Problem 1:
The relationship between height and weight of either males or females.

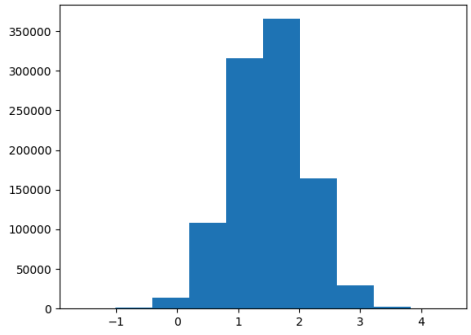
A) X would be the input height
B) Y would be the output weight
C) Measure the height of a subject and the weight and record the data
D) There is no perfect correlation and there might be a wide range of outliers and other variables acting such as age

Problem 2:
Product reviews

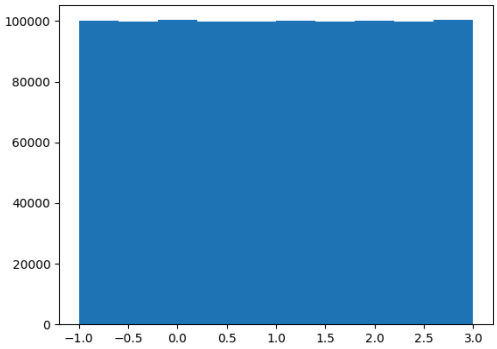
A) Key words in the review
B) Number of stars the review had (1,2,3,4,5)
C) Go on amazon and scan for keywords and input the key words along with the number of stars the product received
D) Hard to get and match every word in a review. Reviews can be convoluted and use a large variety of words

Problem 3:

A) In code
B) In code
C) Plots in code



ps1-3-c-1.png



Ps1-3-c-2.png

X looks like Gaussian Distribution and Z looks like a uniform distribution.

D) In code (including text response). Execution time: 0.7539083957672119 seconds

E) Text in code too. Execution time: 0.002992391586303711 seconds
"Seems that not using a loop is more efficient to add a constant"

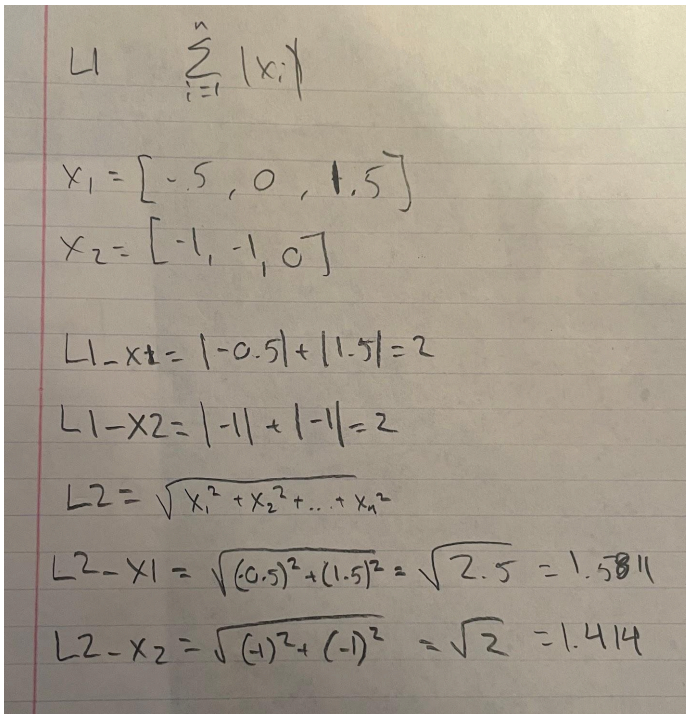
F) Text in code too ""There is a difference between the numbers each time the code is run since Z is randomly generated each time the code is run"

Problem 4:

A) In code

B) In code, including text response: x = 0.3, y = 0.4, z = 0

C)



Problem 5:

A)

```
80  #Problem 5
81
82  #Part A
83  ones = np.ones((3,10))
84  count = np.arange(1,11)
85  X = np.transpose(ones * count)
86  Y = count.reshape(10,1)
87  print("Matrix X: \n" + str(X))
88
89  #Part B
90  rand_rows = np.random.permutation(10)
```

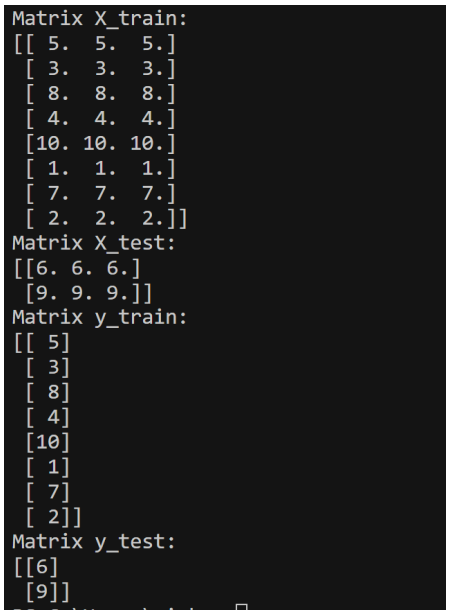
PROBLEMS OUTPUT DEBUG CONSOLE TERMINAL PORTS

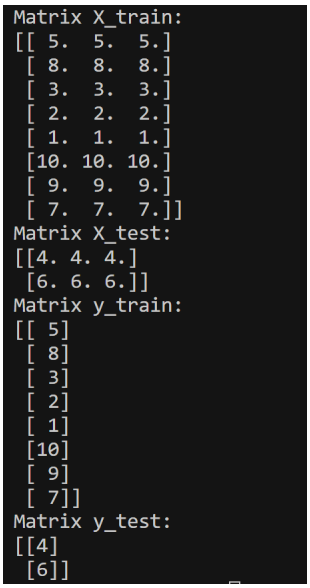
```
Matrix X:
[[ 1.  1.  1.]
 [ 2.  2.  2.]
 [ 3.  3.  3.]
 [ 4.  4.  4.]
 [ 5.  5.  5.]
 [ 6.  6.  6.]
 [ 7.  7.  7.]
 [ 8.  8.  8.]
 [ 9.  9.  9.]
 [10. 10. 10.]]
```

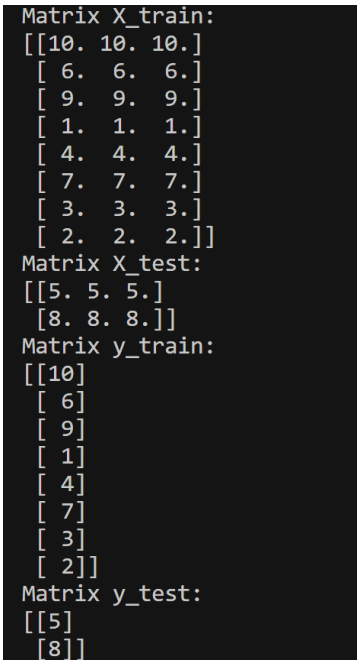
B) In Code

C) In Code

D)







Here are the screenshots of the code run three times and the outputs for the test and train matrices. The matrices are different every time as the row value that is used to pick which rows are in the test and train matrices is randomized each time the code is run.