# Curve Fitting

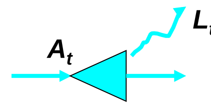
CE 311 K - Introduction to Computer Methods

Daene C. McKinney

## Curve Fitting

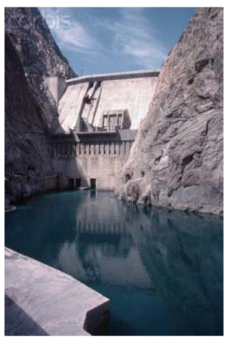
- Linear Regression – Normal Equations
- Polynomial Regression
- Nonlinear Transformations

## Evaporation from Reservoir

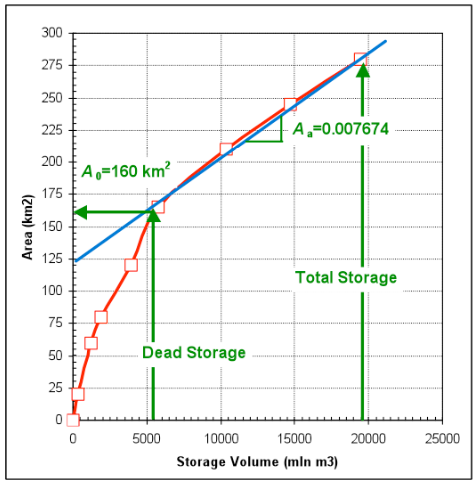



$$S_{t+1} = S_t + Q_t - R_t - L_t$$

- $L_t$ **Losses from reservoir**
- **A Surface area of reservoir**
- $e_t$ **ave. evaporation rate**

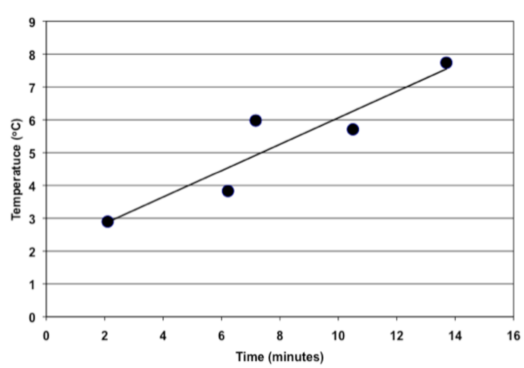


**Toktogul on the Naryn River in Kyrgyzstan**



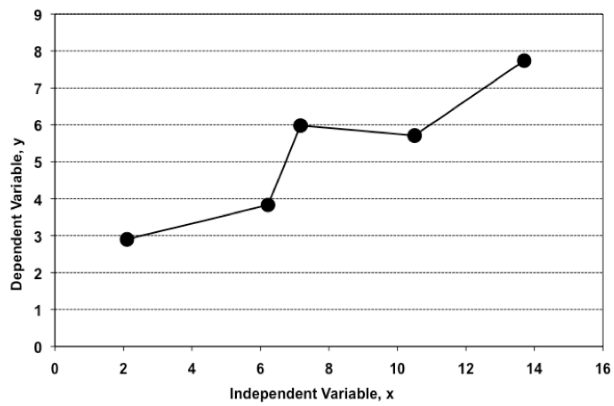

## Curve Fitting

- Data at discrete points or times
- Want estimates at points between measurements
- Fit curve to data to estimate intermediate values



