

LINEAR REGRESSION / LEAST SQUARES

- try to understand the way things behave in the physical world

- try to produce useful products or services from the discoveries of science

THEY BOTH do lots of experiments

THEY'RE ALWAYS talking about

What is data?

Information

RESULTS from Experiments!

HERE IS SOME DATA from an experiment.

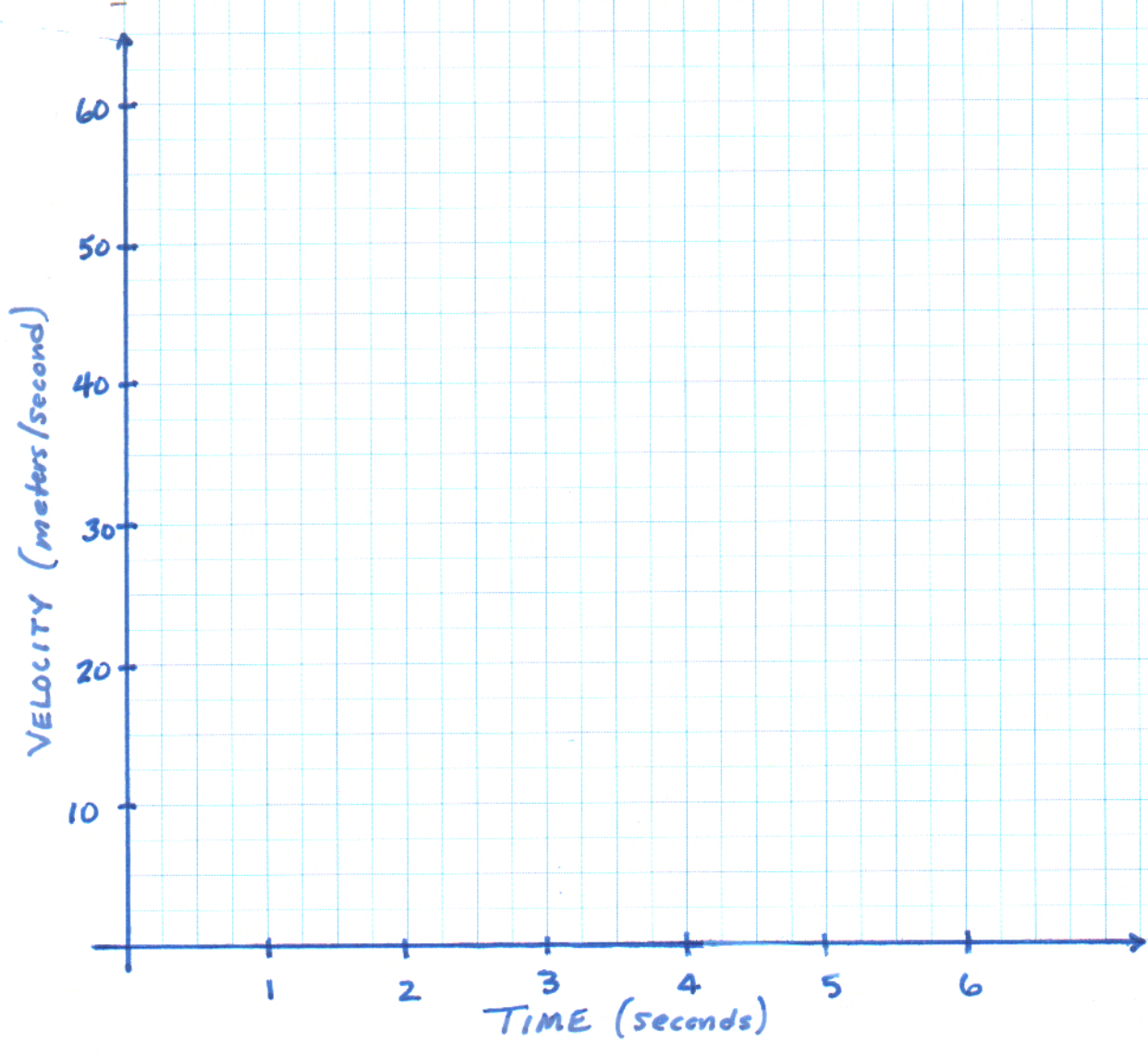
X	Y
TIME (Sec)	VELOCITY (m/s)
1	9
2	21
3	28
4	41
5	47

IT IS VERY IMPORTANT to data

ALWAYS plot your data so you can visualize it.

