1. Fitting a curve to a set of data points
2. Fitting a curve for the Extrapolation of the data points (Trend Analysis).
3. Fitting a curve to a set of points to obtain a continuous smooth curve

Arithmetic mean
$$\bar{y} = \frac{\sum y_i}{n}$$

Standard Deviation
$$s_y = \sqrt{\frac{S_t}{n-1}}$$
where
$$S_t = \sum \left(y_i - \bar{y} \right)^2$$

S_t is the sum of square of the residuals

$$s_y^2 = \frac{S_t}{n-1}$$

Variance:
$$s_y^2 = \frac{\sum y_i^2 - \left(\sum y_i \right)^2 / n}{n-1}$$

Coefficient of variation (c.v.)
$$c.v. = \frac{s_y}{\bar{y}} 100\%$$

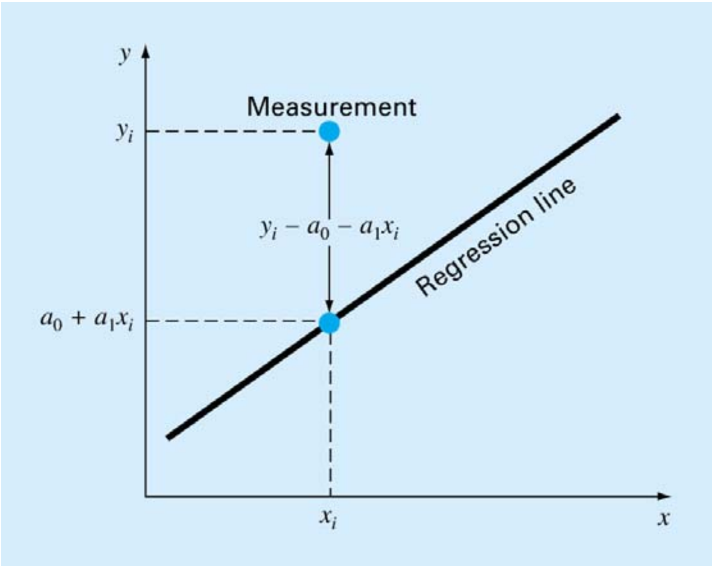
A set of paired observations: $(x_1, y_1);(x_2, y_2);(x_3, y_3);......(x_n, y_n)$

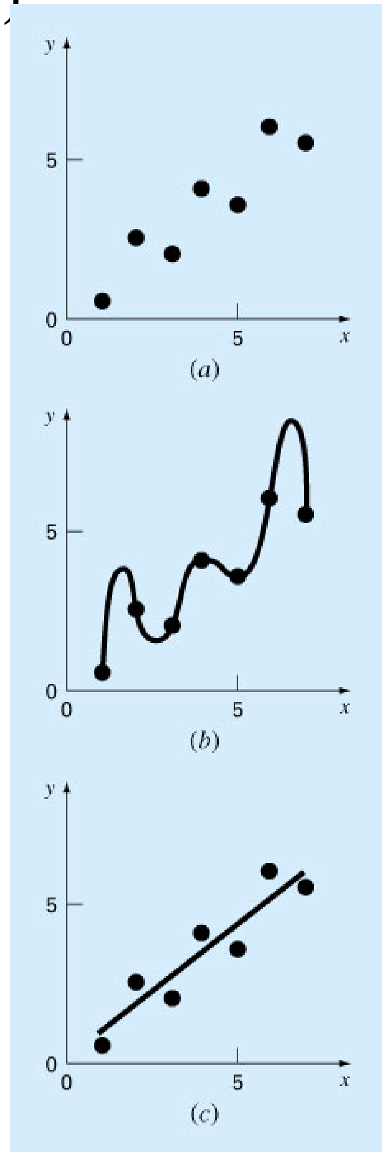
$$y = a_0 + a_1 x + e$$

a_0 and a_1 are coefficients, e is the error or residual

$$e = y - a_0 - a_1 x$$

Difference between the true value and the approximate value obtained from the straight line





"Best Fit"

Minimize:

