自然科学の歩き方
第３回

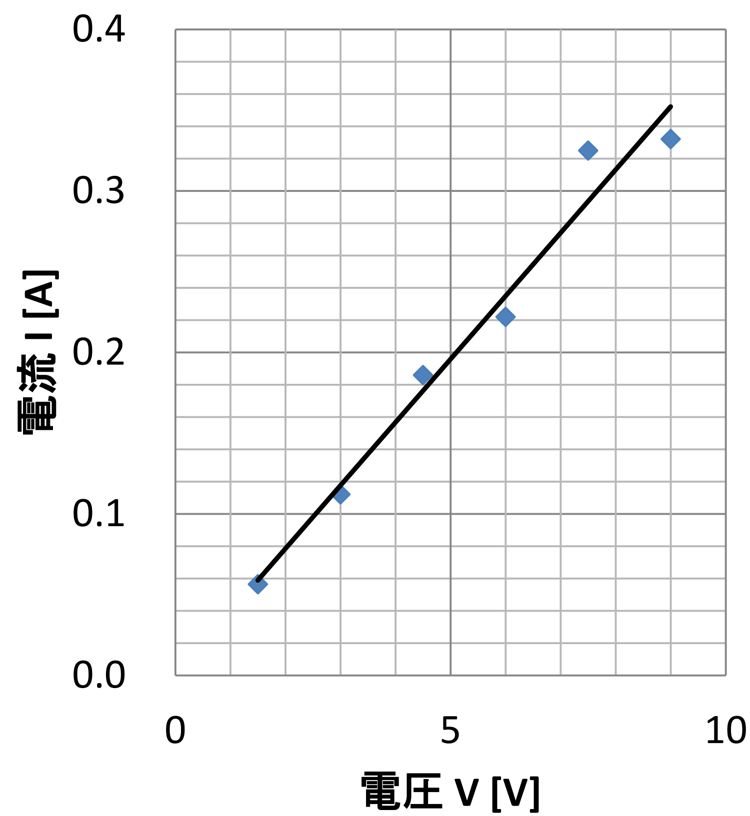
2019年

第1クォーター

金曜4クラス

前回の実習

- グラフに、モデルの「線」を引いた
- 引いた線とデータの間の二乗誤差を計算した
- …で、何がしたいの？



$$(\text{二乗誤差}) = \sum\left[(\text{モデルの値}) - (\text{データの値})\right]^2$$

モデルとは

- ここでは，なんらかの現象を記述する「数式モデル」に限定して考える。
- 例１ ：質点の自由落下での落下時間tと落下距離x　・・・$x = \frac{1}{2}gt^2$
- 例２： １モルの理想気体の圧力p，体積V，温度Tの関係 ・・・ $pV = RT$

変数とパラメータ

- 数式の中で，変化する量･･･ 変数
- 数式の中で，一定と考えられるもの ･･･ 定数あるいはパラメータ（parameter）

このあたりは，人によって感覚が違うが，例えば，円の面積の式で $S=\pi r^2$ で π は定数であろう。パラメータとは呼ばない。

パラメータとは，例えばあるマシンにいくつか調整用のダイヤルがついていて，それらをうまくチューニングして，快適な動作をするように設定したときのダイヤルの目盛りの値のようなイメージである。

