# Curvilinear Relations

Correlation and regression are based on linear relationships.

Often a limited range of values can be fit by a straight line, but for a wider range of values the relationship may curve.

To select a curve to fit the data
- look for natural or logical relationships
- use a minimum number of variables

Many natural laws, however,  have been discovered by fitting a curve and then explaining the relationship.

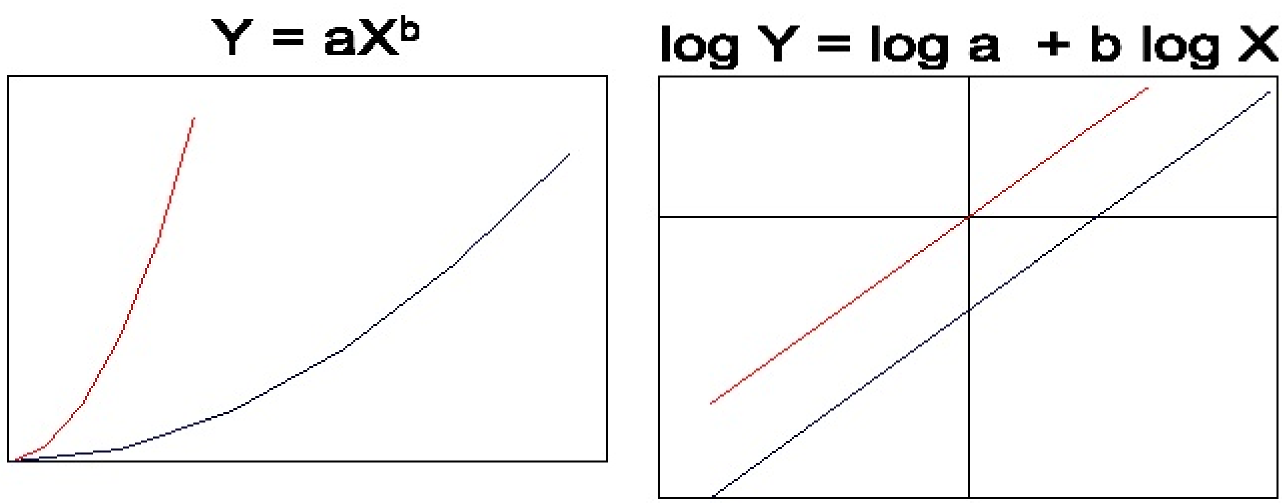
A large number of variables can fit any relationship but explain nothing.