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

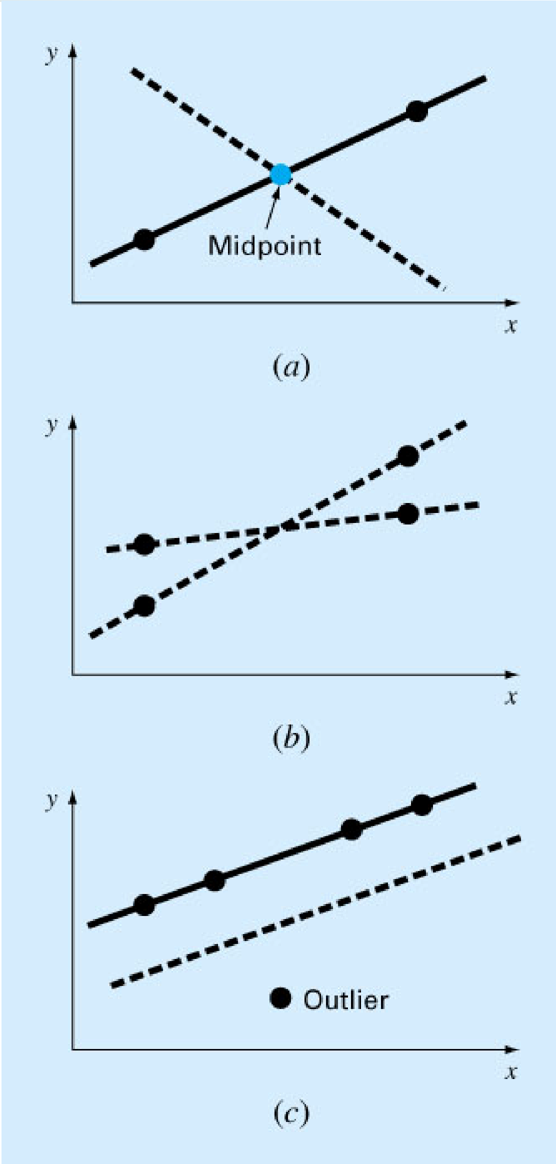
Sum of absolute value of errors $$\sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |(y_i - a_0 - a_1 x_i)|$$

"Minimax"-Minimization of the maximum

Square of the Residuals

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

1. Each x has a certain value. It is not random and known without error
2. The y values are independent random variables and all have the same variance.
3. The y values for a given x must be normally distributed.



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

Necessary Condition:

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_i) x_i\right] = 0$$

Or:

$$n a_0 + \left(\sum x_i\right) a_1 = \sum_{i=1}^{n} y_i$$

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

$$a_1 = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$$

$$a_0 = \overline{y} - a_1 \overline{x}$$

$$\overline{y} = \sum y_i, \qquad \overline{x} = \sum x_i$$

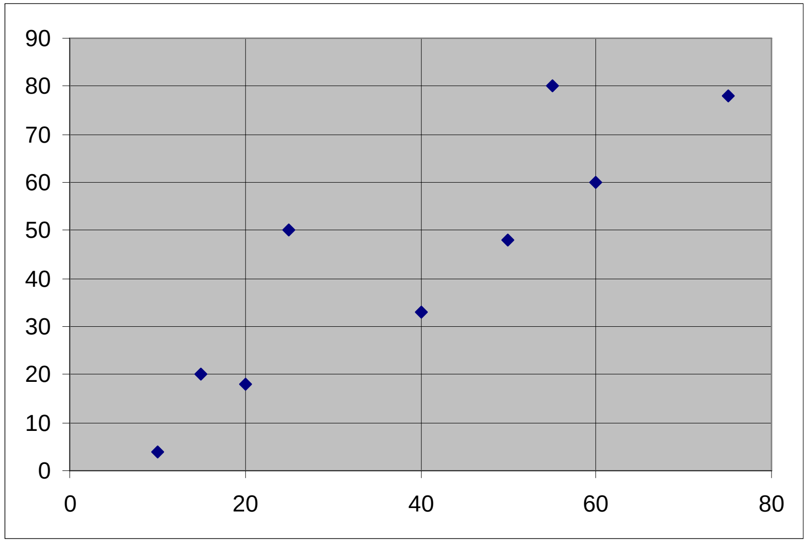
Time	Tensile Strength
10	4
15	20
20	18
25	50
40	33
50	48
55	80
60	60
75	78

