Mathematical Modeling – Least Squares

- Section 2.3

- Three Modeling Methods
  - Known Relationship – underlying mathematical setting is known
  - Finite Differences – theoretical data or hard science data with little scatter
  - Least Squares – modeling data with scatter

Model: Global Warming

Global warming is partly the result of burning fuels, which increases the amount of carbon dioxide in the air. One of the major sources of fuel consumption are cars. Let's examine the number of cars in the U.S. (in millions) as one variable of global warming.

Year	Cars
1940	27.5
1950	40.3
1960	61.7
1970	89.3
1980	121.6
1990	150.5

Linear or Curvilinear?

- Numeric Method – use Finite Differences method to determine if data has a near linear trend.
  - Is the first difference nearly constant?
  - Solution: Nearly so after 1960.

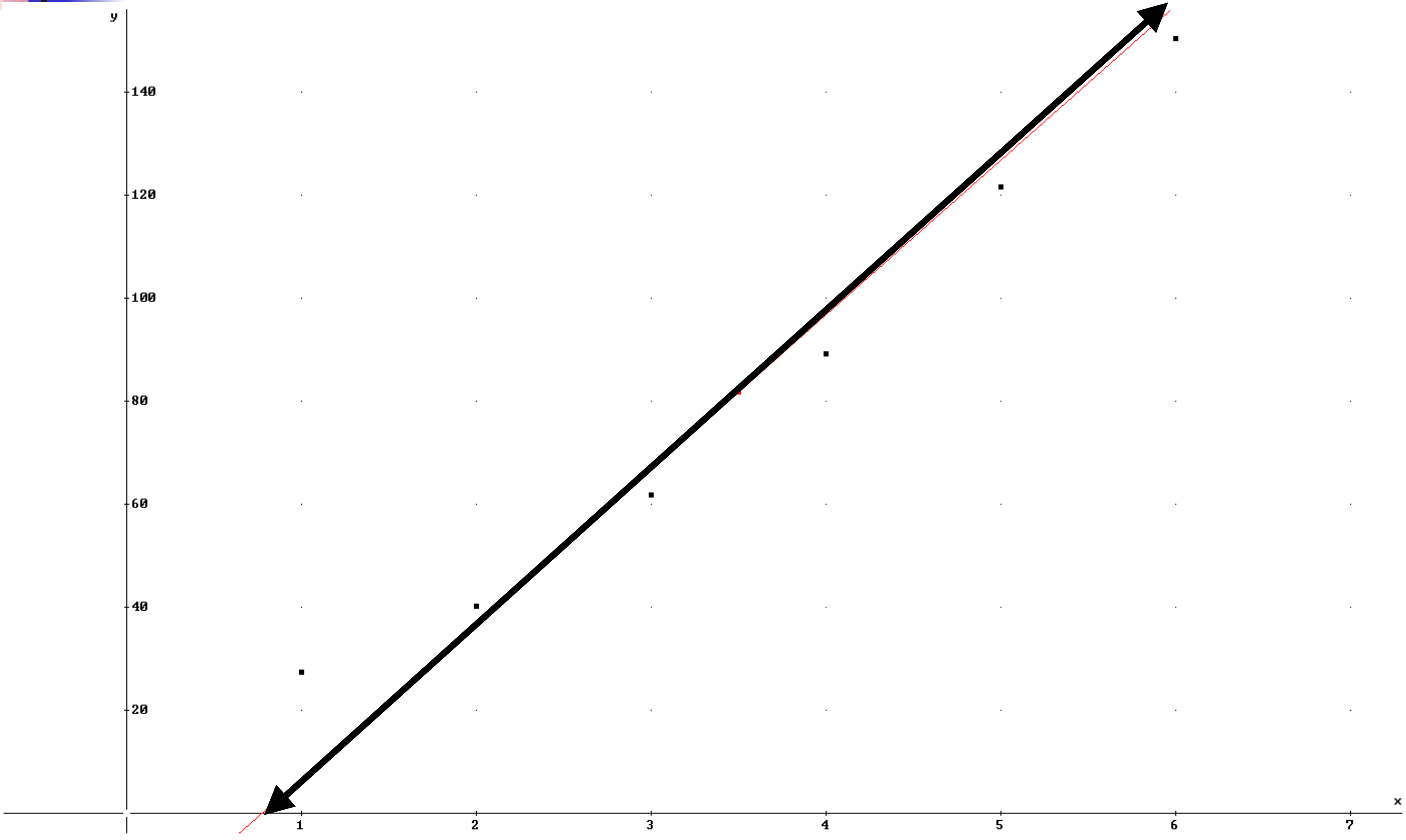
Year	Cars
1940	27.5
1950	40.3
1960	61.7
1970	89.3
1980	121.6
1990	150.5

Linear or Curvilinear?