良いモデルの条件

- 理論的な根拠がある
  ・・・ ただし，いつもそうではない。この話はとりあえずおいておく。

- 実験データをそれなりに再現する
  ・・・ そして，「予測」にも使える

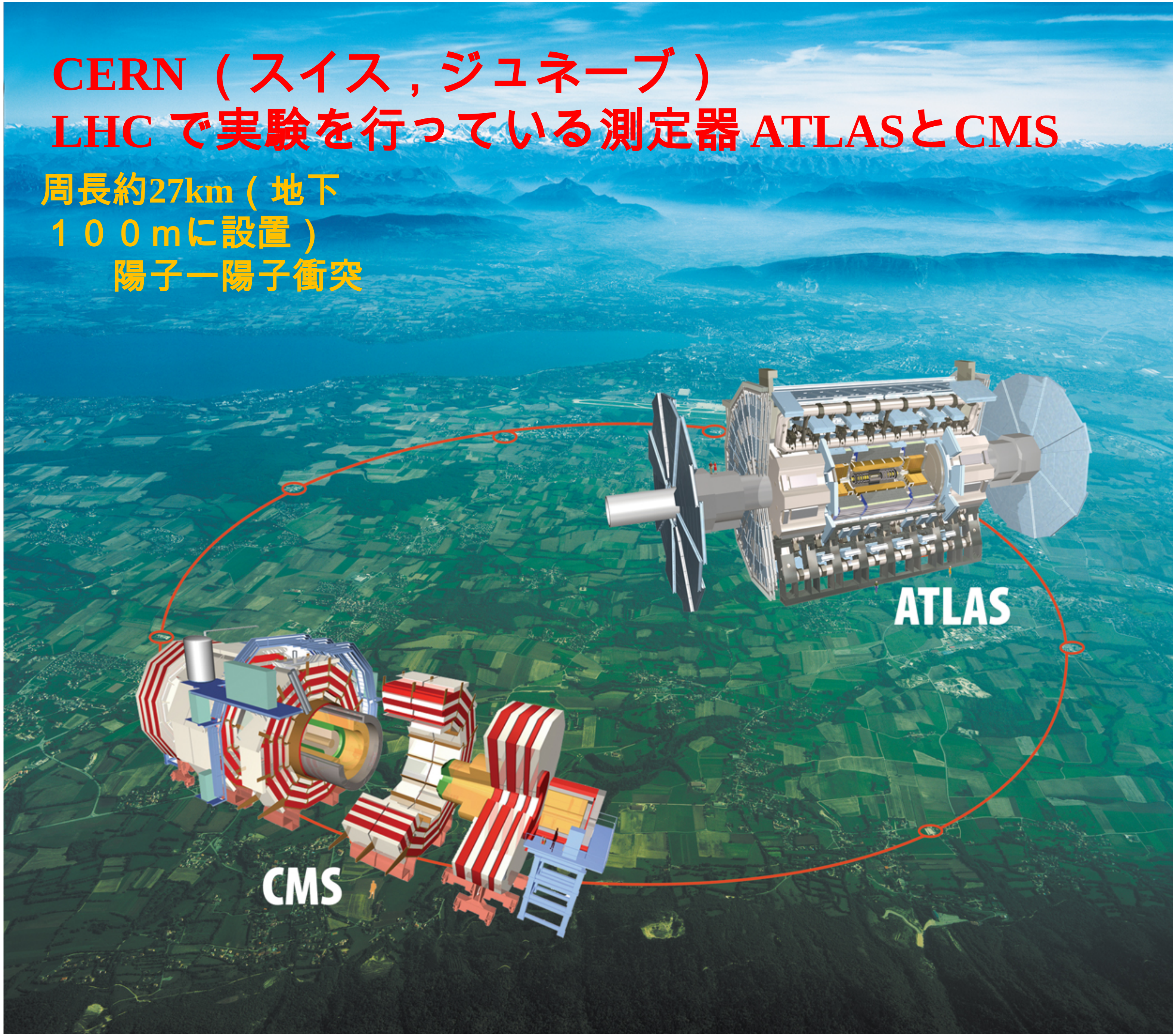
- パラメータが「ほどほどに」少ない
  ・・・オッカムの剃刀（Occam's razor）「ある事柄を説明するためには、必要以上に多くを仮定するべきでない」。

研究現場の一例

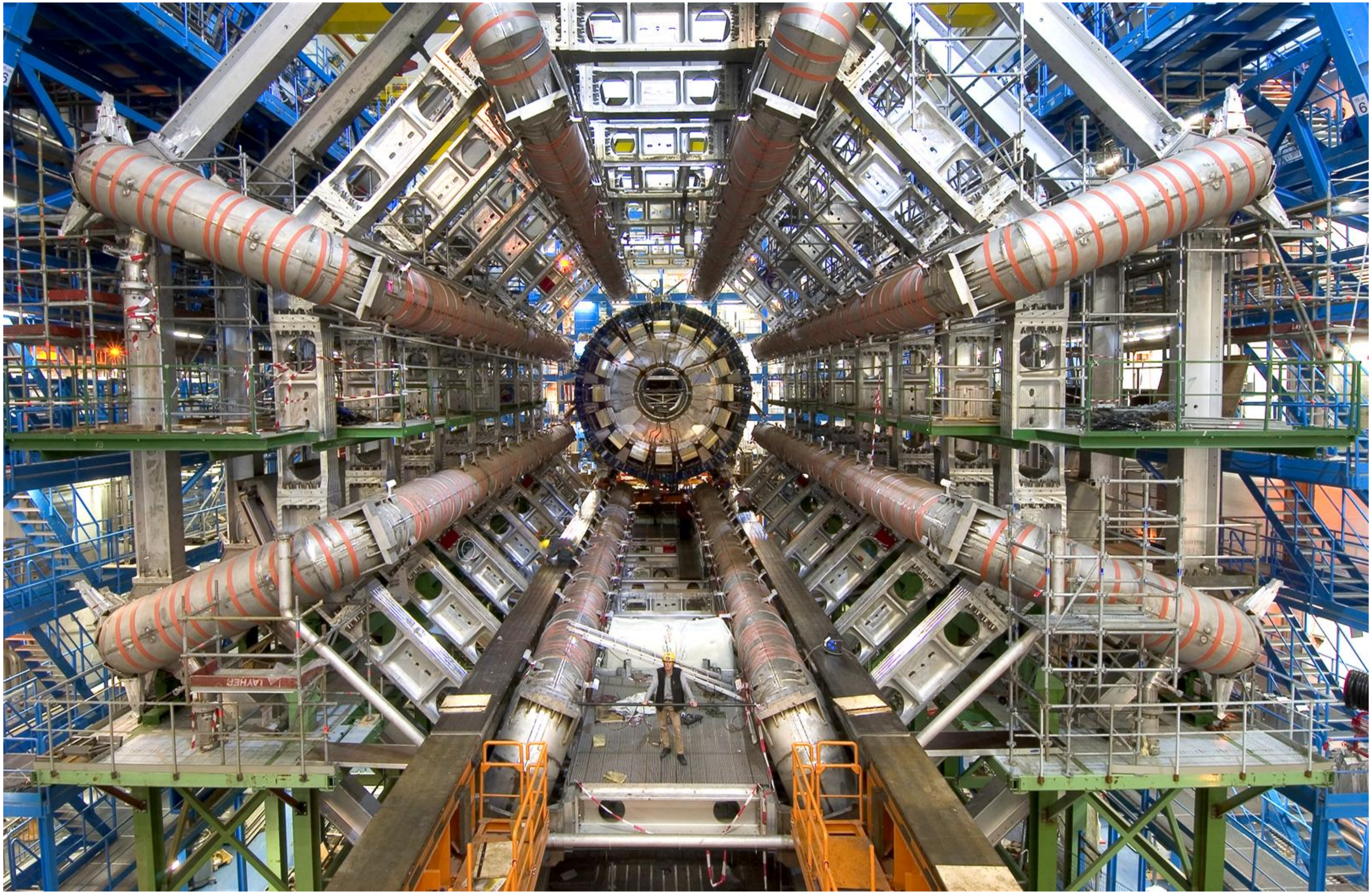
- スイス，ジュネーブにある，世界最大の加速器LHCでの実験
- １０００人を超える研究者チームが数年以上にわたって測定
- ２０１２年 ヒッグス粒子の発見
- ２０１３年ノーベル物理学賞

