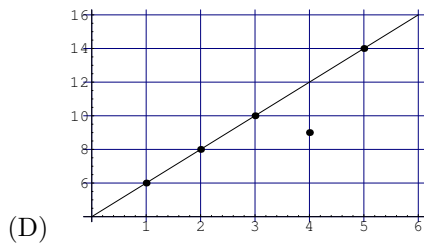## Math 216
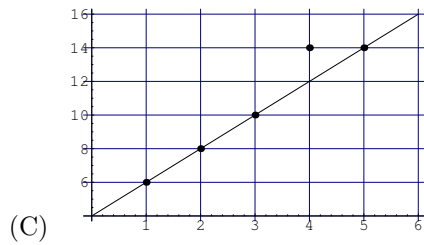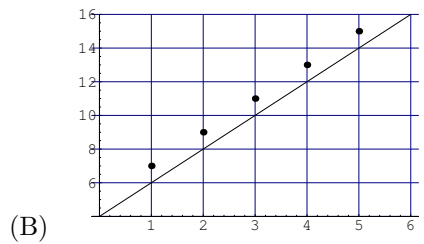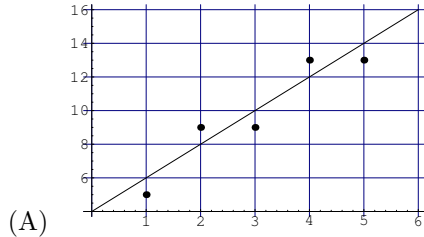## Chapter 7 Clicker Questions

1. The line $y = 4 + 2x$ has been proposed as a line of best for the following four sets of bivariate data. For which data set is this line the *best* fit?
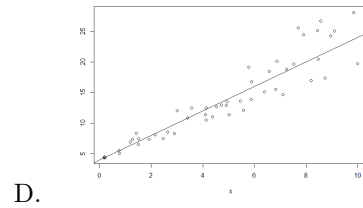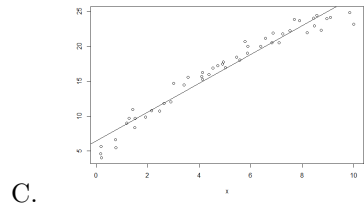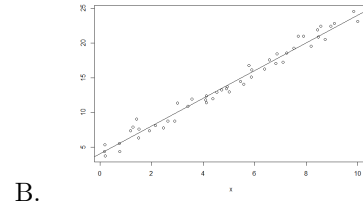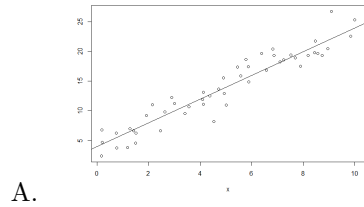
(A)

(B)

(C)

(D)

2. Suppose you use the linear model $\hat{y} = \beta_0 + \beta_1 x$ to fit bivariate data. Which of the following is e_i, the *residual* of the observation (x_i, y_i)?
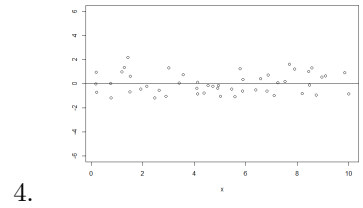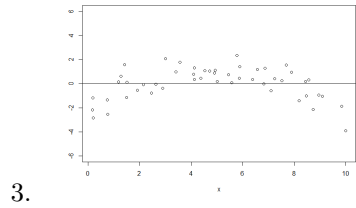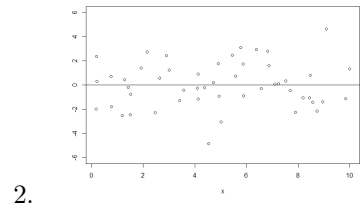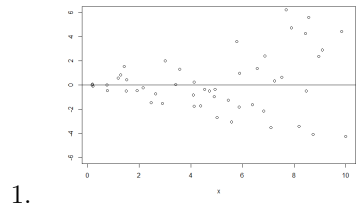
   (A) $\beta_0 + \beta_1 x_i - \hat{y}_i$

