# Experimental Modeling

**Institute of Technology of Cambodia**

Departerment of Applied Mathematics and Statistics.

**Tepmony SIM & Sokea LUEY**

December , 2022

# Course outline

# Curve fitting vs Interpolation

**Curve fitting** is to find a curve that could best indicate the trend of a given set of data. It allows some deviations between model and the collected data.

**Interpolation** is to connect discrete data points so that one can get reasonable estimates of data points between the given points. It connects all the data points using a (nonlinear) curve.

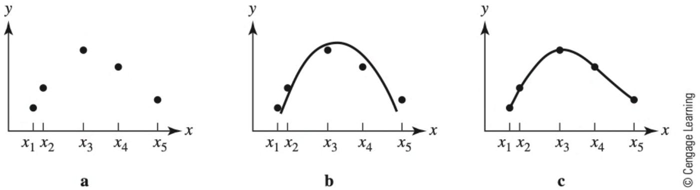



**Figure 4.1**

If the modeler expects a quadratic relationship, a parabola may be fit to the data, as in b. Otherwise, a smooth curve may be passed through the points, as in c.

# Curve fitting vs Interpolation

**Situation.** If the modeler is unable to construct a tractable model form (curve fitting) that satisfactorily explains the behavior of the data, that is, the modeler does not know what kind of curve actually describes the behavior and if it is necessary to **predict** the behavior nevertheless, the modeler may conduct experiments (or otherwise gather data) to investigate the behavior of the dependent variable(s) for selected values of the independent variable(s) within some range. In essence, the modeler desires to construct an **empirical model** based on the collected data rather than select a model based on certain assumptions. In such cases the modeler is strongly influenced by the data that have been carefully collected and analyzed, so he or she seeks a curve (interpolation) that captures the trend of the data to predict in between the data points.

# Harvesting in the Chesapeake Bay and Other One-Term Models

Let's consider a situation in which a modeler has collected some data but is unable to construct an explication model. In 1992, the Daily Press (a newspaper in Virginia) reported some observations (data) collected during the past 50 years on harvesting sea life in the Chesapeake Bay. We will examine several scenarios using observations from (a) harvesting bluefish and (b) harvesting blue crabs by the commercial industry of the Chesapeake Bay.

**Table 4.1** Harvesting the bay, 1940–1990

| Year | Bluefish (lb) | Blue crabs (lb) |
|---|---|---|
| 1940 | 15,000 | 100,000 |
| 1945 | 150,000 | 850,000 |
| 1950 | 250,000 | 1,330,000 |
| 1955 | 275,000 | 2,500,000 |
| 1960 | 270,000 | 3,000,000 |
| 1965 | 280,000 | 3,700,000 |
| 1970 | 290,000 | 4,400,000 |
| 1975 | 650,000 | 4,660,000 |
| 1980 | 1,200,000 | 4,800,000 |
| 1985 | 1,500,000 | 4,420,000 |
| 1990 | 2,750,000 | 5,000,000 |

# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

**Figure 4.2**

Scatterplot of harvesting bluefish versus base year (5-year periods from 1940 to 1990)

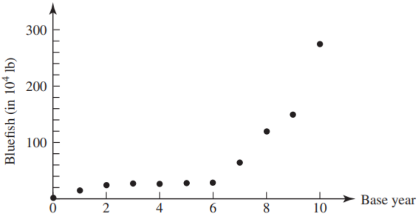
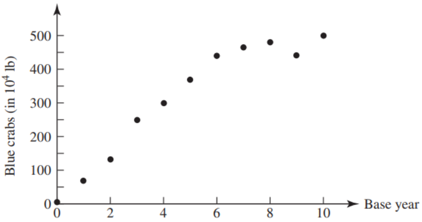



**Figure 4.3**

Scatterplot of harvesting blue crabs versus base year (5-year periods from 1940 to 1990)



# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

- Figure 4.2 clearly shows a tendency to harvest more bluefish over time, indicating or suggesting the availability of bluefish. A more precise description is not so obvious.
- In Figure 4.3, the tendency is for the increase of harvesting of blue crabs. Again, a precise model is not so obvious.
- How we might begin to predict the availability of bluefish over time? Our strategy will be to transform the data of Table 4.1 in such a way that the resulting graph approximates a line, thus achieving a working model.
- But how do we determine the transformation?

# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

Figure 4.4 shows a set of five data $(x, y)$ with $y = x,\ x > 1$.

Suppose we change the $y$ value of each point to $\sqrt{y},\ \log y, and\ \frac{1}{y}$

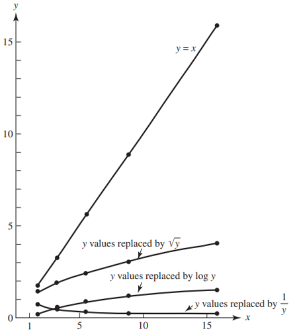



**Figure 4.4:** Relative effects of three transformation

**Table 4.2** Ladder of Transformations

| |
|---|
| ⋮ |
| $z^3$ |
| $z^2$ |
| $z$ (no change) |
| * $\sqrt{z}$ |
| * $\log z$ |
| * $\frac{-1}{\sqrt{z}}$ |
| * $\frac{-1}{z}$ |
| $\frac{-1}{z^2}$ |
| ⋮ |

*The transformations most often used.

# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

## Example 1: Harvesting Bluefish

Recall from the scatterplot in Figure 4.2 that the trend of the data appears to be increasing and concave up. Using the ladder of powers to squeeze the right-hand tail downward, we can change $y$ values by replacing y with $\log y$ or other transformations down the ladder. Another choice would be to replace $x$ values with $x^2$ or $x^3$ values or other powers up the ladder.

- We fit with the least squares model of the form $\log y = mx + b$
- We obtain the model $\log y = 0.7231 + 0.1654x$ to data in **Table 4.3**
- The model can be written as $y = 5.2857(1.4635)^x$

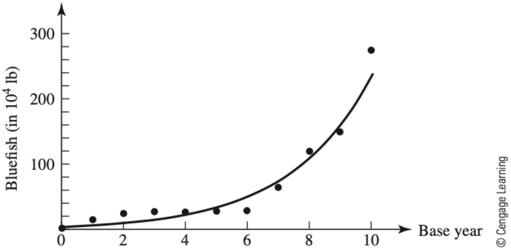
**Table 4.3** Harvesting the bay: Bluefish, 1940–1990

| Year | Base year $x$ | Bluefish (lb) $y$ |
|---|---|---|
| 1940 | 0 | 15,000 |
| 1945 | 1 | 150,000 |
| 1950 | 2 | 250,000 |
| 1955 | 3 | 275,000 |
| 1960 | 4 | 270,000 |
| 1965 | 5 | 280,000 |
| 1970 | 6 | 290,000 |
| 1975 | 7 | 650,000 |
| 1980 | 8 | 1,200,000 |
| 1985 | 9 | 1,550,000 |
| 1990 | 10 | 2,750,000 |

# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

**Figure 4.5**

Superimposed data and model $y = 5.2857(1.4635)^x$



# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

## Example 2: Harvesting Blue Crabs

Recall from our original scatterplot, Figure 4.3, that the trend of the data is increasing and concave down. With this information, we can utilize the ladder of transformations. We will use the data in Table 4.4, modified by making 1940 (year x = 0) the base year, with each base year representing a 5-year period.

