

1. By running the python code included in file *Question1.lbai5.py*, we get the output as following:

---

```

count    1000.00000
mean      76.71500
std       13.16355
min       37.00000
25%      68.00000
50%      77.00000
75%      87.00000
max       100.00000
Name: Midterm, dtype: float64
-----
Mode: [77 83]
-----
Variance:173.279054054

```

---

This result indicate that

- a. Max = 100, Min = 37
  - b. Q1 = 68, median = 77, Q3 = 87
  - c. mean = 76.715
  - d. mode1 = 77 and mode2 = 83
  - e. empirical variance = 173.279
2. By running the python code included in file *Question2.lbai5.py*, we get the output as following:

---

```

Original empirical variance: 173.279054054
Normalized empirical variance: 1.001001001
-----
Normalized 90:1.00922621869
-----
Pearson's Correlation Coefficient:
      Midterm    Final
Midterm 1.000000 0.544425
Final   0.544425 1.000000
-----
Covariance:
      Midterm    Final
Midterm 173.279054 78.254194
Final   78.254194 119.232176

```

---

- a. Original empirical variance > Normalized empirical variance
- b. Normalized 90 = 1.009
- c. Pearson's Correlation Coefficient = 0.544
- d. Covariance = 78.254

3. a.

$$\text{Jaccard Coefficient} = \frac{q}{q + r + s} = \frac{58}{58 + 120 + 2} = 0.322$$

- b. By running the python code included in file *Question3.lbai5.py*, we get the output as following:

---

```

b.
Norm-1:6152.0
Norm-2:715.327896842
Norm-infty:170.0
-----
c.
Cosine similarity:[[ 0.84140403]]
-----
d.
K-L Divergence:0.207080937332

```

---

This result indicate that

- 1. When  $h = 1$ , dist = 6152.0
  - 2. When  $h = 2$ , dist = 715.328
  - 3. When  $h = \infty$ , dist = 170.0
  - c. Cosine similarity = 0.841
  - d. K-L Divergence = 0.207
4. By running the python code included in file *Question4.lbai5.py*, we get the output as following:

---

```
chisquare: 2468.183
```

---

This result indicate that  $\chi^2 = 2468.183$