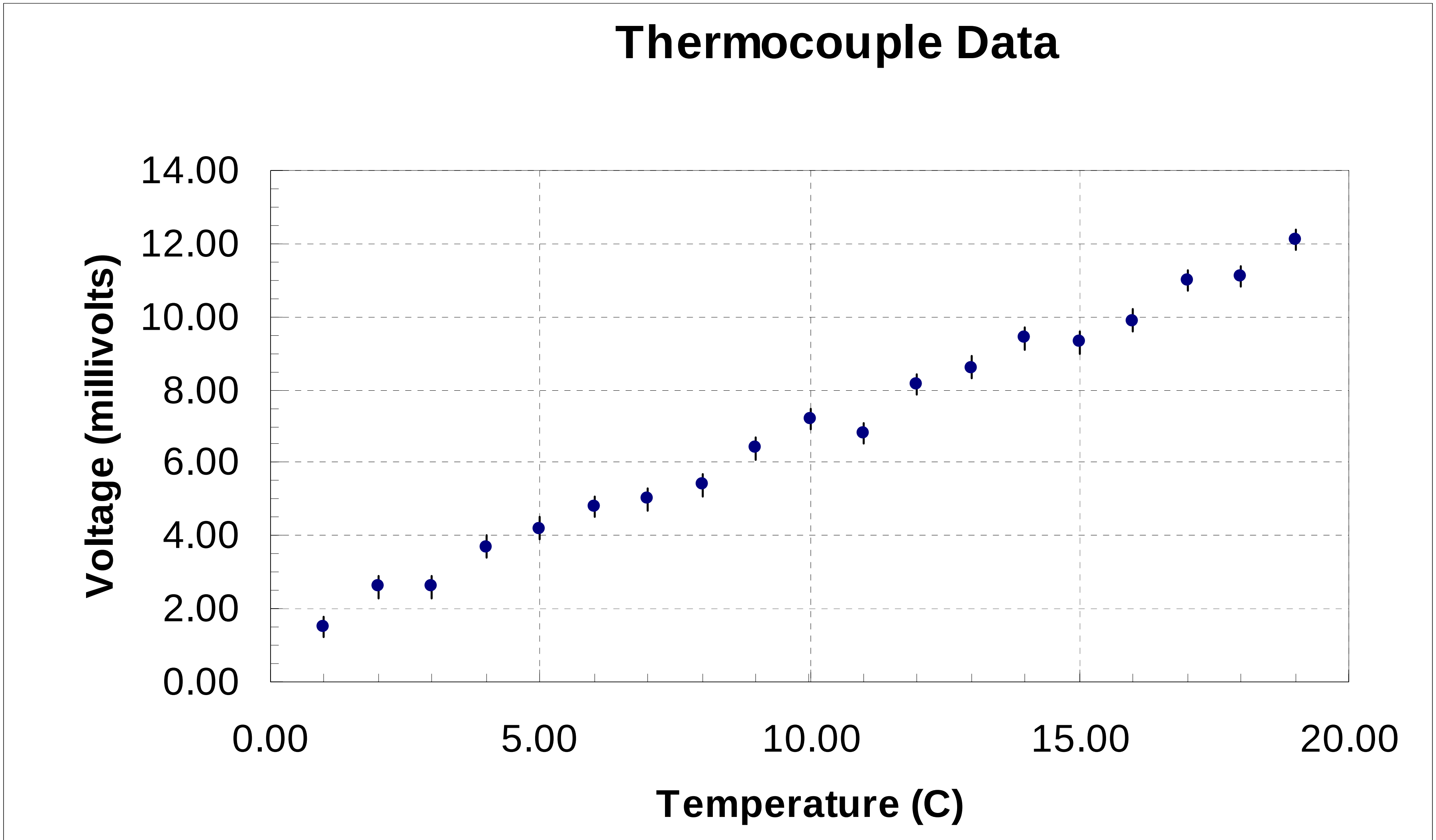
Week 3

Fitting "Theory" Curves to Data

Curve Fitting of Theory to Data- the Science and the Art of it

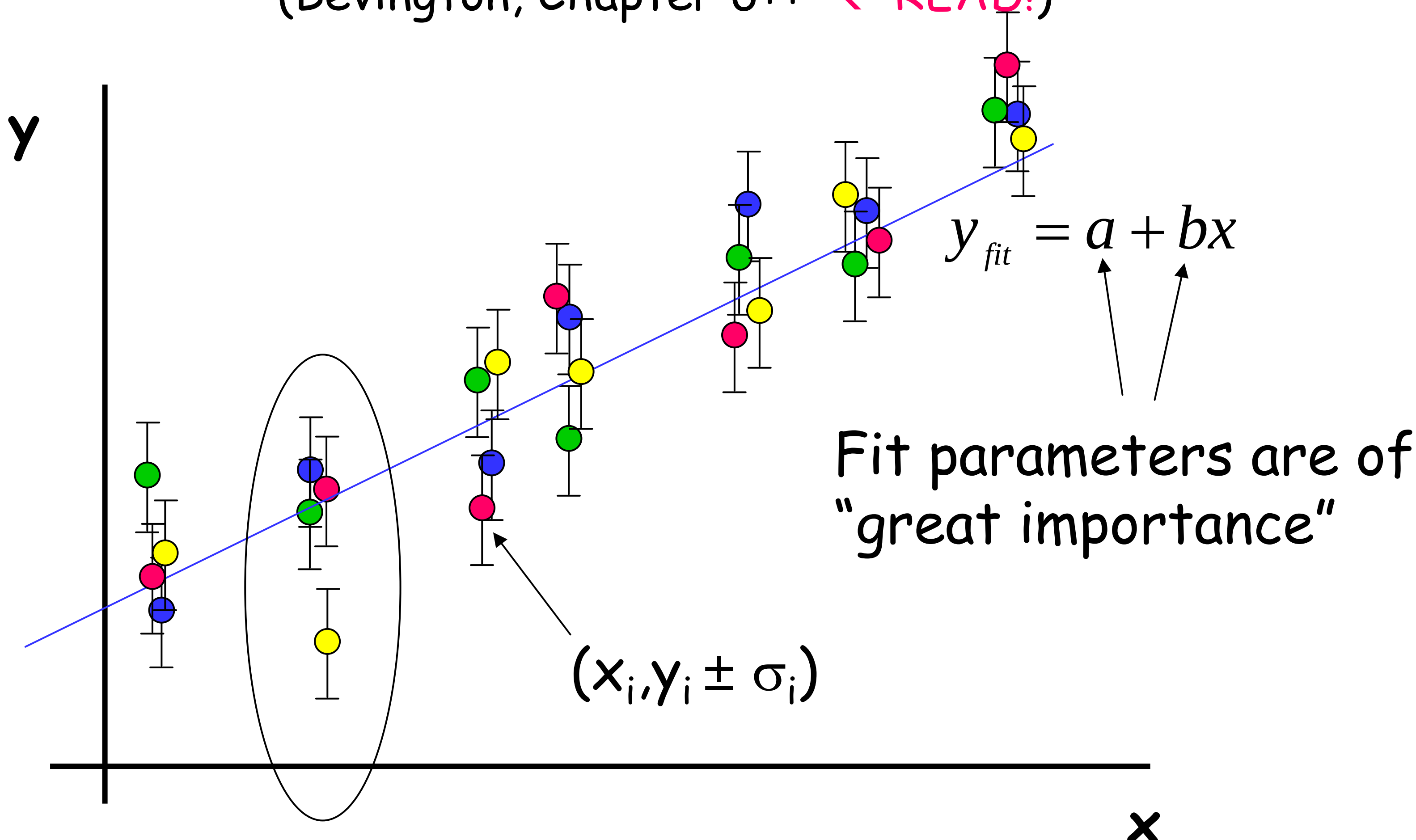
- "Science" part
  - Analytic formulas
  - Precise statements of goodness of fits
  - Algorithms for computation
- "Art" part
  - Judgment re parent functions & parameters
  - Judgment re goodness of results
  - Judgment re when to quit

