Ajuste de Curvas

PhD. Alejandro Paredes

Motivación:

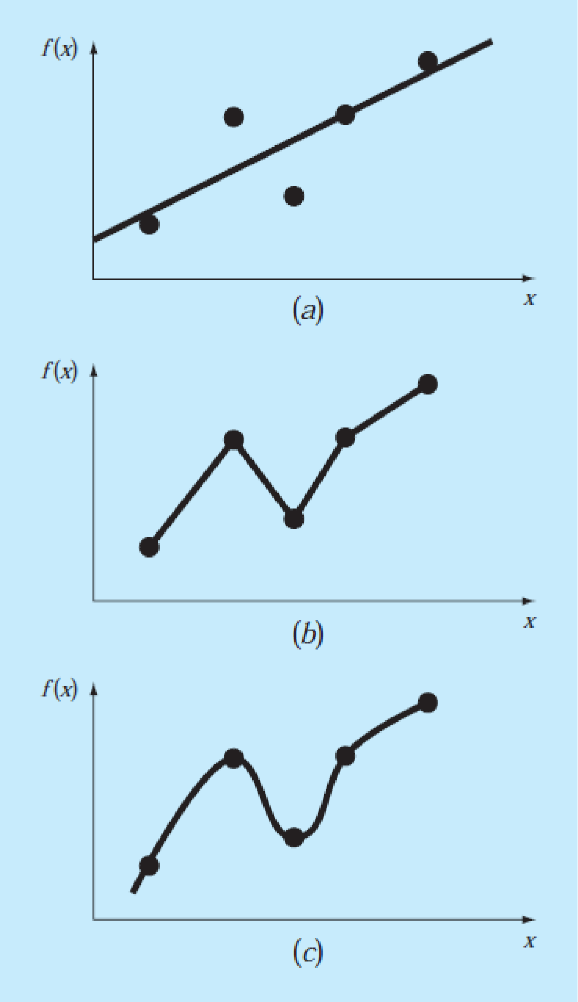
Se desea interpolar o extrapolar información a partir de datos experimentales

Se tienen dos opciones:

- Buscar una función (compleja) que pase por todos los puntos.
- Buscar una función simple que pase solo por algunos .

Los datos contienen mucha incertidumbre:
Curva pasa solo por algunos puntos y que expresa la tendencia.
Regression por minimos cuadrados.

Si los datos contienen poca inceridumbre:
Curva que pase exactamente por los puntos.
Interpolación polinomial.



Motivación:
Aplicaciones del ajsute de curvas

- Test de hypotesis :
  Se desea es comparar un modelo matemático con los datos experiemntales.
  Se propone un modelo y se buscan los parámetros del modelo.

- Adicionalmente el ajuste de curvas es importante para la integración numérica, solución aproxiamda de ecuaciones diferenciales . Tambien se puede utilizar para aproximar funciones complejas a funciones simples.

Estadística inferacional

Mediciones del coeficiente de expansión térmica del acero (10^{-6} F^{-1})

6.495	6.595	6.615	6.635	6.485	6.555
6.665	6.505	6.435	6.625	6.715	6.655
6.755	6.625	6.715	6.575	6.655	6.605
6.565	6.515	6.555	6.395	6.775	6.685