$$n\bar{y} = \sum y_i = 391$$

$$n\bar{x} = \sum x_i = 350$$

$$\sum x_i y_i = 19520$$

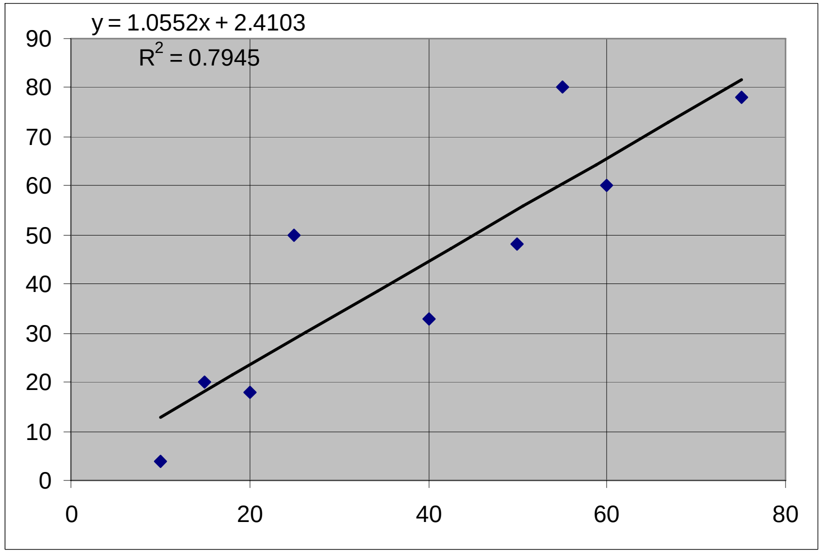
$$\sum x_i^2 = 17700$$

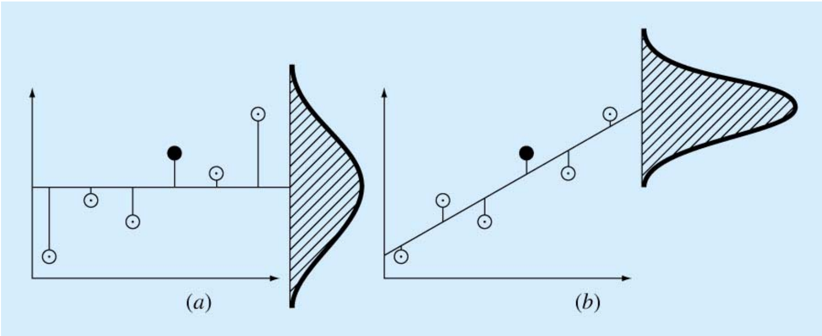


$$a_1 = 2.4103$$

$$a_0 = 1.0552$$

$$y = 1.0552 + 2.4103x$$

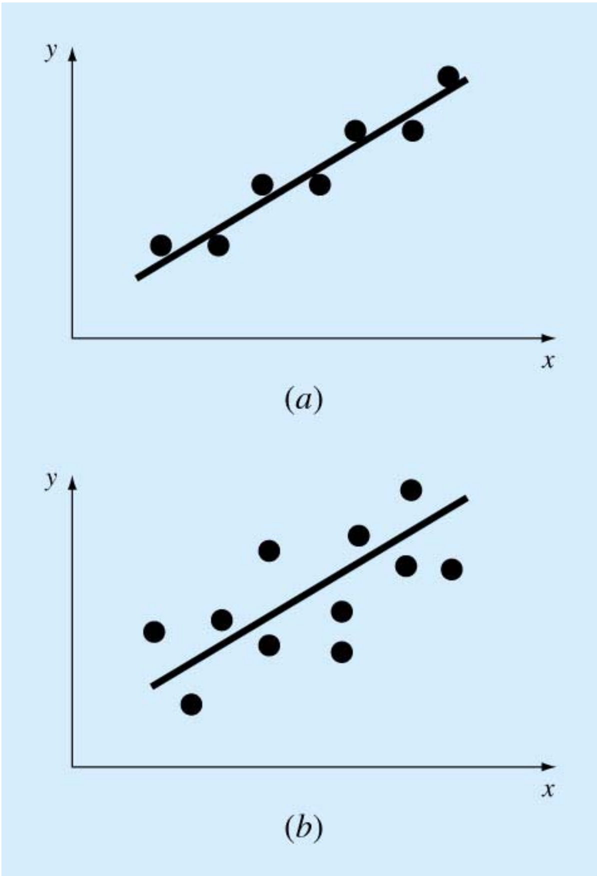




$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Standard error of the estimate

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