CERN（スイス，ジュネーブ）
LHCで実験を行っている測定器ATLASとCMS

周長約27km（地下
１００ｍに設置）
陽子ー陽子衝突

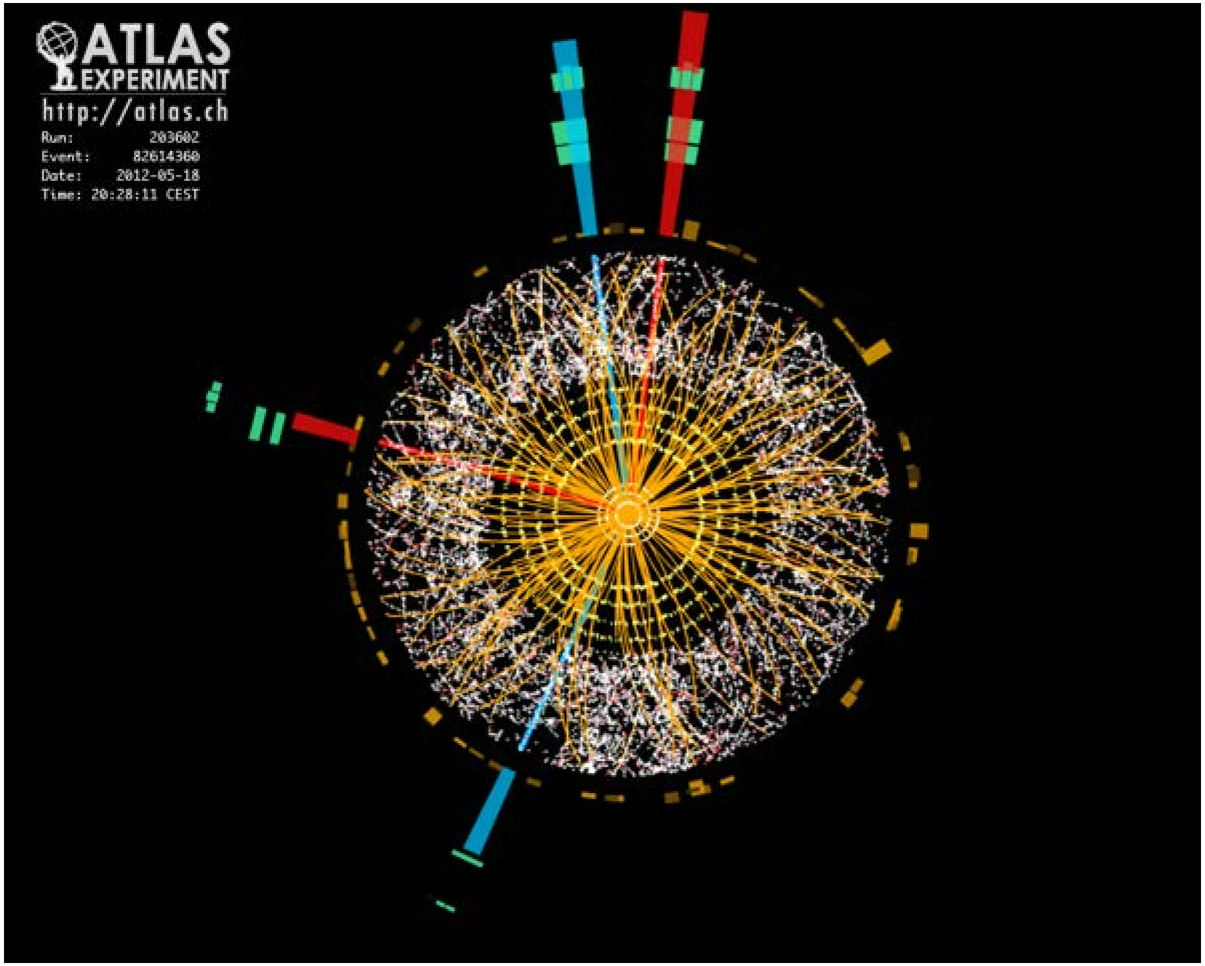


LHC:ATLAS測定器



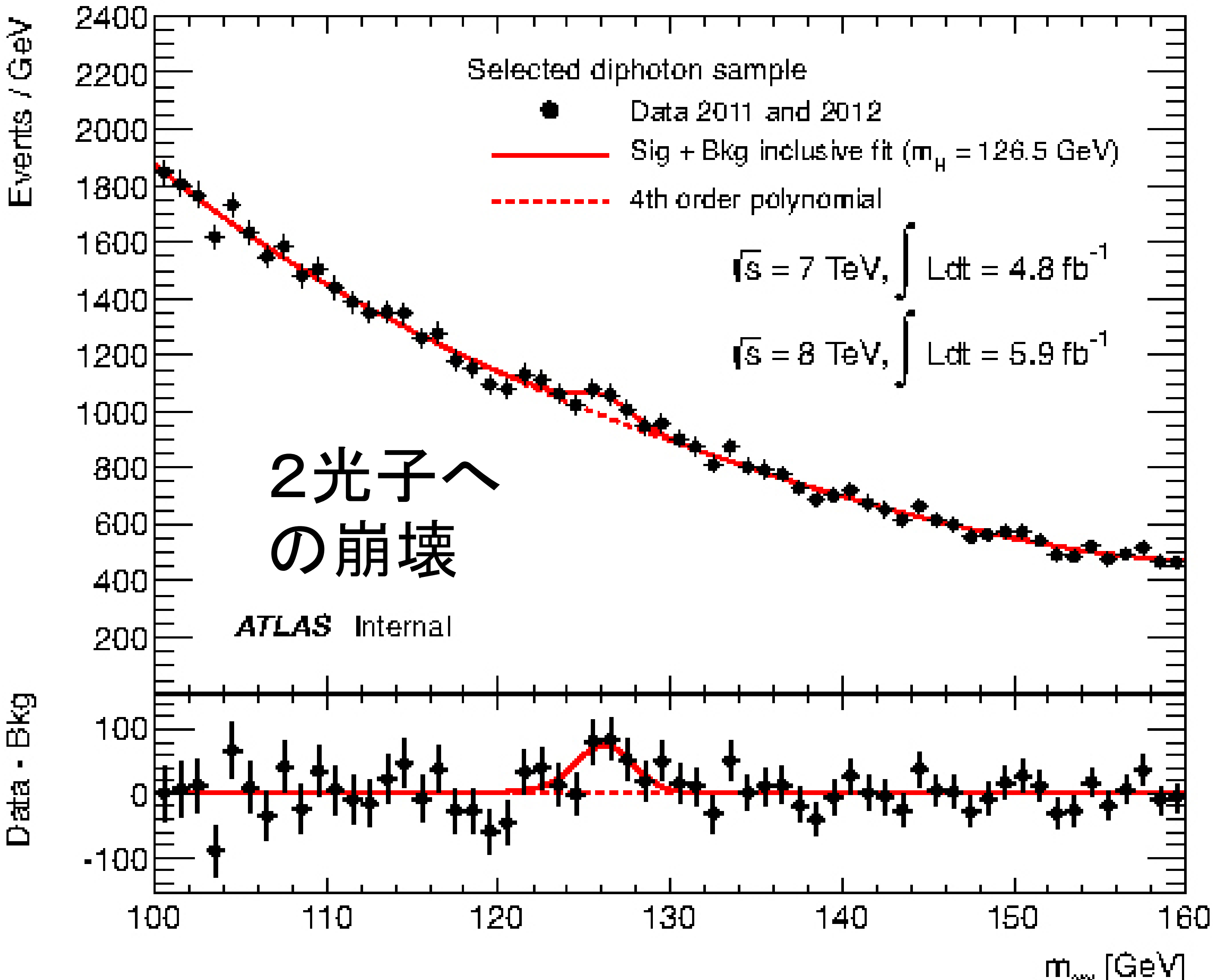
Installing the ATLAS calorimeter. The eight torodial magnets can be seen on the huge ATLAS detector with the calorimeter before it is moved into the middle of the detector. This calorimeter will measure the energies of particles produced when protons collide in the centre of the detector. November 2005

LHC:ATLAS測定器 で測定されたヒッグス粒子生成事象の一例



Event display of a H -> 4e candidate event with m(4l) = 124.5 (124.6) GeV without (with) Z mass constraint. The masses of the lepton pairs are 70.6 GeV and 44.7 GeV. The event was recorded by ATLAS on 18-May-2012, 20:28:11 CEST in run number 203602 as event number 82614360. The tracks and clusters of the two electron pairs are colored red and blue, respectively.