Media= 6.6
S_y= 0.0971
c.v.=1.47%

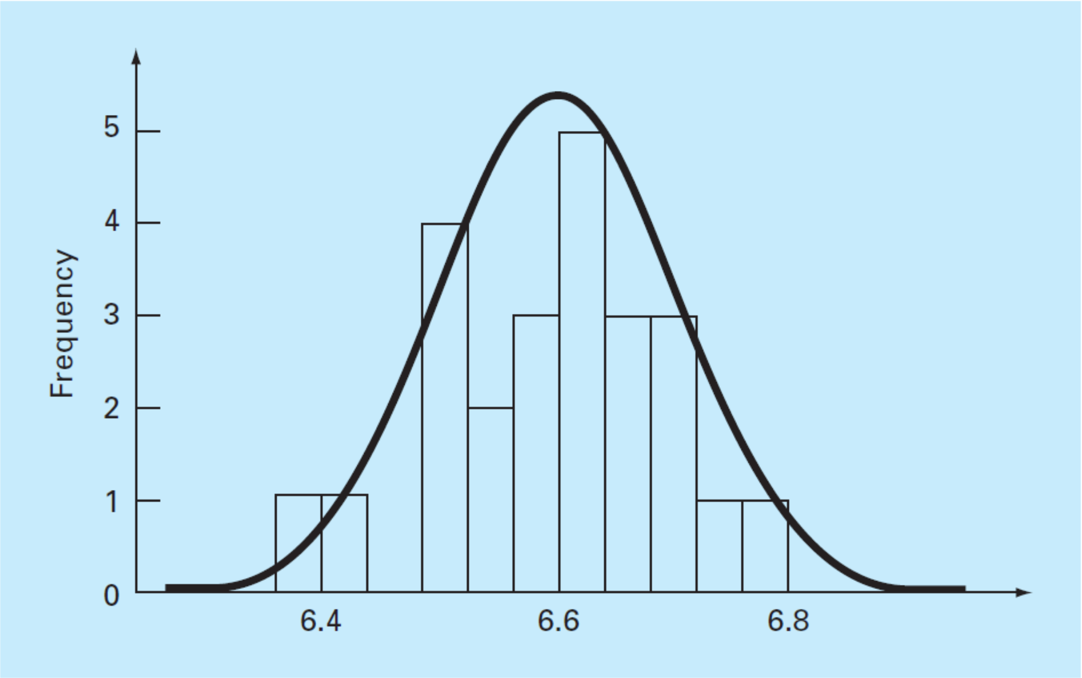
Información adicional se tiene con la media aritmética del conjunto, el grado de esparcimiento del conjunto llamado desviación estandard S_y (variancia) y el coeficiente de varaición CV (error porcentual relativo)

$$\bar{y} = \frac{\Sigma y_i}{n} \qquad s_y = \sqrt{\frac{S_t}{n-1}} \qquad S_t = \Sigma (y_i - \bar{y})^2 \qquad s_y^2 = \frac{\Sigma y_i^2 - (\Sigma y_i)^2 / n}{n-1} \qquad \text{c.v.} = \frac{s_y}{\bar{y}} 100\%$$

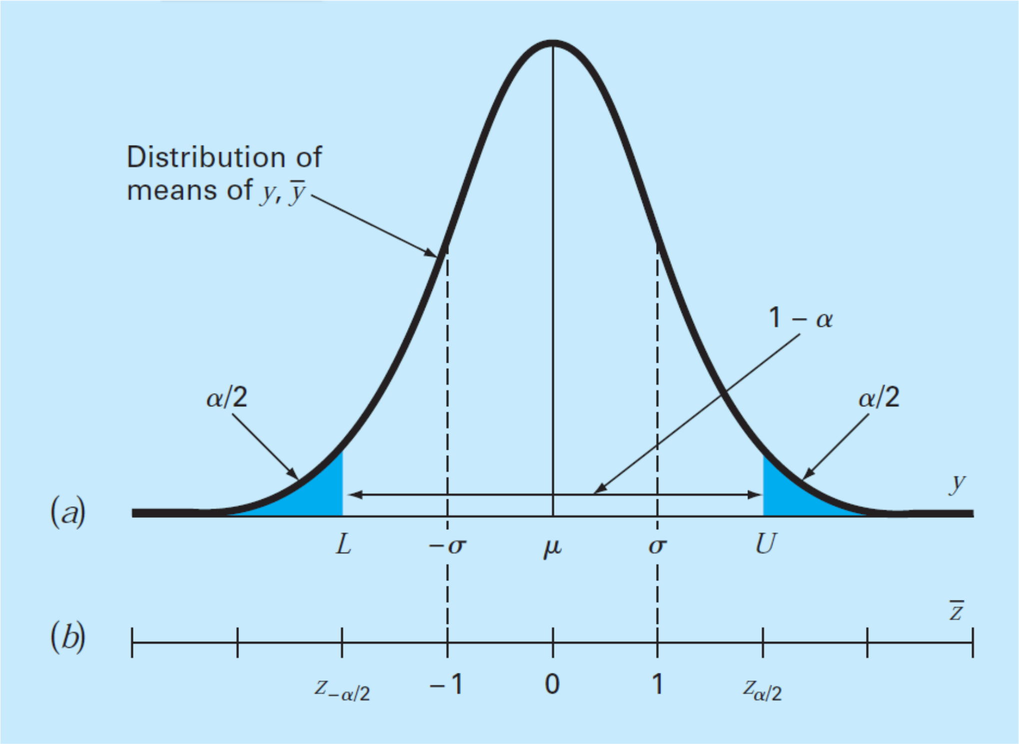
Estadística inferacional

- La idea es poder inferir(estimar) propiedades de la poblacion a partir de una muestra de la población.
- Mientras más datos se tomen el histograma se acerca a una distribución normal.

