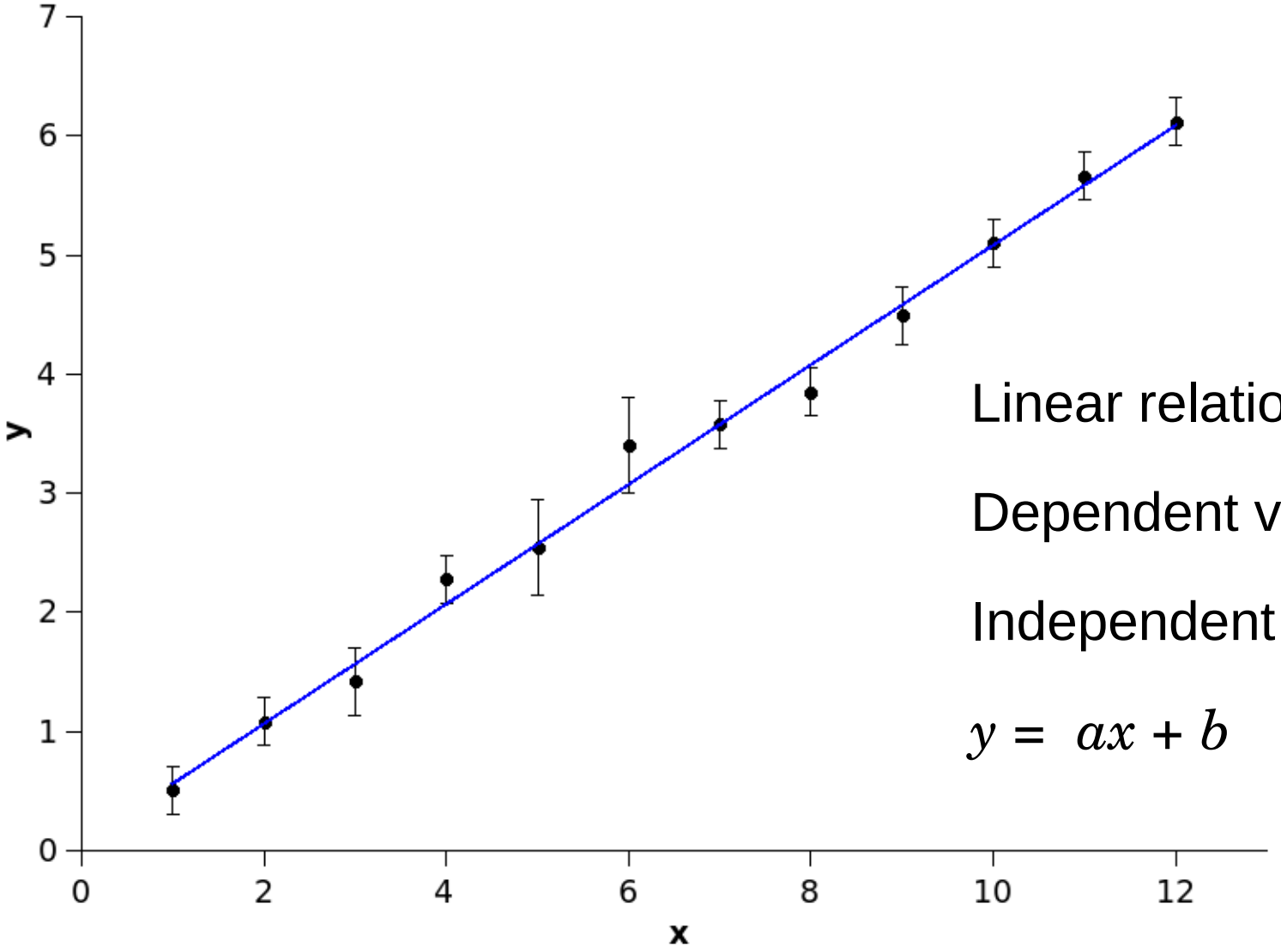
LECTURE 5: Least Squares Fit

Least Squares Fit



Linear relation between x and y

Dependent variable: y

Independent variable: x

$y = ax + b$

Least Squares Fit:
Most common method for Linear Regression

Simple Linear Regression:

Statistical analysis of a data set

x: independent variable

y: dependent variable

Generate a linear function $y = \alpha x + b$ that describes the data

Minimize the (distance)^2 between the line and data points

Least Squares Fit

Assumptions:

Uncertainty only of the dependent variable: $y \pm \Delta y$

x is known with much higher certainty than y

x error bars negligible

Uncertainty Δy for each data point is identical

Data distribution is Gaussian

Result:

Linear fit obtained with just the coordinates of data points: x_i, y_i

Least Squares Fit

$$y = ax + b$$

$$a = \frac{1}{\Delta}\left(\sum x_i^2 \sum y_i - \sum x_i \sum x_i y_i\right)$$

$$b = \frac{1}{\Delta}\left(N \sum x_i y_i - \sum x_i \sum y_i\right)$$

$$\Delta = N \sum x_i^2 - \left(\sum x_i^2\right)^2$$

Graphing software can perform this fit automatically

Least Squares Fit: General Case

Each data point y_i has a unique uncertainty $\Delta y_i = \sigma_i \neq 0$

Uncertainty Δy for each data point is **NOT** identical

$$y = ax + b$$

$$a = \frac{1}{\Delta}\left(\sum \frac{x_i^2}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2}\right)$$

$$b = \frac{1}{\Delta}\left(\sum \frac{1}{\sigma_i^2} \sum \frac{x_i y_i}{\sigma_i^2} - \sum \frac{x_i}{\sigma_i^2} \sum \frac{y_i}{\sigma_i^2}\right)$$

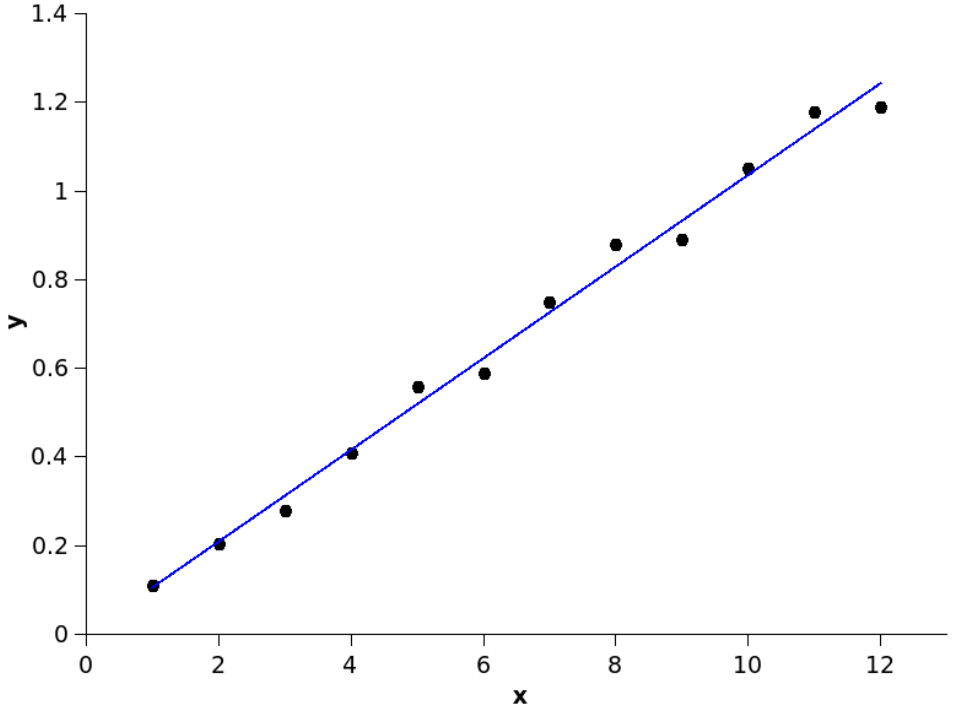
$$\Delta = \sum \frac{1}{\sigma_i^2} \sum \frac{x_i^2}{\sigma_i^2} - \left(\sum \frac{x_i^2}{\sigma_i^2}\right)^2$$