Power curve

$Y = aX^b$

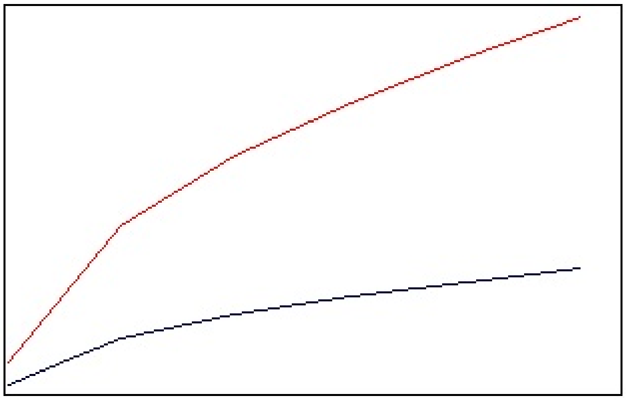
$\log Y = \log a + b \log X$

X and Y must be positive to take the log

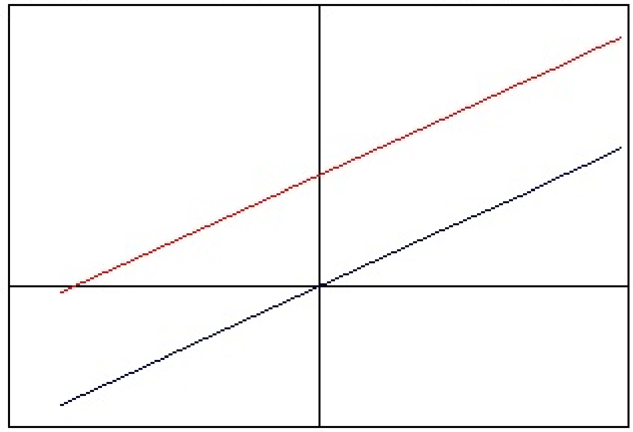



Red: $Y = X^2$ Blue: $Y = 0.1\,X^2$

$Y = aX^b$

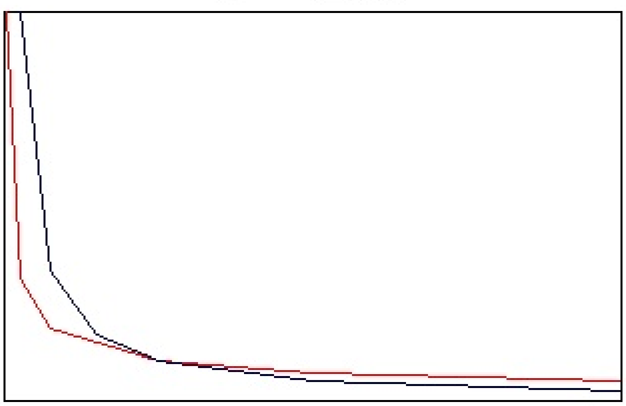


$\log Y = \log a + b \log X$

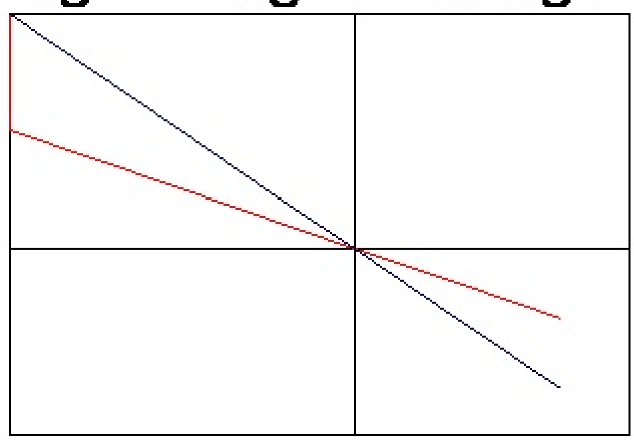


Red: $Y = 3X^{1/2}$ Blue: $Y = X^{1/2}$

$Y = aX^b$



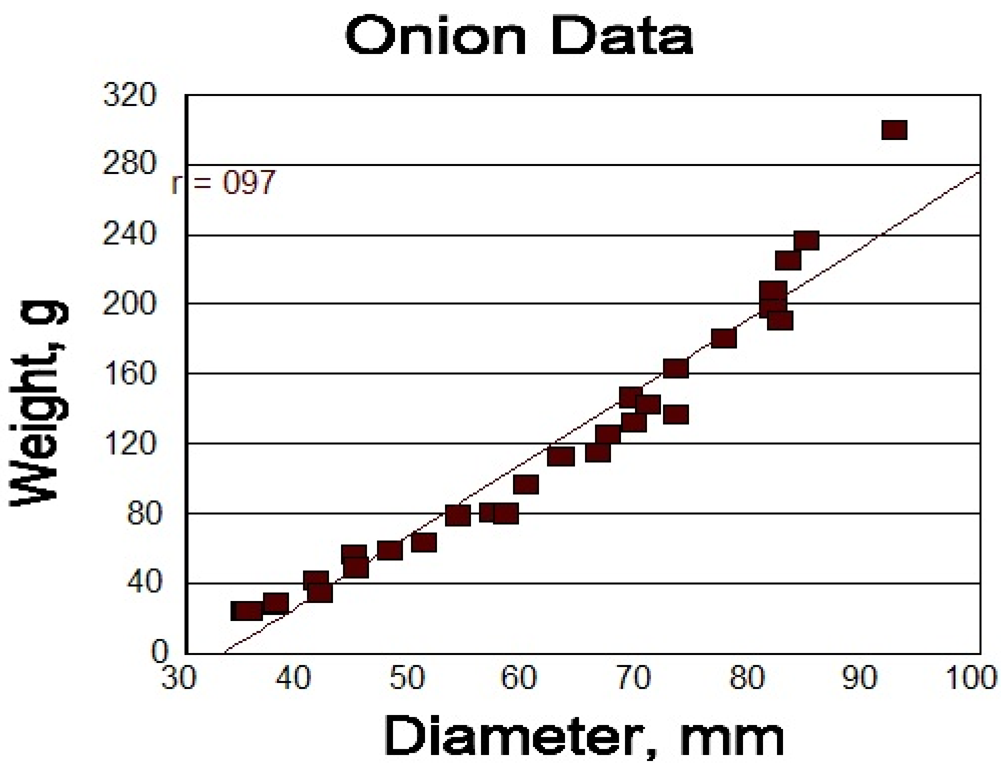
$\log Y = \log a + b \log X$

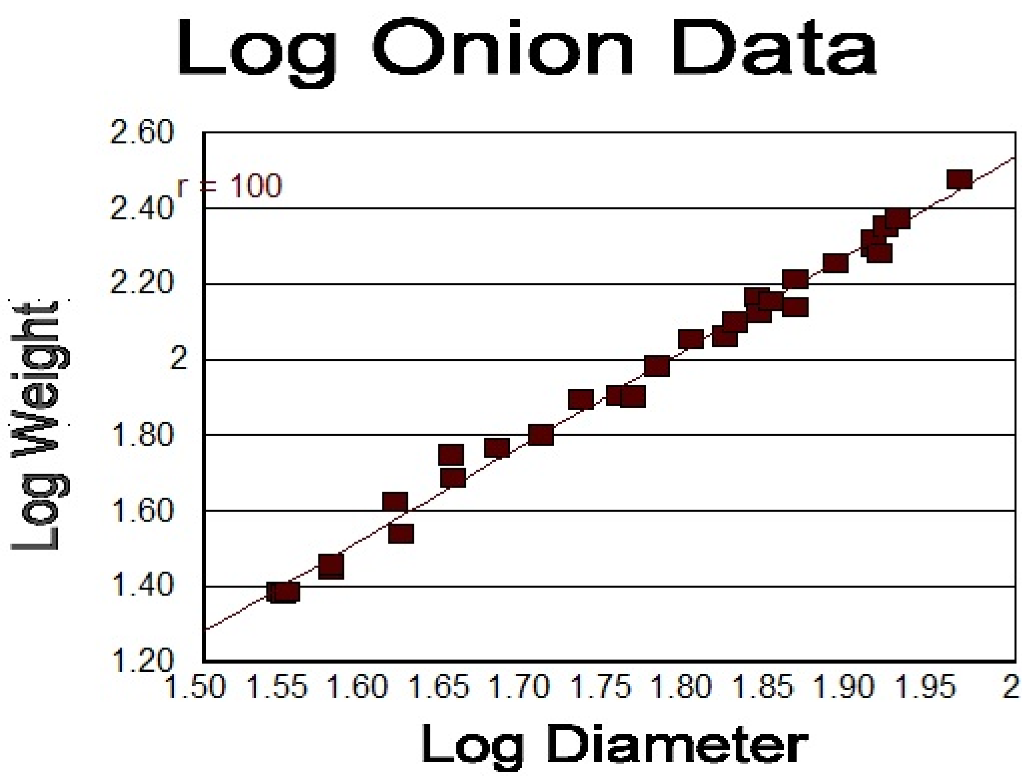


