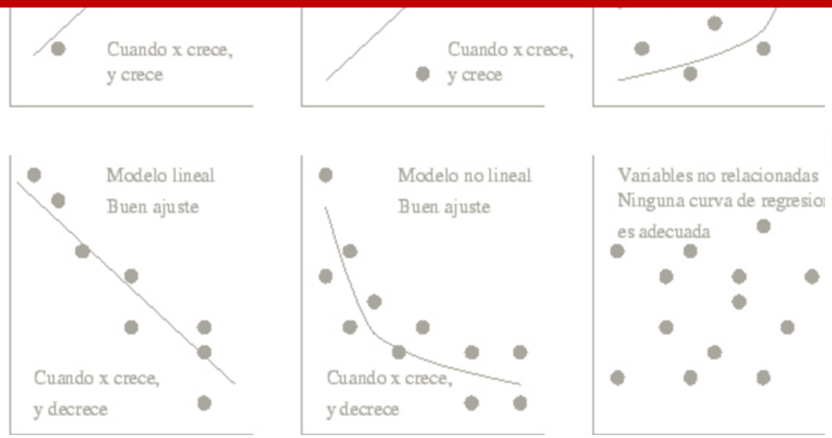
Unidad 5: Series de Fourier

•Introducción a los Modelos de Regresión•



Introducción

El ***ajuste de curvas*** posibilita la descripción de un conjunto de ***datos experimentales*** a partir de una ***ecuación matemática***:

$$\boxed{y = f(x)}$$

donde x es la variable independiente (controlada por el experimentador), f es la función de ajuste e y es la variable dependiente (medida durante el experimento)

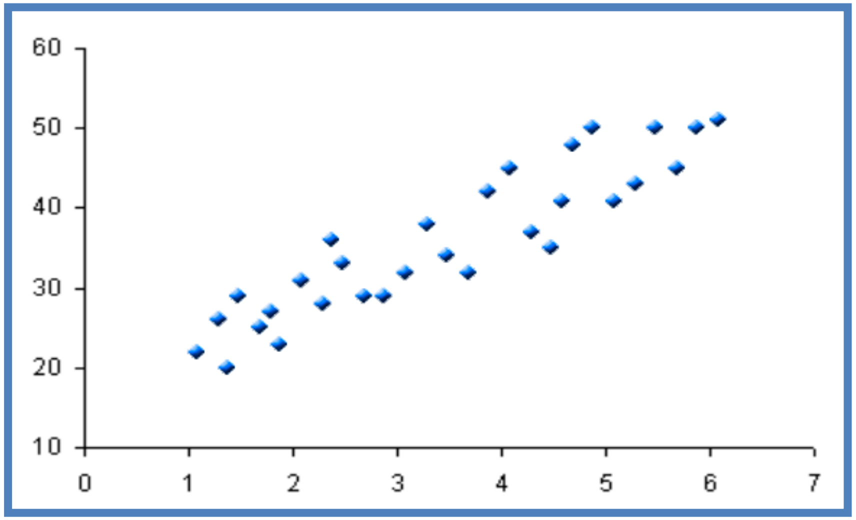
Cuanto mejor es el ajuste, mejor resulta la descripción

Introducción

Formalmente, dicho análisis se denomina ***Regresión***. Se propone una hipótesis en términos de la relación entre las variables (***un modelo***) de modo de hallar los parámetros de una ***Ecuación de regresión***. No obstante, la existencia de dicha ecuación ***no implica una relación causa-efecto*** entre dichas variables (ambas pueden estar influenciadas por una tercera que las afecta simultáneamente) .

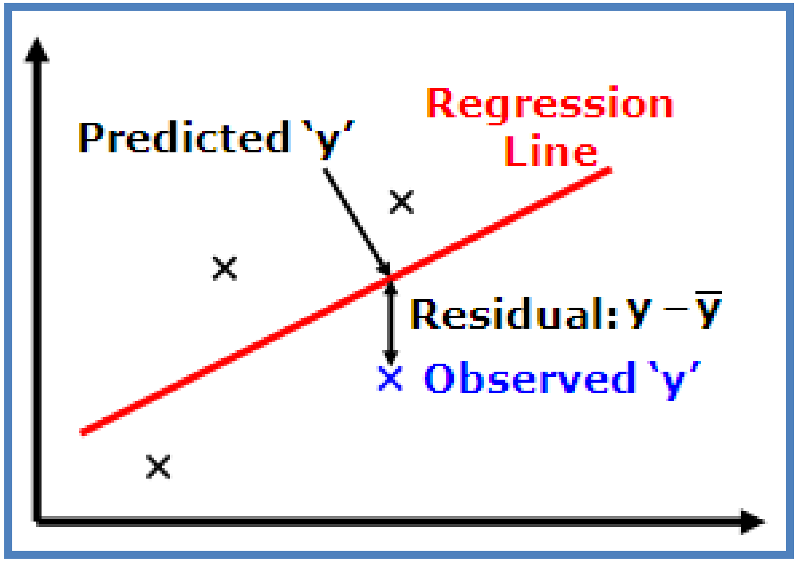
Gráfico de dispersión de los datos medidos en pares ordenados