   (B) $|y_i - \hat{y}_i|$

   (C) $y_i - \hat{y}_i$

   (D) $(y_i - \hat{y}_i)^2$

3. Match each scatterplot with the corresponding residual plot.

Scatterplots

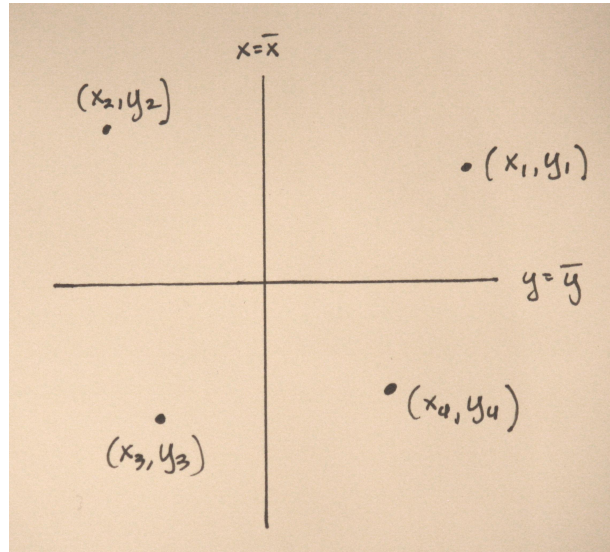A.

B.

C.

D.

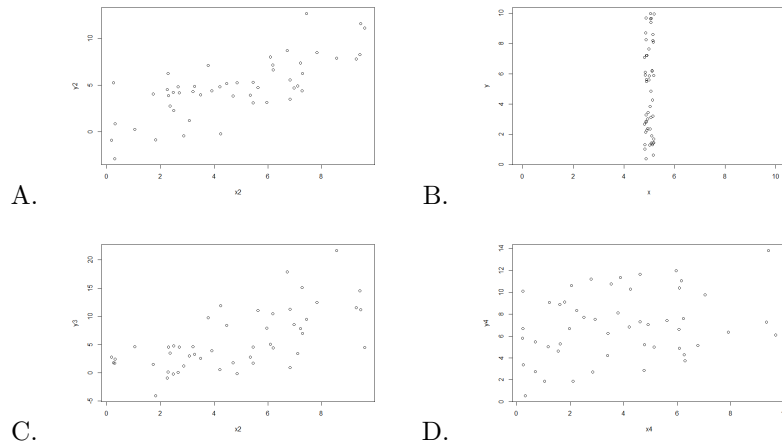**Residual Plots**

1.

2.

3.

4.

4. For which of the points (x_i, y_i) plotted on the coordinate axes below is the quantity $(x_i - \bar{x})(y_i - \bar{y})$ positive? *(Multiple Mark)*
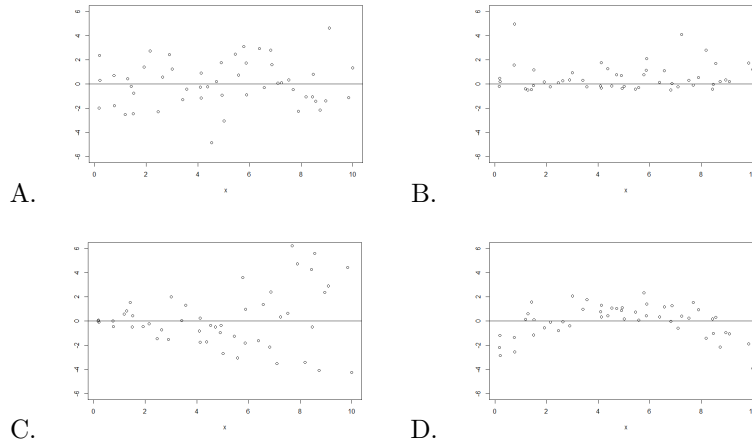

(A) (x_1, y_1)

(B) (x_2, y_2)

(C) (x_3, y_3)

(D) (x_4, y_4)

5. Which of the following sets of data has the correlation coefficient closest to 0?

A.

B.

C.

D.

6. Given the residual plots below, which corresponding set or sets of data are good candidates for fitting by least squares lines? *(Multiple Mark)*

A.

B.

C.

D.

7. Which of the following scatterplots likely has the greatest value of R^2?
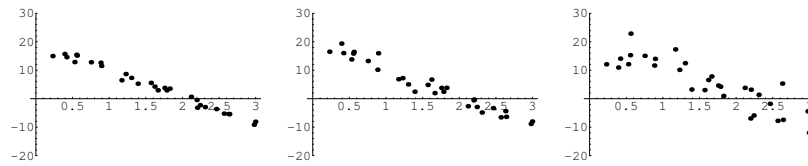
8. Which of the following scatter plots likely has the greatest value of the total sum of squares?

