_k + C - f_K)(x_k) = 0$$

$$\frac{\partial I}{\partial C} = \sum_{k=1}^{n} 2(Ax_k^2 + Bx_k + C - f_K)(1) = 0$$

Generando el sistema normal:

$$\begin{cases} A\sum_{k=1}^{n} x_k^4 + B\sum_{k=1}^{n} x_k^3 + C\sum_{k=1}^{n} x_k^2 = \sum_{k=1}^{n} f_k x_k^2 \\ A\sum_{k=1}^{n} x_k^3 + B\sum_{k=1}^{n} x_k^2 + C\sum_{k=1}^{n} x_k = \sum_{k=1}^{n} f_k x_k \\ A\sum_{k=1}^{n} x_k^2 + B\sum_{k=1}^{n} x_k + C*n = \sum_{k=1}^{n} f_k \end{cases}$$

Del cual obtenemos los valores A, B y C.

Retomando el ejemplo anterior, ajústemelo a una parábola de mínimos cuadrados:

t	T	t^2	tT	t^3	t^4	t^2*T
0	20	0	0	0	0	0
2	40	4	80	8	16	160
3	70	9	210	27	81	630
5	100	25	500	125	625	2500
10	230	38	790	160	722	3290

Quedando el sistema normal:

$$\begin{cases} 722A + 160B + 38C = 3290 \\ 160A + 38B + 10C = 790 \\ 38A + 10B + 4C = 230 \end{cases}$$

Lo cual nos da como resultado A=0.83333, B=12.37179, y C=18.65384

Por tanto la parábola de mínimos cuadrados quedara:

$$T(t) = 0.83333t^2 + 12.37179\text{t} + 18.65384$$

De donde al realizar las estimaciones tendremos:

$$T(1) = 31.85896, \qquad T(6) = 122.88446$$

POLINOMIOS N-ESIMOS DE MINIMOS CUADRADOS

$$P_N(x) = A_N x^N + A_{n-1}x^{N-1} + \cdots + A_1 x + A_0$$

El sistema normal será:

$$\begin{cases} A_N \sum_{k=1}^{n} x_k^{2N} + A_{N-1}\sum_{k=1}^{n} x_k^{2N-1} + \cdots + A_0 \sum_{k=0}^{n} x_k^N = \sum_{k=0}^{n} f_k x_k^N \\ A_N \sum_{k=1}^{n} x_k^{2N-1} + A_{N-1}\sum_{k=1}^{n} x_k^{2N-2} + \cdots + A_0 \sum_{k=0}^{n} x_k^{N-1} = \sum_{k=0}^{n} f_k x_k^{N-1} \\ \vdots \\ A_N \sum_{k=1}^{n} x_k^{N} + A_{N-1}\sum_{k=1}^{n} x_k^{N-1} + \cdots + A_0 * n = \sum_{k=0}^{n} f_k \end{cases}$$

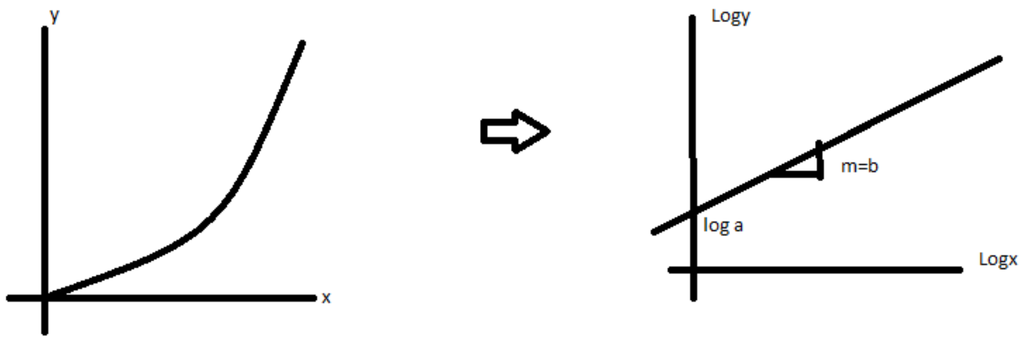
Para este polinomio de mínimos cuadrados de grado N tendremos que $s_{y/x} = \sqrt{\frac{I}{n-(N+1)}}$,