GET OUT:

- some ENGR paper
- plain paper with ruler to accurately make axes



NOTE: Axes labels
Axes Units

Plot versus ALWAYS!
X is the CONTROL or variable
Y is the Observed or variable

REFER to Plot as Y vs X ALWAYS!

An [redacted] that represents this behavior would be valuable for [redacted].

SCIENTISTS & ENGINEERS call this equation a "[redacted]"

LET'S CALCULATE an equation for these data (NOTE: data is a plural noun)

- It Appears to be [redacted]
- Pick 2 points
- Calculate Slope ($\frac{\text{rise}}{\text{run}}$ or $\frac{\Delta y}{\Delta x}$)
- Write EQUATION using point-slope form

THEN WE'LL COMPARE answers!

Look around @ your table; use different points from neighbors!!

MINE

- Endpoints

$m =$ [redacted]

- EQUATION $y = mx + b$

- What's b ? DO NOT ASSUME it's ZERO!

→ EASIER $y = mx + b$ [redacted]

Solve simultaneously to get m & b

Ⓐ Subtract ② from ①

Ⓑ Plug in m into either ① or ②

[redacted] Where T is in seconds and V is in meters/second

So THE "MATHEMATICAL MODEL" I get is

$$V = 9.5 T - 0.5$$

EVERYONE who chose the same 2 pts should get the
ANSWER

HOWEVER, if you chose pts # 2 & #5, your answer is

$$V = 8\frac{2}{3} T + 3\frac{2}{3}$$

Which ANSWER IS THE BEST?

LET'S use a MATHEMATICAL technique to find THE BEST
AND then we'll all get the SAME ANSWER!

WE'LL USE

WE'LL TAKE the derivative & set equal to zero!

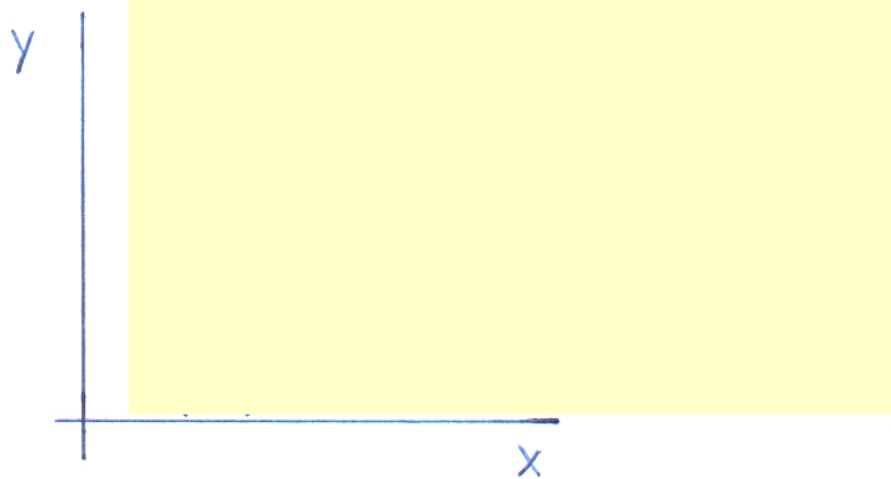
WE'LL just use the result; We'll learn about DERIVATIVES later.

THE CONCEPT to get the BEST ANSWER is IMPORTANT!

between the line and all the data points!

- behavior follows form of $y = mx + b$ (slope - intercept)

- study of relationship between variables ($x \neq y$)



This technique will generate the line (everyone will get the same answer!)

How do you minimize something mathematically?

the derivative is zero!

Earlier we said line, but some are positive & some Negative!

So we change to **MINIMIZE** the line of the distances, which is where the term line comes from.

MINIMIZE this



Using calculus we find the following equations for m & b by minimizing the square of the distances between the points & the line

$$m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad b = \frac{\sum y_i - m \sum x_i}{n}$$

Notice that b contains m , so you must solve for m 1st.

EXAMPLE

Calculate the BEST line to represent the behavior of the velocity vs. time data.