Graphing software can perform this fit automatically

Every data point has same y-error

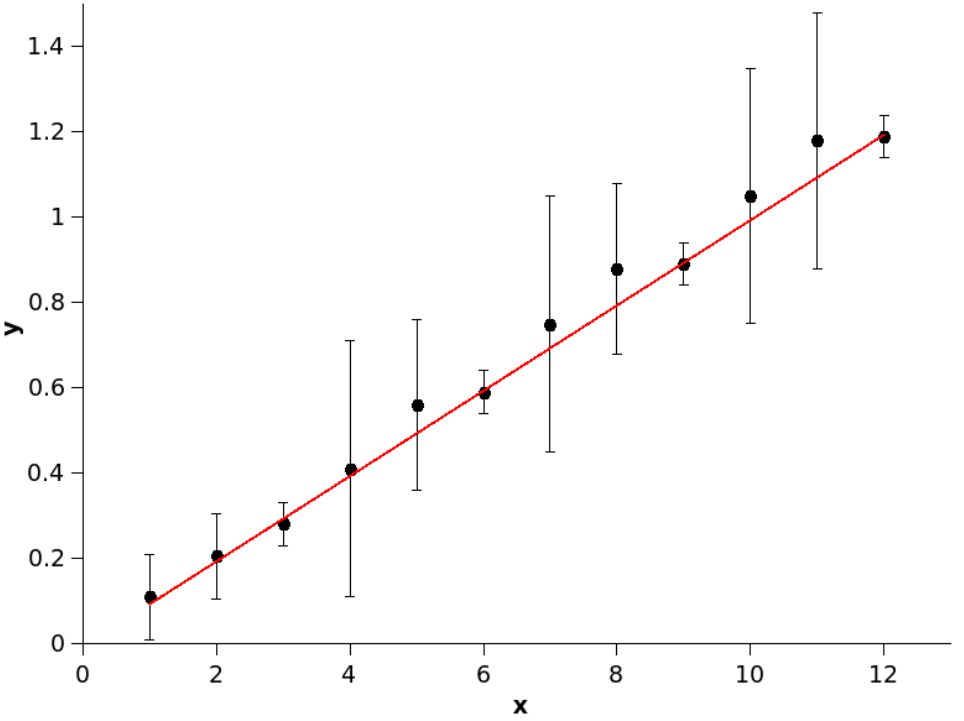
Least squares fit:
a = 0.103; b = +2.65 x 10^{-3}

$$y = ax + b$$



Identical data but with varying y-error bars

Least squares fit:
a = 0.10; b = –7.2 x 10^{-3}



Coefficient of Determination: R^2

Variation of the data (y) from the linear regression line (f)

How well does the calculated line fit the data?

Total variation between the y-data and line

$$R^2 = 1 - \frac{\sum (y_i - f_i)^2}{\sum (y_i - \overline{y})^2}$$

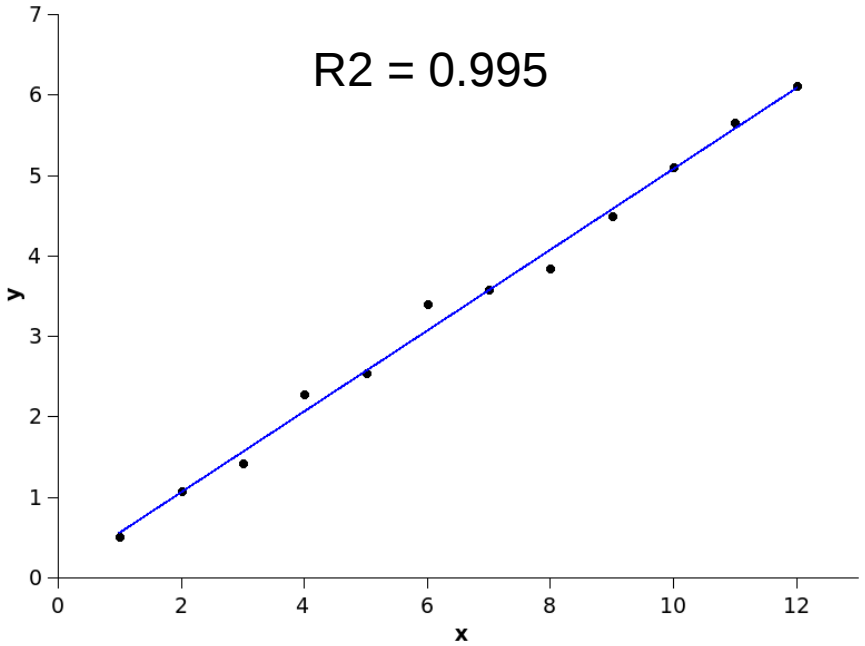
Total variation of the y-data values from the average y