LHC:ATLAS測定器 で測定されたヒッグス粒子生成の分布の一例



例はここまで

パラメータ数 vs 再現性

- 実験データを再現するモデルは、原理的には、いくらでもある

例）　次のデータがある

$$x = 0 \Rightarrow y = 1 \quad x = 1 \Rightarrow y = 2$$

モデル1

$$\boxed{y = ax + b}$$

パラメータは a, b

モデル2

$$\boxed{y = a\sin x + b\cos x}$$

パラメータは a, b

$$\begin{cases} 1 = a \cdot 0 + b \\ 2 = a \cdot 1 + b \end{cases}$$

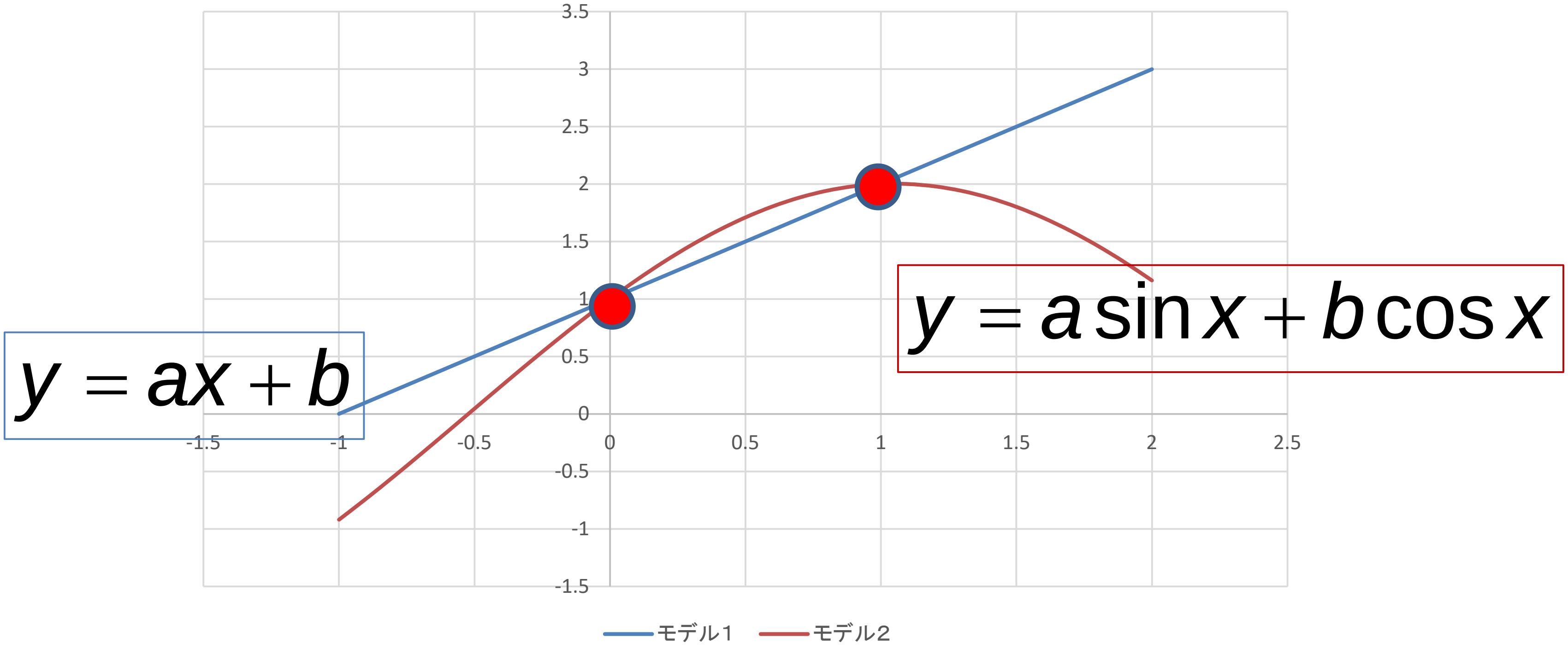
結果 $a = 1$

$b = 1$

$$\begin{cases} 1 = a \cdot \sin 0 + b \cdot \cos 0 \\ 2 = a \cdot \sin 1 + b \cdot \cos 1 \end{cases}$$

結果 $a = 1.735$

$b = 1$



パラメータ数 vs 再現性

- N個のデータ点に対し、（N－１）次関数を用いれば、必ずデータに合うモデルを作れる

ｎ 個の数値のデータがあるとする。

$$(x_1, y_1),\ (x_2, y_2),\ (x_3, y_3), \cdots, (x_n, y_n)$$

これに対して以下の式を「モデル」とする。

$a_0, a_1, a_2, \cdots, a_{n-1}$ は「未知の」パラメータである。

$$\boxed{y = a_0 + a_1 x + a_2 x^2 + \cdots + a_{n-1} x^{n-1}}$$

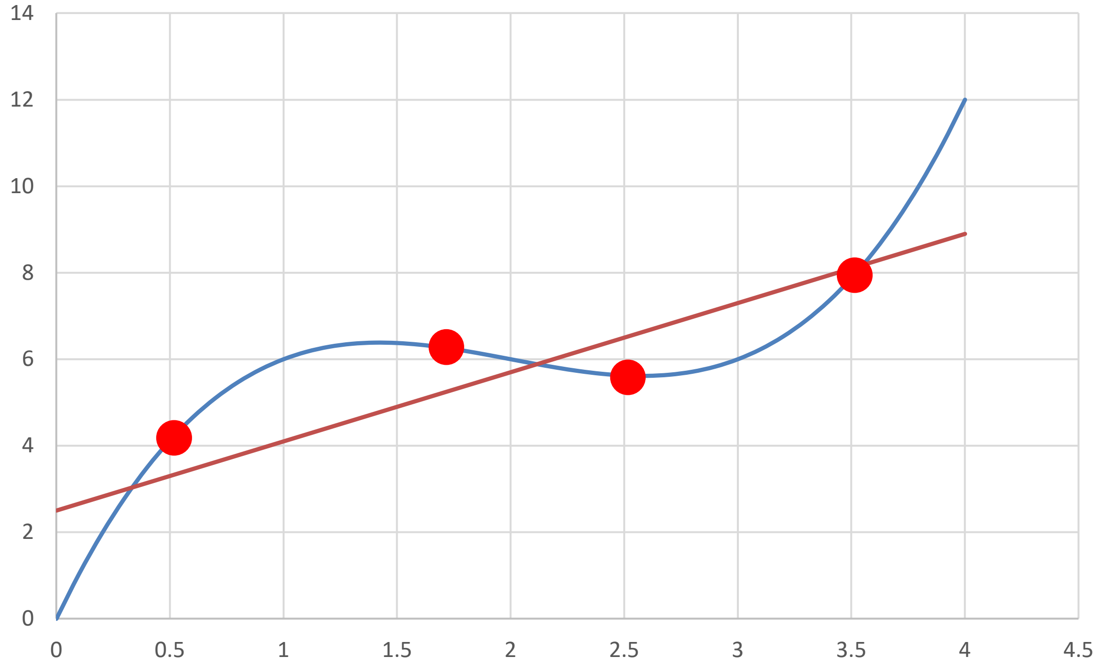
以下の連立方程式を解き $a_0, a_1, a_2, \cdots, a_{n-1}$ を求める。

