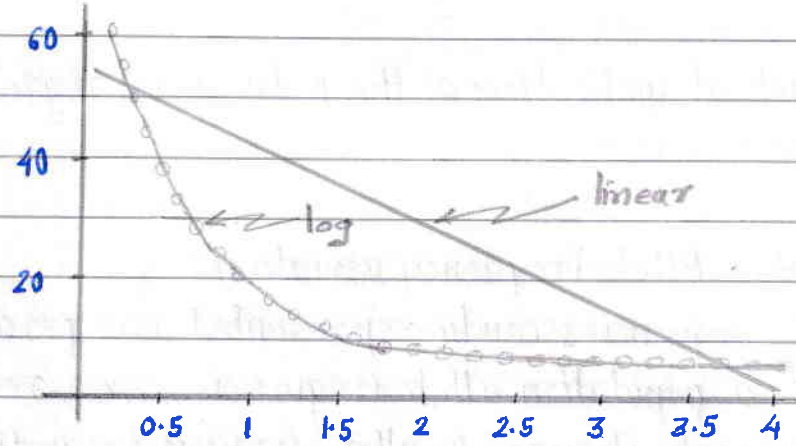
Non-linear regression

Many times the linear regression model does not yield the best predictor model for Y.

Often we need to use a non-linear equation relating Y to X of the general form $Y = f(X, p)$ where $f(\ )$ is a function and p are parameters.



Attenuation of light with depth in the water column of an estuary

For example, an exponential model, with parameters k and Y_0 is a good explainatory model for light attenuation

$$Y = Y_0 \exp(-kX)$$

This is known as Beer-Lambert law. K is called extinction coefficient and Y_0 is the light just below the surface.

This equation, which is a non-linear model can be transformed into linear model as

$$\ln(Y) = \ln(Y_0) - KX.$$

Use linear regression, substituting $\ln(Y)$ for Y, $\ln(Y_0)$ for β_0 and $-K$ for β_1. Once Y_0 and K are determined, better regression response can be obtained.

Polynomial Regression

In this case, the predictor is a linear combination of the increasing powers of X. In general,

$$Y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots\ldots + \beta_m x^m.$$

The polynomial order 'm' is chosen as per the computational need. For most of the work, terms of **x** no higher than cubic are used.

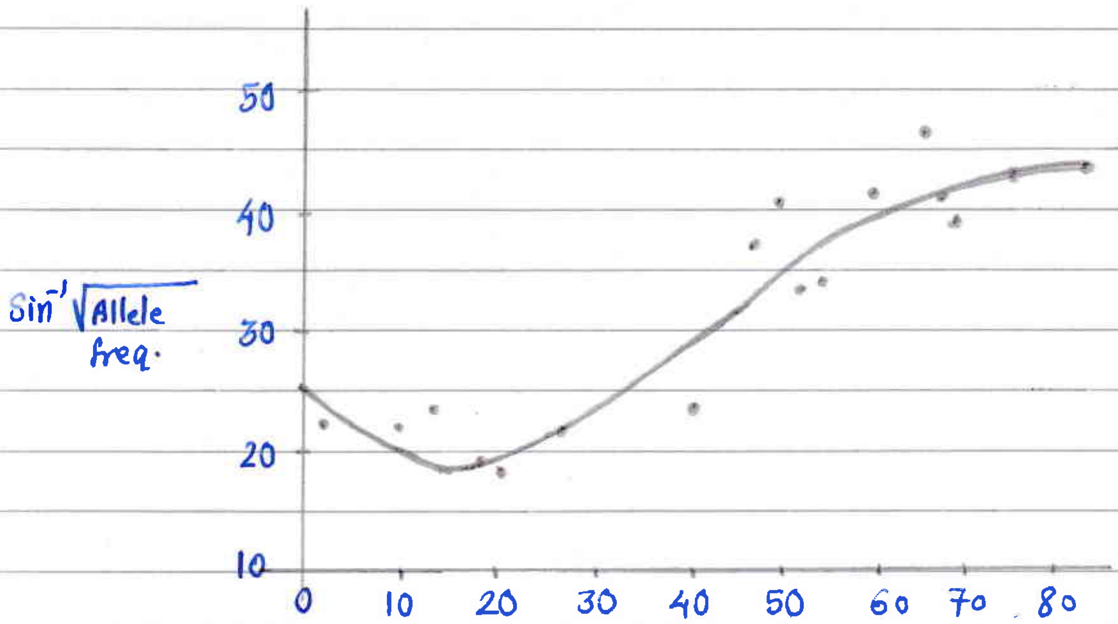
A practical application of the polynomial regression is the calibration of sensors.

Case study – Allele frequency variation.

Allele frequencies can be represented as a percentage or a fraction. In a population, allele frequencies are a reflection of genetic diversity. Changes in allele frequencies indicate that genetic drift is occurring or that new mutations have been introduced into the population.

Oceanographers have observed that the intervention of humans at the seashore or nearby have a direct impact on some of the marine species. In a mussel population, a sensitive & fragile marine species, the allele frequencies are impacted as a function of distance from shoreline

Allele frequency	Miles from Southport CN	Allele frequency	Miles from Southport CN
0·155	1	0·305	44
0·15	9	0·315	45
0·165	11	0·46	51
0·115	17	0·545	54
0·11	18·5	0·47	55·5
0·16	25	0·45	57
0·17	36	0·51	61
0·355	40·5	0·525	67
0·43	42		



The polynomial regression equation fitted with using cubic function is

$$Y = 26.22325 - 0.94408\,x + 0.042145\,x^2 - 0.00035\,x^3$$