Math 10 – Homework 2 **ANSWER KEY**

Name:

Lecturer:

Instructions

- Type your answers and paste images directly into this document.
- Answers are usually short, with 1-3 sentences.
- Print out and hand in homework in class on Tuesday.
- You are free to use any scatter/dot plot and correlation coefficient tool. Suggestion: Excel, or the online tool here: http://www.alcula.com/calculators/statistics/scatter-plot/
- You are also free to use a standard deviation calculator such as: http://www.calculator.net/standard-deviation-calculator.html
- Note: in the SD calculator above, we want what they call the ***sample standard deviation*** (scroll down on that page for formula).
- You may collaborate on the homework but you must write it up yourselves.

Problem 1 – Summarizing Distributions, Bivariate Data

Dataset A		Dataset B		Dataset C	
x	y	x	y	x	y
5	8.2	8	89.4	3.4	4.3
4.7	7	1	23.4	6.3	5.1
4.6	6.5	4	38.2	6.2	5.2
4.4	5.3	2	30.4	5.3	3.3
4.3	5.2	5	52.2	5.2	6.6
4.1	4	6	62.3	7.9	7.7
4.2	4.6	3	36.8	3.9	6.2
4.2	4.5	7	109.6	3.0	4.3
4.2	4.2			3.1	4.1

1. Write down the mean of x, and the mean of y for all three datasets, to 2

decimal places if possible. (*3 points*)

To 2 decimal places. We ignored rounding errors/differences.

A) Mean of x = 4.41, y = 5.5
B) Mean of x = 4.5, y = 55.28
C) Mean of x = 4.92, y = 5.2

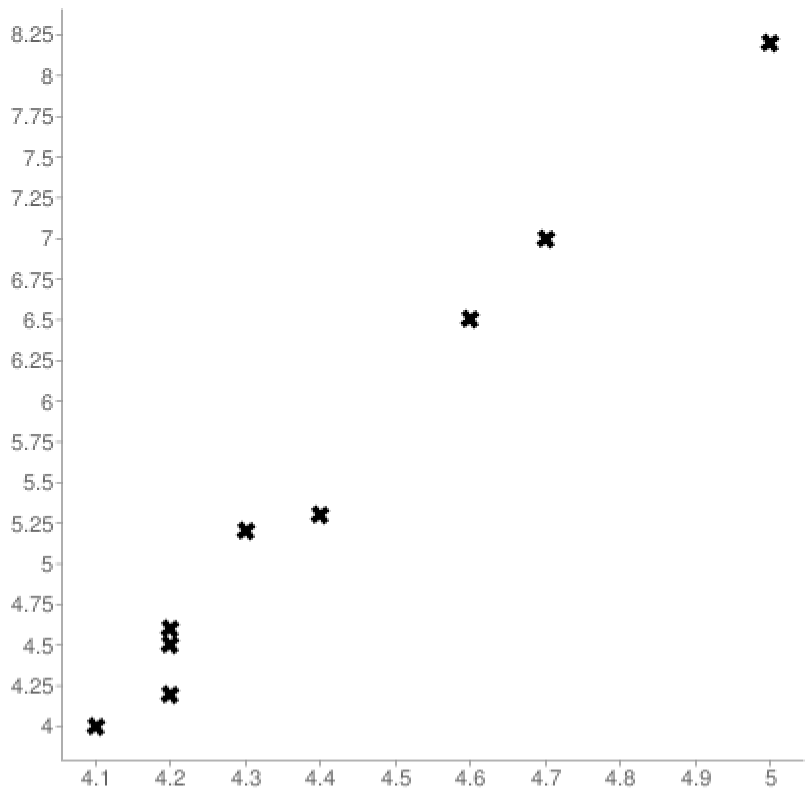
2. Compute the standard deviation of x, and the standard deviation of y, using the estimate of the variance s^2, for all three datasets, to 4 decimal places if possible. (*3 points*)

To 4 decimal places. We ignored rounding errors/differences.

