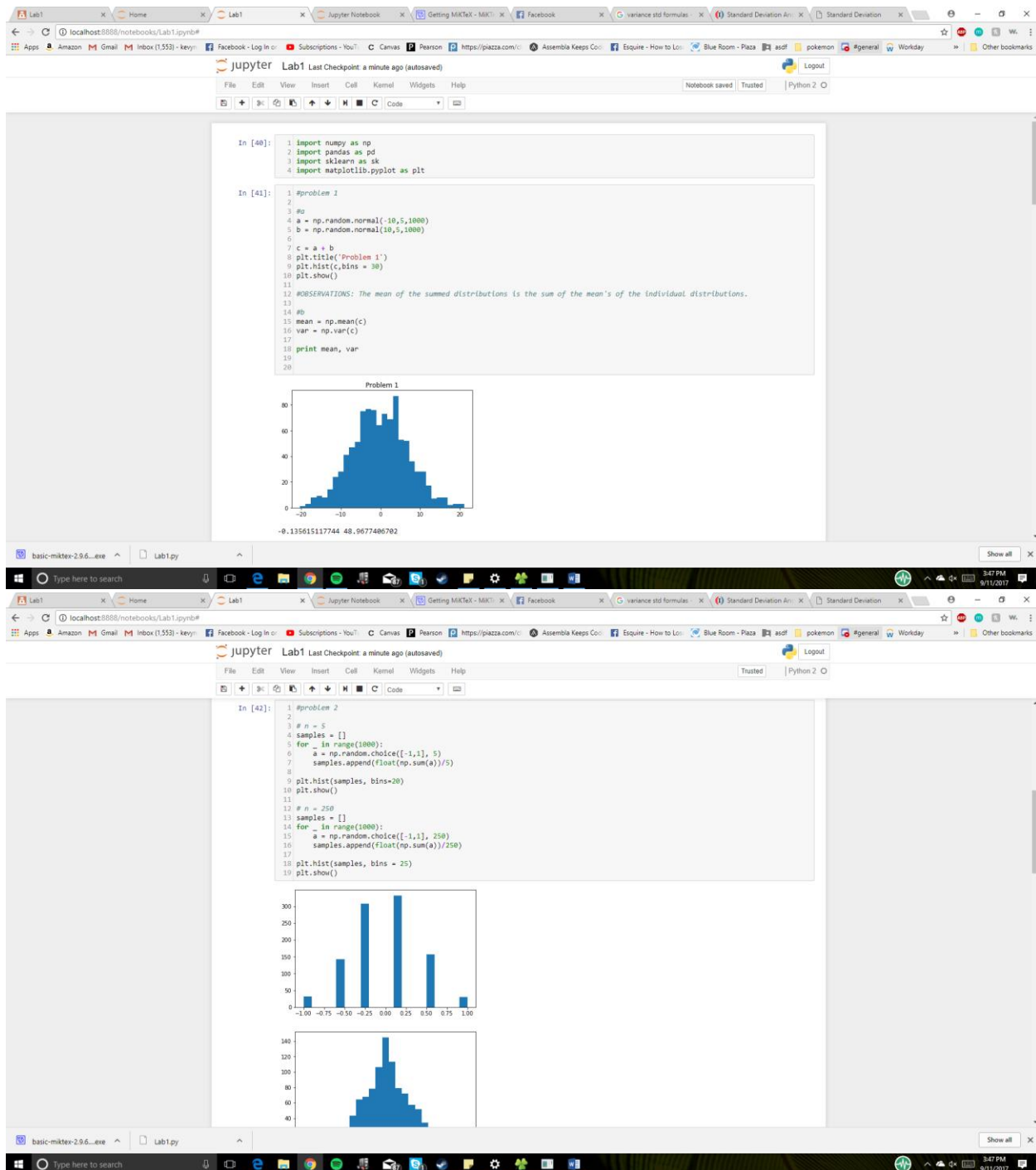