- we can attempt to linearize these data by changing $y$ values to $y^2$ or $y^3$ values or to others moving up the ladder.
- Replace the $x$ values with $\sqrt{x}$
- Fit the model $y = k\sqrt{x}$ to data in **Table 4.4**
- Use least squares to find $k$, yielding

$$y = 158.344\sqrt{x}$$

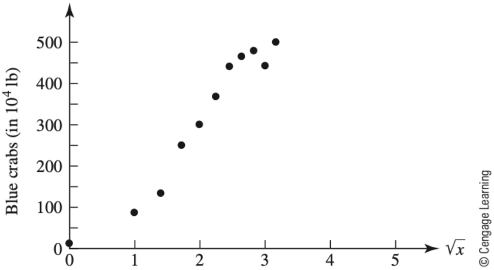
**Table 4.4** Harvesting the bay: Blue crabs, 1940–1990

| Year | Base year $x$ | Blue crabs (lb) $y$ |
|---|---|---|
| 1940 | 0 | 100,000 |
| 1945 | 1 | 850,000 |
| 1950 | 2 | 1,330,000 |
| 1955 | 3 | 2,500,000 |
| 1960 | 4 | 3,000,000 |
| 1965 | 5 | 3,700,000 |
| 1970 | 6 | 4,400,000 |
| 1975 | 7 | 4,660,000 |
| 1980 | 8 | 4,800,000 |
| 1985 | 9 | 4,420,000 |
| 1990 | 10 | 5,000,000 |

# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

**Figure 4.6**

Blue crabs (in $10^4$ lb) versus $\sqrt{x}$



# Harvesting in the Chesapeake Bay and Other One-Term Models Cont.

**Figure 4.7**

The line $y = 158.344\sqrt{x}$

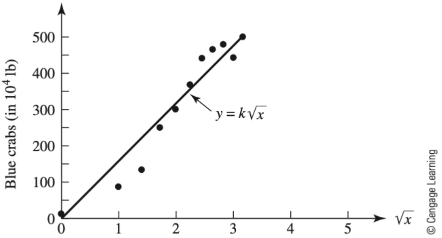
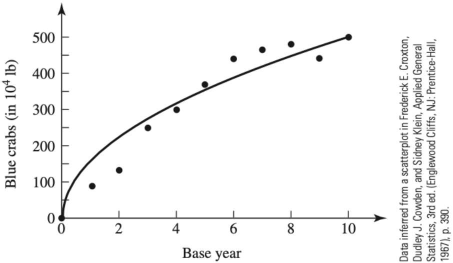



**Figure 4.8**

Superimposed data and model $y = 158.344\sqrt{x}$



# High-Order Polynomial Models

- In some cases, models with a few terms may not be sufficient, hence models with more terms must be considered.
- Polynomial Model is one of the popular models as it is analytically easy to deal with.
- The polynomial models that pass through each point in a data set that includes only one observation for each value of the independent variable is called **polynomial interpolation**.

# High-Order Polynomial Models

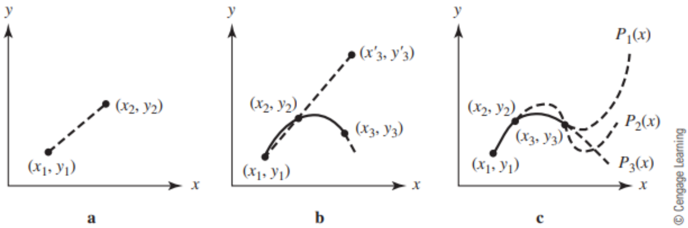



**Figure 4.10**

A unique polynomial of at most degree 2 can be passed through three data points (**a** and **b**), but an infinite number of polynomials of degree greater than 2 can be passed through three data points (**c**)

# High-Order Polynomial Models Cont.

Consider the data

- In Figure 4.10 a, A unique line of $y = a_0 + a_1 x$ can be passed through $(x_1, y_1)$ and $(x_2, y_2)$. Determine the constant $a_0$ and $a_1$.
  By the conditions we obtain

$$y_1 = a_0 + a_1 x_1$$

  and

$$y_2 = a_0 + a_1 x_2$$

- In Figure 4.10 b, A unique polynomial function of (at most) degree 2, $y = a_0 + a_1 x + a_2 x^2$, can be passed through $(x_1, y_1)$, $(x_2, y_2)$ and $(x_3, y_3)$. Determine the constant $a_0$, $a_1$ and $a_2$.
  By condition, we can find $a_0$, $a_1$ and $a_2$ by solving the following system of linear equations:

$$y_1 = a_0 + a_1 x_1 + a_2 x_1^2$$
$$y_2 = a_0 + a_1 x_2 + a_2 x_2^2$$
$$y_3 = a_0 + a_1 x_3 + a_2 x_3^2$$

