

Homework Week 2

Part 1

- 1) What **best** describes a one-dimensional object that contains data
- A. Numpy array
 - B. List
 - C. Pandas Dataframe
 - D. Pandas Series

Consider the following dataframe df:

	Points	name	year
Day1	4	Joe	2012
Day2	24	Cat	2012
Day3	31	Mike	2013
Day4	2	Kim	2014
Day5	3	Amy	2014

- 2) What would the following line output: `df['Points']`
- A. The column Points as a Pandas Dataframe
 - B. The column Points as a Pandas Series
 - C. A numpy Array
 - D. A list
- 3) The data type of the column year can be
- A. int
 - B. float
 - C. string
 - D. all of the above
- 4) if you did a Group by year and applied the mean how many rows would your output have
- A. 1
 - B. 2
 - C. 3
- 5) True or False the index of the dataframe is 0,1,2,3,4

Part 2

Consider the following sets of data A=[1,2,3], B=[100,0,-100], C=[10000, 10000, 10000]

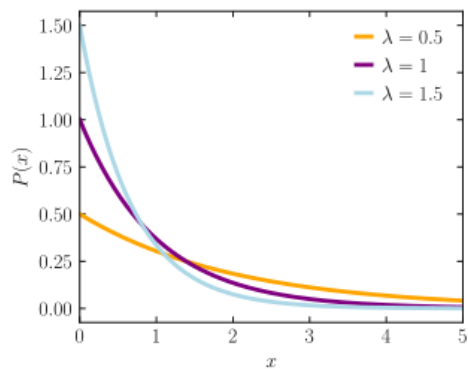
- 6) What data has the highest mean

- A.
- B.
- C.

7) What data has the highest standard deviation

- A.
- B.
- C.

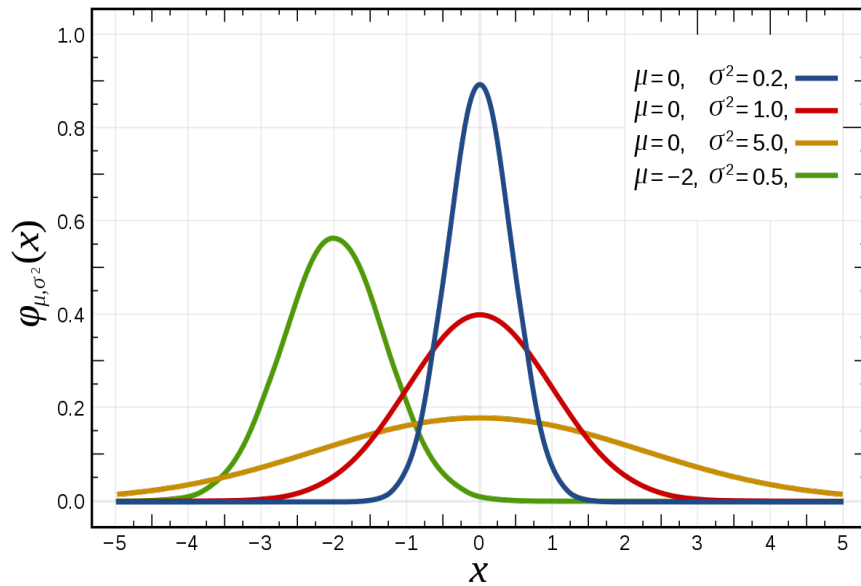
8) True or False the following data is normally distributed



9) What minimum factors do you need to characterize the normal distribution

- A. All the data
- B. Mean
- C. standard deviation
- D. mean and standard deviation

10) Select the distribution that is standardized by color



- A. Blue
- B. Yellow
- C. Red
- D. Green

11) What is the range of the Pearson correlation?

- A. Between -1 and 1
- B. Between 0 and 1
- C. Between all positive and negative numbers
- D. All positive numbers

12) What would you expect to have the **highest** correlation

- A. Gas consumption and miles driven in a car
- B. Gas consumption and car price
- C. car price and weight of driver

Solutions

- 1. D
- 2. B
- 3. D
- 4. C
- 5. False
- 6. C
- 7. B

- 8. False
- 9. D
- 10. C
- 11. A
- 12. A