(valores de y_i correspondientes a cada x_i)



El método más utilizado para la estimación de los parámetros de la ecuación de regresión es el de ***cuadrados mínimos***. Esencialmente, se busca ***minimizar*** la suma de los ***cuadrados de las diferencias (error)*** entre el resultado proporcionado por la ecuación propuesta (y) y los valores medidos (y_i). Dicha diferencias se denominan formalmente ***residuos*** (r)

Si los datos pueden ajustarse a una recta $y=mx+b$, ***el análisis de regresión se torna lineal***



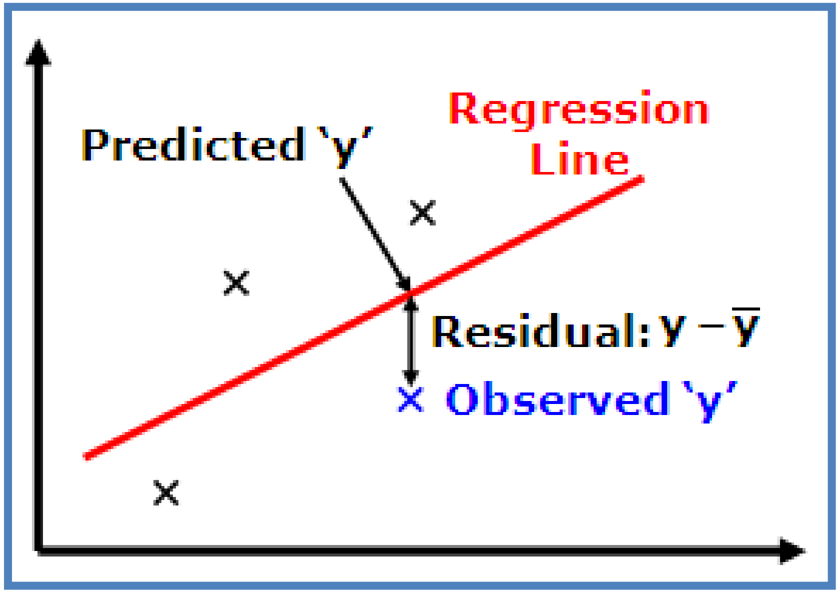
Cuadrados Mínimos

De esta manera, se define como la suma de los residuos (***r***r_i) al cuadrado (***SS***):

$$\boxed{SS_y = \sum_{i=1}^{n} (r_i)^2 = \sum_{i=1}^{n} (y - y_i)^2}$$

El mejor ajuste a la recta $y=mx+b$ se ***obtiene con el menor valor de SS***

(Nótese que al elevar al cuadrado se eliminan los efectos de compensación de valores positivos y negativos en la sumatoria)



El objetivo entonces consiste en la obtención de la mínima ***SS*** que ajusta a la recta de regresión:

$$\boxed{SS = \sum_{i=1}^{n} r_i^2 = \sum_{i=1}^{n} (y - y_i)^2 = \sum_{i=1}^{n} (mx_i + b - y_i)^2 = Mínimo}$$

Derivando entonces ***SS*** parcialmente respecto de los paráme-tros de la recta e igualando a cero, se podrán hallar los valores de ***m*** y ***b*** que cumplen dicha premisa:

$$\boxed{\frac{\partial}{\partial m}\left[\sum_{i=1}^{n} (mx_i + b - y_i)^2\right] = \sum_{i=1}^{n} \frac{\partial}{\partial m}(mx_i + b - y_i)^2 = 0}$$

$$\boxed{\frac{\partial}{\partial b}\left[\sum_{i=1}^{n} (mx_i + b - y_i)^2\right] = \sum_{i=1}^{n} \frac{\partial}{\partial b}(mx_i + b - y_i)^2 = 0}$$