$$\begin{cases} y_1 = a_0 + a_1 x_1 + a_2 x_1^2 + \cdots + a_{n-1} x_1^{n-1} \\ y_2 = a_0 + a_1 x_2 + a_2 x_2^2 + \cdots + a_{n-1} x_2^{n-1} \\ \vdots \\ y_n = a_0 + a_1 x_n + a_2 x_n^2 + \cdots + a_{n-1} x_n^{n-1} \end{cases}$$

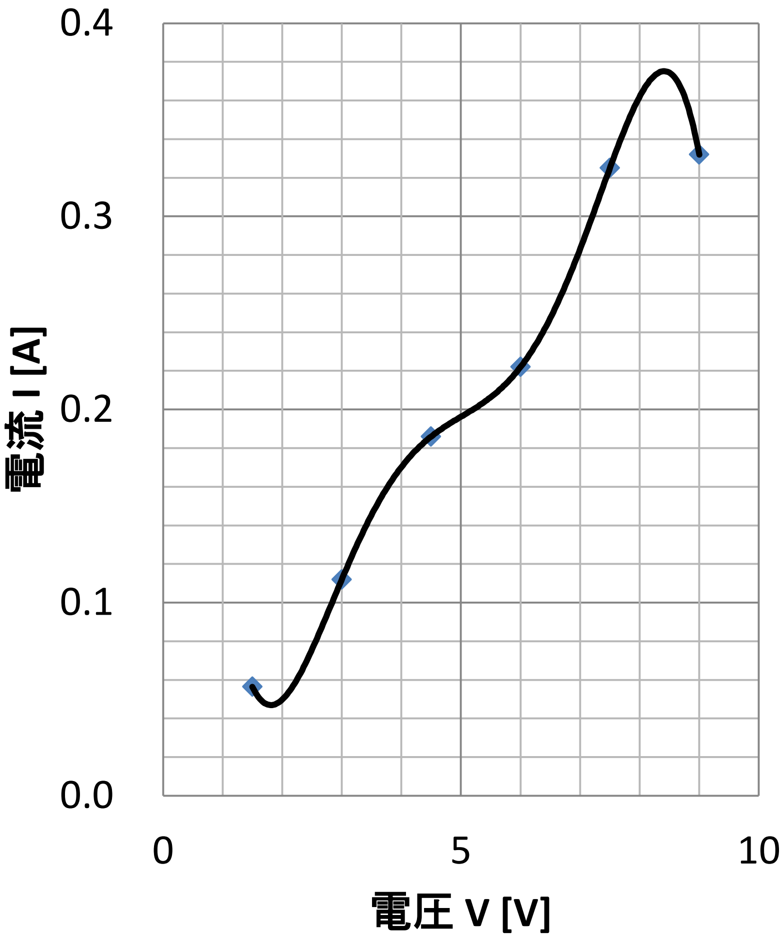
→ パラメータ $a_0, a_1, a_2, \cdots, a_{n-1}$ が決まった。