# High-Order Polynomial Models Cont.

## Example. Elapsed Time of a Tape Recorder

Let's construct an empirical model to predict the amount of elapsed time of a tape recorder as a function of its counter reading. Let $c_i$ represent the counter reading and $t_i$ (sec) the corresponding amount of elapsed time. Consider the following data:

| $c_i$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| $t_i$ (sec) | 205 | 430 | 677 | 945 | 1233 | 1542 | 1872 | 2224 |

- One empirical model is a polynomial that pass through each of the data point. Since we have 8 data point, so a unique polynomial of at most 7 degree as following

$$P_7(c) = a_0 + a_1 c + a_2 c^2 + a_3 c^3 + a_4 c^4 + a_5 c^5 + a_6 c^6 + a_7 c^7$$

- The eight data points require that the constants $a_i$ satisfy the following system of linear algebraic equations:

# High-Order Polynomial Models Cont.

$$205 = a_0 + 1a_1 + a_1^2 2 + 1^3 a_3 + 1^4 a_4 + 1^5 a_5 + 1^6 a_6 + 1^7 a_7$$
$$430 = a_0 + 2a_1 + 2^2 a_2 + 2^3 a_3 + 3^4 a_4 + 5^5 a_5 + 2^6 a_6 + 2^7 a_7$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$222 = a_0 + 8a_1 + 8^2 a_2 + 8^3 a_3 + 8^4 a_4 + 8^5 a_5 + 8^6 a_6 + 8^7 a_7$$

- We divide each counter reading of above system of linear algebraic equation by 100 to lessen the numerical difficulties. We get:

$$\begin{aligned} a_0 &= -13.9999923 & a_4 &= -5.354166491 \\ a_1 &= 232.9119031 & a_5 &= 0.8013888621 \\ a_2 &= -29.08333188 & a_6 &= -0.0624999978 \\ a_3 &= 19.78472156 & a_7 &= 0.00198412222269 \end{aligned}$$

- Let's see how well the empirical model fits the data. Denoting the polynomial prediction by $P_7(c_i)$, we find

| $c_i$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| $t_i$ (sec) | 205 | 430 | 677 | 945 | 1233 | 1542 | 1872 | 2224 |
| $P_7(c_i)$ | 205 | 430 | 677 | 945 | 1233 | 1542 | 1872 | 2224 |

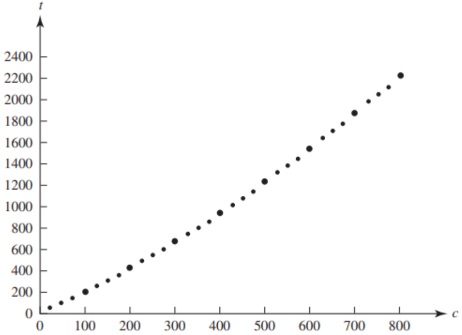
# High-Order Polynomial Models Cont.

Can we really consider this model to be better than other model we could propose?

Let's see how well this new model $P_7(c_i)$ captures the trend of the data. The model is graphed in Figure 4.11.