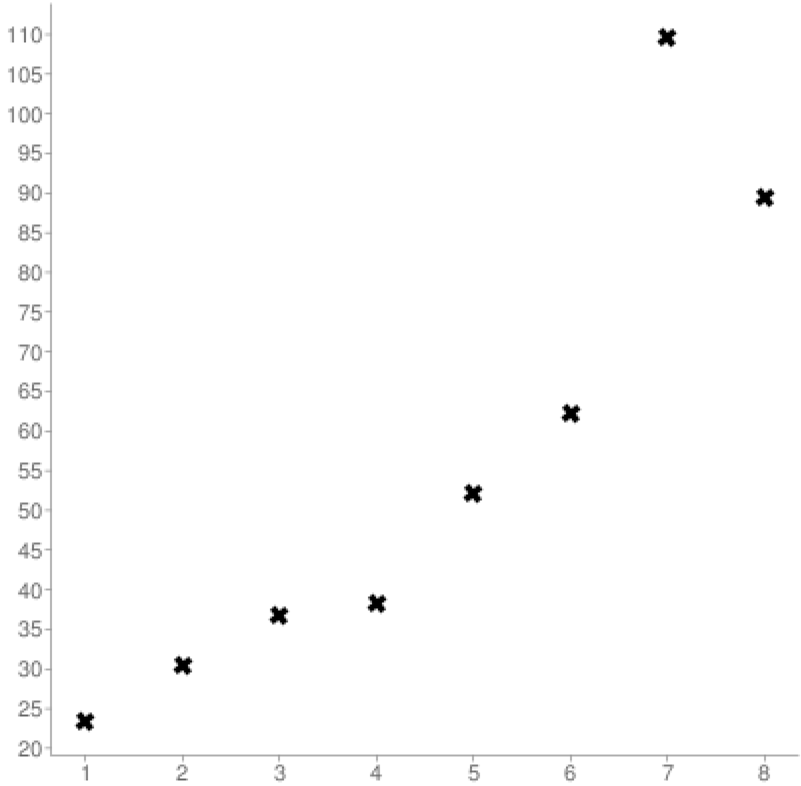
A) SD of x = 0.2977, y = 1.4327
B) SD of x = 2.4494, y = 30.3367
C) SD of x = 1.6954, y = 1.3991

3. For each of the three datasets: plot the scatter/dot plot and state the Pearson correlation coefficient r, to 4 decimal places if possible. (*6 points*)