完全に元のデータを再現する「モデル」が出来上がった

４つのデータ点（●）にぴったり合う3次関数（４パラメータ）と，だいたい傾向があう1次関数（２パラメータ）



今回の実験データ



5次関数を使うと「ぴったりと」あわせることができるが，どうみても不自然。

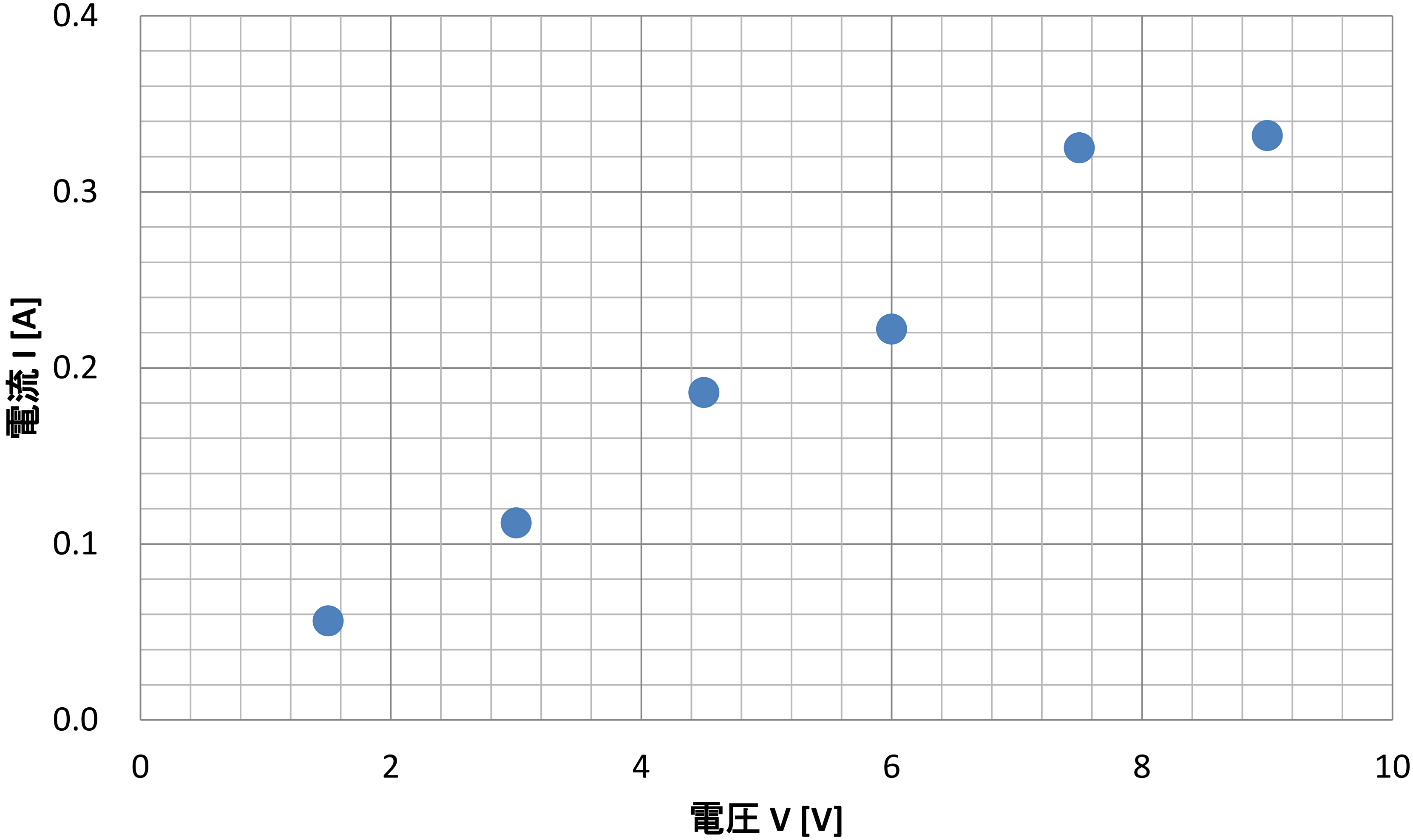
「合いすぎる」のは何か異常がある兆候

モデルの合わせ方

- モデルを一つ決める
  - ここは、知識＋センス＋経験…
  - 「職人の技」・「秘伝のたれ」など

- 一つ決めたモデルの中で、パラメータを変える
- 二乗誤差が小さくなるパラメータが、最も良さそう
  - この計算は、（効率を考えなければ）機械的な作業
  - まずは、この計算をしっかりできるようになりましょう