Consecuentemente:

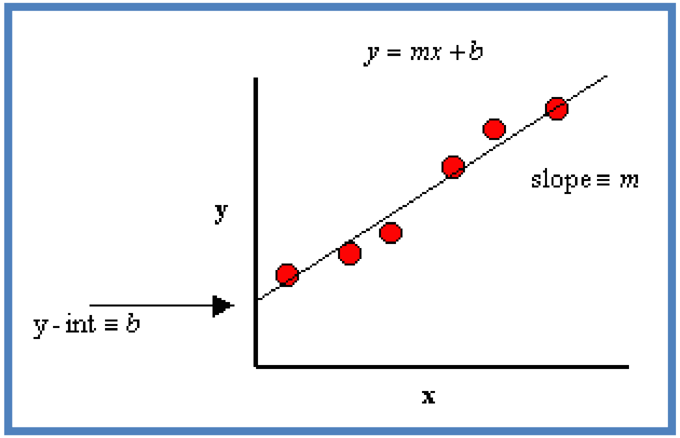
$$\boxed{\sum_{i=1}^{n} 2(mx_i + b - y_i)x_i = 0 \rightarrow b\sum_{i=1}^{n} x_i + m\sum_{i=1}^{n} x_i^{\,2} = \sum_{i=1}^{n} y_i x_i}$$

$$\boxed{\sum_{i=1}^{n} 2(mx_i + b - y_i) = 0 \rightarrow nb + m\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i}$$

Se obtienen dos Ecuaciones con dos incógnitas

Y aplicando la regla de Cramer:

$$\boxed{m = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2}}$$



$$\boxed{b = \frac{\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^{\,2} - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^{\,2} - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n} y_i - m\sum_{i=1}^{n} x_i}{n} = \bar{y} - m\bar{x}}$$

donde $\bar{x}$ e $\bar{y}$ son los ***valores medios (esperados)*** de las observaciones $\boldsymbol{x}$ e $\boldsymbol{y}$

Coeficiente de Correlación

El ***coeficiente de determinación*** (***R^2***) es una medida entre ***0*** y ***1*** que expresa la ***proporción (porcentaje) de la varianza de la variable dependiente explicada por la variable independiente***:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - y\right)^2}{\sum_{i=1}^{n} \left(y_i - \bar{y}_i\right)^2} = 1 - \frac{Error}{Dist\ valores\ a\ la\ media}$$

$$R^2 = 0 \rightarrow x\ no\ ayuda\ a\ predecir\ \ y$$

$$R^2 \rightarrow 1 \rightarrow el\ ajuste\ es\ cada\ vez\ mejor$$

Coeficiente de Correlación

La raíz cuadrada del coeficiente de determinación se denomina ***coeficiente de correlación de Pearson*** (***r***):

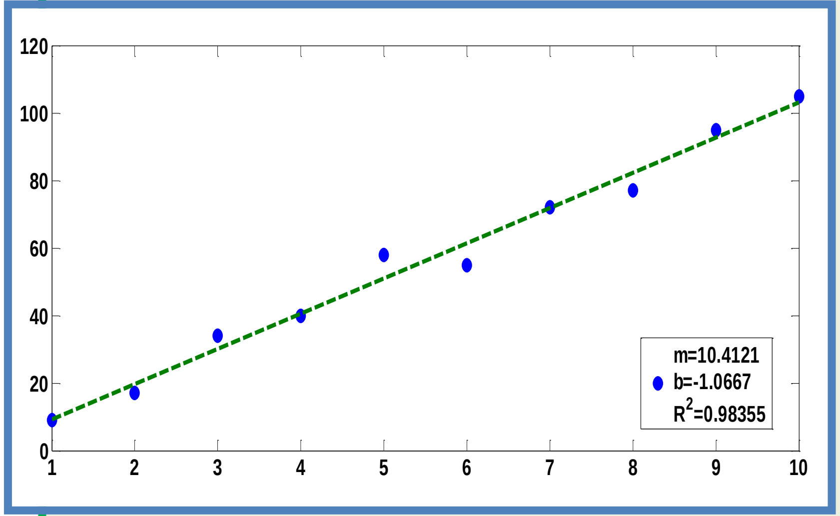
$$\boxed{\pm r = \sqrt{R^2}}$$

En el ***ajuste lineal***, el signo de ***r*** estará determinado por el signo de la pendiente de la recta de regresión. Formalmente puede decirse que ***r*** mide el ***"grado de asociación"*** entre las variables. Si $-1<r<0$, las mismas se mueven en direcciones opuestas (si una aumenta, la otra disminuye y viceversa).