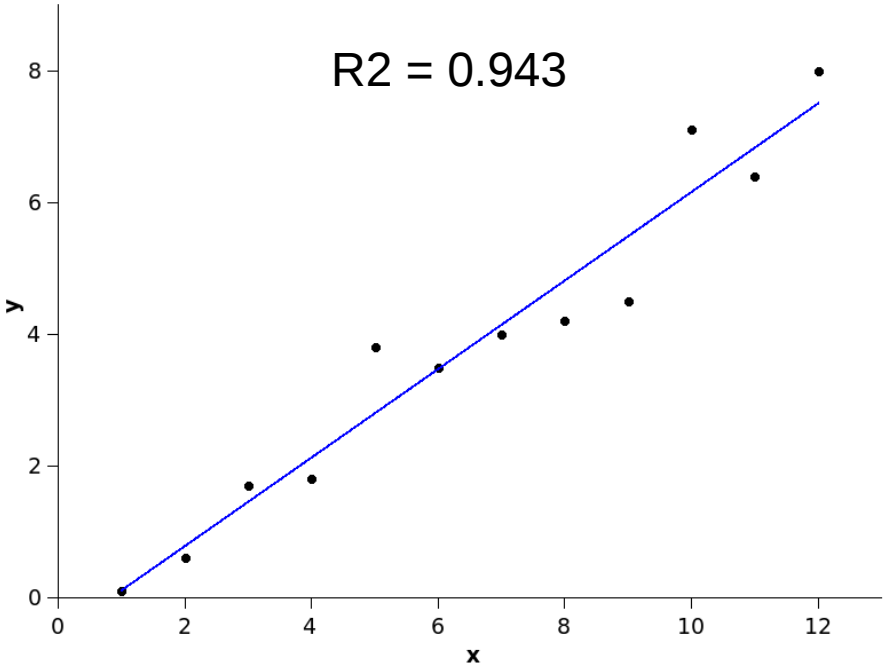
y_i: Data points

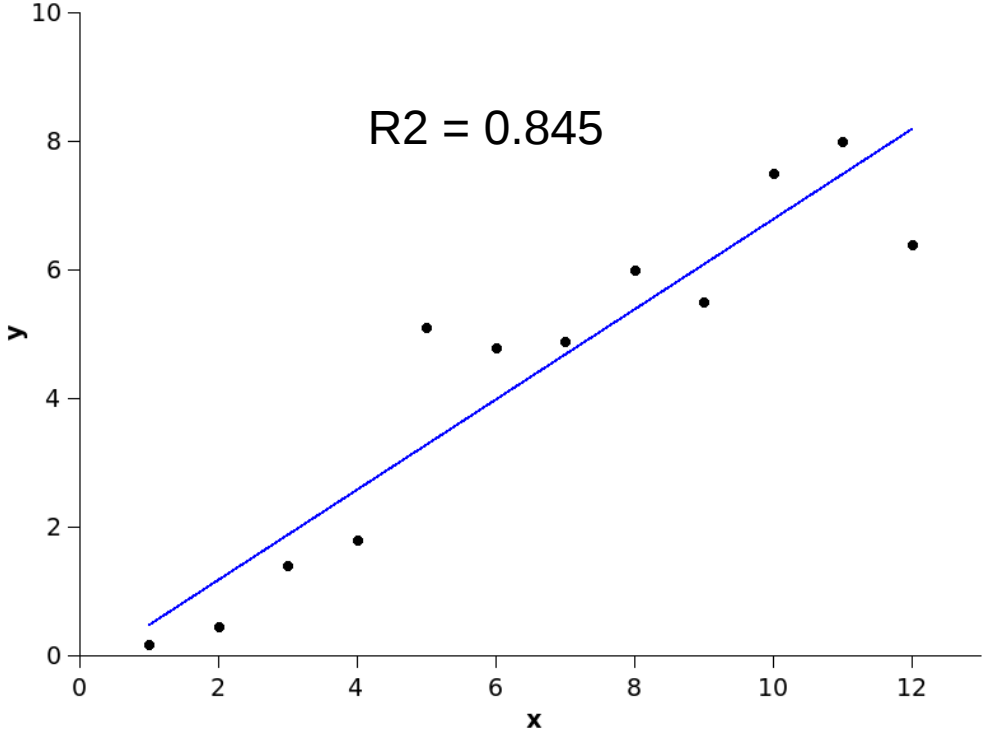