Distribución Gaussiana



Estadística inferacional



Media de la muestra : $\bar{y}$
Desviación standard de la muestra : S_y

Media de la población : μ
Desviación estandar de la población : σ

$$t = \frac{\bar{y} - \mu}{s_y/\sqrt{n}}$$

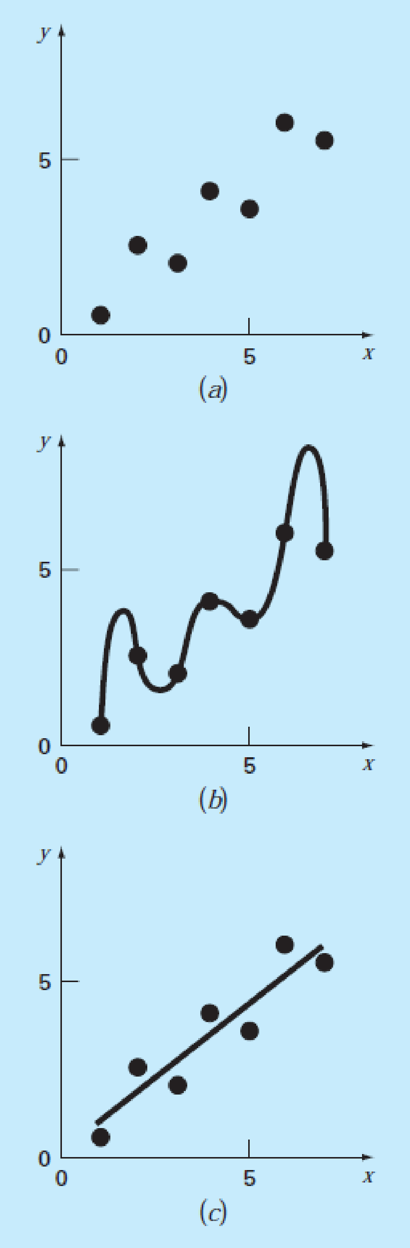
$$L = \bar{y} - \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1} \qquad U = \bar{y} + \frac{s_y}{\sqrt{n}} t_{\alpha/2,n-1}$$

Regresión lineal

Considere que se tiene un conjunto de puntos con alta incertidumbre .

El caso mas simple es ajustar los datos $(x_1, y_1)(x_2, y_2) \ldots (x_n, y_n)$ a una línea recta de la forma

$$y(x) = a_0 + a_1 x + e; \quad e = \text{error o residuo}$$



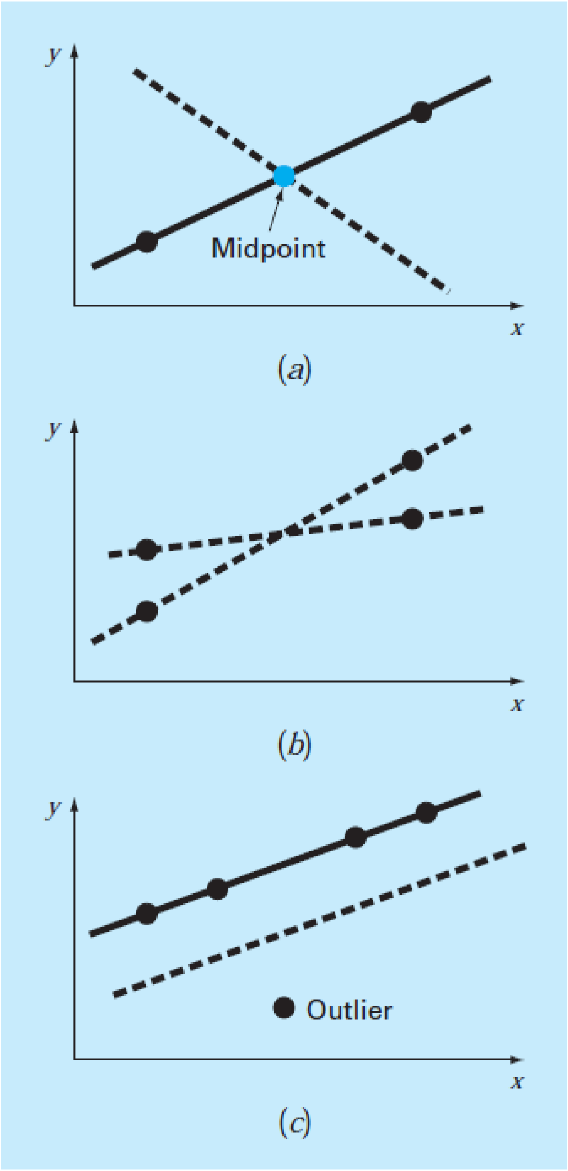
Regresión lineal

Opciones para encontrar los coeficientes:

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)$$

$$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - a_0 - a_1 x_i|$$

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{measured}} - y_{i,\text{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$



Regresion lineal por mínimos cuadrados

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{measured}} - y_{i,\text{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum (y_i - a_0 - a_1 x_i)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum [(y_i - a_0 - a_1 x_i) x_i]$$

$$0 = \sum y_i - \sum a_0 - \sum a_1 x_i$$

$$0 = \sum y_i x_i - \sum a_0 x_i - \sum a_1 x_i^2$$

$$n a_0 + \left(\sum x_i\right) a_1 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 = \sum x_i y_i$$

$$\boxed{a_1 = \frac{n\Sigma x_i y_i - \Sigma x_i \Sigma y_i}{n \Sigma x_i^2 - (\Sigma x_i)^2}}$$

$$\boxed{a_0 = \bar{y} - a_1 \bar{x}}$$

Regresion lineal por mínimos cuadrados

Ejemplo

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1 x_i)^2$
1	0.5	8.5765	0.1687
2	2.5	0.8622	0.5625
3	2.0	2.0408	0.3473
4	4.0	0.3265	0.3265
5	3.5	0.0051	0.5896
6	6.0	6.6122	0.7972
7	5.5	4.2908	0.1993
Σ	24.0	22.7143	2.9911

$$y = 0.07142857 + 0.8392857x$$

$$n = 7 \qquad \sum x_i y_i = 119.5 \qquad \sum x_i^2 = 140$$

$$\sum x_i = 28 \qquad \bar{x} = \frac{28}{7} = 4$$

$$\sum y_i = 24 \qquad \bar{y} = \frac{24}{7} = 3.428571$$

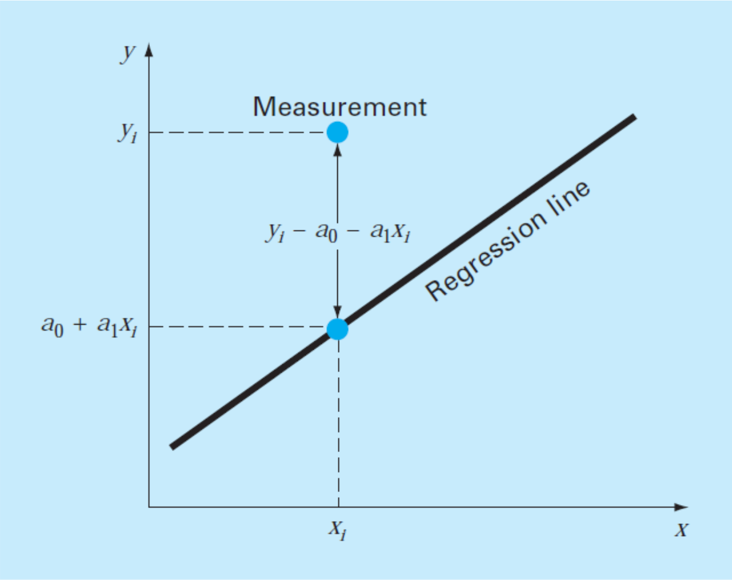
$$a_1 = \frac{7(119.5) - 28(24)}{7(140) - (28)^2} = 0.8392857$$

$$a_0 = 3.428571 - 0.8392857(4) = 0.07142857$$

Cuantificación del error

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,\text{measured}} - y_{i,\text{model}})^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Se puede calcular



