

CSE 103

Homework #6 Solution Fall 2019

Due: Monday, November 11, 2019 at 11:00PM on Gradescope

1 Directions

You may work with one other student. If working with a partner, **submit only one submission per pair** : one partner uploads the submission and adds the other partner to the Gradescope submission. You can post public questions about the assignment to Piazza, discuss the questions and their answers with at most one other student, and ask questions in office hours

Your answers have to be typeset, not handwritten. This is for two reasons: (a) to reduce ambiguity of the answers, and (b) to be kind to the TA's eyesight. We recommend you use latex, but you can also use word-processors that support mathematical formulas. More directions are available here: <https://tinyurl.com/y2gv9bn9>.

You will submit this assignment via Gradescope (<https://www.gradescope.com>) in the assignment called "Homework 6". You can submit each question as many times as you like. You should solve the problems and ask questions about them offline first, then try submitting once you are confident in your answers.

No late submissions are accepted.

2 Problems

1. (15 points) Suppose that a study were done on married couple in a small town. We learned that one person in each couple is always exactly 7% shorter than their partner. What would the correlation between the heights of the married couple be?

Since the height of the shorter partner is always $0.93 \times$ height of the taller partner. The points would fall perfectly on one straight line.

Correlation is 1.

2. (30 points + Additional 5 points for providing legible and organized answers) In each of the three questions below, compute the mean and std of x and of y and the correlation of x and y . Using the grid provided, create a plot of the means, std and SD lines as shown in the example.

Table a:

x	y
6	0
5	2
5	1
2	4
-2	7
7	0
7	2
1	3

$$\text{mean}(x) = 3.88, \text{std}(x) = 3.02, \text{mean}(y) = 2.38, \text{std}(y) = 2.18, \text{correlation} : -0.91$$

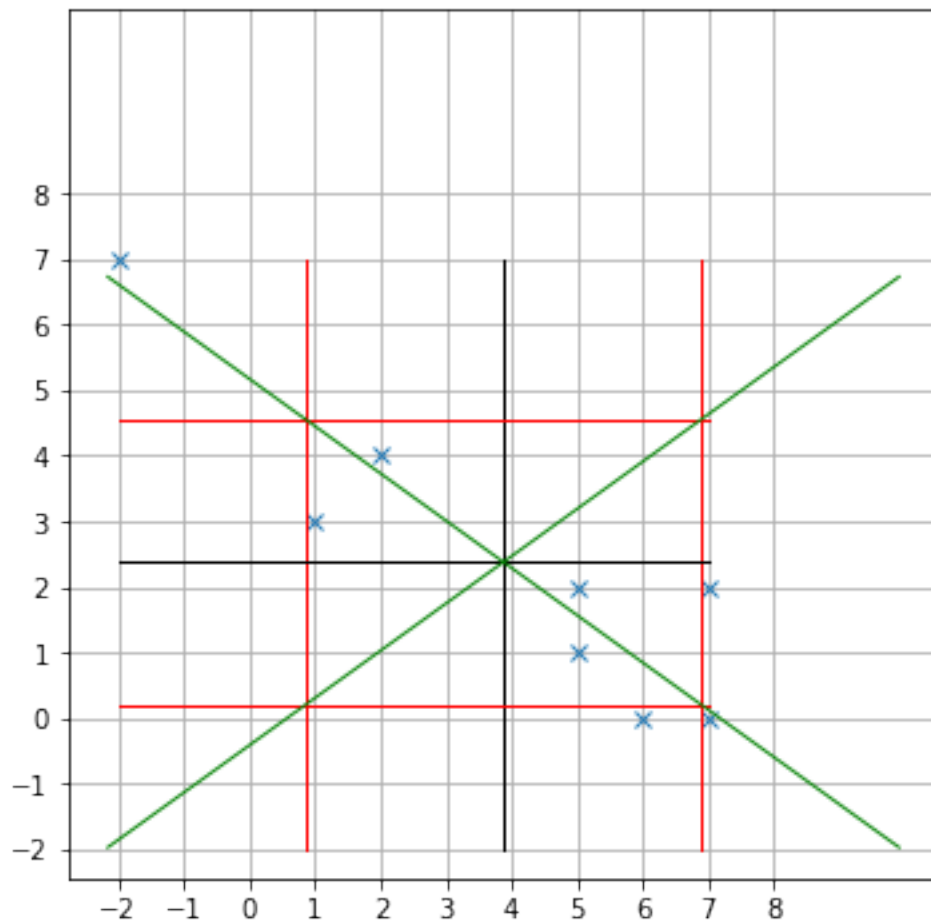


Table b:

x	y
-2	4
-3	7
0	-1
-3	5
-3	-3
-2	4
-2	-3
-1	-2

$mean(x) = -2.00, std(x) = 1.00, mean(y) = 1.38, std(y) = 3.77, correlation : -0.43$