**Figure 4.11**

An empirical model for predicting the elapsed time of a tape recorder



# High-Order Polynomial Models Cont.

## Theorem 1: Lagrangian Form of the Polynomial

If $x_0, x_1, \cdots, x_n$ are $(n+1)$ distinct points and $y_0, y_1, \cdots, y_n$ are corresponding observations at these points, then there exists a unique polynomial $P(x)$, of at most degree $n$, with the property that

$$y_k = P(x_k), \text{ for each } k = 0, 1, \cdots, n.$$

This polynomial is given by

$$P(x) = y_0 L_0(x) + \cdots + y_n L_n(x) \tag{4.3}$$

where

$$L_k(x) = \frac{(x-x_0)(x-x_1)\cdots(x-x_{k-1})(x-x_{k+1})\cdots(x-x_n)}{(x_k-x_0)(x_k-x_1)\cdots(x_k-x_{k-1})(x_k-x_{k+1})\cdots(x_k-x_n)}$$

(4.3) passes through each of the data points, the resultant sum of absolute deviation is zero. Considering the various criteria of best fit presented in Chapter 3, we are tempted to use high-order polynomials to fit larger sets of data.

# High-Order Polynomial Models Cont.

After all, the fit is precise. Let's examine both the advantages and the disadvantages of using high-order polynomials.
Suppose the following data have been collected:

| $x$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ |
|---|---|---|---|---|
| $y$ | $y_1$ | $y_2$ | $y_3$ | $y_4$ |

By Theorem 1, we obtain cubic polynomial function:

$$
\begin{aligned}
P_3(x) =& \frac{(x-x_2)(x-x_3)(x-x_4)}{(x_1-x_2)(x_1-x_3)(x_1-x_4)}y_1 + \frac{(x-x_1)(x-x_3)(x-x_4)}{(x_2-x_1)(x_2-x_3)(x_2-x_4)}y_2 \\
&+ \frac{(x-x_1)(x-x_2)(x-x_4)}{(x_3-x_1)(x_3-x_2)(x_3-x_4)}y_3 + \frac{(x-x_1)(x-x_2)(x-x_3)}{(x_4-x_1)(x_4-x_2)(x_4-x_3)}y_4
\end{aligned}
$$

Convince yourself that the polynomial is indeed cubic and agrees with the value $y_i$ when $x = x_i$. Notice that the $x_i$ values must all be different to avoid division by zero. Observe the pattern for forming the numerator and the denominator for the coefficient of each $y_i$. This same pattern is followed when forming polynomials of any desired degree.

# High-Order Polynomial Models Cont.

## Advantages and Disadvantages of High-Order Polynomials

1. **Advantages**
   - Easy to estimate the coefficients of the polynomial
   - Calculus (computing derivative and integral) is easy with polynomial
2. **Disadvantages**
   - High-order polynomials may oscillate severely near the endpoints of the interval, a serious disadvantage to using them.
   - the polynomial can change quickly from increasing to decreasing, also making interpolation questionable.

See an example below.

# High-Order Polynomial Models Cont.

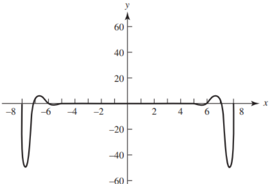
Consider some of the disadvantages of higher-order polynomials. For the 17 data points presented in Table 4.10, it is clear that the trend of the data is $y = 0$ for all $x$ over the interval $-8 \leq x \leq 8$

**Table 4.10**

| $x_i$ | $-8$ | $-7$ | $-6$ | $-5$ | $-4$ | $-3$ | $-2$ | $-1$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y_i$ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Suppose Equation (4.3) is used to determine a polynomial that passes through the points. Because there are 17 distinct data points, it is possible to pass a unique polynomial of degree at most 16 through the given points. The graph of a polynomial passing through the data points is depicted in Figure 4.12.

**Figure 4.12**
Fitting a higher-order polynomial through the data points in Table 4.10



# Smoothing: Low-Order Polynomial Models

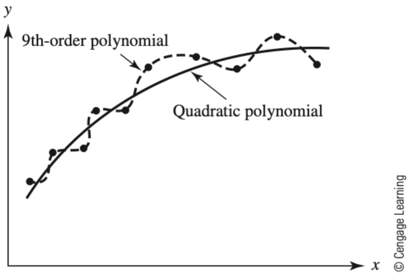
- We seek methods that retain many of the conveniences found in high-order polynomials without incorporating their disadvantages.
- One popular technique is to choose a low-order polynomial regardless of the number of data points.
- However, this choice normally results in a situation in which the number of data points exceeds the number of constants necessary to determine the polynomial.
- Low-order polynomials may not pass through all the data points.
- The process of finding low-order polynomial that best fits the data is call **polynomial smoothing**.
- This type of smoothing reduces both the tendency of the polynomial to oscillate and its sensitivity to small changes in the data.

