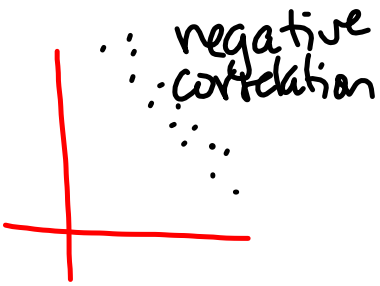
Unit 3 Day 7: Linear Regression

Best Fit Line

You have learned how to find and write equations for lines of fit by hand. Many calculators use complex algorithms that find a more precise line of fit called the best-fit line.

One algorithm is called linear regression. We can find the linear regression using the graphing calculator.



To enter the data: [STAT] EDIT L_1 is independent variable; L_2 is dependent variable

Calculator Steps: [STAT] CALC 4: LinReg [VARS] Y-VARS FUNCTION 1: Y_1 [Enter]

Your calculator may also compute a number called the correlation coefficient. This number will tell you if your correlation is positive or negative and how closely the equation is modeling the data. The closer the correlation coefficient is to 1 or -1, the more closely the equation models the data.

To turn the correlation coefficient on: [2nd] [0] DiagnosticOn [Enter]

- If the correlation coefficient is close to 1 or -1, the fit is Strong (points are close to a line)
- The farther away from 1 or -1, the weaker the fit. (points are spread out)
- If the scatterplot appears random, there is no correlation
- If the correlation coefficient is positive, the slope will be positive.
- If the correlation coefficient is negative, the slope will be negative.

We will often need to interpret the slope and y-intercept in the context of the problem. Descriptions of slope can often follow this pattern:

_____ _____ _____ _____ per _____.
(Topic of data) *(increases/decreases)* *(slope)* *(y-units)* *per* *(x-units).*

+ −

The y-intercept is the starting value, or what the dependent variable is when the independent variable is 0.

y x

EXAMPLE: The average lifespan of American women has been tracked, and the model for the data is $y = 0.2t + 73$, where $t = 0$ corresponds to 1960.

INTERPRETATION:

Real World Example 1: Revenue

The table shows the yearly revenue (in millions) of the Baltimore Ravens.

Year	Revenue (in millions)
2001	148
2002	155
2003	172
2004	192
2005	201
2006	205
2007	226
2008	240
2009	255
2010	262
2011	279
2012	292
2013	304
2014	345

www.statista.com

1. Enter the data into a list using the graphing calculator. Let x = the number of years after 2000.

STAT Edit

2. Find the best fit line using the graphing calculator.

$y = 13.97x + 129.20$

STAT → Calc 4:LinReg

3. State the correlation coefficient and describe the type of fit.

r

$r = 0.99$

Close to 1, meaning, the correlation is positive and strong.

$y = ax + b$ Slope y-int.

4. Interpret the slope in the context of the problem using a complete sentence.

$m = \frac{13.97}{1} = \frac{\Delta y}{\Delta x}$ revenue / time

Ravens earn 13.97 million in revenue per year.

ZOOM

5. Interpret the y-intercept.

(0, 129.20)

time t | revenue y

They made about 129.20 million in 2000 (at start)

9: ZOOMSTAT

6. Use your model to predict what the revenue will be in 2016.

$x = 16$ (16 years after 2000)

$y = 13.97x + 129.20$

$y = 13.97(16) + 129.20$

$y = 352.72$ million

REAL WORLD EXAMPLE 2: HOCKEY

The table below shows the number of goals scored by the Mustang Girls Hockey Team per year. Let x represent the number of years after 2003.

Year	2003	2004	2005	2006	2007	2008	2009	2010
Goals	63	44	55	63	81	85	93	84

L1: 0 1 2 3 4 5 6 7

write $y = mx + b$ (m: pick 2 points to find slope; b: y-int)

1. Find the best fit line using the graphing calculator.

$y = 5.95x + 50.17$

2. State the correlation coefficient and describe its fit.

$r = 0.85$ Correlation is positive, but weaker, so it's not the best line to predict future data.

3. Interpret the slope in the context of the problem.

5.95 On average the team scores an additional 5.95 goals per year

4. Interpret the y-intercept in the context of the problem.

(0, 50.17) using 2003 goals

In 2003 team made about 50.17 goals.

x	y
0	63
1	44

$m = \frac{-19}{1} = -19$

$y = -19x + 63$

5. ANALYSIS: If this were your hockey team, would you want to use this model to predict the number of goals expected in the future? Why or why not?

NO. not very accurate because the y-intercept was off from real one.

the correlation coefficient (r) shows the fit is weak

points don't form a line they're scattered