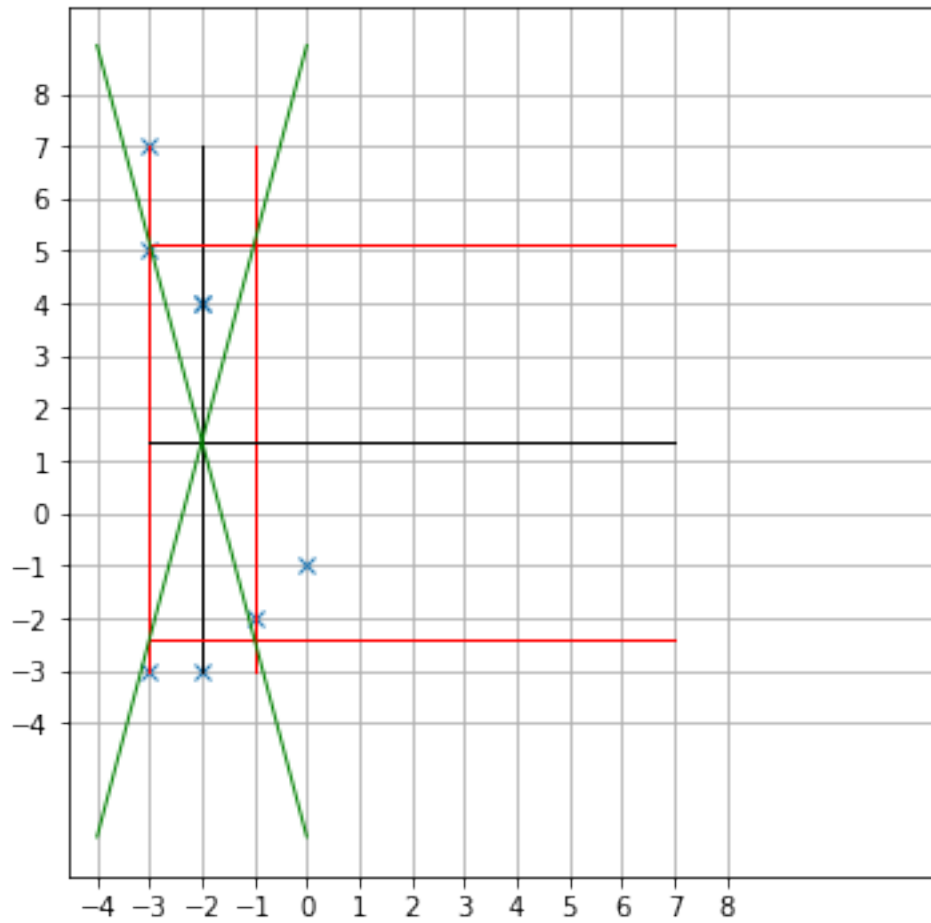
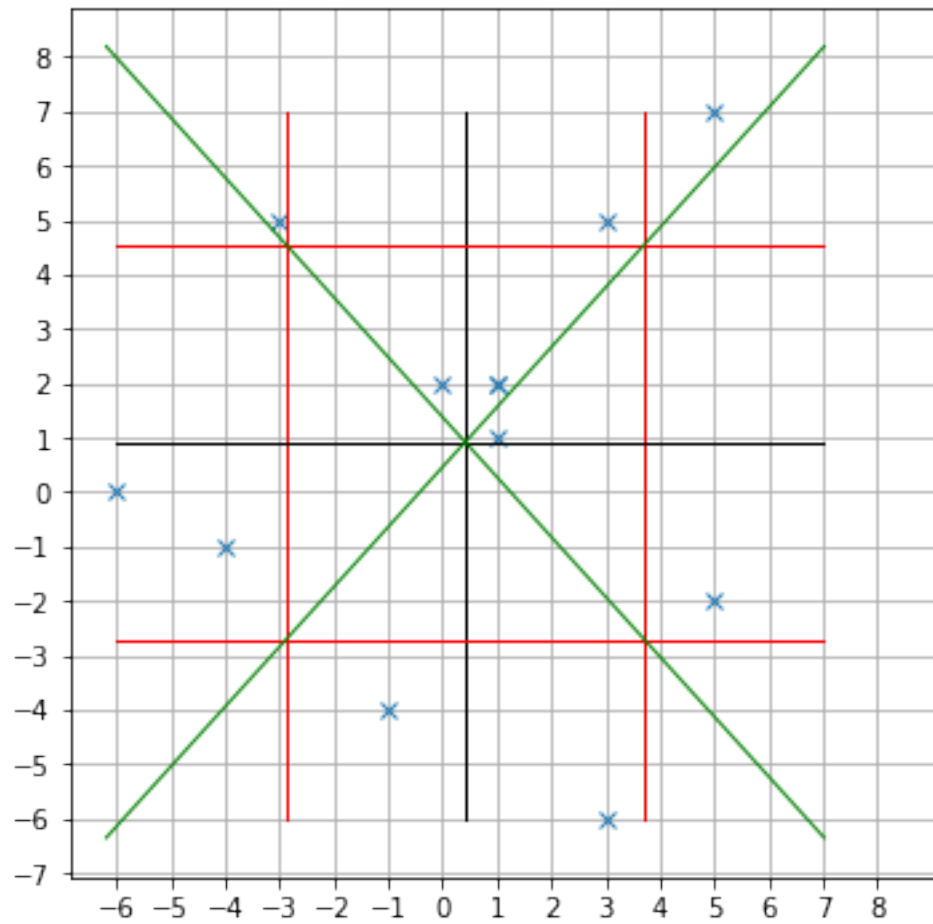