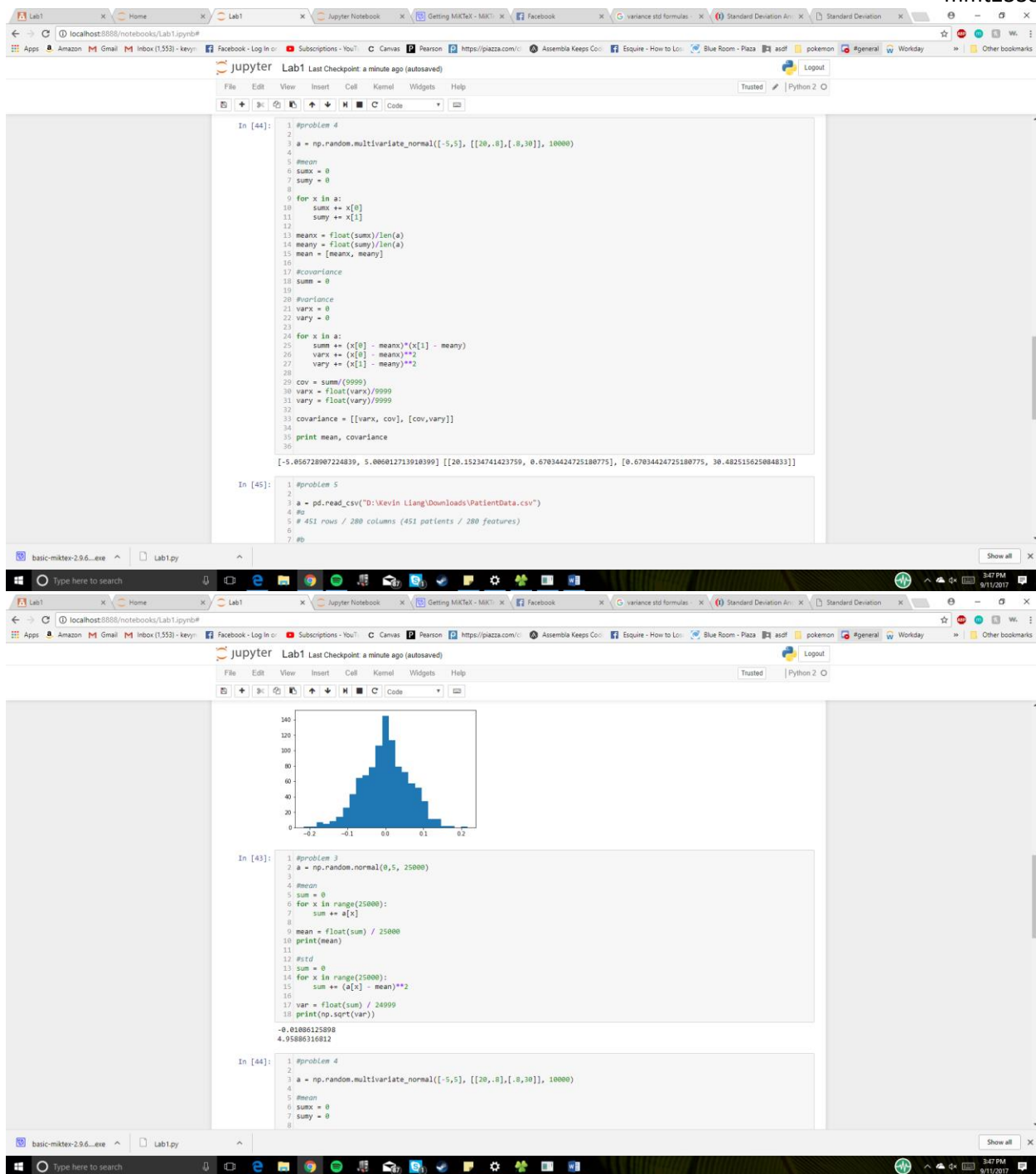