Red: $Y = X^{-1/2}$ Blue: $Y = X^{-1}$

Expected when Y and X involve different dimensions, such as height and weight
Weight is related to volume,
volume is related to $ht^3$

E.g. Onion weight and diameter,
for a sphere $V = (4/3)\,\pi\, r^3$


Onion Data
r = 097
320
280
240
200
160
120
80
40
0
30
40
50
60
70
80
90
100
Weight, g
Diameter, mm

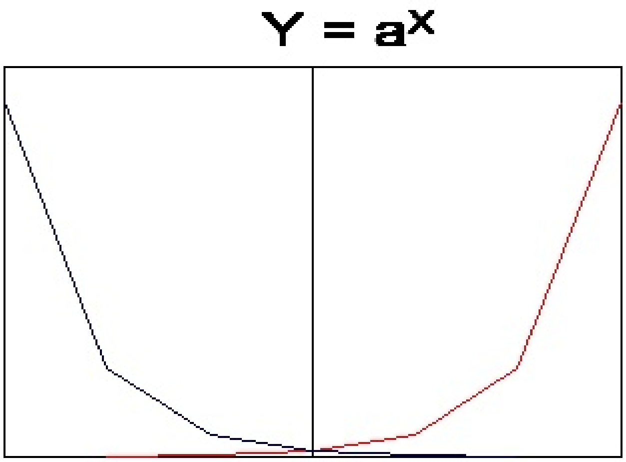

Log Onion Data
r = 100
2.60
2.40
2.20
2
1.80
1.60
1.40
1.20
1.50
1.55
1.60
1.65
1.70
1.75
1.80
1.85
1.90
1.95
2
Log Weight
Log Diameter

