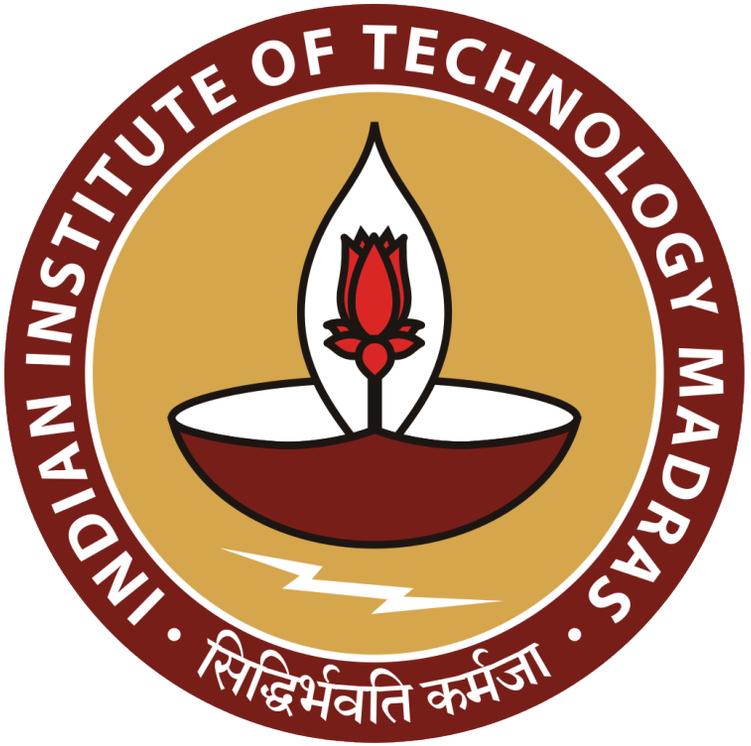


IIT Madras

ONLINE DEGREE

Example:Real-World

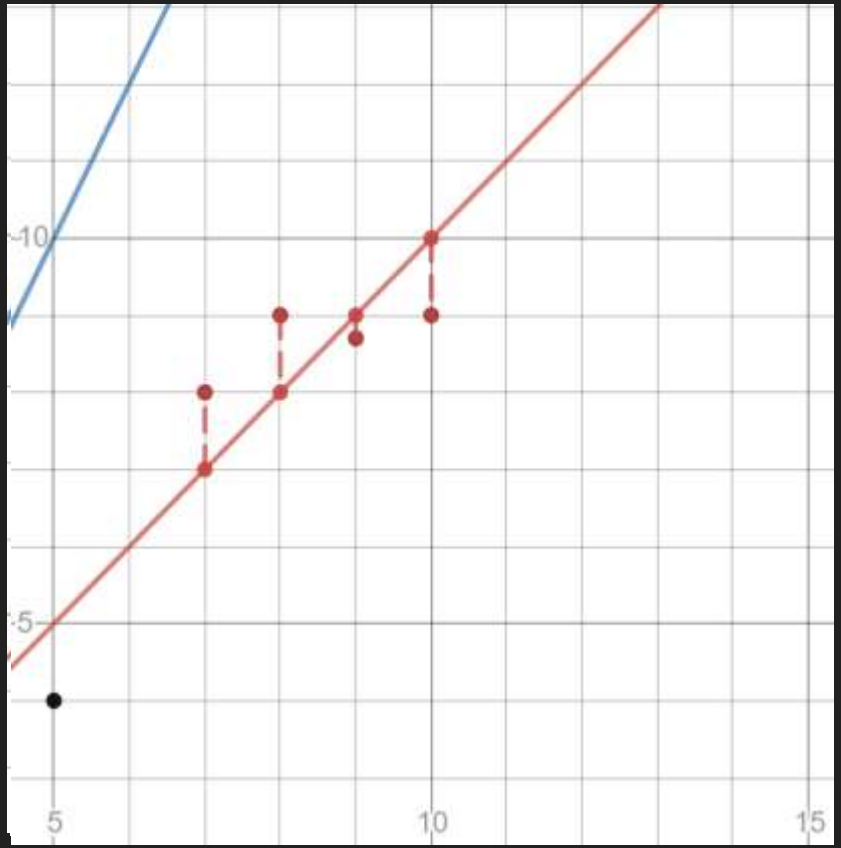
It is known that V=IR.

That is, Voltage= current x Resistance

In our context, this represents a line passing through the origin.

That is, y = m x, where y is the voltage, x is the current and m is the resistance.

You have been asked to perform an experiment to verify this phenomenon.



Example: Real-World (Contd)

How to say mathematically which line is better?

Let the equation of two lines be y = x and y = 2x.

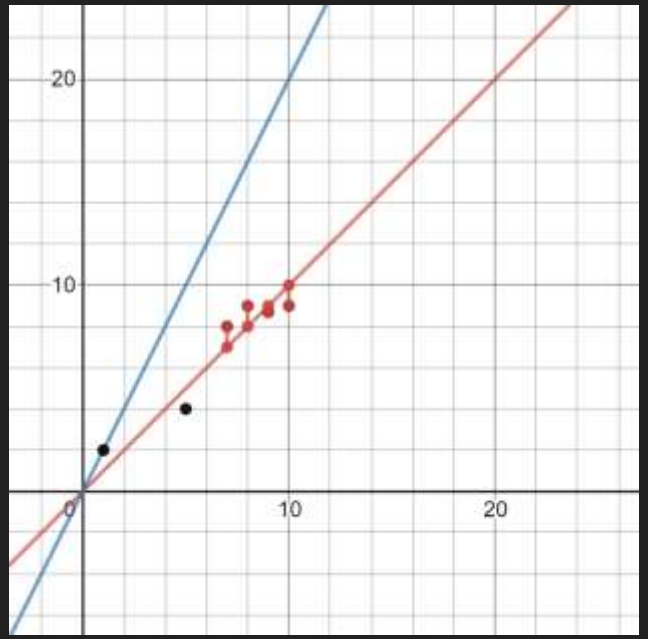
From the set of observations, (x_i, y_i), i=1,2,3,4,5,6.

We can consider the square of the differences

$$\sum_{i=1}^{6}(y_i - x_i)^2 \quad and \sum_{i=1}^{6}(y_i - 2x_i)^2$$

The first difference is 5.09 and the second difference is 328.49.

Therefore, the first line is better than the second line.



x_i	y_i
1	2
5	4
7	8
8	9
9	8.7
10	9

Distance of a Set of Points from a Line

Apart from perpendicular distance, we can also talk about the distance which is parallel to Y-axis.

Consider the set of points $\{(x_i,y_i)|\ i = 1,2,\ldots,n\}$ and a line with equation $y = mx + c$.

Then the **squared sum** of the distance of set of points from the line is defined as

$$SSE = \sum_{i=1}^{n} (y_i - mx_i - c)^2.$$

Least Squares Motivation

- In general, this raises the following question
- Given a set of points, how to find the line that fits the given set of points?
- In other words, what is the equation of the best fit line for given set of points?

In other words, if I need to find the equation of line $y = mx + c$, then the question can be reframed into two questions.

- What is the value of m and c that best fits the given set of points.
- What is a meaning of best fit?

Best Fit: Given a set of *n* points, $\{(x_i,y_i)|\ i = 1,2,...,n\}$, define

$$SSE = \sum_{i=1}^{n} (y_i - mx_i - c)^2.$$

Find the value of *m* and *c* that minimizes *SSE*.