# Smoothing: Low-Order Polynomial Models

Figure 4.16 illustrates **quadratic smoothing** for 10 data points. This quadratic function smooths the data because it is not required to pass through all the data points.

**Figure 4.16**

The quadratic function smooths the data because it is not required to pass through all the data points.



# Smoothing: Low-Order Polynomial Models

## Example. Elapsed Time of a Tape Recorder Revisited

Find the quadratic smoothing polynomial: $P_2(c) = a + bc + dc^2$ for the following data

**Table 4.13** Data collected for the tape recorder problem

| $c_i$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| $t_i$ (sec) | 205 | 430 | 677 | 945 | 1233 | 1542 | 1872 | 2224 |



**Solution.** To find the coefficients $a$, $b$, $d$, we use least squares method:

$$\text{minimize} \quad S = \sum_{i=1}^{m} \left[t_i - \left(a + bc_i + dc_i^2\right)\right]^2$$

# Smoothing: Low-Order Polynomial Models

## Example. Elapsed Time of a Tape Recorder Revisited

The necessary conditions for a minimum to exist ($\partial S/\partial a = \partial S/\partial b = \partial S/\partial d = 0$) yield the following equations:

$$\begin{aligned} ma + \left(\sum c_i\right) b + \left(\sum c_i^2\right) d &= \sum t_i \\ \left(\sum c_i\right) a + \left(\sum c_i^2\right) b + \left(\sum c_i^3\right) d &= \sum c_i t_i \\ \left(\sum c_i^2\right) a + \left(\sum c_i^3\right) b + \left(\sum c_i^4\right) d &= \sum c_i^2 t_i \end{aligned}$$

For the data given in Table 4.13, the preceding system of equations becomes

$$\begin{aligned} 8a + 3600b + 2{,}040{,}000d &= 9128 \\ 3600a + 2{,}040{,}000b + 1{,}296{,}000{,}000d &= 5{,}318{,}900 \\ 2{,}040{,}000a + 1{,}296{,}000{,}000b + 8.772 \times 10^{11} d &= 3{,}435{,}390{,}000 \end{aligned}$$

Solution of the preceding system yields the values $a = 0.14286$, $b = 1.94226$, and $d = 0.00105$, giving the quadratic

$$P_2(c) = 0.14286 + 1.94226c + 0.00105c^2$$

# Smoothing: Low-Order Polynomial Models

## Example. Elapsed Time of a Tape Recorder Revisited

We can compute the deviation between the observations and the predictions made by the model $P_2(c)$:

| $c_i$ | 100 | 200 | 300 | 400 | 500 | 600 | 700 | 800 |
|---|---|---|---|---|---|---|---|---|
| $t_i$ | 205 | 430 | 677 | 945 | 1233 | 1542 | 1872 | 2224 |
| $t_i - P_2(c_i)$ | 0.167 | −0.452 | 0.000 | 0.524 | 0.119 | −0.214 | −0.476 | 0.333 |

Note that the deviations are very small compared to the order of magnitude of the times.

When we are considering the use of a low-order polynomial for smoothing, two issues come to mind:

1. Should a polynomial be used?
2. If so, what order of polynomial would be appropriate?

The derivative concept can help in answering these two questions.