NOTA: Si bien pueden obtenerse valores elevados de r, el carácter explicativo del modelo lo proporciona R^2

EN MATLAB…

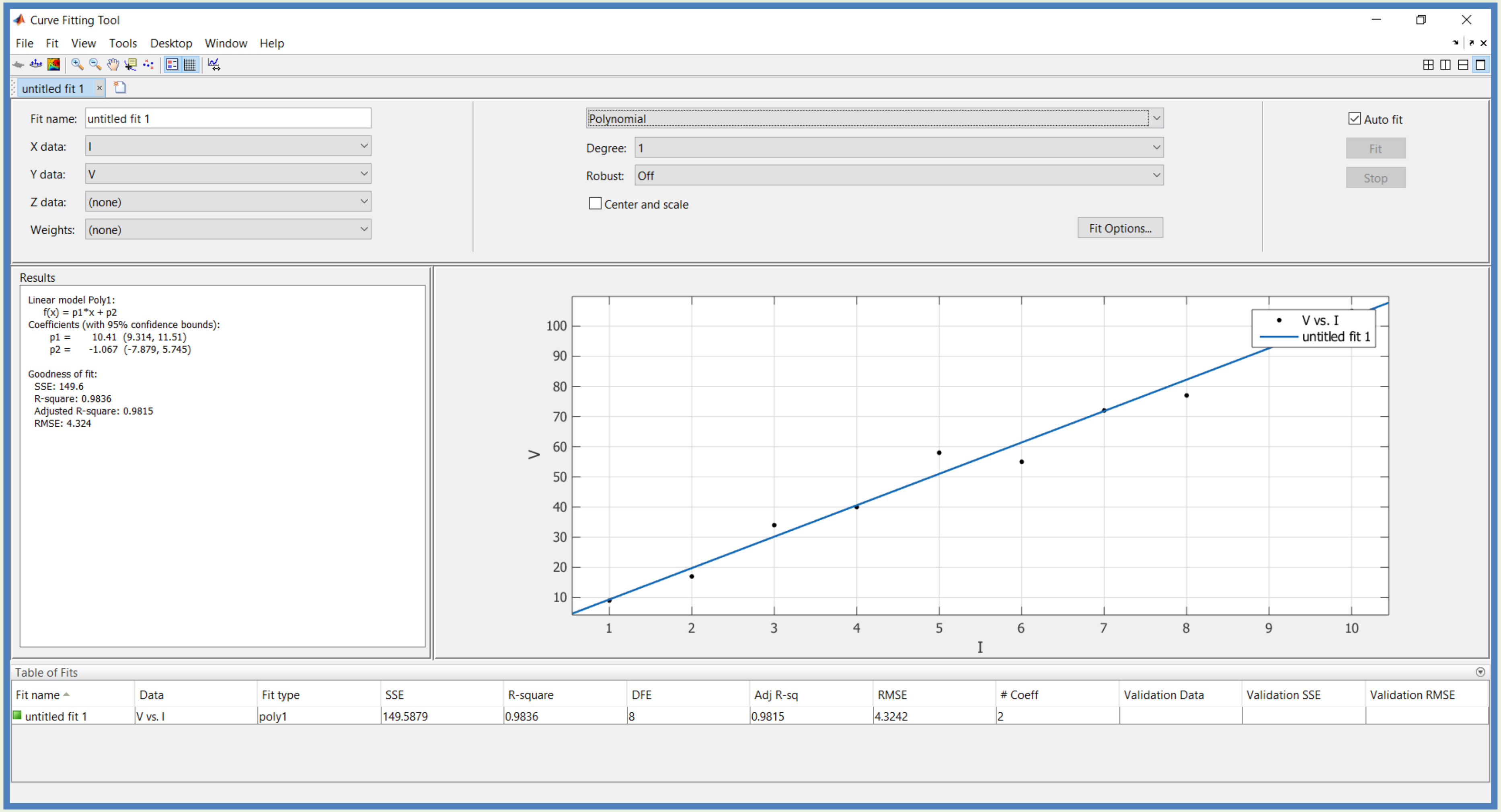
```
%ANÁLISIS DE REGRESIÓN APLICADA A VALORES DE
%TENSIÓN Y CORRIENTE

%Pares ordenados de tensión (V) y corriente
%(I) medidos
I=[1 2 3 4 5 6 7 8 9 10];
V=[9 22 29 45 56 60 72 78 94 105];
%Obtención de los coeficientes de un
%polinomio de primer orden (recta)
Coefs=polyfit(I,V,1);
%Obtención de los valores de V conforme la
%recta de ajuste
Vaj=polyval(Coefs,I);
%Coeficiente de correlación (R^2)
R2=(corrcoef(V,Vaj)).^2;
%Visualización
plot(I,V,'o',I,Vaj);
legend(['m=' num2str(Coefs(1)) char(10)...
        'b=' num2str(Coefs(2)) char(10)...
        'Ajuste=' num2str(R2(1,2))]);
```