Desviacion estandar de estimacion del error debido al ajuste

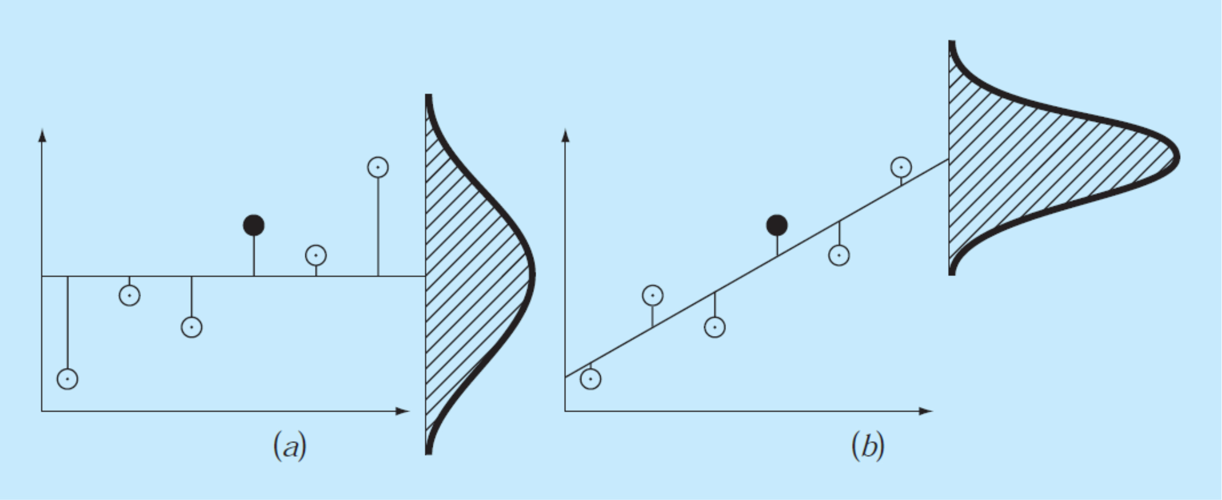
$$s_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

Error standard dela estimación

Desviacion estandar de estadsitica

$$s_y = \sqrt{\frac{S_t}{n-1}}$$

$$S_t = \Sigma(y_i - \bar{y})^2$$



Cuantificación del error

Coeficiente de determinación

$$r^2 = \frac{S_t - S_r}{S_t}$$

- r=1 implica S_r=0 y el ajuste pasa por todos los puntos
- r=0 implica S_t=S_r el error del ajsute es igual al error estadistico y no se ha ganado nada.

Coeficiente de correlación

$$r = \frac{n\Sigma x_i y_i - (\Sigma x_i)(\Sigma y_i)}{\sqrt{n\Sigma x_i^2 - (\Sigma x_i)^2}\sqrt{n\Sigma y_i^2 - (\Sigma y_i)^2}}$$

$$s_t = \sqrt{\frac{22.7143}{7-1}} = 1.9457 \qquad s_{y/x} = \sqrt{\frac{2.9911}{7-2}} = 0.7735$$

$$r^2 = \frac{22.7143 - 2.9911}{22.7143} = 0.868 \qquad r = \sqrt{0.868} = 0.932$$

Algoritmo

```
SUB Regress(x, y, n, a1, a0, syx, r2)

  sumx = 0: sumxy = 0: st = 0
  sumy = 0: sumx2 = 0: sr = 0
  DOFOR i = 1, n
    sumx = sumx + xi
    sumy = sumy + yi
    sumxy = sumxy + xi*yi
    sumx2 = sumx2 + xi*xi
  END DO
  xm = sumx/n
  ym = sumy/n
  a1 = (n*sumxy − sumx*sumy)/(n*sumx2 − sumx*sumx)
  a0 = ym − a1*xm
  DOFOR i = 1, n
    st = st + (yi − ym)^2
    sr = sr + (yi − a1*xi − a0)^2
  END DO
  syx = (sr/(n − 2))^0.5
  r2 = (st − sr)/st

END Regress
```

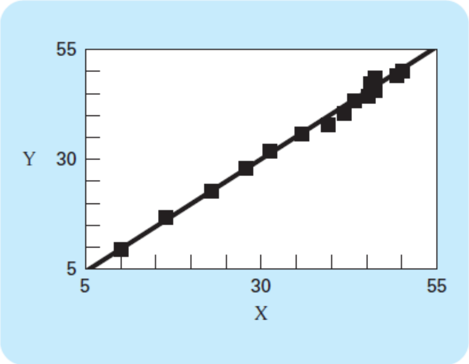
Ejemplo: Caida de un paracaidista

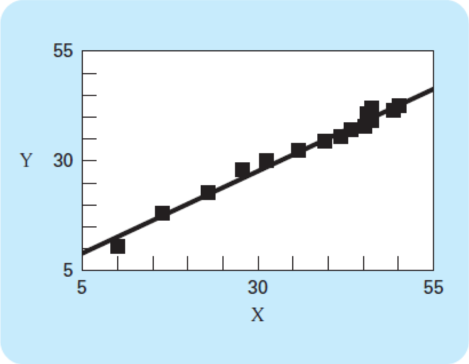
$$v(t) = \frac{gm}{c}\left(1 - e^{(-c/m)t}\right)$$