Table c:

x	y
5	7
5	-2
1	2
1	1
0	2
3	5
1	2
3	-6
-1	-4
-6	0
-4	-1
-3	5

$mean(x) = 0.42, std(x) = 3.30, mean(y) = 0.92, std(y) = 3.64, correlation : 0.11$

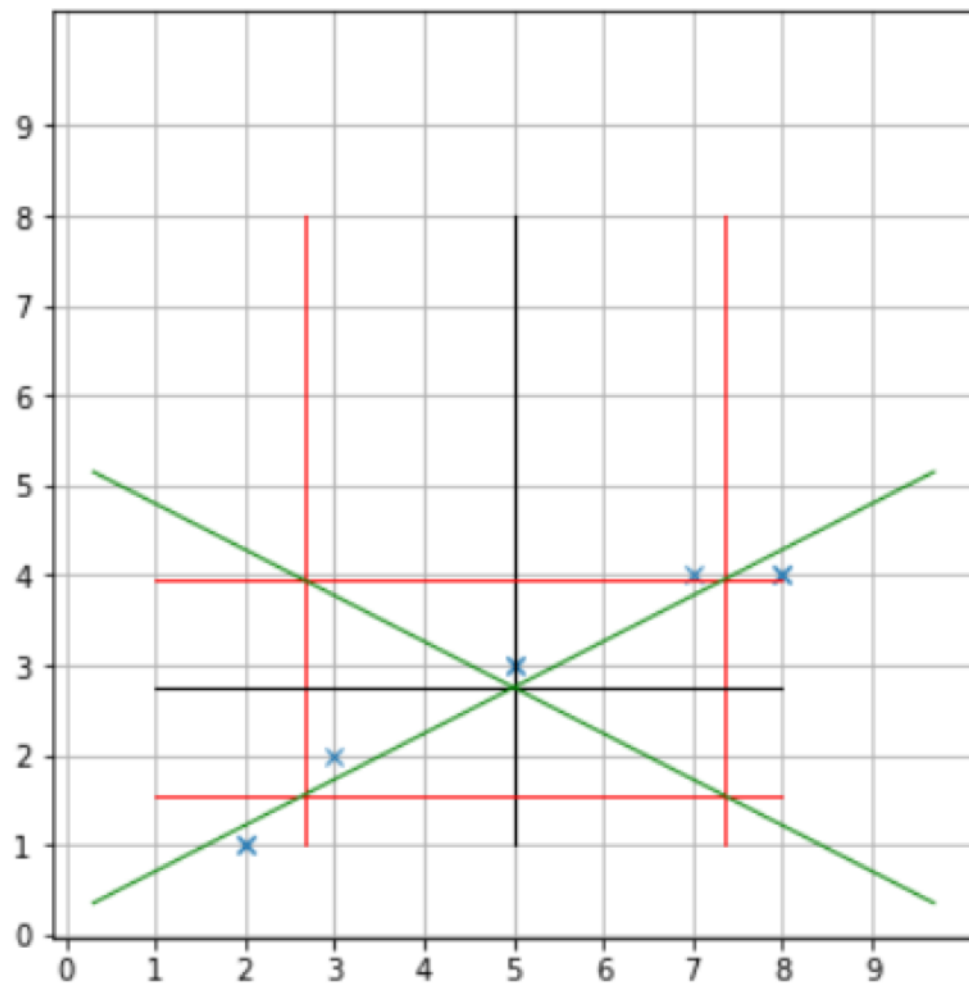


Example:

x	y
5	3
8	4
7	4
3	2
2	1
5	3
2	1
8	4

Expected answer:

```
mean(x)=5.00, std(x)=2.35,  
mean(y)=2.75, std(y)=1.20,  
correlation: 0.98
```