TIME (sec)	VELOCITY (m/s)
1	9
2	21
3	28
4	41
5	47

WE SHOULD ALWAYS PLOT the DATA FIRST! (we already did)

Let's use a TABLE to help manage the calcs

x T	y V	x y (T)(V)	x ² T ²
1	9		
2	21		
3	28		
4	41		
5	47		

Now the SUMMATIONS

Σ =			
Σ x _i	Σ y _i	Σ x _i y _i	Σ x _i ²

Now the EQUATIONS

$$m = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}$$

=

$$b = \frac{\sum y_i - m \sum x_i}{n}$$

b =

So the BEST FIT LINE is
V = T +

→ PLOT IT to see what it looks like!

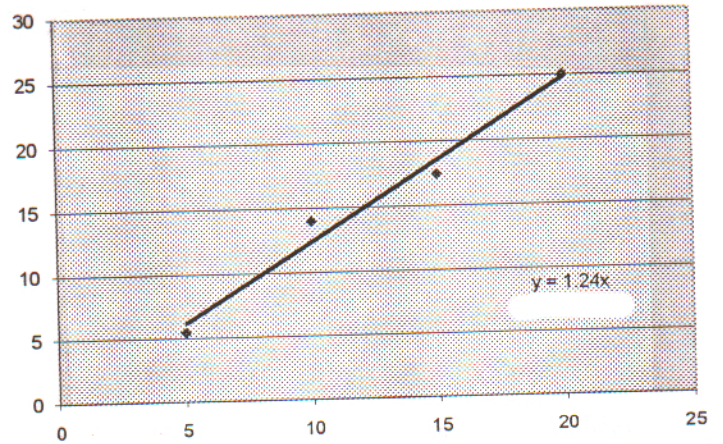
Now you try one...

Class Problem

Here are the data from a spring experiment. The force required to stretch a spring a certain amount is recorded in the table.

Find the best fit line for these data. (THE SOLN IS IN RED)

X (cm)	F (N)		
0	0		
5	5.5		
10	14		
15	17.5		
20	25		



8
DOING A problem by hand gives you good practice
but its time consuming.

LET'S get Excel to help us!

ENTER DATA

PLOT DATA

GENERATE Σ

Calculate m & b using Σ equations

SEE FILE AND ATTACHED PRINTOUT

Use EXCEL to help organize the calcs

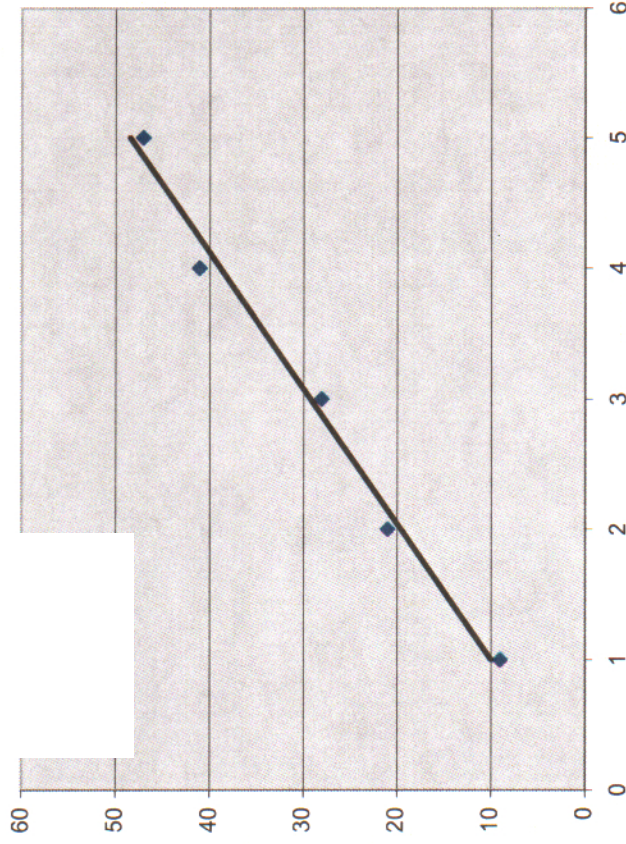
X	Y
1	9
2	21
3	28
4	41
5	47

A Convenient Arrangement

	X	Y	xy	x ²
n	5			
	1	9	9	1
	2	21	42	4
	3	28	84	9
	4	41	164	16
	5	47	235	25
	ΣX	ΣY	ΣXY	ΣX²
	15	146	534	55
			225	(ΣX)²

m 9.6
b 0.4

Get the slope and y-intercept using the summation equations



Can you use Excel to find m and b for the class problem?

LinRegres ClassProb 1

X	Y
0	0
5	5.5
10	14
15	17.5
20	25