$$v(t) = \frac{gm}{c}\left(\frac{t}{3.75 + t}\right)$$

$$v_{\text{model}} = -0.859 + 1.032 v_{\text{measure}}$$

Time, s	Measured v, m/s (a)	Model-calculated v, m/s (b)	Model-calculated v, m/s (c)
1	10.00	8.953	11.240
2	16.30	16.405	18.570
3	23.00	22.607	23.729
4	27.50	27.769	27.556
5	31.00	32.065	30.509
6	35.60	35.641	32.855
7	39.00	38.617	34.766
8	41.50	41.095	36.351
9	42.90	43.156	37.687
10	45.00	44.872	38.829
11	46.00	46.301	39.816
12	45.50	47.490	40.678
13	46.00	48.479	41.437
14	49.00	49.303	42.110
15	50.00	49.988	42.712





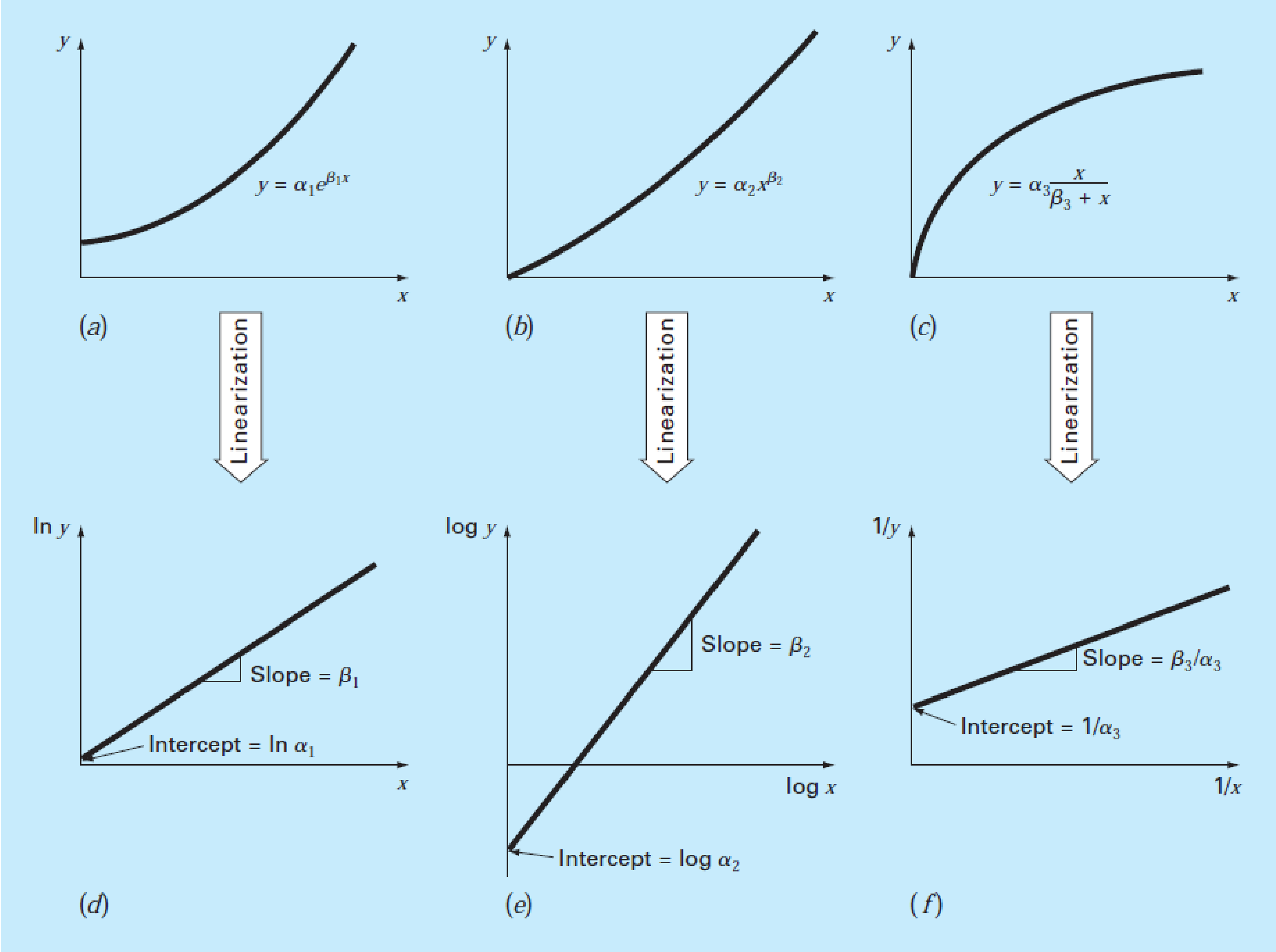
$c = 12.5\ kg/s$

r>0.99

r>0.99

$$v_{\text{model}} = 5.776 + 0.752 v_{\text{measure}}$$

Linealización de dependencia no lineales



- Solo algunas relaciones se pueden linealizar.
- Una vez que se han linealizado, se puede utilizar la regresión lineal por mínimos cuadrados.

$$y = \alpha_1 e^{\beta_1 x} \qquad \ln(y) = \ln(\alpha_1) + \beta_1 x$$

$$y = \alpha_2 x^{\beta_1} \qquad \ln(y) = \ln(\alpha_2) + \beta_2 \ln(x)$$

$$y = \alpha_3 \frac{x}{\beta_3 + x} \qquad \frac{1}{y} = \frac{\beta_3}{\alpha_3}\frac{1}{x} + \frac{1}{\alpha_3}$$

Regresión polinomial

Deseamos ajustar un conjuntos de datos a un polinomio de orden 2

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_m x^m + e$$

$$y = a_0 + a_1 x + a_2 x^2 + e$$

$$S_r = \sum_{i=1}^{n} \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)^2$$