En el ejemplo anterior de la parábola de mínimos cuadrados tenemos que I=94,2307701 de donde $r^2 = \frac{3675-94,2307701}{3675} = 0,9743$ lo cual indica que el modelo lineal explica el 97.43% de la incertidumbre original la cual es ligeramente mayor que la que nos dio en la recta de mínimos cuadrados.

4.3. LINEALIZACION

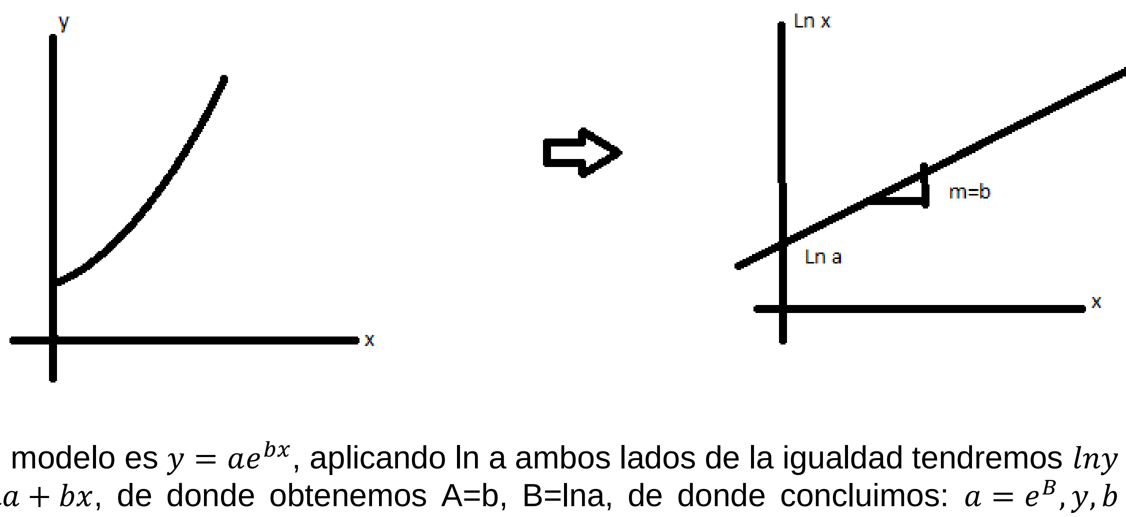
En esta sección veremos otro tipo de ajuste de curvas llamado linealización, el cual como su nombre lo indica consiste en ver los modelos como si fueran rectas de mínimos cuadrados $Y = AX + B$, analizaremos particularmente tres modelos: potencial, exponencial y saturación.

POTENCIAL.:



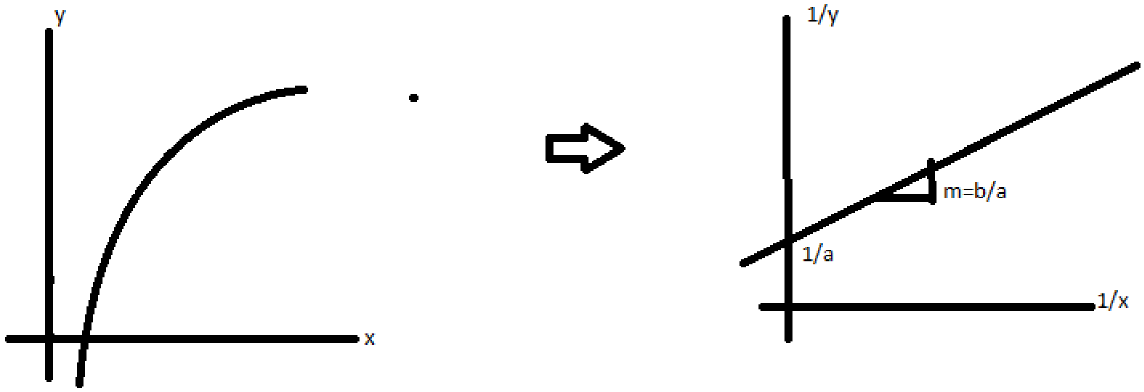
El modelo es $y = ax^b$, la idea es relacionar los valores a y b con los valores A y B de la recta de mínimos cuadrados:
Si aplicamos log a ambos lados de la igualdad obtendremos. $logy = loga + b * logx$, el cual al comparar con la recta de minimos cuadrados obtendremos. $A = b\ , y\ , B = loga$, de donde concluimos: $a = 10^B, y\ , b = A$. Tomando Y=logy y X=logx

EXPONENCIAL:



El modelo es $y = ae^{bx}$, aplicando ln a ambos lados de la igualdad tendremos $lny = lna + bx$, de donde obtenemos A=b, B=lna, de donde concluimos: $a = e^B, y, b = A$.tomando Y=lny y X=x.

SATURACION:



El modelo es $y = \frac{ax}{b+x}$, en este caso invertimos los dos lados de la igualdad $\frac{1}{y} = \frac{b+x}{ax}$, separando términos $\frac{1}{y} = \frac{b}{a}\frac{1}{x} + \frac{1}{a}$, donde $A = \frac{b}{a}$, y , $B = \frac{1}{a}$. De donde obtendremos $a = \frac{1}{B}, y, b = \frac{A}{B}$ en el cual asignamos $X = \frac{1}{x}, y, Y = \frac{1}{y}$.

Ejemplo.: ajustar la siguiente tabla a un modelo potencial y un modelo exponencial

x	1	3	7	10
y	10	52	98	120

Solución:

Modelo potencial:

x	y	X=logx	Y=logy	X^2	X*Y
1	10	0	1	0	0
3	52	0,47712125	1,71600334	0,22764469	0,81874167
7	98	0,84509804	1,99122608	0,7141907	1,68278125
10	120	1	2,07918125	1	2,07918125
	sumas=	2,32221929	6,78641067	1,94183539	4,58070417

Sistema normal:

$$\begin{cases} 1.9418A + 2.32221B = 4.5807 \\ 2.32221A + 4B = 6.78641 \end{cases}$$

De donde A=1.07944, y , B=1.06992.

$$a = 10^{1.06992} = 11.746, \qquad b = 1.0794$$

Quedando el modelo $y = 11.746x^{1.0794}$.

Modelo exponencial:

x=X	y	Y=lny	X^2	X*Y
1	10	2,30258509	1	2,30258509
3	52	3,95124372	9	11,8537312
7	98	4,58496748	49	32,0947724
10	120	4,78749174	100	47,8749174
21	sumas=	15,626288	159	94,126006

Sistema normal.

$$\begin{cases} 159A + 21B = 94.126006 \\ 21A + 4B = 15.626288 \end{cases}$$

De donde: A=0.24795, y, B=2.60478

$$a = e^{2.60478} = 13.5282, \quad b = 0.24795$$

Quedando el modelo $y = 13.5282e^{0.24795\mathrm{x}}$.

BIBLIOGRAFIA

- Chapra. Steven C. Métodos Numéricos Para Ingenieros. Quinta Edición. Editorial Mc Graw Hill.2007
- Burden. Richard L. Análisis Numérico. Séptima Edición. Editorial Thomson. 2002
- George. Thomas. Calculo Una Variable. Undécima edición. Editorial Pearson. 2006
- Domínguez Nieves. Métodos Numéricos Aplicados a la Ingeniería. Primera Edición. Grupo Editorial Patria. 2014.