r^2= Coefficient of Determination

r= correlation coefficient

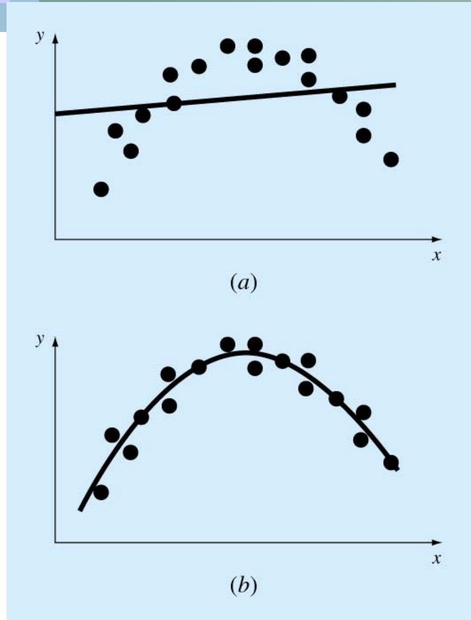
$$r^2 = \frac{S_t - S_r}{S_t}$$

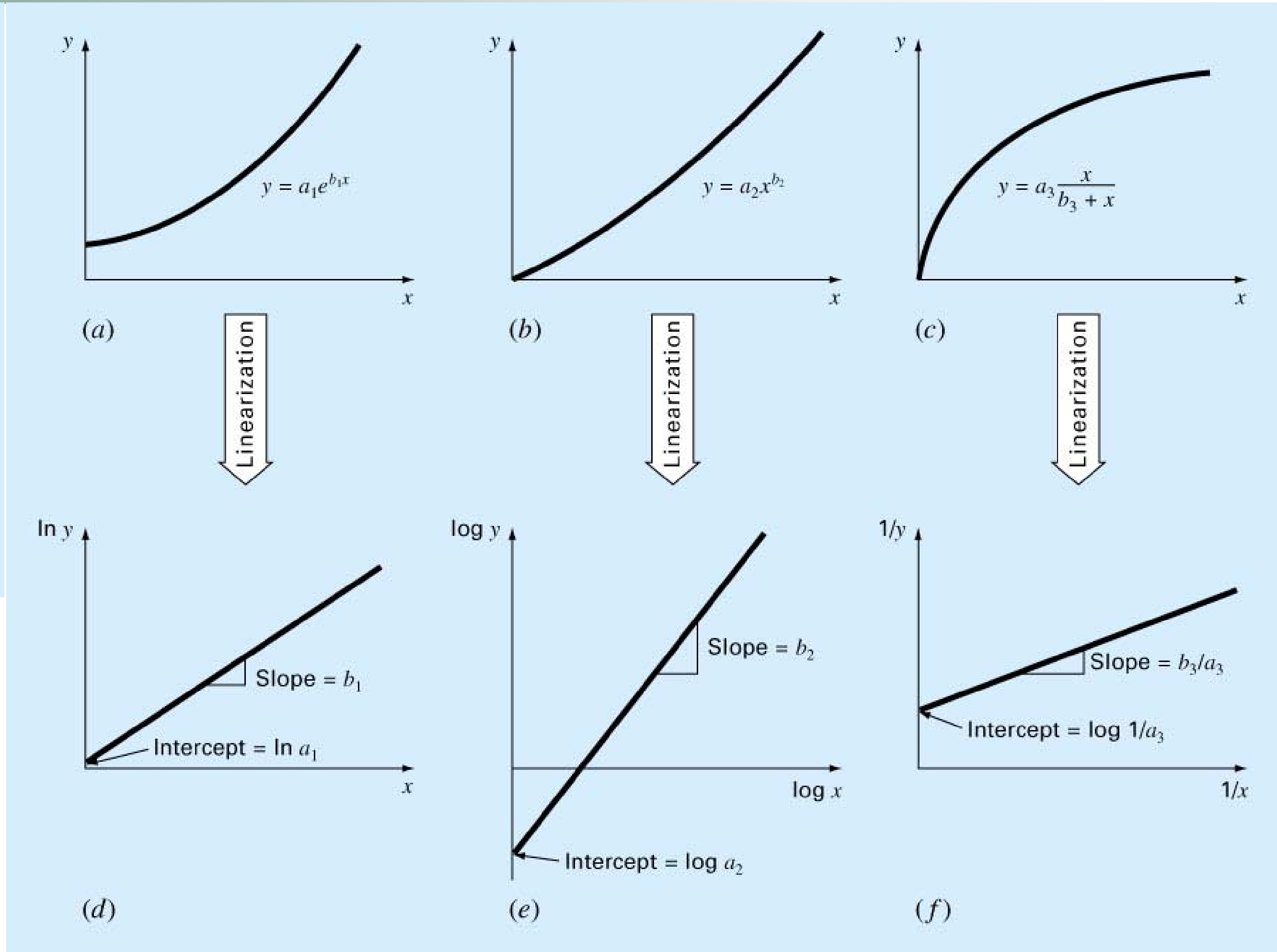
$$r^2 = \frac{n\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)}{\left(\sqrt{\left(n\sum x_i^2 - \left(\sum x_i\right)^2\right)}\right)\left(\sqrt{\left(n\sum y_i^2 - \left(\sum y_i\right)^2\right)}\right)}$$