$$(n)a_0 + \left(\sum x_i\right) a_1 + \left(\sum x_i^2\right) a_2 = \sum y_i$$

$$\left(\sum x_i\right) a_0 + \left(\sum x_i^2\right) a_1 + \left(\sum x_i^3\right) a_2 = \sum x_i y_i$$

$$\left(\sum x_i^2\right) a_0 + \left(\sum x_i^3\right) a_1 + \left(\sum x_i^4\right) a_2 = \sum x_i^2 y_i$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum x_i \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum x_i^2 \left(y_i - a_0 - a_1 x_i - a_2 x_i^2\right)$$

$$s_{y/x} = \sqrt{\frac{S_r}{n - (m+1)}} \qquad m = 2$$

Algoritmo para regresión polinomial

La matriz simétrica a_{ij} es la matriz aumentada
Correspondiente al sistema lineal

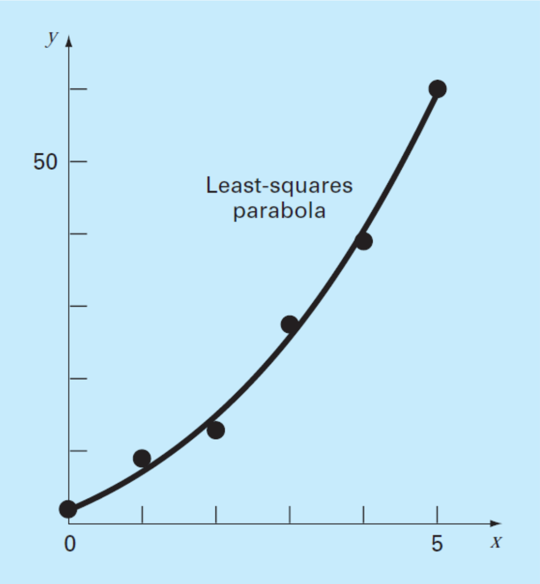
$$(n)a_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 = \sum y_i$$
$$\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 = \sum x_i y_i$$
$$\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 = \sum x_i^2 y_i$$

```
DOFOR i = 1, order + 1
  DOFOR j = 1, i
    k = i + j − 2
    sum = 0
    DOFOR ℓ = 1, n
      sum = sum + x_ℓ^k
    END DO
    a_i,j = sum
    a_j,i = sum
  END DO
  sum = 0
  DOFOR ℓ = 1, n
    sum = sum + y_ℓ · x_ℓ^(i−1)
  END DO
  a_i,order+2 = sum
END DO
```

Regresión polinomial

Ejemplo

x_i	y_i	$(y_i - \bar{y})^2$	$(y_i - a_0 - a_1x_i - a_2x_i^2)^2$
0	2.1	544.44	0.14332
1	7.7	314.47	1.00286
2	13.6	140.03	1.08158
3	27.2	3.12	0.80491
4	40.9	239.22	0.61951
5	61.1	1272.11	0.09439
Σ	152.6	2513.39	3.74657



