The Line of Best Fit

0:00 / 0:00

1.0x

SPEAKER: MICHAEL J. MAHOMETA, Ph.D.

We've learned that we can actually fit a line - a linear model or function -

to some data.

But how does this actually happen?

How do we decide on a model?

Download transcript .srt

Comprehension Check

1. What is a residual? Select all that apply.

(1/1 point)

- [ ] the average distance between any data point and the regression line
- [ ] the difference you get when you subtract a data point from the next closest data point

Lecture Videos
Comprehension Check due Mar 15, 2016 at 18:00 UTC

R Tutorial Videos

Pre-Lab
Pre-Lab due Mar 15, 2016 at 18:00 UTC

Lab
Lab due Mar 15, 2016 at 18:00 UTC

Problem Set
Problem Set due Mar 15, 2016 at 18:00 UTC

☑ the distance between a data point in a scatterplot and the line of best fit ✔

☑ e=y-ŷ ✔

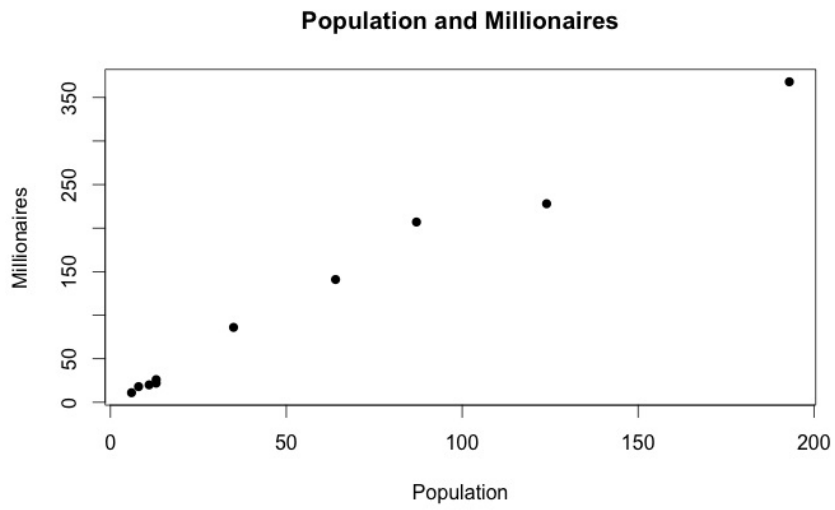
✔

2. Do states with higher populations have more millionaires? Here is data from 2008. The variable labeled "Population" in the table and scatterplot will be referred to as "State.Population" in the questions that follow in order to avoid confusion with the meaning of "population" as a concept in statistics.

State	**Millionaires** *(in thousands)*	**Population** *(in hundreds of thousands)*
Connecticut	86	35
Delaware	18	8
Maine	22	13
Massachusetts	141	64
New Hampshire	26	13
New Jersey	207	87
New York	368	193
Pennsylvania	228	124
Rhode Island	20	11
Vermont	11	6

Using linFit(), the following linear model is found:

Millionaires = 6.296 + (1.921 * State.Population)



(2/2 points)

2a. What is the correlation between Millionaires and State.Population? *(Round to 3 decimal places.)*

- -0.454
- 0.763
- 0.992 ✔
- 1.921

2b. What is the coefficient of determination? *(Round to 3 decimal places.)*

- 0.015
- 0.763
- 0.984 ✔
- 0.992