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

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

END Regress
```





$$y = a_1 e^{b_1 x}$$

$$\ln y = \ln a_1 + b_1 x$$

$$y = a_2 x^{b_2}$$

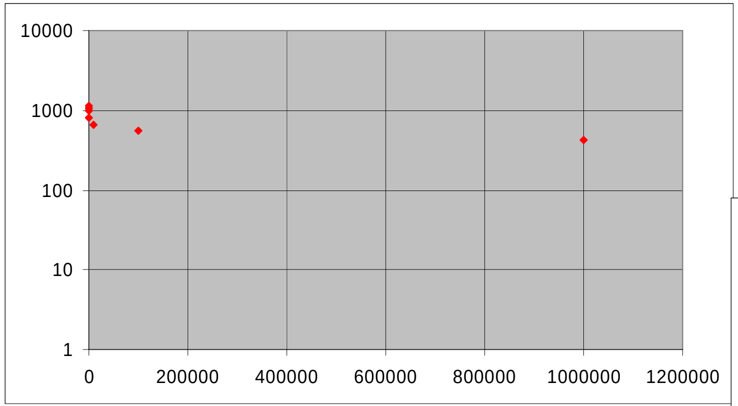
$$\ln y = \ln a_2 + b_2 \ln x$$

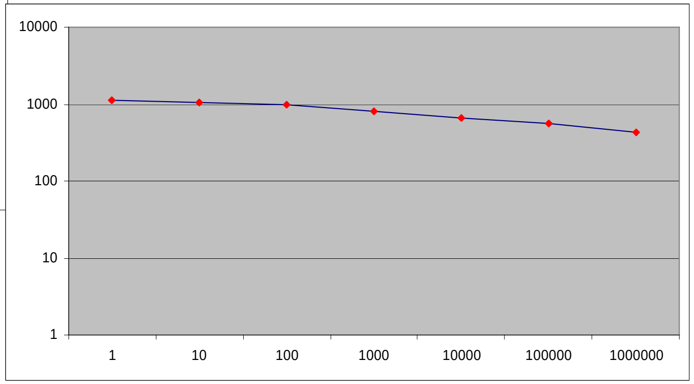
$$y = a_3 \frac{x}{b_3 + x}$$

$$\frac{1}{y} = \frac{b_3}{a_3}\frac{1}{x} + \frac{1}{a_3}$$

Example (problem 17.20)

N (Cycles)	Stress(Mpa)
1	1131
10	1058
100	993
1000	801
10000	651
100000	562
1000000	427





Ln(N)	Ln(Stress)
0	7.030857
2.302585	6.964136
4.60517	6.900731
6.907755	6.685861
9.21034	6.47851
11.51293	6.331502
13.81551	6.056784

$$y = a_2 x^{b_2}$$

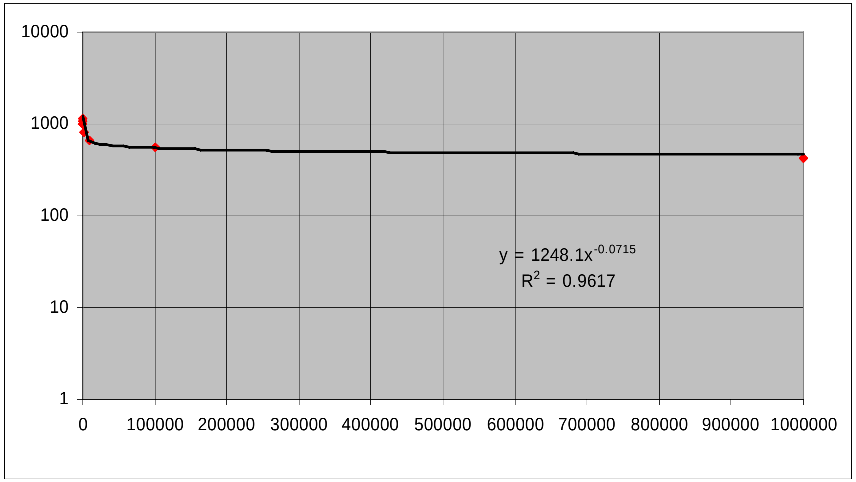
$$\ln y = \ln a_2 + b_2 \ln x$$

$\ln a_2$=7.12938

$a_{2=}$1248.103

b_2=-0.0715

$$\sigma = 1248.103 N^{-0.0715}$$



$$y = a_0 + a_1 x + a_2 x^2 .... + a_k x^k + e$$

$$e = y - a_0 - a_1 x - a_2 x^2 .... - a_k x^k$$

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2 ...... - a_k x_i^k)^2$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... - a_k x_i^k) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... - a_k x_i^k) x_i \right] = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... - a_k x_i^k) x_i^2 \right] = 0$$

....

$$\frac{\partial S_r}{\partial a_k} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_i - a_2 x_i^2 - .... - a_k x_i^k) x_i^n \right] = 0$$

$$na_0 + \left(\sum x_i\right)a_1 + \left(\sum x_i^2\right)a_2 \ldots\ldots\ldots\ldots\ldots\ldots + \left(\sum x_i^k\right)a_k = \sum_{i=1}^{n} y_i$$

