

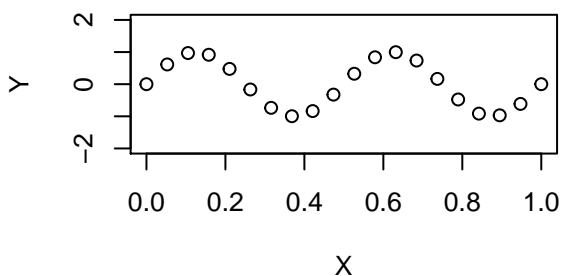
# Homework 2

Due: Thursday 2/13/20 by 8:30am

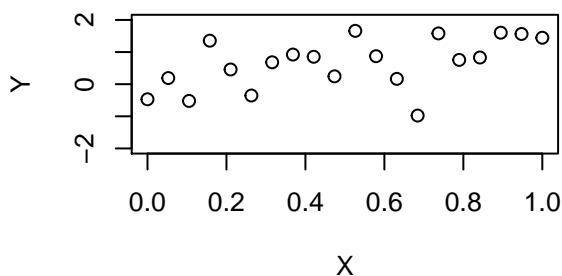
This homework assignment focuses on material covered in Chapter 1 of the textbook.

- For each of these figures, indicate whether a functional or statistical relationship is depicted between  $Y$  and  $X$ .

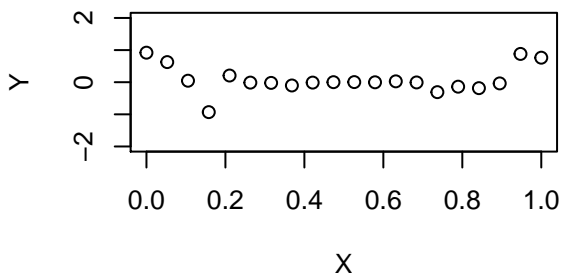
**(a)**



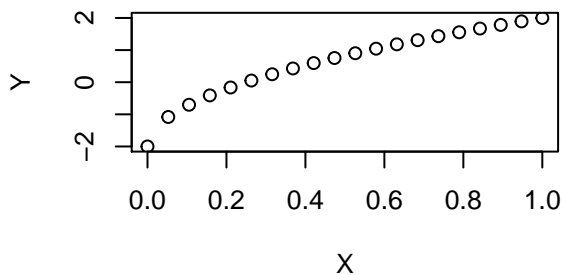
**(b)**



**(c)**



**(d)**



- Suppose we collected data on the age and shoe size of  $n = 3$  people. Let  $Y_i$  refer to the  $i$ -th subject's shoe size, and  $X_i$  refer to subject  $i$ 's age.
  - Suppose the first two subjects were the same age, i.e.  $X_1 = X_2$ , but had different shoe sizes,  $Y_1 \neq Y_2$ . What feature of the simple linear regression model described in Equation 1.1 of the text is illustrated by this?
  - Suppose that we only collected data on high school students. If we assume the simple linear regression model described in Equation 1.1 of the text, what is the scope of the model?
- Suppose Instagram magically knew that every time the number of times user  $i$  purchases a product, denoted by  $Y_i$ , is related to the number of times the product has been advertised to user  $i$ , denoted by  $X_i$ , as follows:

$$Y_i = 1 + 2X_i + \epsilon_i$$

where  $\epsilon_i$  is a random error term with mean  $E\{\epsilon_i\} = 0$  and variance  $\sigma^2\{\epsilon_i\} = 0.1$ ;  $\epsilon_i$  and  $\epsilon_j$  are uncorrelated so that their covariance is zero (i.e.,  $\sigma\{\epsilon_i, \epsilon_j\} = 0$  for all  $i \neq j$ ) for  $i = 1, \dots, n$ .

- (a) Which value corresponds to the intercept,  $\beta_0$ ? In at most one sentence, interpret it, assuming that the scope of the model includes  $X_i = 0$ .
  - (b) Which value corresponds to the intercept,  $\beta_1$ ? In at most one sentence, interpret it.
  - (c) What do you expect the regression function to look like?
  - (d) **Do not do this question, we will return to it next week!** Using R, make a plot with three panels. Plot the density of the errors  $\epsilon_i$  for  $X_i = 0$ ,  $X_i = 1$ , and  $X_i = 10$ , using a separate panel for each value of  $X_i$ . Ensure that the axes are the same across all three plots.
  - (e) Based on the assumed model and the information provided, can we conclude that the number of times user  $i$  purchases a product  $Y_i$  is uncorrelated with the number of times user  $j$  purchases a product  $Y_j$ ?
  - (f) Based on the assumed model and the information provided, can we state the exact probability that a single value  $Y_i$  will be greater than 4 given that  $X_i = 1$ ?
  - (g) **Do not do this question, we will return to it next week!** Simulate  $n = 100$  observations from the model, with  $X_i = i$ . Using R, make a scatter plot of the data and overlay the regression function on the scatterplot.
  - (h) **Do not do this question, we will return to it next week!** Repeat (g), but instead of assuming that  $\sigma^2\{\epsilon_i\} = 0.1$ , assume that  $\sigma^2\{\epsilon_i\} = 10$ . In at most one sentence, describe how increasing  $\sigma^2\{\epsilon_i\}$  changes how the regression function relates to the scatter plot.
4. Problem 1.28 from the .pdf version of the textbook. Requires use of the `crime` data that has been posted on the Homework page.