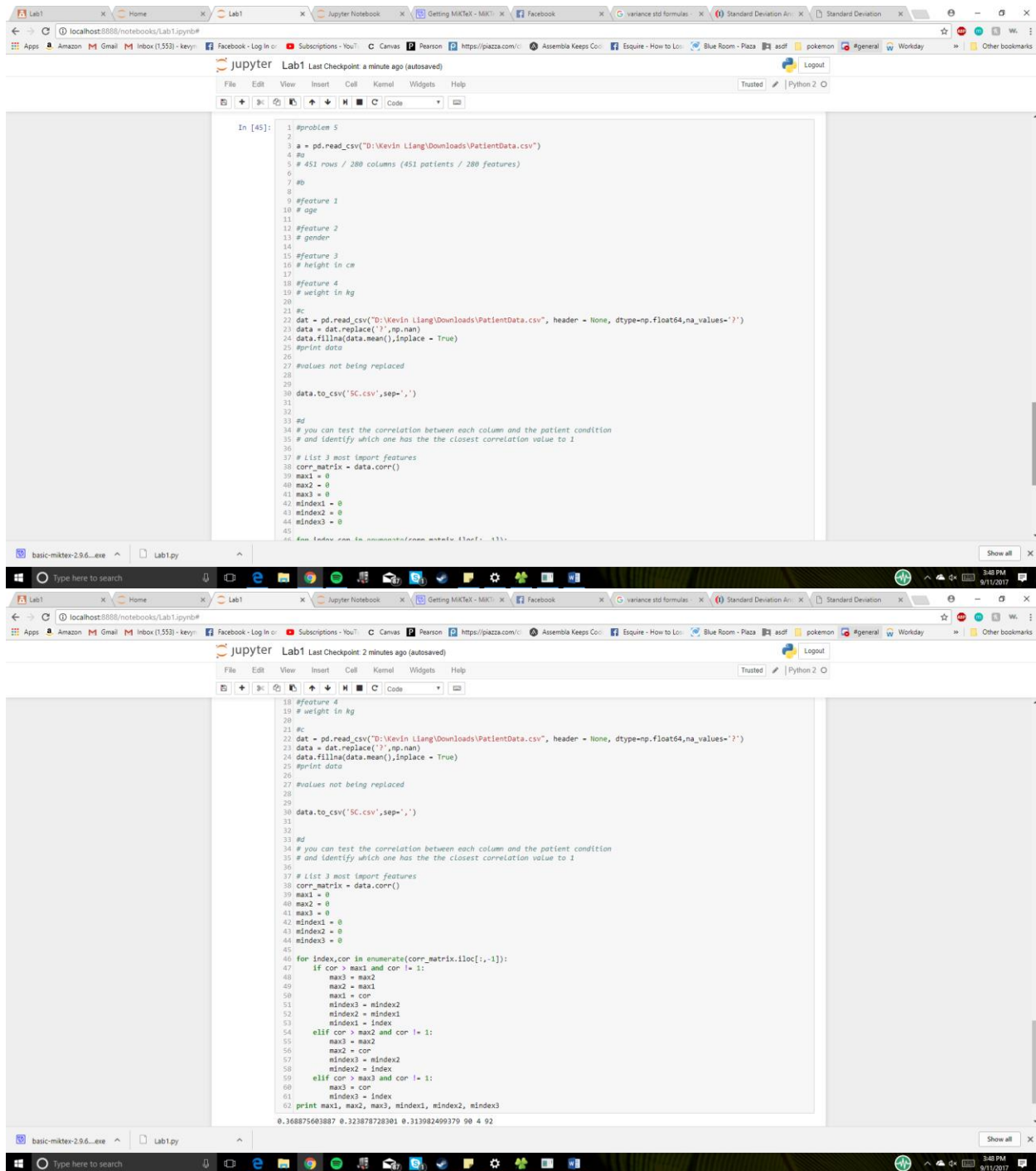
Lab1



Kevin Liang
kgl392
Matthew Tan
mmt2338



Kevin Liang
kgl392
Matthew Tan
mmt2338



The image displays two screenshots of a Jupyter Notebook interface, likely running on a local host. The browser's address bar shows the URL `localhost:8888/notebooks/Lab1.ipynb#`. The notebook is titled "Lab1 Last Checkpoint 2 minutes ago (autosaved)".

The first screenshot shows the initial code in the notebook, which includes comments and Python code for loading and processing a CSV file:

```
1 #Problem 5
2
3 a = pd.read_csv("D:\Kevin Liang\Downloads\PatientData.csv")
4 #a
5 # 451 rows / 280 columns (451 patients / 280 features)
6
7 #b
8
9 #feature 1
10 # age
11
12 #feature 2
13 # gender
14
15 #feature 3
16 # height in cm
17
18 #feature 4
19 # weight in kg
20
21 #c
22 dat = pd.read_csv("D:\Kevin Liang\Downloads\PatientData.csv", header = None, dtype=np.float64, na_values='')
23 data = dat.replace('', np.nan)
24 data.fillna(data.mean(), inplace = True)
25 #print data
26
27 #values not being replaced
28
29
30 data.to_csv('5C.csv', sep=',')
31
32
33 #d
34 # you can test the correlation between each column and the patient condition
35 # and identify which one has the closest correlation value to 1
36
37 # List 3 most import features
38 corr_matrix = data.corr()
39 max1 = 0
40 max2 = 0
41 max3 = 0
42 minindex1 = 0
43 minindex2 = 0
44 minindex3 = 0
45
46 #m index can be enumerated from matrix.iloc[:, 1:]
```