# Smoothing: Low-Order Polynomial Models

## Divided Differences

Notice that a quadratic function is characterized by the properties that its second derivative is constant and its third derivative is zero. That is, given

$$P(x) = a + bx + cx^2$$

we have

$$\begin{aligned} P'(x) &= b + 2cx \\ P''(x) &= 2c \\ P'''(x) &= 0 \end{aligned}$$

However, the only information available is a set of discrete data points. How can these points be used to estimate the various derivatives? Refer to Figure 4.17, and recall the definition of the derivative:

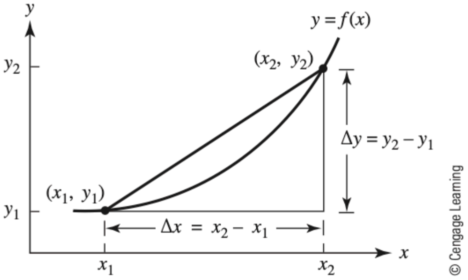
$$\frac{dy}{dx} = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x}$$

# Smoothing: Low-Order Polynomial Models

## Divided Differences

**Figure 4.17**

The derivative of $y = f(x)$ at $x = x_1$ is the limit of the slope of the secant line.




If $\Delta x = x_2 - x_1$ is small, then

$$\frac{dy}{dx}(x_1) \approx \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

# Smoothing: Low-Order Polynomial Models

## Divided Differences

If we have 3 data points, we can then estimate $d^2y/dx^2$ at $x = x_1$ by the second divided differences as illustrated in the following table:

**Table 4.16** The first and second divided differences estimate the first and second derivatives, respectively

| Data | | First divided difference | Second divided difference |
|---|---|---|---|
| $x_1$ | $y_1$ | | |
| | | $\dfrac{y_2 - y_1}{x_2 - x_1}$ | |
| $x_2$ | $y_2$ | | $\dfrac{\dfrac{y_3 - y_2}{x_3 - x_2} - \dfrac{y_2 - y_1}{x_2 - x_1}}{x_3 - x_1}$ |
| | | $\dfrac{y_3 - y_2}{x_3 - x_2}$ | |
| $x_3$ | $y_3$ | | |

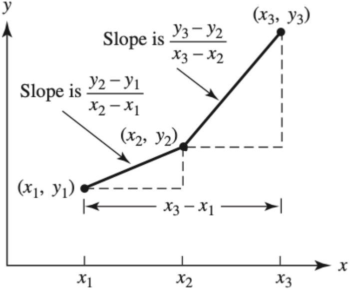
# Smoothing: Low-Order Polynomial Models

## Divided Differences

For graphic representation of the estimation of second derivative, see Figure 4.18 below.

**Figure 4.18**

The second divided difference may be interpreted as the difference between the adjacent slopes (first divided differences) divided by the length of the interval over which the change has taken place.



# Smoothing: Low-Order Polynomial Models

## Divided Differences: Example 1

**Notation.** Denote $\Delta^n$, the $n$th divided difference that is used to estimate $d^n y/dx^n$. For the following data:

**Table 4.14** A hypothetical set of collected data

| $x_i$ | 0 | 2 | 4 | 6 | 8 |
|---|---|---|---|---|---|
| $y_i$ | 0 | 4 | 16 | 36 | 64 |



then, we have the following estimate:

**Table 4.17** A divided difference table for the data of Table 4.14

| Data $x_i$ | Data $y_i$ | Divided differences $\Delta$ | $\Delta^2$ | $\Delta^3$ |
|---|---|---|---|---|
| 0 | 0 | $4/2 = 2$ | $4/4 = 1$ | $0/6 = 0$ |
| 2 | 4 | $12/2 = 6$ | $4/4 = 1$ | $0/6 = 0$ |
| 4 | 16 | $20/2 = 10$ | $4/4 = 1$ | |
| 6 | 36 | $28/2 = 14$ | | |
| 8 | 64 | | | |

$\Delta x = 6$



For this data, we will choose polynomial of degree 2 as smoothing model (as the 3rd derivative is 0).

# Smoothing: Low-Order Polynomial Models

## Divided Differences: Example 2. Elapsed Time of a Tape Recorder Revisited Again

Recall the data from Table 4.13. The divided differences are displayed in Table 4.18.