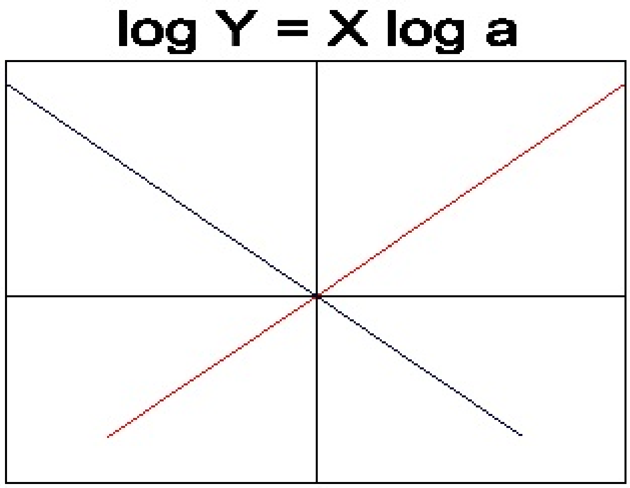
Exponential curve

$Y = ab^x$

$\log Y = \log a + X \log b$

Semilog relationship since we have log Y but not log X

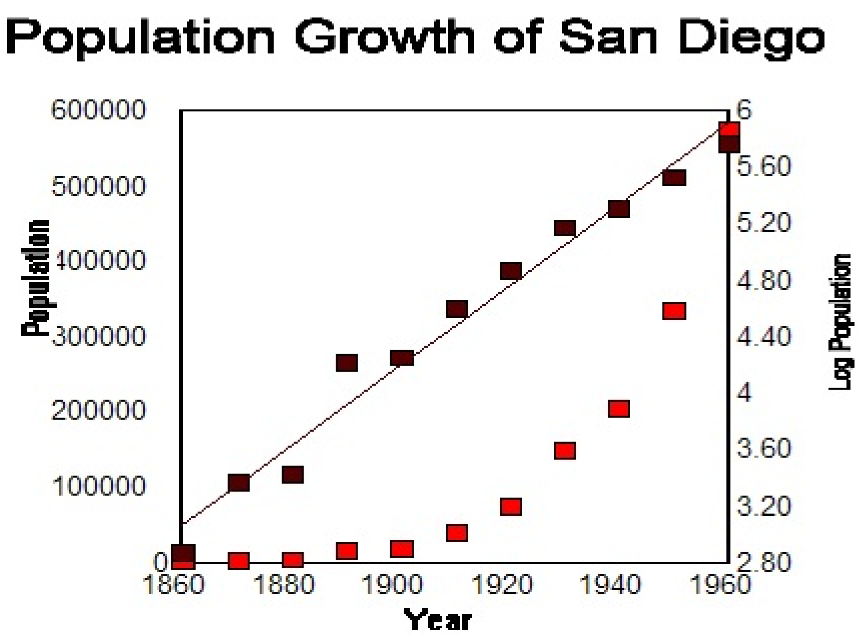



Red: $Y = 2^X$ Blue: $Y = ½^X$

E.g. interest rates, growth, chemical reactions, radioactive decay e.g. annual payments for compounded interest, $A = P(1 + r)^t$

E.g. growth of population of San Diego



Polynomial curve

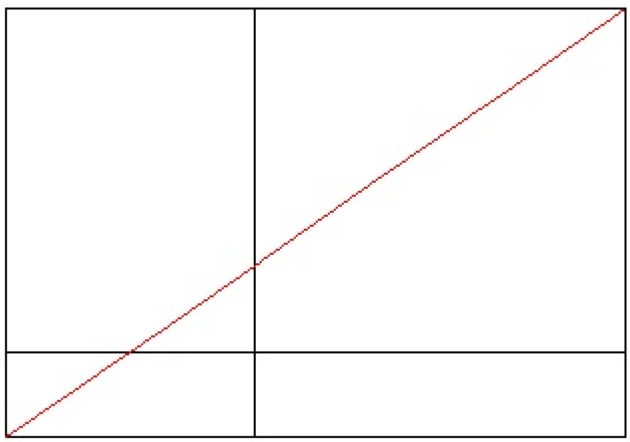
$Y = a^0 + a_1X + a_2X^2 + a_3X^3 + ...$