Unidad 5: Series de Fourier

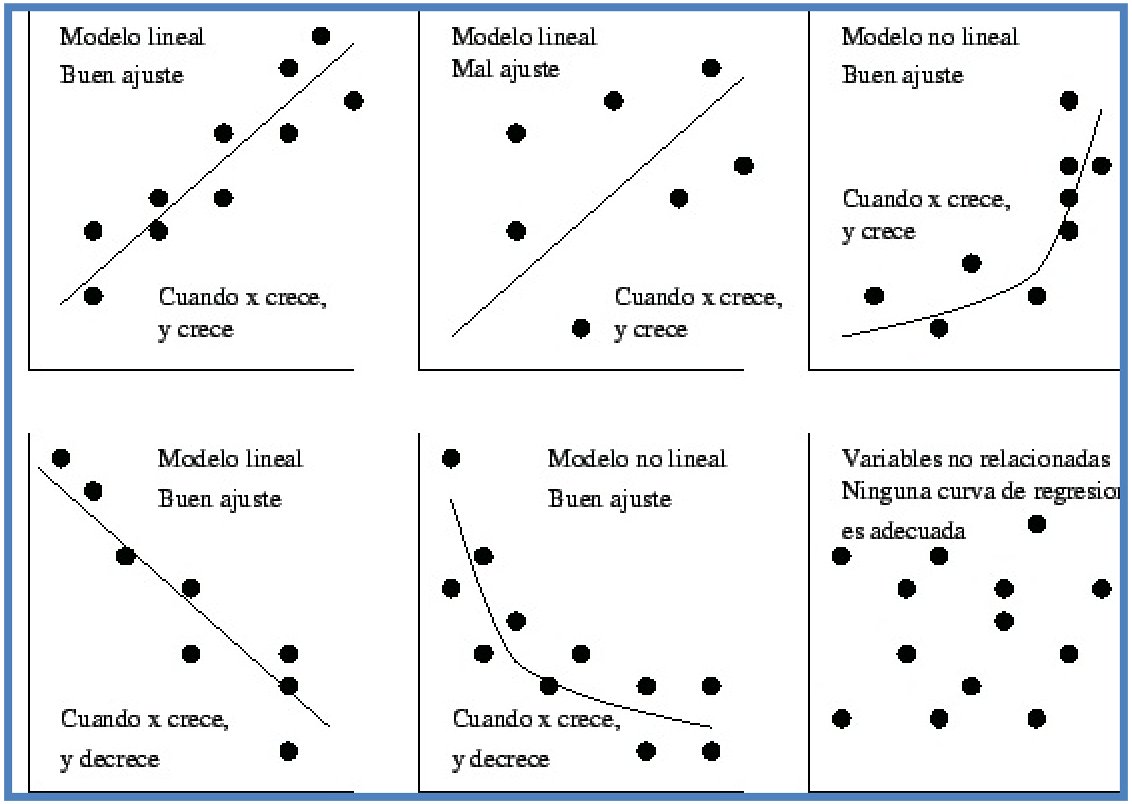
EN MATLAB…(CFTOOL)