$$\left(\sum x_i\right)a_0 + \left(\sum x_i^2\right)a_1 + \left(\sum x_i^3\right)a_2 \ldots\ldots\ldots + \left(\sum x_i^{k+1}\right)a_k = \sum y_i x_i$$

$$\left(\sum x_i^2\right)a_0 + \left(\sum x_i^3\right)a_1 + \left(\sum x_i^4\right)a_2 \ldots\ldots\ldots + \left(\sum x_i^{k+2}\right)a_k = \sum y_i x_i^2$$

....

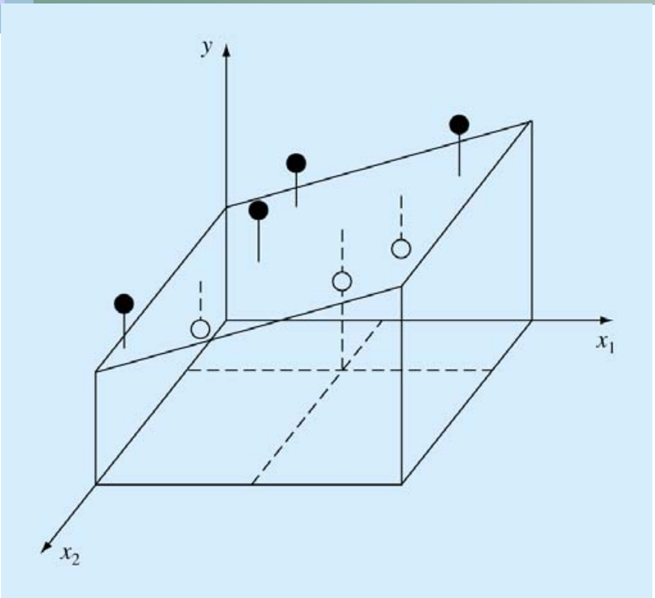
$$\left(\sum x_i^k\right)a_0 + \left(\sum x_i^{k+1}\right)a_1 + \left(\sum x_i^{k+2}\right)a_2 \ldots + \left(\sum x_i^{2k}\right)a_k = \sum y_i x_i^k$$

k+1 (linear) equations in k+1 unknowns. k cannot be larger than the number of data points. Do not use very high order polynomials.

Step 1: Input order of polynomial to be fit, m.
Step 2: Input number of data points, n.
Step 3: If $n < m + 1$, print out an error message that regression is impossible and terminate the process. If $n \geq m + 1$, continue.
Step 4: Compute the elements of the normal equation in the form of an augmented matrix.
Step 5: Solve the augmented matrix for the coefficients $a_0, a_1, a_2, \ldots, a_m$, using an elimination method.
Step 6: Print out the coefficients.

Pseudocode for determining the coefficients of the Augmented Matrix A :

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



$$y = a_0 + a_1 x_1 + a_2 x_2 \ldots\ldots\ldots + a_k x_k + e$$

$$e = y - a_0 - a_1 x_1 - a_2 x_2 \ldots\ldots\ldots - a_k x_k$$

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} \ldots\ldots - a_k x_{ki})^2$$

