

Name:

7-A The Line of Best Fit

1. Write the prediction you made about the relationship between wealth and Trump support by state.

State	Wealth*	Trump**
Alabama	43	62
California	57	32
Georgia	48	51
Indiana	46	57
Louisiana	39	58
Michigan	50	48
Montana	45	56
New Jersey	67	41
North Dakota	56	63
Pennsylvania	52	59
Tennessee	43	61
Virginia	65	44
Wyoming	58	67

2. To test this relationship, I collected data on 13 random states. What operational definitions did I use for my two variables?

a) wealth

b) Trump support

3. Find the equation of the line of best fit for the data.

* 2016 U.S. median household income
source: *Huffington Post*—http://www.huffingtonpost.com/moneytips/infographic-us-median-hou_b_10351650.html
** % of votes for Trump in general election
source: *Wikipedia*—*United States presidential election, 2016*

4. Use the line of best fit equation to predict the level of Trump support in the following states.

a) Massachusetts (\$64K/year)

b) Arizona (\$47K/year)

5. Calculate the residual for the following states.

a) California

b) Michigan

7-B Statistically Significant Correlations

6. Calculate the following values.

a) the correlation coefficient r

b) the p value p

7. Which direction is the correlation in the sample, and how strong is it?

8. Are the data statistically significant?

9. What does your answer to #8 mean in the context of the variables being studied?

10. Explain why someone might have a different answer to #8 and #9 if she had used a different random selection of states for her sample.

11. Explain how someone could have a different answer to #8 and #9 even if he had used the same states as his sample.

12. State how the following considerations affect your confidence in the accuracy in predictions from the line of best fit for these data.

a) the value of r

b) the value of p

c) extrapolation

7-C Causal Relationships

13. Explain how the correlation found in the sample could be due to each of the following.

a) coincidence

b) causation

c) reverse causation

d) confounding variables

14. Data snooping (making conclusions from sample data about unhypothesized claims) should be avoided.

a) What problem does data snooping tend to cause?

b) Why does this happen?