(A)

(B)

(C)

(D)

9. Which of the following scatter plots likely has the greatest value of the *error sum of squares*?
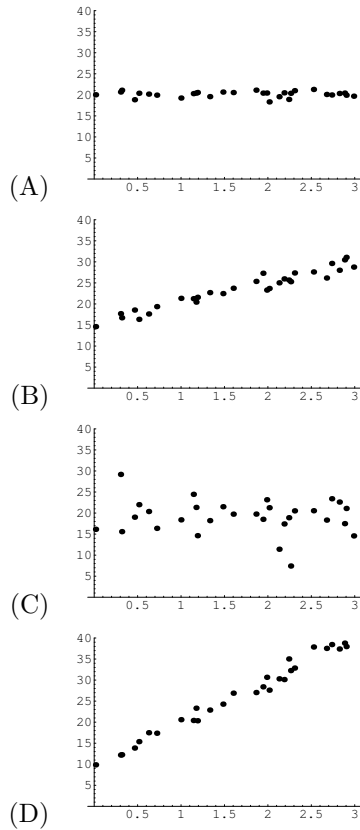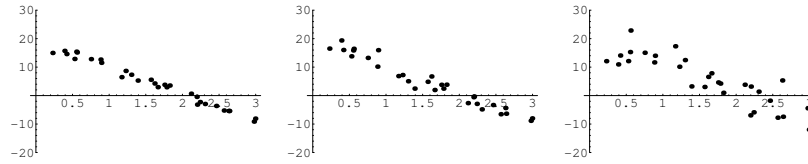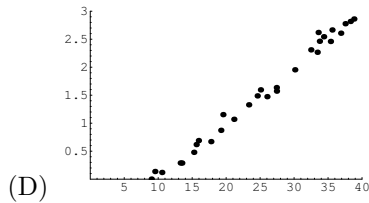

10. Which of the following scatter plots likely has the greatest value of $\sum_{i=1}^{n} (x_i - \bar{x})^2$?

(A)

(B)

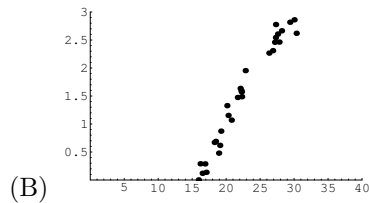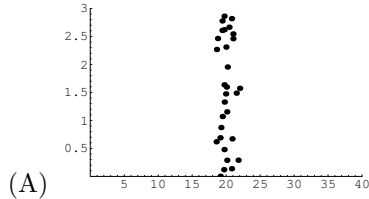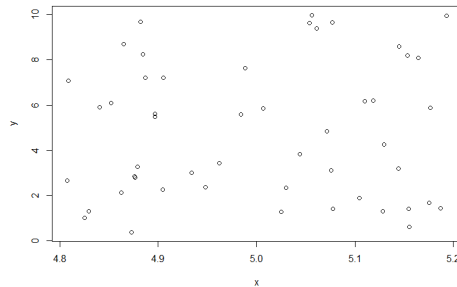(C)

(D)
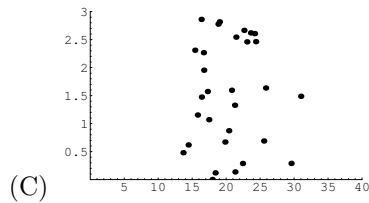
11. If we hold everything else constant, what happens to the confidence interval for β_1 if there's an increase in the variance of the observed x-values in the data set?

(A) It gets wider.

(B) It gets narrower.

Answers

1. C

2. C

3. A-2, B-4, C-3, D-1

4. A and C

5. B. Here's that same scatterplot, with a different "window."

   There's no linear relationship (positive or negative) here.

6. A is the only clear candidate here. Recall from the reading that, when fitting a least squares lines, we want:

   (A) Linearity: The data should show a linear trend. This knocks out D, which seems more parabolic than linear.

   (B) Constant variability: The variability of points around the least squares line remains roughly constant. This knocks out C, in which the variable increases along with x.

   (C) Nearly normal residuals: Generally, the residuals must be nearly normal. This knocks out B, since the residuals in B have a very short left tail. That is, they are not symmetric about any line. (The fact that they are mostly above the center line isn't an issue. It could be that our regression line is just off.)

7. A (the one on the left)

8. D

9. C (the one on the right)

10. D

11. B