f_i: Points defining fitting line

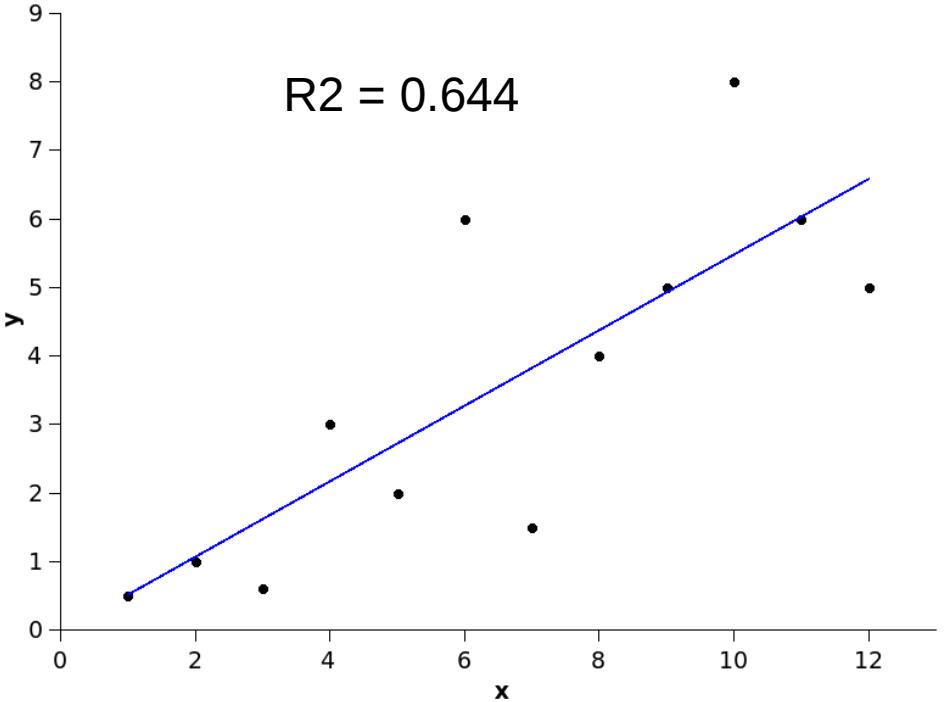
$\overline{y}$: Average of all the data points