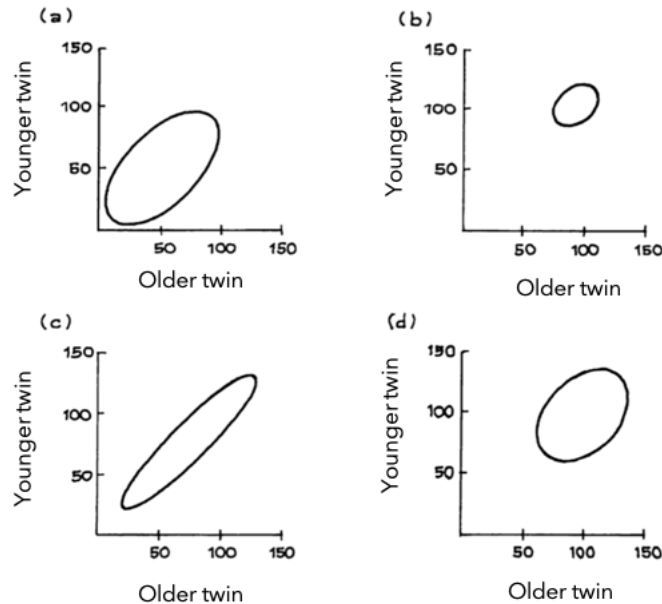
3. (20 points) A study was conducted on identical twins. The IQs of the twins were obtained:

Older twin: average IQ= 100, SD = 15

Younger twin: average IQ= 100, SD = 15

$$r = 0.6$$

One of the following is a scatter diagram for the data. Which one? Explain briefly why the others are rejected.



Not Diagram a because the average IQ of that graph is 50.

Not Diagram b because the $2 \times \text{SD}$ is a lot smaller than 30, which is the 2SD of the question.

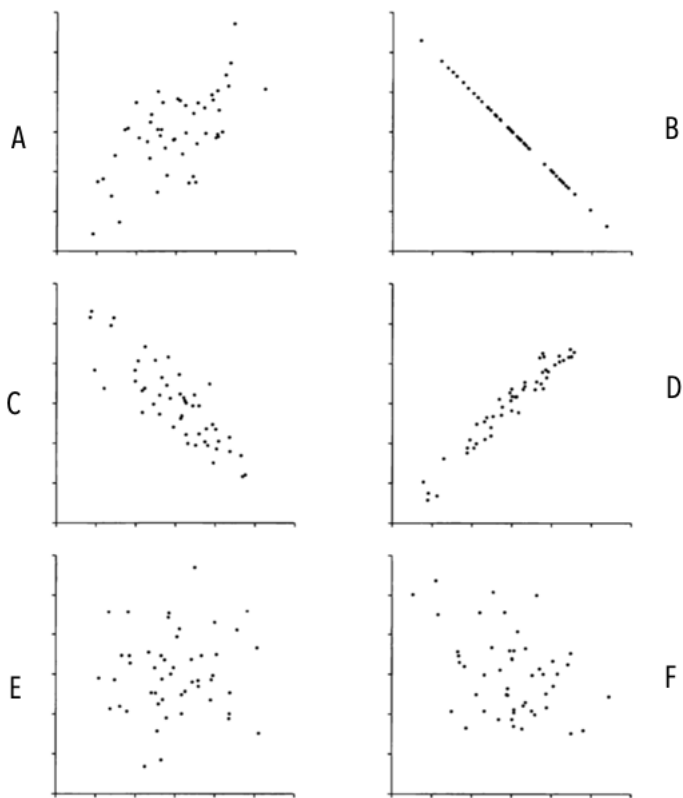
Not Diagram c because the average is less than 100, and correlation is closer to 1.

Answer is Diagram D. Average is (100,100). The "football" spans about 30 in x and y direction ($2 \times \text{SD}$). The correlation is 0.6 so the football is a bit wider.

4. (30 points) The figure below has six scatter diagrams for hypothetical data. The correlation coefficients, in scrambled order, are:

-0.85 -0.38 -1.00 0.06 0.97 0.62

Match the scatter diagrams with the correlation coefficients.



A:0.62

B:-1

C:-0.85

D:0.97

E:0.06

F:-0.38