

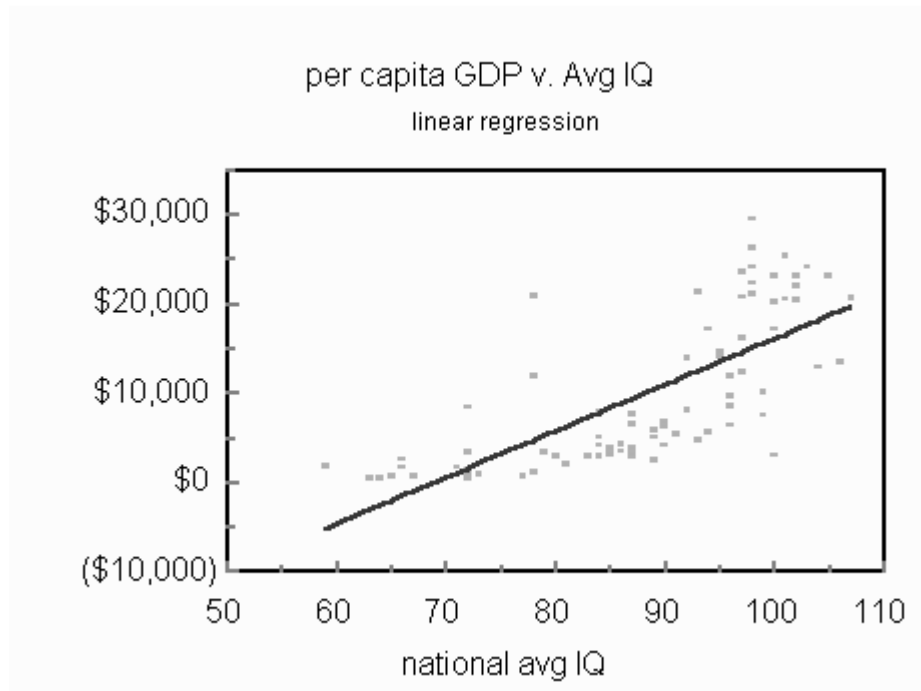
Classroom Voting Questions: Statistics

Cautions about Correlation and Regression

1. Gas mileage and weight were recorded for each automobile in a sample of 20 compact cars. There was a strong negative correlation, with $r = -.87$. Based on the value of r , it is reasonable to conclude that increasing the weight of a compact car causes a decrease in gas mileage.
 - (a) True, and I am very confident.
 - (b) True, and I am not very confident.
 - (c) False, and I am not very confident.
 - (d) False, and I am very confident.

2. Which of the following characteristics in a residual plot are indicative of potential problems?
 - (a) A strong pattern in the residual plot
 - (b) Isolated points in the residual plot
 - (c) A lack of any strong pattern in the residual plot
 - (d) Both (a) and (b) above are indicative of potential problems
 - (e) (a), (b), and (c) above are all indicative of potential problems

3. Which phrase best describes the scatterplot?



- (a) strong $+r$
- (b) strong $-r$
- (c) weak $+r$
- (d) weak $-r$
- (e) influential outliers
- (f) non-linearity
- (g) Two from (A)-(F) are true.
- (h) Three from (A)-(F) are true.

4. Why is it important to look for outliers in data prior to applying regression?

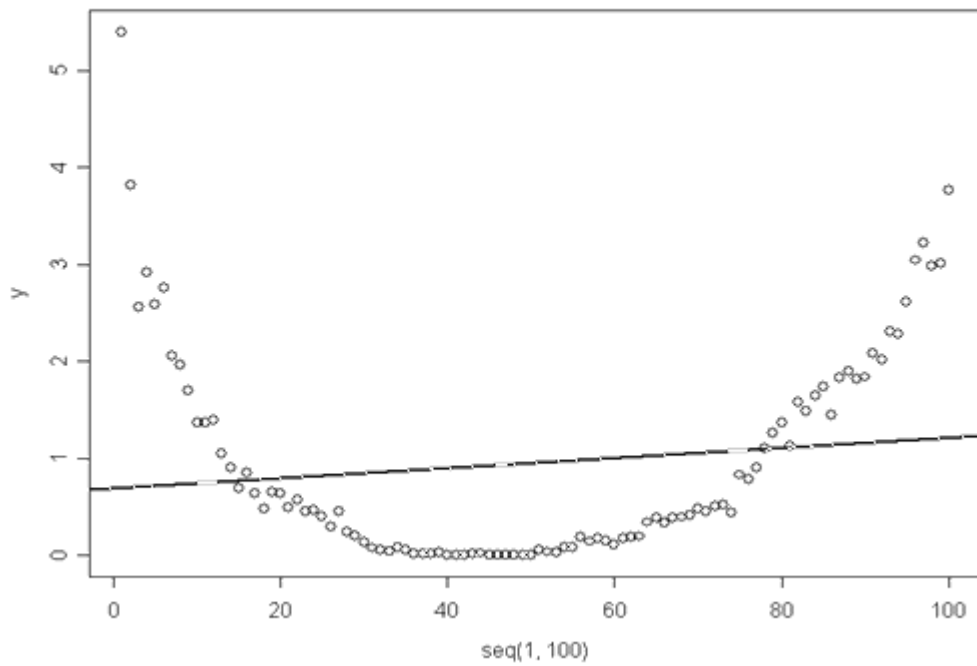
- (a) Outliers always affect the magnitude of the regression slope.
- (b) Outliers are always bad data.
- (c) Outliers should always be eliminated from the data set.
- (d) Outliers should always be considered because of their potential influence.
- (e) We shouldn't look for outliers, because all the data must be analyzed.

5. Which of the following factors is *NOT important* to consider when interpreting a correlation coefficient?

- (a) restriction of range

- (b) problems associated with aggregated data
- (c) outliers
- (d) lurking variables
- (e) unit of measurement

6. What is the greatest concern about the regression below?



- (a) It has a small slope.
- (b) It has a high R^2 .
- (c) The investigator should not be using a linear regression on these data.
- (d) The residuals are too large.
- (e) The regression line does not pass through the origin.