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_k x_{ki}) = 0$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_k x_{ki}) x_{1i} \right] = 0$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_k x_{ki}) x_{2i} \right] = 0$$

....

$$\frac{\partial S_r}{\partial a_k} = -2 \sum_{i=1}^{n} \left[(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i} - \ldots - a_k x_{ki}) x_{ki} \right] = 0$$

$$
\begin{bmatrix}
n & \sum x_{1i} & \sum x_{2i} & \cdots & \sum x_{ki} \\
\sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i} x_{2i} & \cdots & \sum x_{1i} x_{ki} \\
\cdots & \cdots & \cdots & \cdots & \cdots \\
\sum x_{ki} & \sum x_{ki} x_{1i} & \sum x_{ki} x_{2i} & & \sum x_{ki}^2
\end{bmatrix}
\begin{bmatrix}
a_0 \\ a_1 \\ \cdots \\ a_k
\end{bmatrix}
=
\begin{bmatrix}
\sum y_i \\ \sum x_{1i} y_i \\ \cdots \\ \sum x_{ki} y_i
\end{bmatrix}
$$

$$y = a_0 z_0 + a_1 z_1 + a_2 z_2 \ldots\ldots\ldots + a_k z_k + e$$

Where $z_0, z_1, \ldots\ldots z_k$ are different functions

$[z_0, z_1, \ldots\ldots z_k] = [1, x_1, x_2, \ldots\ldots x_k]$ or $= [1, x, x^2, \ldots. x^k]$

or $= [1, \cos(\omega t), \sin(\omega t)]$;

$$[Y] = [Z][A] + [E]$$

A: unknown coefficients

$$\left[[Z]^T [Z] \right] [A] = \left\{ [Z]^T [Y] \right\}$$

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

Example

$$y_i = f(x_i) = a_0(1 - e^{-a_1 x_i}) + e_i$$

Gauss Newton Method

Expand $f(x_i, a_0, a_1, ...a_m)$ in Taylor Series

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

At j=0, assume values for $a_0, a_1, ...a_m$, Using the above equation solve for Δa_k values, repeat as in the Newton Raphson Method.