

AMS 102: QUIZ 2

SOLUTIONS

1. The population of long-tailed hamster in eastern Texas has been declining for years. One possible reason is the growth of human settlement; another, drying up of the hamster's natural habitat.

Two statistical studies were conducted that, on the year-by-year basis

- (1) compared hamster population with the area of human settlement,
- (2) compared hamster population with annual precipitation.

(a) What are the response and explanatory variables in these studies?

Response variable: size of the hamster population. Proposed explanatory variables: area of human settlement; levels of annual precipitation.

(b) The first study yielded the correlation of -0.78 , the second had the correlation of 0.32 . On this basis, which explanation for the hamster's decline seems better?

The first because it show correlation with the bigger value. (That correlation is negative implies only that association is negative: the more people, the less hamsters.)

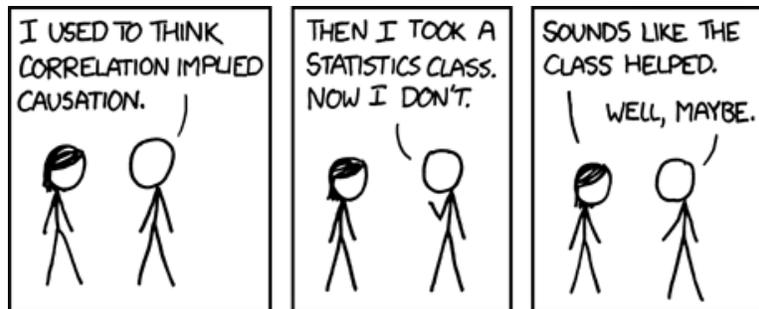
(c) The least-square regression line obtained in the first study has the equation $\hat{y} = 73.1 - 0.12x$. (x is measured in square miles; \hat{y} in thousands.) If the area of human settlement is 100 square miles, what hamster population does regression predict?

$$\hat{y} = 73.1 - 0.12(100) = 73.1 - 12 = 61.1 \text{ thousand hamsters.}$$

(d) The least-square regression line obtained in the first study has the equation $\hat{y} = 73.1 - 0.12x$. (x is measured in square miles; \hat{y} in thousands.) What is the statistical meaning of the term "73.1" in this formula?

None. It corresponds to the hamster population in absence of human settlement. But since there are always people living somewhere, this number has no basis in reality.

2. Explain the joke:



©xkcd

If after taking a statistics class, you stop believing that correlation is causation, was it because of the class? There is correlation between two events (the class and the change in belief) but it does not mean the statistics class was the cause.