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.

   Answer: (FALSE). There is a correlation that suggests that heavier vehicles get lower gas mileage, but no causal relationship has been proven.

   by Roxy Peck for the textbooks: Roxy Peck and Jay Devore, Statistics: The Exploration and Analysis of Data, 6th Edition, Brooks/Cole Cengage Learning 2008 and Roxy Peck, Chris Olsen and Jay Devore, Introduction to Statistics and Data Analysis, 3rd Edition, Brooks/Cole Cengage Learning 2008.

   STT.02.04.010

   CC HZ MA207 F09: 95/**5** time 1:00
   CC KC MA207 F09: 97/**3** time 1:00
   AS DH MA3321 Su12: 83/**17** time 1:10
   CC KC MA207 F15: 89/11/0/**0**
   CC KC MA315 F18: 29/51/6/**14**
   CC KC MA207 S19: 15/20/35/**30**
   CC KC MA315 S19: 39/39/6/**17**
   CC KC MA315 S20: 32/36/18/**14**

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

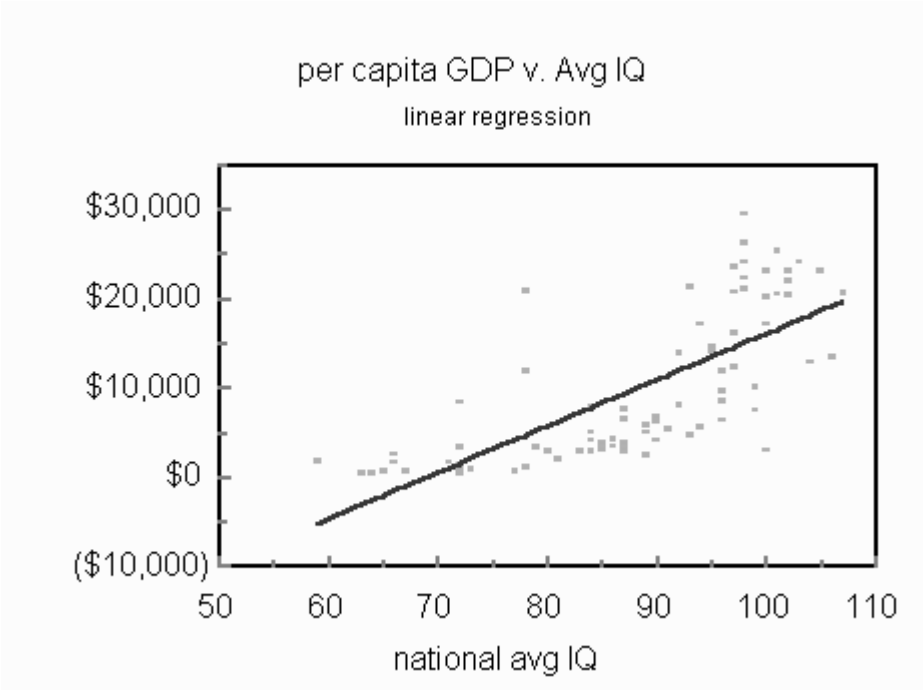
Answer: (d).

by Roxy Peck for the textbooks: Roxy Peck and Jay Devore, Statistics: The Exploration and Analysis of Data, 6th Edition, Brooks/Cole Cengage Learning 2008 and Roxy Peck, Chris Olsen and Jay Devore, Introduction to Statistics and Data Analysis, 3rd Edition, Brooks/Cole Cengage Learning 2008.

STT.02.04.020

AS DH 3321 010 S14: 5/24/0/**71**/0 time 3:10 ,
CC KC MA315 F18: 3/10/53/**3**/30
CC KC MA315 S19: 11/11/39/**22**/17
CC KC MA315 S20: 4/30/39/**22**/4

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.

Answer: (g). Note: The line is not part of the scatterplot; it is there to help students recognize the non-linearity.

(A) This statement is true, but this is not the best answer because (F) is also true.

(B) The relationship is positive but this correlation is negative.

(C), (D) Because the points are relatively close to the line, the relationship appears to be strong rather than weak.

(E) Although there may be outliers, they occur in the center and so they are not particularly influential.

(F) This statement is true, but this is not the best answer because (A) is also true.

(G)* correct (A) strong $+r$ and (F) non-linearity are both true.

(H) Only (A) and (F) are true.

by Murphy, McKnight, Richman, and Terry

STT.02.04.030

CC HZ MA207 F09: 5/0/10/0/0/16/**63**/5 time 1:30
AS DH MA3321 Su12: 0/0/0/0/0/17/**67**/17 time 2:20
CC KC MA207 S19: 0/0/42/0/0/0/**58**/0

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.