Análisis de Señales y Sistemas R2041



Unidad 5: Series de Fourier
Coeficiente de Correlación

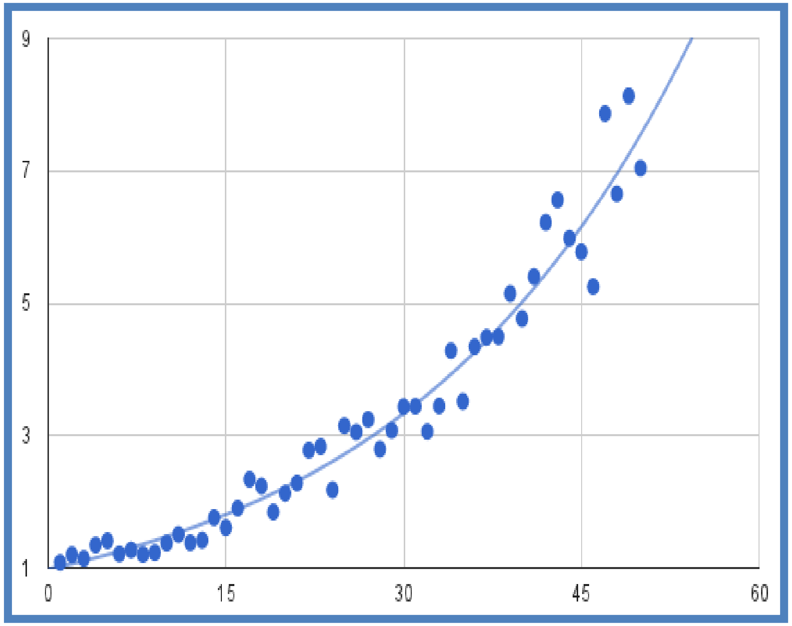
¿Cuán bien ajusta la ecuación de regresión elegida a los datos observados?¿Y si el ajuste no es lineal?



Ajuste no Lineal

En los casos donde no es posible proponer una recta de regresión lineal, se utiliza una ***función no lineal*** para efectuar el ajuste. Una primera opción es incorporar términos de mayor orden al polinomio de primer grado (***regresión polinomial***).

Si bien pueden obtenerse mejores ajustes, el problema redunda en que ***los polinomios no manifiestan leyes físicas,*** como las ***exponenciales*** o los ***logaritmos***. En este último caso las ***curvas se linealizan*** y se aplica la misma metodología.



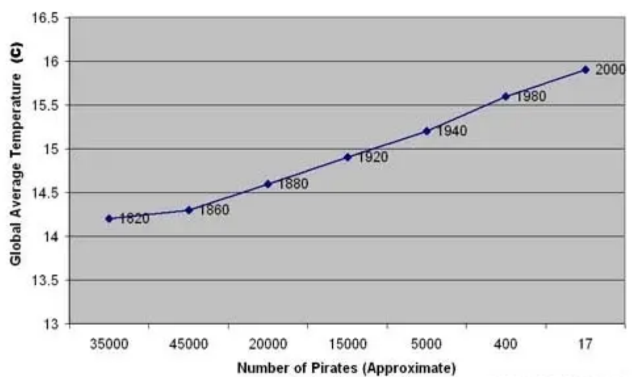
Ajuste No Lineal

Para otro tipo de curvas, pueden aplicarse procedimientos de ajuste por ***cuadrados mínimos de forma iterativa***. Esencialmente, se proponen valores iniciales y luego se busca el mínimo de ***SS*** variando levemente los parámetros. Entre los distintos métodos pueden mencionarse:

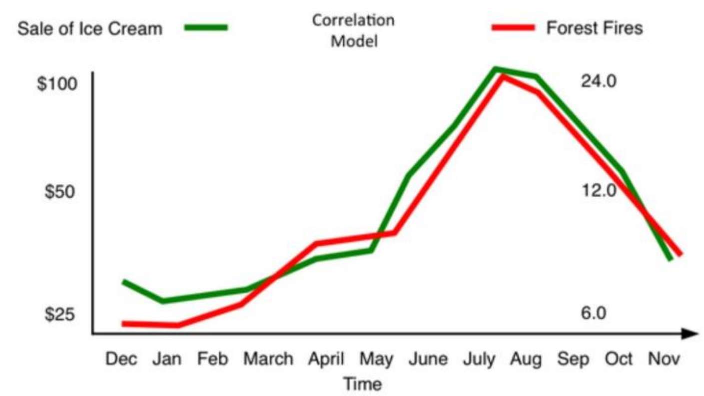
- Gauss-Newton
- Marquardt – Levenberg
- Nelder-Mead
- Descenso de Gradiente

Correlaciones Espurias (bizarras)

Temperatura Promedio Global vs. Número de Piratas



Venta de Helados vs Incendios Forestales



Número de obtención de doctorados en ingeniería civil vs el Consumo de queso mozarella

Facebook cancela la disminución en los niveles de colesterol iniciada por el nacimiento de Justin Bieber