実習で使っているデータ



モデル

1次関数でどうか。

$$I = aV + b$$

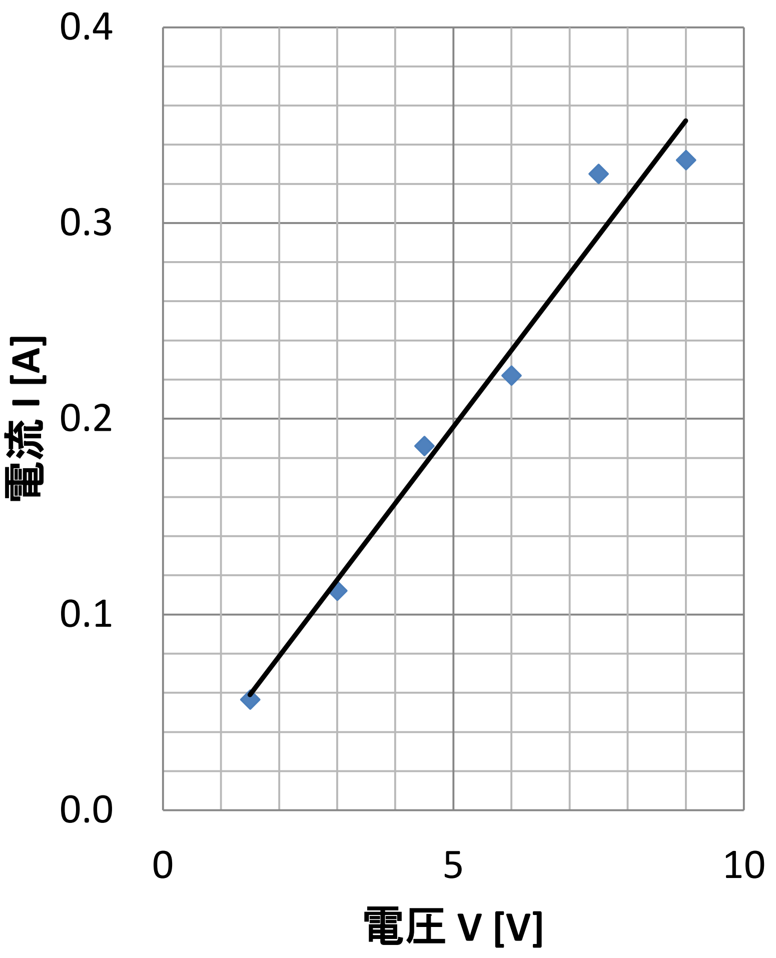
理論的考察

$V=0$ なら $I=0$ になるんじゃないか？

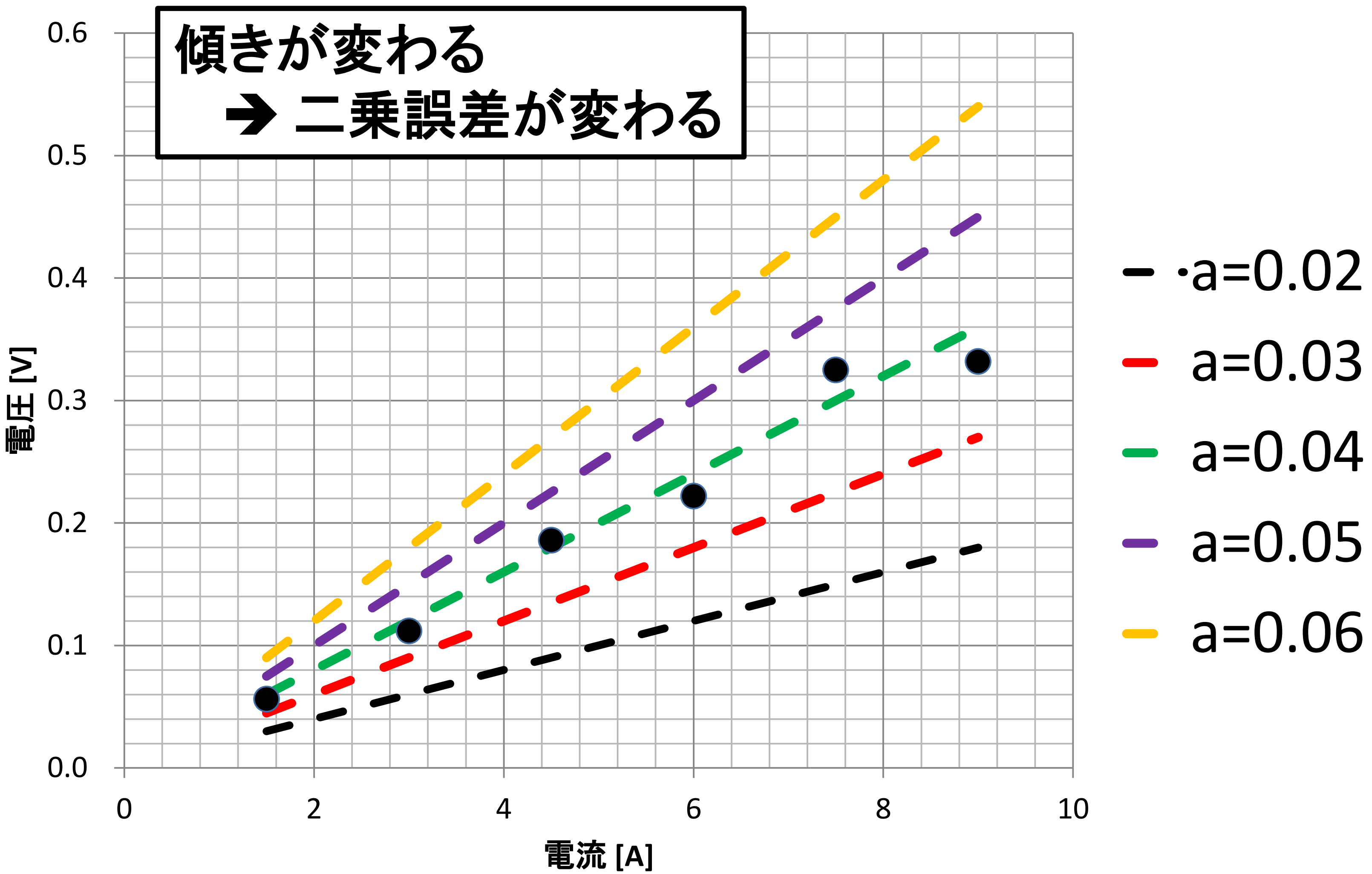
モデル（1パラメタ）

$$\boxed{I = aV}$$

最適なパラメタ（a）は？



傾きaを変えてみる



傾きが変わる
➔ 二乗誤差が変わる
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0
2
4
6
8
10
電圧 [V]
電流 [A]
a=0.02
a=0.03
a=0.04
a=0.05
a=0.06

実習：最適パラメータの推定

- 傾きを何通りか変え、それぞれについて二乗誤差を計算 a=0.02, 0.03, 0.04, 0.05, 0.06

- 二乗誤差が最も小さくなったａについて、I=aVの線を、グラフに**実線で**記入

以下の表に数値を記入していき，5つのaの値の場合に二乗誤差Eを計算する。

電圧 [V]	1.50	3.00	4.50	6.00	7.50	9.00	E 二乗誤差
電流 [A]	0.0564	0.112	0.186	0.222	0.325	0.332	
モデルでの電流 $I = aV$							
a=0.02							
a=0.03							
a=0.04							
a=0.05							
a=0.06							

$$\overset{E}{(\text{二乗誤差})} = \sum \left[(\text{モデルの値}) - (\text{データの値}) \right]^2$$

測定点の和