R^2: Percentage of data that can be described by the regression line


R2 = 0.995
y
x
0
1
2
3
4
5
6
7
0
2
4
6
8
10
12


R2 = 0.943
y
x
0
2
4
6
8
0
2
4
6
8
10
12


R2 = 0.845
y
x
0
2
4
6
8
10
0
2
4
6
8
10
12


R2 = 0.644
y
x
0
1
2
3
4
5
6
7
8
9
0
2
4
6
8
10
12

GOODNESS OF FIT

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

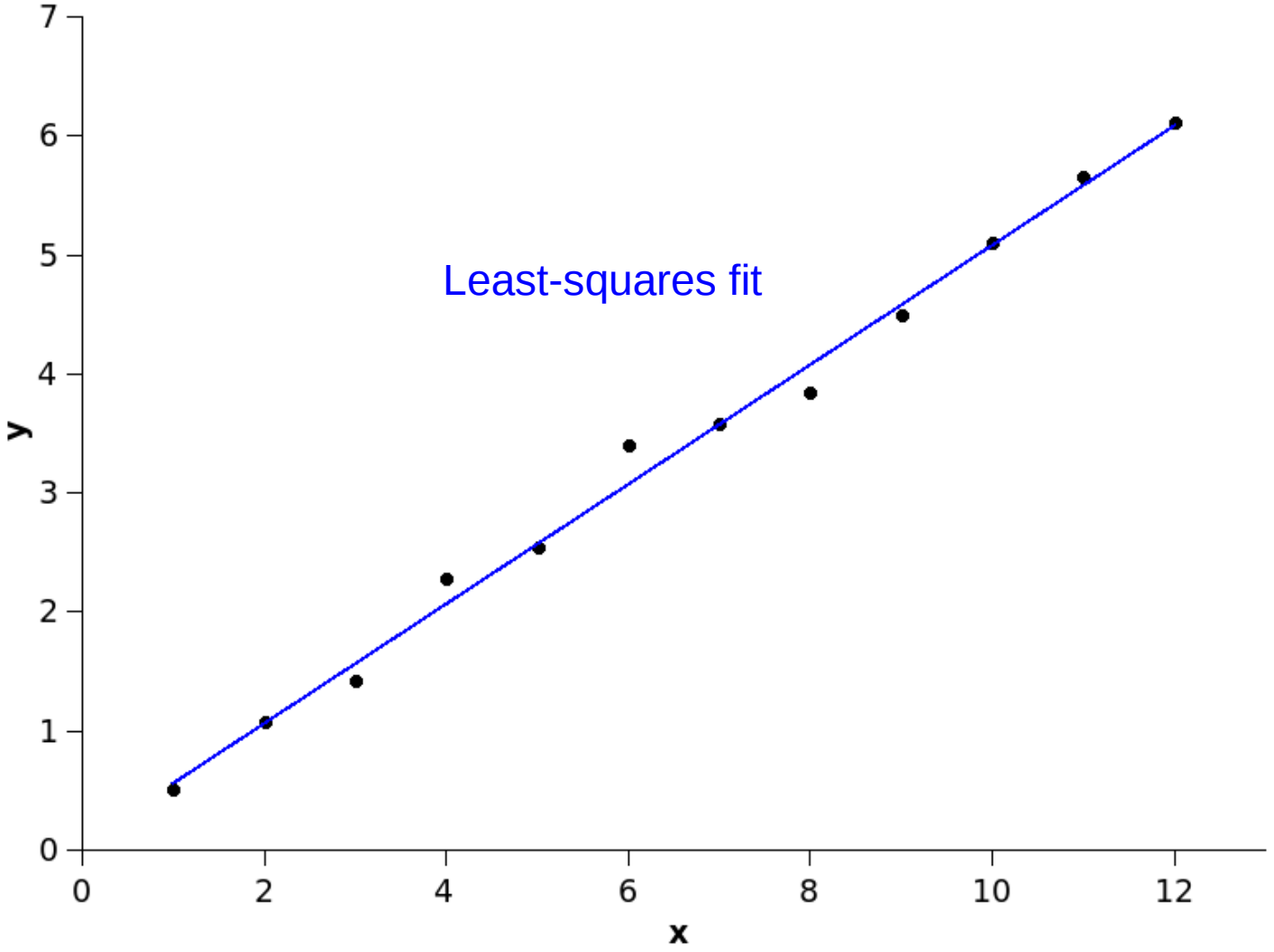
O: Observed value, data

E: Expected value from theory/model

Used with Binomial/Poisson distributions

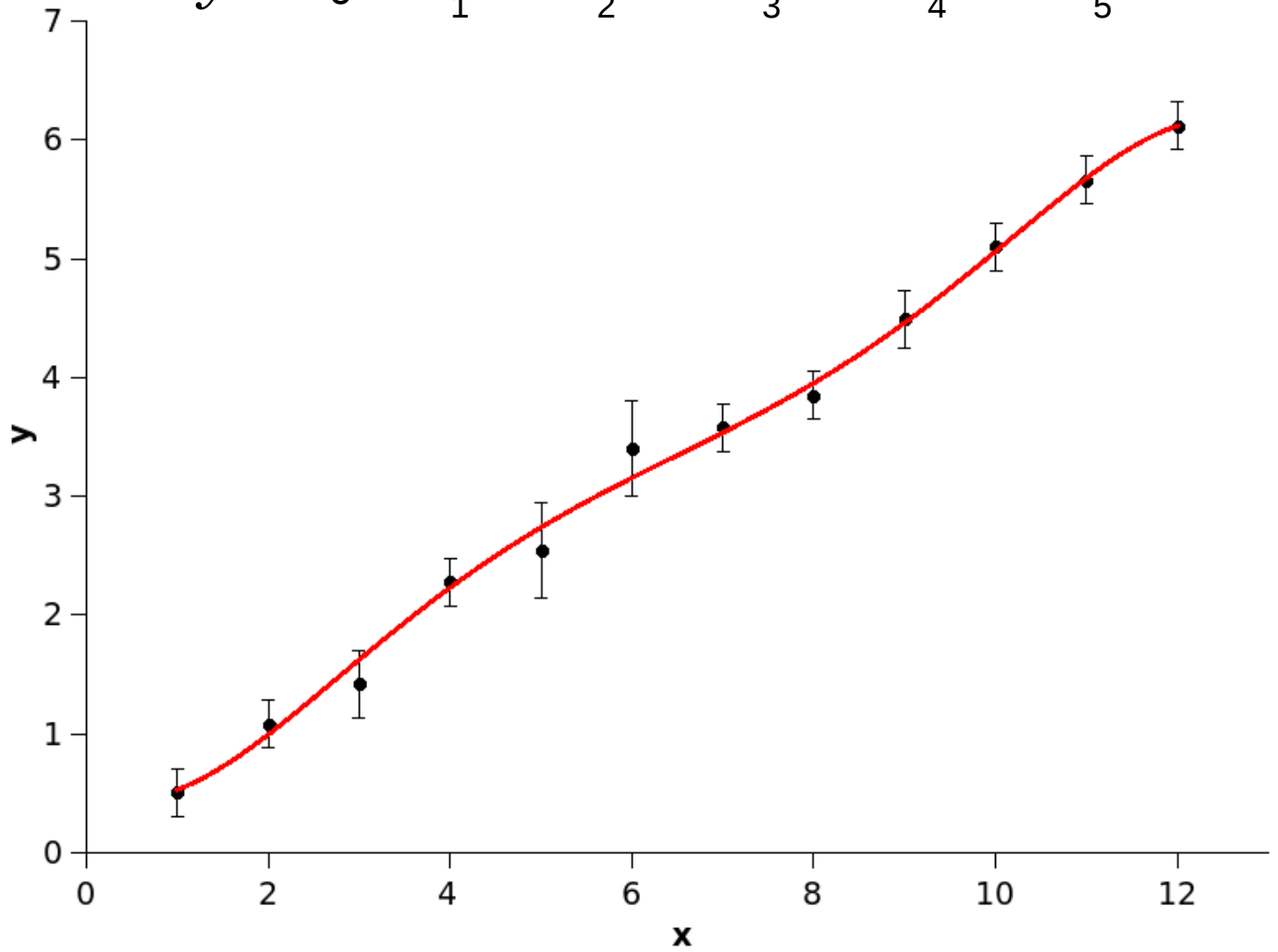
How many degrees of freedom in the experiment? (Histogram bins -1)

Better fits with more complicated functions?