The second screenshot shows the same code with additional logic for finding the top 3 features with the highest correlation to the patient condition:

```
18 #feature 4
19 # weight in kg
20
21 #c
22 dat = pd.read_csv("D:\Kevin Liang\Downloads\PatientData.csv", header = None, dtype=np.float64, na_values='')
23 data = dat.replace('', np.nan)
24 data.fillna(data.mean(), inplace = True)
25 #print data
26
27 #values not being replaced
28
29
30 data.to_csv('5C.csv', sep=',')
31
32
33 #d
34 # you can test the correlation between each column and the patient condition
35 # and identify which one has the closest correlation value to 1
36
37 # List 3 most import features
38 corr_matrix = data.corr()
39 max1 = 0
40 max2 = 0
41 max3 = 0
42 minindex1 = 0
43 minindex2 = 0
44 minindex3 = 0
45
46 for index, cor in enumerate(corr_matrix.iloc[:, 1:]):
47     if cor > max1 and cor != 1:
48         max3 = max2
49         max2 = max1
50         max1 = cor
51         minindex3 = minindex2
52         minindex2 = minindex1
53         minindex1 = index
54     elif cor > max2 and cor != 1:
55         max3 = max2
56         max2 = cor
57         minindex3 = minindex2
58         minindex2 = index
59     elif cor > max3 and cor != 1:
60         max3 = cor
61         minindex3 = index
62 print max1, max2, max3, minindex1, minindex2, minindex3
```

The output of the final print statement is displayed at the bottom of the code cell:

```
0.368875603887 0.323878728301 0.313982499379 90 4 92
```

$$1. a). \frac{1}{4} + \frac{1}{2} = \sqrt{\frac{7}{12}}$$

$$b). \frac{1}{3} / \frac{2}{6} = \sqrt{\frac{2}{3}}$$

$$c). \text{Var}(X) = E[X^2] - E[X]^2 = 1\left(\frac{1}{4} + \frac{1}{2}\right) - \left(1\left(\frac{1}{4} + \frac{1}{2}\right)\right)^2 \\ = \frac{7}{12} - \frac{49}{144} = \sqrt{\frac{35}{144}}$$

$$d). \text{Var}(X|Y=1) = E[X^2|Y=1] - (E[X|Y=1])^2 \\ = \frac{2}{3} - \frac{4}{9} = \sqrt{\frac{2}{9}}$$

$$e). E[X^3 + Y^2 + 3Y^3|Y=1] = 3 \cdot \frac{1}{3} + 5 \cdot \frac{2}{3} = \sqrt{\frac{13}{3}}$$

$$2. v_1 = [1, 1, 1], v_2 = [1, 0, 0]$$

$$v_1 \times v_2 = [0, 1, -1]$$

$$P_1 = \text{proj}_N P_1 = \frac{P_1 \cdot N}{N \cdot N} = \frac{0+3-3}{1+1} [0, 1, 1] = 0$$

$$\text{proj}_N P_1 = [3, 3, 3]$$

$$P_2 = \text{proj}_N P_2 = \frac{2-3}{2} [0, 1, -1] = [0, -\frac{1}{2}, \frac{1}{2}]$$

$$\text{proj}_N P_2 = P_2 - \text{proj}_N P_2 = [1, 2, 3] - [0, -\frac{1}{2}, \frac{1}{2}] = [1, \frac{5}{2}, \frac{5}{2}]$$

$$P_3 = \text{proj}_N P_3 = \frac{-1}{2} [0, 1, -1] = [0, -\frac{1}{2}, \frac{1}{2}]$$

$$\text{proj}_N P_3 = P_3 - \text{proj}_N P_3 = [0, 0, 1] - [0, -\frac{1}{2}, \frac{1}{2}] = [0, \frac{1}{2}, \frac{1}{2}]$$

3. Bin CDF: $\sum_{i=0}^x \binom{100}{i} p^i (1-p)^{(100-i)} I_{(0,1)}(i)$

binom CDF (50, 100, 0.666) in MATLAB

$\Rightarrow 4.419 \times 10^{-4}$