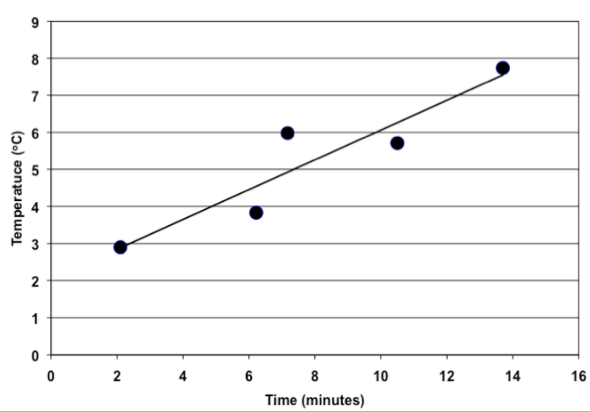
# Interpolation

- Precise data – no error in y
- Force curve through each point



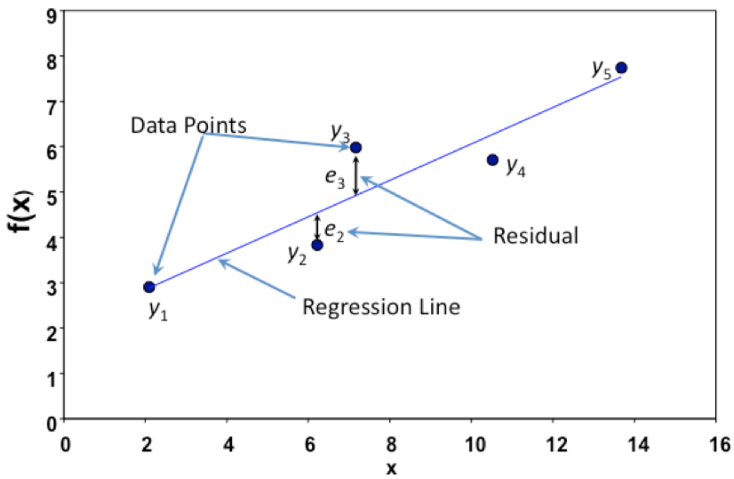

# Regression

- Experimental Data
  - Noisy (contains errors or inaccuracies)
  - *x* values are accurate, *y* values are not
- Find general trend (relationship) between *x* and $y = f(x)$
  - Without passing through any specific point



# Noisy Data From Experiment

| i→ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| X | 2.10 | 6.22 | 7.17 | 10.5 | 13.7 |
| Y | 2.90 | 3.83 | 5.98 | 5.71 | 7.74 |




# Least Squares Regression

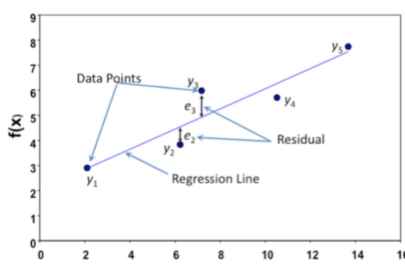
- Minimize the residual between the data points and the line

Model: $y = a_0 + a_1 x$

Estimate $a_0$ and $a_1$:

$$y_i = a_0 + a_1 x_i$$

$$e_i = y_i - a_0 - a_1 x_i$$




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

Find the values of $a_0$ and $a_1$ that minimize $S_r$

# Least Squares Regression

- Minimize $S_r$ by taking derivatives WRT $a_0$ and $a_1$,

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

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

$$n a_0 + \left[\sum_{i=1}^{n} x_i\right] a_1 = \sum_{i=1}^{n} y_i$$

$$\left[\sum_{i=1}^{n} x_i\right] a_0 + \left[\sum_{i=1}^{n} x_i^2\right] a_1 = \sum_{i=1}^{n} x_i y_i$$

Normal Equations

# Normal Equations - Solution

$$a_0 = \frac{\frac{1}{n}\sum_{i=1}^{n} y_i \sum_{i=1}^{n} x_i^2 - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2}$$

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left[\sum_{i=1}^{n} x_i\right]^2}$$

Need

$$\sum_{i=1}^{n} x_i$$

$$\sum_{i=1}^{n} x_i^2$$

$$\sum_{i=1}^{n} y_i$$

$$\sum_{i=1}^{n} x_i y_i$$

# Example

| $i$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $x_i$ | 2.10 | 6.22 | 7.17 | 10.5 | 13.7 |
| $y_i$ | 2.90 | 3.83 | 5.98 | 5.71 | 7.74 |

$$\sum_{i=1}^{5} x_i = 39.69$$

$$\sum_{i=1}^{5} x_i^2 = 392.3201$$

$$\sum_{i=1}^{5} y_i = 26.16$$

$$\sum_{i=1}^{5} x_i y_i = 238.7416$$

$$a_0 = \frac{\frac{1}{5}(26.16)(392.3) - \frac{1}{5}(39.69)(238.7)}{392.3 - \frac{1}{5}[39.69]^2} = 2.038$$

$$a_1 = \frac{238.7 - \frac{1}{5}(39.69)(26.16)}{392.3 - \frac{1}{5}[39.69]^2} = 0.4023$$

$$\boxed{y = 2.038 + 0.4023x}$$

# Example