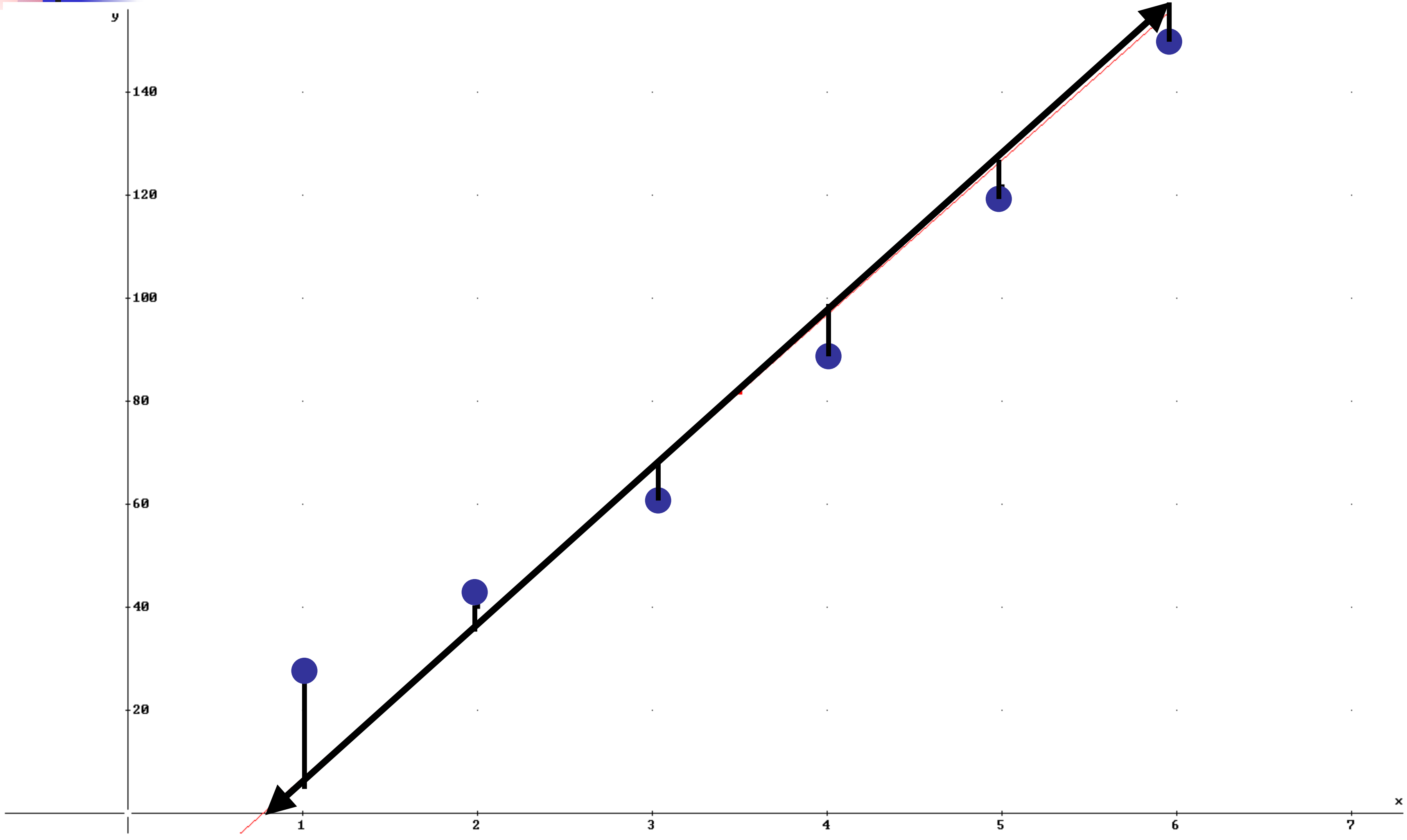
- Graphic Method – plot the data to see the trend
- Estimate a line of best fit
  - What point should the line pass through?
  - Solution: Midpoint (3.5,81.8)
  - Estimate the trend with approximate slope of m = 30

Year Scaled	Cars
1	27.5
2	40.3
3	61.7
4	89.3
5	121.6
6	150.5

Eyeball line of fit



Error in Model
Error = observed - predicted



Finding Total Error

- Why not just add the errors to get total error?
  - Solution: + and – values cancel out
- Square the differences to make them positive
- Sum the squares to find the total error
- The best fit line makes this error as small as possible

Mean Squared Error

- For a set of data (x,y) with n elements modeled by a line $\hat{y} = mx + b$

$$MSE = \frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2}{n}$$

- Standard Deviation: a measure of error found by square rooting the MSE to get a measure of error in the same dimension as the original data.