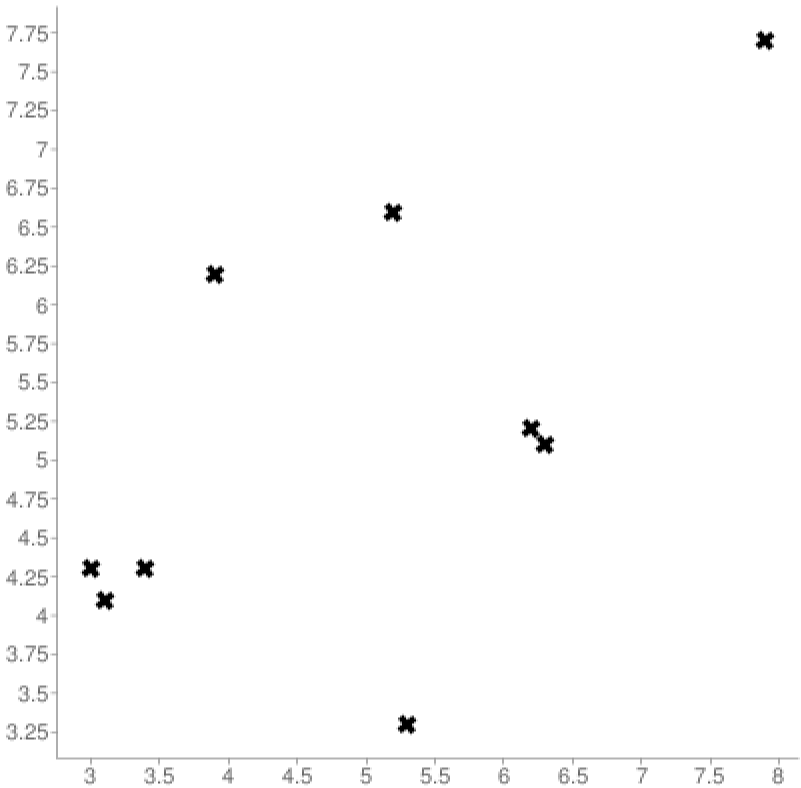
A) r = 0.9936



B) r = 0.9117



C) r = 0.5828



For question 4-6, consider your scatter plots and Pearson's r for each dataset. Describe the relationship between variables x and y. Also describe the strength of this relationship and/or how each dataset differs. (*3 points*)

4. Dataset A

   Very high r. Very strongly positively related. Strongest correlation.

5. Dataset B

   High r. Strong positive relation, but not as strong as in dataset A.

6. Dataset C

   r is not as high as in dataset A and B. Positive correlation.

Dataset A		Dataset B		Dataset C	
x	y	x	y	x	y
Divorce Rate in Maine (per 100 people)	Per Capita Consumption of Margarine in the USA (in pounds)	Level of Education Attained	Average Income (in US$1000)	Both are independent random variables.	

Now, you are given what the variables x and y actually are in each dataset.

For questions 7-8, comment on our use of statistics for each dataset. Is Pearson's r appropriate? Why? Were the results misleading? (*6 points*)

7. Dataset A

   No. Variables are quite possibly unrelated. Spurious correlation. Misleading results.

8. Dataset B

   Yes. Supported by some plausible economic reasoning.

9. Dataset C

   No. Variables are known to have nothing to do with each other. Results could be misleading since there appears to be a linear association.

Problem 2 - Probability

Warren Buffett (famous stock investor and one of the richest man in the world) has been known to trick people into a playing a version of the game described below. Bill Gates (a founder of Microsoft) almost became one of his victims but managed to figure out the trick in time.

There are three dice with different faces.

a) Dice A has faces 2,2,4,4,9,9
b) Dice B has faces 1,1,6,6,8,8
c) Dice C has faces 3,3,5,5,7,7

You and a friend decide to bet $10 on this game: one person gets to choose one of the three dice. Then, the other person chooses one of the two remaining dice. Both players can see the faces of the three dice always. The dice are only picked once at the start and are never re-picked again for the rest of the game.

The game then proceeds with 100 rounds. In each round, both players simultaneously roll their chosen dice. The person who rolled the highest wins the round. The winner of the $10 is the person who won the most of the 100 rounds (nothing happens if tied).

Please show your work/calculations for the questions below. No points will be awarded otherwise.

1. If your friend chose dice B, and you chose dice A, what is the probability that you will roll the larger number in a single round? (*2 points*)

   Hint: list the nine ways this chance process can happen using a table.

   Answer: 5/9

You ->	2	4	9
1	Win	Win	Win
6	Lose	Lose	Win
8	Lose	Lose	Win

2. If your friend chose dice C, and you chose dice B, what is the probability that you will roll the larger number in a single round? (*2 points*)

   Answer: 5/9

You ->	1	6	8
3	Lose	Win	Win
5	Lose	Win	Win
7	Lose	Lose	Win

3. If your friend chose dice A, and you chose dice C, what is the probability that you will roll the larger number in a single round? (*2 points*)

   Answer: 5/9

You ->	3	5	7
2	Win	Win	Win
4	Lose	Win	Win
9	Lose	Lose	Lose

4. If you want to win the $10, would you want to choose a dice first? Or would you let your friend choose a dice first? Why? (*2 points*)

   From previous parts, it is clear that you can always find a dice that lets you win with greater than 50% probability if you go second.