Answer: (d). (A) Outliers don't always affect the regression slope.

(B), (C) Outliers may be the data of most interest and are certainly not always bad data.

(D)* correct - Outliers should always be considered but are not always influential.

(E) Even if one analyzes all the data, one should be aware of outliers because of their impact.

by Murphy, McKnight, Richman, and Terry

STT.02.04.040

AS DH MA3321 Su12: 36/0/0/**64**/0 time 1:40
AS DH MA1333 010 F12: 0/0/0/**100**/0 time 1:30
AS DH MA1333 020 F12: 0/0/0/**100**/0 time 1:30
AS DH 3321 010 F14: 0/0/0/**100**/0 time 1:50 ,
CC KC MA207 S19: 0/0/0/**100**/0

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

   Answer: (e). (A) In general, assuming the data are measured on interval or ratio scales, a restricted range reduces correlation and so this factor should be considered.

   (B) Aggregation results in fewer data points, which is important in interpreting the correlation coefficient.

   (C) Outliers can affect the value of the correlation and so this factor should be considered.

   (D) A lurking variable may affect the interpretation of the relationship that is implied by the correlation coefficient and so this factor should be considered.

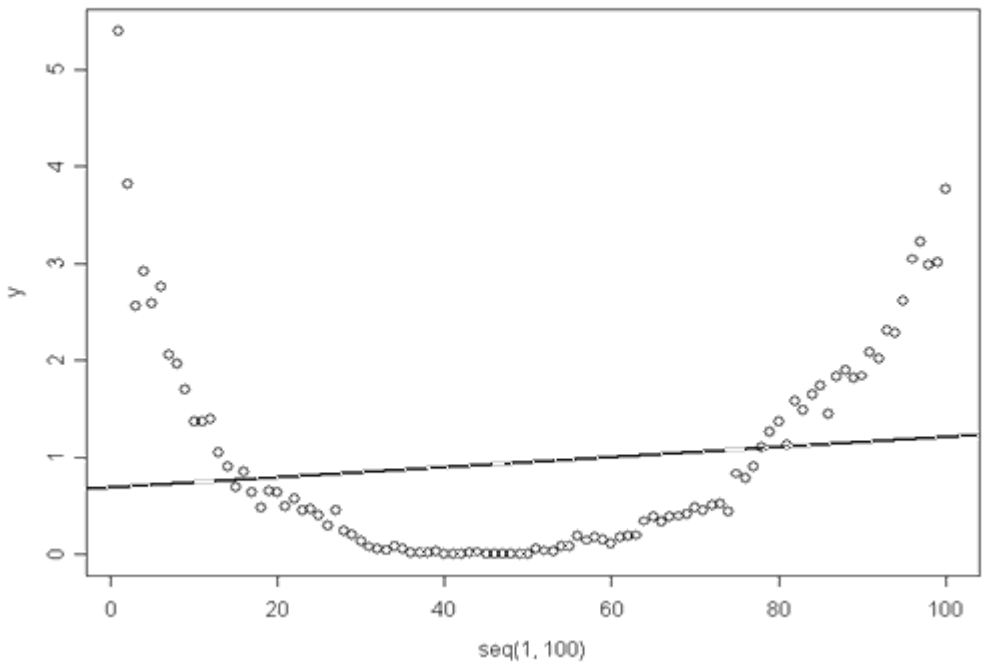
   (E)* correct The effect of units of measurement are eliminated by the role of standard deviation in correlation.

   by Murphy, McKnight, Richman, and Terry

   STT.02.04.050

   CC KC MA207 S19: 0/5/5/15/**75**

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.

Answer: (c). (A) It does have a small slope, but that is not the greatest concern.

(B) The correlation is weak, so R^2 is not high.

(C)* correct The relationship appears to be nonlinear so one should use a polynomial regression.

(D) The residuals are to large because the relationship is nonlinear so the nonlinear relationship is a greater concern.

(E) The regression line need not pass through the origin, so this is not a concern.

by Murphy, McKnight, Richman, and Terry

STT.02.04.060

CC HZ MA207 F09: 0/0/**100**/0/0 time 0:45
CC KC MA207 F09: 0/0/**76**/7/17 time 1:30
AS DH MA3321 Su12: 0/0/**100**/0/0 time 1:20
CC KC MA207 F18: 0/0/**100**/0/0