Fit by eye-and-ruler:
determine the slope & intercept



Least-Squares Fitting

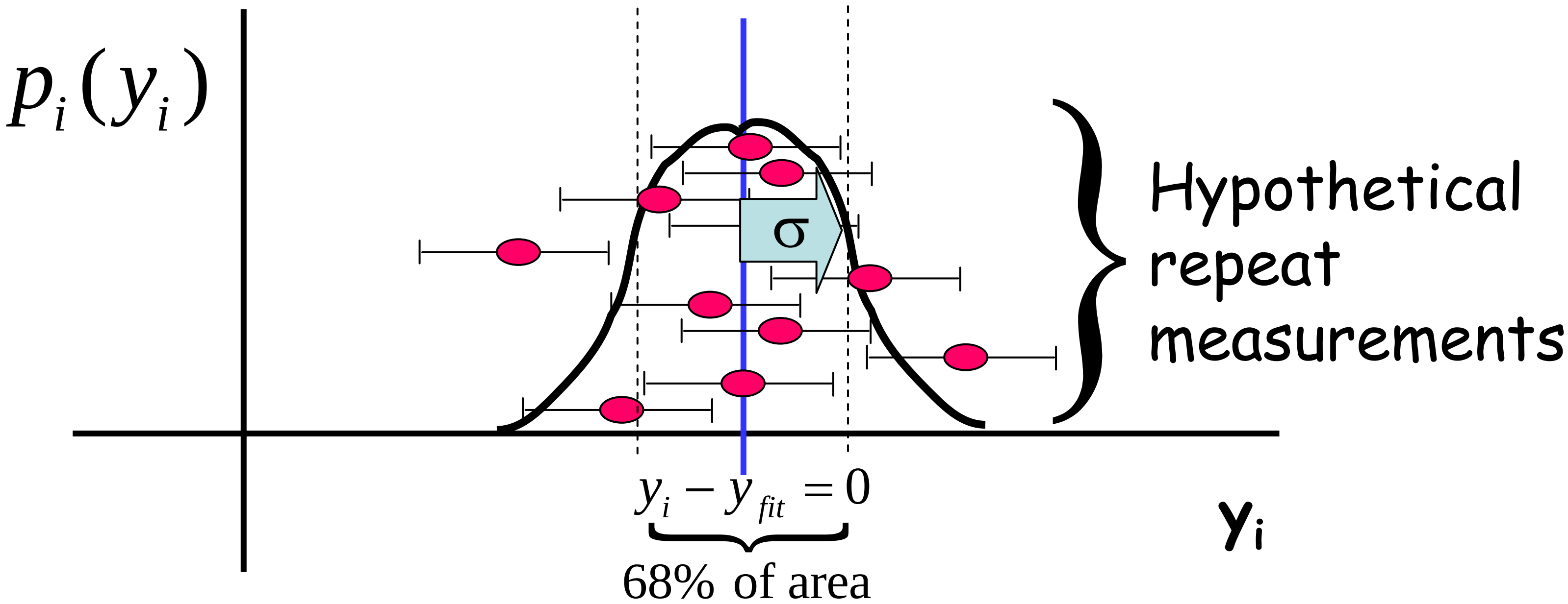
(Bevington, Chapter 6++ ← READ!)



Expect repeat measurement of i^{th} point to fall within $\pm 1\sigma$ about 2/3 times.

"Gaussian" Probability Density

for <u>one</u> data point



$$p_i(y_i) = \frac{1}{\sqrt{2\pi}}\frac{1}{\sigma_i}e^{-\frac{1}{2}\left(\frac{y_i - y_{fit}(x_i)}{\sigma_i}\right)^2}$$

THE probability that the "true" value lies between y_i and y_i+dy is

$$P_i(y_i, y_i + dy) = p(y_i)dy$$

Global Probability that {a,b} are the best parameters

Now if we have N data points stemming from <u>one</u> set of measurements, the overall probability of getting those values, given a and b, is

$$P(a,b) = \Pi_{i=1}^{N} p_i = \Pi_{i=1}^{N} \left\{ \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_i} e^{-\frac{1}{2}\left(\frac{y_i - y_{fit}(a,b,x_i)}{\sigma_i}\right)^2} \right\}$$

We <u>maximize</u> the likelihood of getting the meas-urements we obtained by varying parameters a and b

The trick to understanding "least squares":

$$P = \prod_{k=1}^{10} e^k = e^1 e^2 e^3 e^4 e^5 e^6 e^7 e^8 e^9 e^{10} = e^{\sum_1^{10} k} = e^{55} = huge$$

$$P(a,b) = \prod_{i=1}^{N} p_i = \left(\prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}} \frac{1}{\sigma_i} \right) e^{\underline{-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - y_{fit}(a,b,x_i)}{\sigma_i}\right)^2}}$$

Maximizing P(a,b) is equivalent to minimizing the exponent. We give it a special name: "chi-squared":

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y_{fit}(a,b,x_i)}{\sigma_i} \right)^2$$

Vary parameters to minimize χ^2

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - y_{fit}(x_i)}{\sigma_i} \right)^2$$

General form

$$\chi^2 = \sum_{i=1}^{N} \frac{1}{\sigma_i^{\,2}} \left(y_i - (a + bx_i) \right)^2$$

This case
(straight-line fit)

What are the "best" values?

Must estimate
these correctly
beforehand.

<u>Sometimes</u> an analytic solution exists… like in this case

$$\frac{\partial \chi^2}{\partial a} = 0 = \sum_{i=1}^{N} \frac{1}{\sigma_i^{\,2}} \left(y_i - (a + b x_i) \right)$$

$$\frac{\partial \chi^2}{\partial b} = 0 = \sum_{i=1}^{N} \frac{1}{\sigma_i^{\,2}} x_i \left(y_i - (a + b x_i) \right)$$

2 equations
2 unknowns

Solve via method of determinants

$$a = (\text{big formula involving sum over data points})$$

$$b = (\text{another big formula})$$

See Bevington Eq 6.13 !

Backup Slides…