Same Data: 5th order polynomial fit

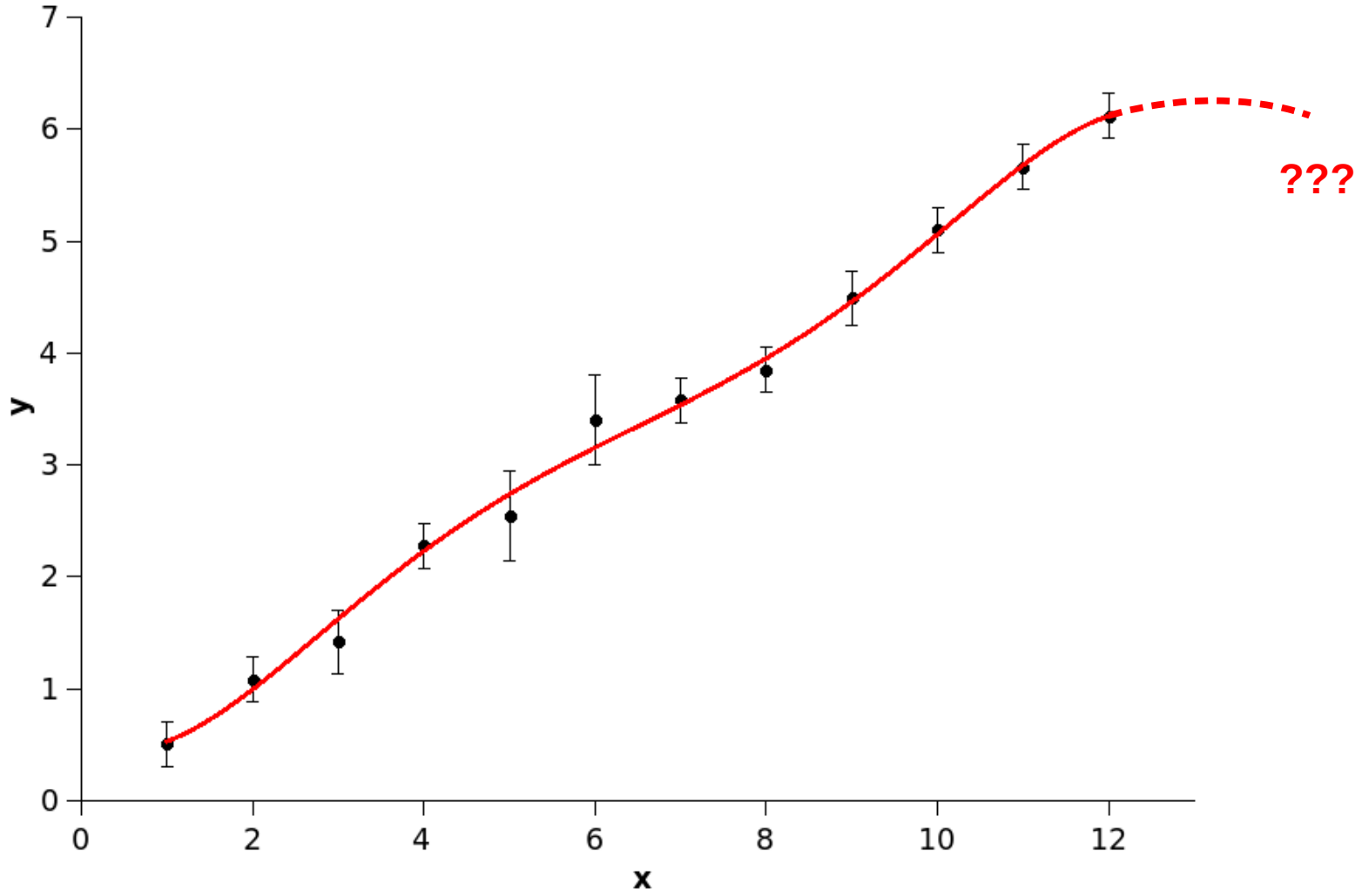
$$y = c_0 + c_1 x + c_2 x^2 + c_3 x^3 + c_4 x^4 + c_5 x^5$$



Does this fit have physical significance?

Do coefficients c_i provide any physical insight?

Extrapolation beyond endpoints may be inaccurate



Lab Report vs Scientific Paper

ABSTRACT

Lab Report: “In this lab, we studied the charge of the electron using the Millikan oil drop apparatus.”

Science Paper: “The elementary charge of an electron has been measured as $1.6 \pm 0.1 \times 10^{-19}$ C.”

EXPERIMENT

Lab Report: “We first tried a multi-meter and then decided it was better to use an oscilloscope to measure the time more accurately.”

Science Paper: “An oscilloscope was used to measure charging times with an estimated resolution of ± 0.1 s.”

Lab Report vs Scientific Paper

Lab Report:

- Chronology
- Following steps in a recipe: “Next we did this...”
- Minutiae included

Science Paper:

- Enough information for understanding
- Only include details if necessary
- Concise, Quantitative, Precise
- Avoid phrasing like:

  ...significantly better, reasonably good, seems, about,

  more or less, somewhat, hopefully...

Lab Report vs Scientific Paper

Writing quality science papers is ***NOT EASY***

Like any skill, scientific writing takes time to develop

Study published physics papers:
Abstracts, Intros, Graphics, References

“My results are awesome. Let my work speak for itself. And I'm not an English major!”

Journals routinely reject manuscripts for poor scientific writing and presentation

Poor communication can hinder career advancement