X is linear or first degree polynomial

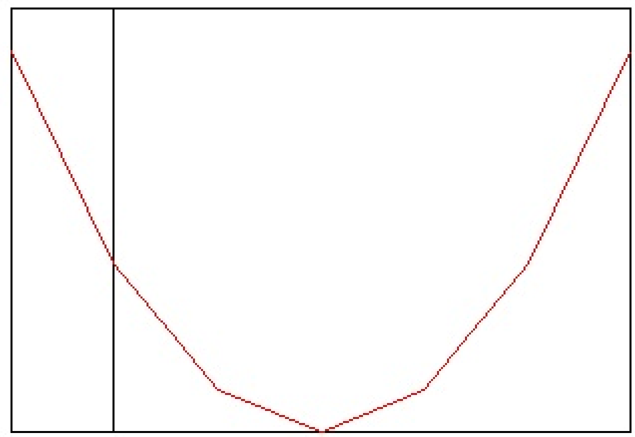
$X^2$ is quadratic or second degree

$X^3$ is cubic or third degree

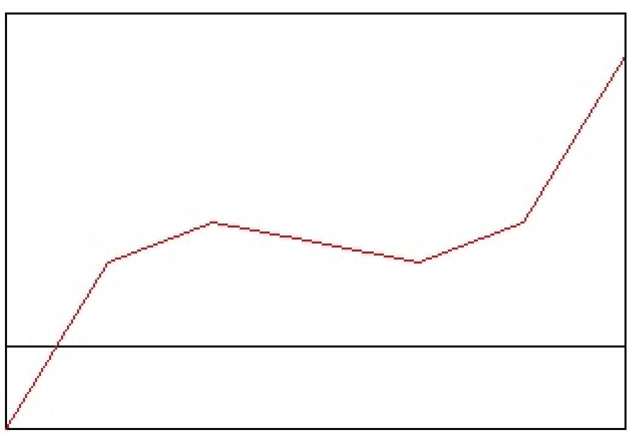
$Y = 1 + X$



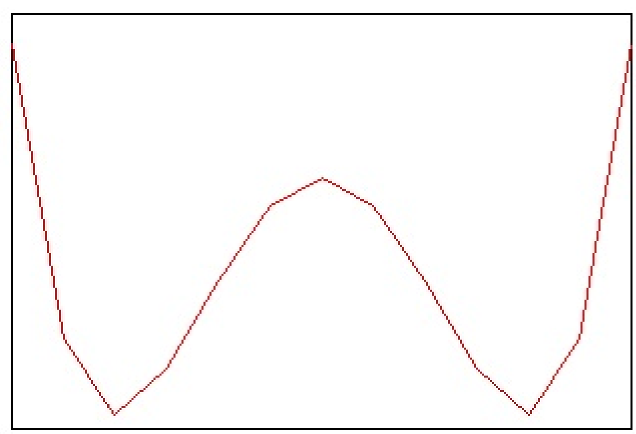
$Y = 16 - 8X + X^2$



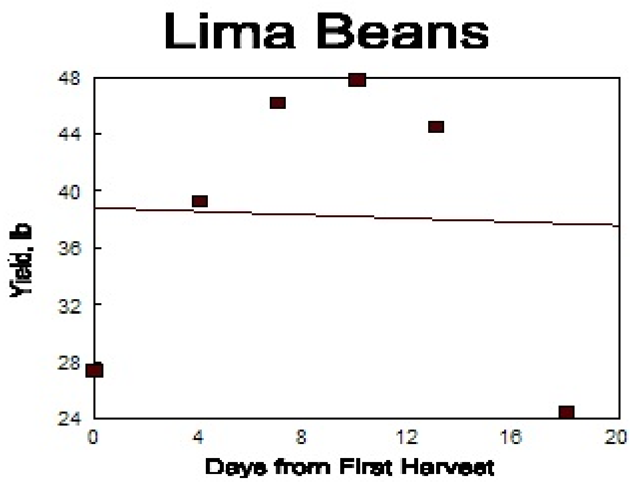
$Y = -4 + 6X - (9/16)X^2 + (1/16)X^3$



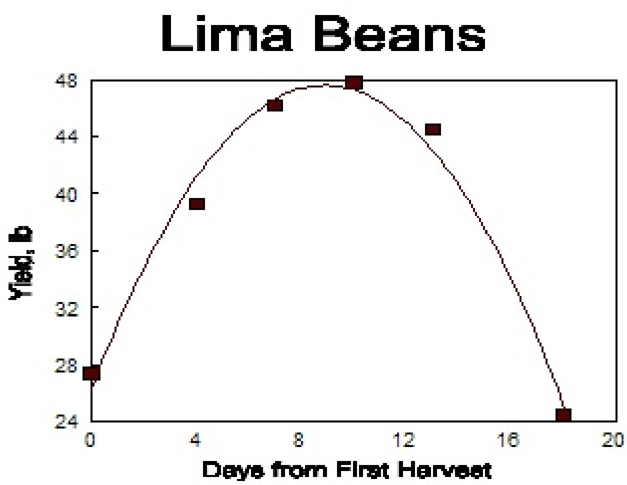
$Y = 13 - 15X + (23/4)X^2 - (3/4)X^3 + (1/32)X^4$



