



LINE OF BEST FIT

Perform Linear Regression to find the Line of BEST Fit

Unit 4
Day 13

Let's Warm It Up!

The table below shows the age of a car (in years) and its corresponding value (in thousands).

Age of car	1	2	3	4	5	6	7	8
Value	24	21	19	18	15	12	8	7

Identify the variables. Be specific!

 $x = \text{age of car (in years)}$ $y = \text{value of car (in thousands)}$

Describe the correlation.

as time goes on (\uparrow), the value of the car decreases...

NEGATIVE!

With your partner, use any two points to write a line of fit to model the data.

ex: (1, 24) (8, 7)

$$\frac{7-24}{8-1} = \frac{-17}{7} \approx -2.43$$

$$y - 24 = -2.43(x - 1)$$

$$y = -2.43x + 26.43$$

Let's COMPARE! Are all of our equations the same?

What if there was a way to create a line of BEST fit? Get those calculators out!

w/calc: $y = -2.48x + 26.64$

* see half-sheet for calculator directions!

Linear Regression:

using the calculator to create a line that best represents the data points of two variables

Correlation Coefficient (r):

LinReg(ax + b)

 $y = ax + b$

a =

b =

 $r^2 =$

r =

how well the line of best fit models the data... you want "r" to be as close to ± 1 as possible! $r = +1 \Rightarrow$ Strong POSITIVE correlation $r = -1 \Rightarrow$ Strong NEGATIVE correlation

Another Example!

The table below shows the ages of several people and their salaries (in thousands).

Age	35	37	41	43	45	47	53	55
Salary	42	44	47	50	52	51	49	45

a. Identify the variables. Be specific!

 $x = \text{age}$ $y = \text{Salary (in thousands)}$

b. Describe the correlation.

as people get older, they make more \$...

POSITIVE

c. Line of Best Fit: $y = .21x + 38.09$ d. Correlation Coefficient: $.42$ NOT a great model!

Give it a shot on YOUR OWN!

The table below shows the attendance (in thousands) at an amusement park from 2005 to 2014 where "0" represents the year 2005.

Year	0	1	2	3	4	5	6	7	8	9
Attendance	850	845	828	798	800	792	785	781	775	760

a. Identify the variables. Be specific!

$x = \text{years since 2005}$

$y = \text{attendance (in thousands)}$

b. Describe the correlation.

as time goes on, the # of people attending the park decreases... **NEGATIVE!**

c. Line of Best Fit: $y = -9.59x + 844.55$

d. Correlation Coefficient: $r = -.96$

a great model!

e. If this trend continues what would you expect the attendance of the amusement park to be in 2020? What recommendations would you have the amusement park based on these findings?

↑
 $x = 15$

700.7 people!

f. Can you find the x-intercept of the line and describe what it represents?

$x = 88.07 \Rightarrow 2093$ no one will come to the park



Is Orange *REALLY* the New Black??

The following table shows the number of sentenced female prisoners under the jurisdiction of state or federal correctional authorities from 2007–2013. Let "0" be the year 2007.

Year	# of Female Prisoners
2007 0	105,763
2008 1	106,358
2009 2	105,335
2010 3	104,903
2011 4	103,706
2012 5	101,289
2013 6	104,134

1. Define your variables. Be specific!

$x = \text{years since 2007}$

$y = \text{\# of female prisoners}$

2. Use your calculator to identify the line of best fit.

$y = -594.79x + 106282.64$

3. Does this model accurately fit the data? Explain why or why not.

$r = -.76$ (OK model... would have liked it closer to -1)

Some EXTRA Vocab surrounding Line of Best Fit

- Linear Extrapolation: Using data to approximate a value outside our data → dangerous!
- Linear interpolation: Using data to approx. within our data - Safe