Mean Squared Error

$$MSE = \frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \cdots + (y_n - \hat{y}_n)^2}{n}$$

- Calculate the MSE for the Global warming data when modeled by the eyeball line of best fit

$$\hat{y} = 30x - 23.2$$

- Solution: MSE 98.3

Year	Cars Observed	Cars Predicted
1	27.5	6.8
2	40.3	36.8
3	61.7	66.8
4	89.3	96.8
5	121.6	126.8
6	150.5	156.8

Line of Best Fit

- Least Squares Fit Method – algebraic method of finding the best fit line
- This method gives a line which has the smallest possible mean squared error.
- Discuss why the method works
  - Verification on Pg 282-283

Line of Best Fit

- The coefficients a and b of the line of best fit $\hat{y} = a + bx$

$$b = \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{\sum x_i^{\,2} - n\,\bar{x}^2} \qquad\qquad a = \bar{y} - b\,\bar{x}$$

- CAS, Grapher, or Graphing Calculator can perform these tedious calculations

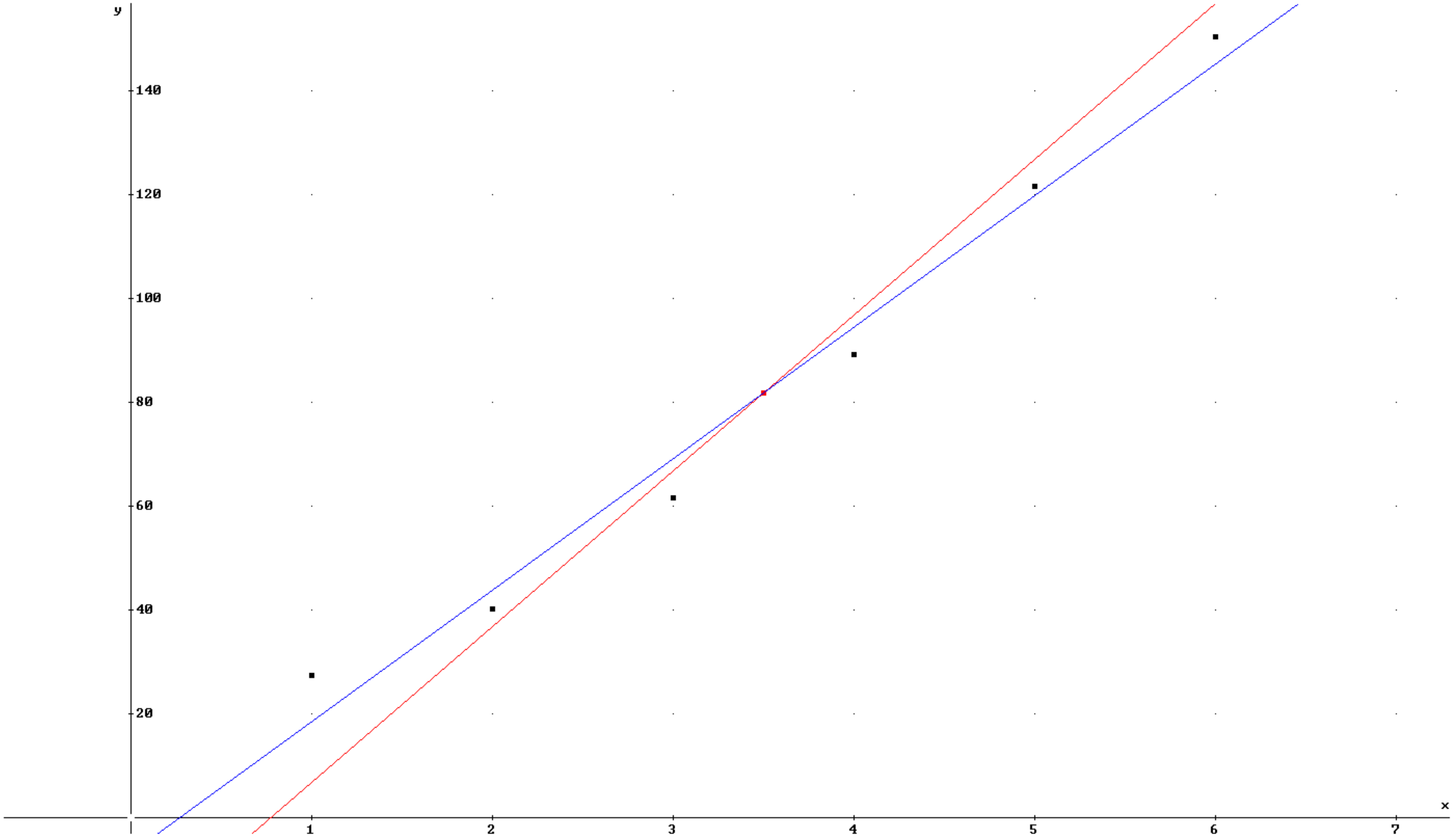
Line of Best Fit

- Derive's calculated line of best fit is

  $$y = 25.3\,x - 6.8$$

- Mean Squared Error is reduced from 98.3 for the eyeballed line of fit to 34.6 for the line of best fit.

Line of Best Fit (blue)



Polynomial of Best Fit

- Is the line of best fit a better model then a quadratic or cubic polynomial?
  - Examine the line of best fit with the data. Is the data curvilinear?
  - Solution: Data starts above line of best fit, then goes below and finishes back above – indication that data is curvilinear
- Extension of Linear Method to Other Polynomial Functions

Polynomial of Best Fit

- Derive calculated quadratic of best fit for the Global Warming Data

$$y(x) = 2.2x^2 + 9.8x + 13.9$$

- The MSE is reduced from 34.6 for the line of best fit to 2.02 for the quadratic of best fit

Polynomial of Best Fit