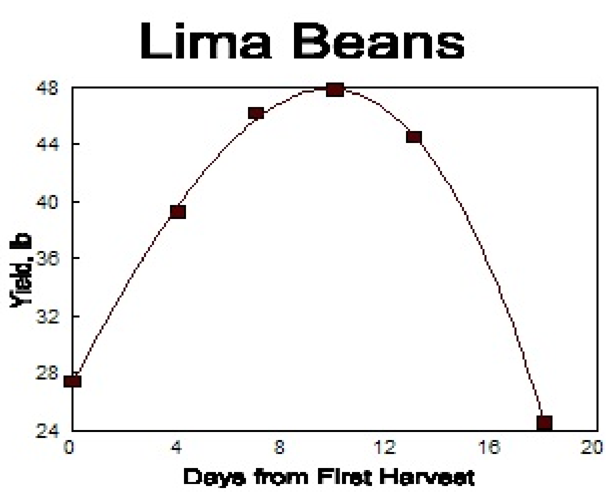
E.g. lima beans versus date of harvest




Linear fit
$r^2 = 0.0015$




Quadratic fit
$r^2 = 0.984$

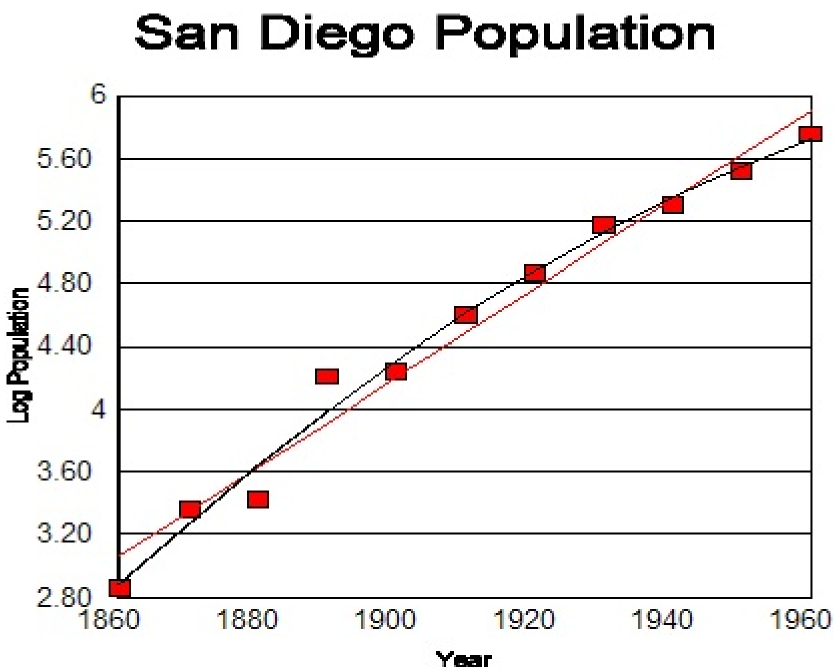



Cubic fit
$r^2 = 0.999$

Combined curves

e.g. $\log Y = \log a + X \log b + X^2 \log c$

E.g. exponential fit to San Diego population growth




Linear: log Pop = 3.06 + 0.285 Yr

$r^2 = 0.974$

Quadratic: log Pop = 2.879 + 0.406 Yr - 0.0121 $\text{Yr}^2$

$r^2 = 0.987$

Periodic curves

Fourier curve

$Y = a_0 + a_1 \cos cx + b_1 \sin cx + a_2 \cos cx + b_2 \sin cx$ ...

Where

x = time from start

c = 360/number of units in cycle

e.g. if units are hours, c = 360/24 = 15 °

for first degree curves

$a_0$ = weighted mean or central value

A = sqrt $(a_1 + b_1)$ = semiamplitude

arctan $(b_1/a_1)$ = phase angle

e.g. mean monthly temperature

e.g. circadian hormonal variation

Pulses

Waves or curves are additive

1. Calculate the overall mean.
2. Pulses or peaks can be distinguished by being > 2 standard deviations from the overall mean.
3. Mean must be recalculated without pulses after they have been identified and step 2 repeated.
4. Keep repeating until no further pulses are identified.