**Table 4.18** A divided difference table for the tape recorder data

| Data | | Divided differences | | | |
|---|---|---|---|---|---|
| $x_i$ | $y_i$ | $\Delta$ | $\Delta^2$ | $\Delta^3$ | $\Delta^4$ |
| 100 | 205 | | | | |
| | | 2.2500 | | | |
| 200 | 430 | | 0.0011 | | |
| | | 2.4700 | | 0.0000 | |
| 300 | 677 | | 0.0011 | | 0.0000 |
| | | 2.6800 | | 0.0000 | |
| 400 | 945 | | 0.0010 | | 0.0000 |
| | | 2.8800 | | 0.0000 | |
| 500 | 1233 | | 0.0011 | | 0.0000 |
| | | 3.0900 | | 0.0000 | |
| 600 | 1542 | | 0.0011 | | 0.0000 |
| | | 3.3000 | | 0.0000 | |
| 700 | 1872 | | 0.0011 | | |
| | | 3.5200 | | | |
| 800 | 2224 | | | | |



For this data, choosing polynomial of degree 2 as smoothing model (as the 3rd derivative is 0) is appropriate.

# Smoothing: Low-Order Polynomial Models

## Divided Differences: Example 3. Vehicular Stopping Distance

Consider the following data:

**Table 4.19** Data relating total stopping distance and speed

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Speed $v$ (mph) | 20 | 25 | 30 | 35 | 40 | 45 | 50 | 55 | 60 | 65 | 70 | 75 | 80 |
| Distance $d$ (ft) | 42 | 56 | 73.5 | 91.5 | 116 | 142.5 | 173 | 209.5 | 248 | 292.5 | 343 | 401 | 464 |



For this data, choosing polynomial of degree 2 as smoothing model (as the 3rd derivative is close to 0) is suggested. See Figure 4.20.

# Smoothing: Low-Order Polynomial Models

## Divided Differences: Example 3. Vehicular Stopping Distance

**Table 4.20** A divided difference table for the data relating total vehicular stopping distance and speed

| Data | | Divided differences | | | |
|---|---|---|---|---|---|
| $v_i$ | $d_i$ | $\Delta$ | $\Delta^2$ | $\Delta^3$ | $\Delta^4$ |
| 20 | 42 | | | | |
| | | 2.2800 | | | |
| 25 | 56 | | 0.0700 | | |
| | | 3.5000 | | −0.0040 | |
| 30 | 73.5 | | 0.0100 | | 0.0006 |
| | | 3.6000 | | 0.0080 | |
| 35 | 91.5 | | 0.1300 | | −0.0007 |
| | | 4.9000 | | −0.0060 | |
| 40 | 116 | | 0.0400 | | 0.0004 |
| | | 5.3000 | | 0.0027 | |
| 45 | 142.5 | | 0.0800 | | 0.0000 |
| | | 6.1000 | | 0.0027 | |
| 50 | 173 | | 0.1200 | | −0.0004 |
| | | 7.3000 | | −0.0053 | |
| 55 | 209.5 | | 0.0400 | | 0.0005 |
| | | 7.7000 | | 0.0053 | |
| 60 | 248 | | 0.1200 | | −0.0003 |
| | | 8.9000 | | 0.0000 | |
| 65 | 292.5 | | 0.1200 | | 0.0001 |
| | | 10.1000 | | 0.0020 | |
| 70 | 343 | | 0.1500 | | −0.0003 |
| | | 11.6000 | | −0.0033 | |
| 75 | 401 | | 0.1000 | | |
| | | 12.6000 | | | |
| 80 | 464 | | | | |

# Cubic Spline Models

## Cubic Spline Interpolation

- A very much popular modern technique
- Preserve continuity
- Preserve smoothness up to second derivative (that is the model is of class $C^2$)
- Unlike polynomial smoothing, cubic spline can deal with oscillation problem.
- But, what is called **Spline**?