$m = 2$ $\quad \sum x_i = 15$ $\quad \sum x_i^4 = 979$

$n = 6$ $\quad \sum y_i = 152.6$ $\quad \sum x_i y_i = 585.6$

$\bar{x} = 2.5$ $\quad \sum x_i^2 = 55$ $\quad \sum x_i^2 y_i = 2488.8$

$\bar{y} = 25.433$ $\quad \sum x_i^3 = 225$

$$\begin{bmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} 152.6 \\ 585.6 \\ 2488.8 \end{Bmatrix}$$

$a_0 = 2.47857$ $\quad a_1 = 2.35929$

$a_2 = 1.86071$

$$y = 2.47857 + 2.35929x + 1.86071x^2 \qquad s_{y/x} = \sqrt{\frac{3.74657}{6-3}} = 1.12 \qquad r^2 = 0.99851$$

Regresión no lineal

- Se minimiza el cuadrado de los residuos entre los datos y las ecuaciones no lineales
- La ecuación no lineal se linealiza con una expansión de Taylor

$$y_i = f(x_i; a_0, a_1, \ldots, a_m) + e_i$$

Proceso iterativo

$$f(x_i)_{j+1} = f(x_i)_j + \frac{\partial f(x_i)_j}{\partial a_0} \Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1} \Delta a_1$$

$$y_i - f(x_i)_j = \frac{\partial f(x_i)_j}{\partial a_0} \Delta a_0 + \frac{\partial f(x_i)_j}{\partial a_1} \Delta a_1 + e_i$$

$$\{D\} = [Z_j]\{\Delta A\} + \{E\}$$

- y_i: valor medido de la variable dependiente
- $f(x_i;a_0,\ldots,a_n)$: ecuación que es función de la variable independiente x_i y función no-lineal de los parámetros $a_1,a_2,\ldots,a_n$ e_i = error aleatorio
- $f(x_i)$: Aproximación de Taylor de la función no-lineal al rededor de los parámetros y truncada a la primera derivada.
- j=estimación inicial
- j+1 = predicción
- $\Delta a_0 = a_{0j+1} - a_{0j}$

Regresión no-lineal

$$[Z_j] = \begin{bmatrix} \partial f_1/\partial a_0 & \partial f_1/\partial a_1 \\ \partial f_2/\partial a_0 & \partial f_2/\partial a_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \partial f_n/\partial a_0 & \partial f_n/\partial a_1 \end{bmatrix}$$

- [Zj] : Matriz de derivadas parciales evaluadas en la j-esima estimación.
- n: numero de datos
- $\frac{\partial f_i}{\partial a_k}$ = derivada parcial de la función f con respecto al k-esimo parámetro evaluada en el i-esimo dato

$$\{D\} = \begin{Bmatrix} y_1 - f(x_1) \\ y_2 - f(x_2) \\ \cdot \\ \cdot \\ \cdot \\ y_n - f(x_n) \end{Bmatrix} \qquad \{\Delta A\} = \begin{Bmatrix} \Delta a_0 \\ \Delta a_1 \\ \cdot \\ \cdot \\ \cdot \\ \Delta a_m \end{Bmatrix}$$

Condición de minimos cuadrados

$$\left[[Z_j]^T[Z_j]\right]\{\Delta A\} = \left\{[Z_j]^T\{D\}\right\}$$

$$a_{0,j+1} = a_{0,j} + \Delta a_0$$

$$a_{1,j+1} = a_{1,j} + \Delta a_1$$

Ejercicio 1

A continuación se muestra un modelo para el conjunto de datos mostrado

$$y = \left(\frac{a+\sqrt{x}}{b\sqrt{x}}\right)^2$$

x	0.5	1	2	3	4
y	10.4	5.8	3.3	2.4	2

- Linearizar la ecuación del modelo.
- Utilizar una regresión lineal para determinar a y *b*

Ejercicio 2

A continuación se muestra un modelo para el conjunto de datos mostrado

$$y = \alpha_4 x e^{\beta_4 x}$$

x	0.1	0.2	0.4	0.6	0.9	1.3	1.5	1.7	1.8
y	0.75	1.25	1.45	1.25	0.85	0.55	0.35	0.28	0.18

- Linearizar la ecuación del modelo y utilizar una regresión lineal para determinar α_4 y β_4.
- Utilizar una regresión no lineal para determinar α_4 y β_4.