# Homework #2

*Instructor:* Dr. Zafeirakis Zafeirakopoulos   *Name: Baran Hasan Bozduman*   *Student Id: 171044036*
*Assistant: Gizem Süngü*

**Course Policy**: Read all the instructions below carefully before you start working on the assignment, and before you make a submission.

- It is not a group homework. Do not share your answers to anyone in any circumstance. Any cheating means at least -100 for both sides.

- Do not take any information from Internet.

- No late homework will be accepted.

- For any questions about the homework, send an email to gizemsungu@gtu.edu.tr.

- Submit your homework (both your latex and pdf files in a zip file) into the course page of Moodle.

- Save your latex, pdf and zip files as "Name_Surname_StudentId".{tex, pdf, zip}.

- The answer which has only calculations without any formula and any explanation will get zero.

- The deadline of the homework is 07/06/20 23:55.

- I strongly suggest you to write your homework on LATEX. However, hand-written paper is still accepted **IFF** your hand writing is **clear and understandable to read**, and the paper is well-organized. Otherwise, I cannot grade your homework.

- You do not need to write your Student Id on the page above. I am checking your ID from the file name.

---

**Problem 1:**                                                    (10+10+10+10+10+10+40 = 100 points)

**WARNING:** Please show your OWN work. Any cheating can be easily detected and will not be graded.

For the question, please follow the file called manufacturing_defects.txt while reading the text below.

In each year from 2000 to 2019, the number of manufacturing defects in auto manufacturers were counted. The data was collected from 14 different auto manufactory companies. The numbers of defects for the companies are indicated in 14 columns following the year column. Assume that the number of manufacturing defects per auto company per year is a random variable having a Poisson($\lambda$) and that the number of defects in different companies or in different years are independent.
(Note: You should implement a code for your calculations for each following subproblem. You are free to use any programming languages (Python, R, C, C++, Java) and their related library.)

**(a)** Give a table how many cases occur for all companies between 2000 and 2019 for each number of defects (# of Defects).
Hint: When you check the file you will see: # of Defects = {0, 1, 2, 3, 4}.

**(b)** Estimate $\lambda$ from the given data.
$$\frac{(defects1 * cases1) + (defects2 * cases2)...}{cases1 + cases2...} = 0.7$$

**(c)** $Update Table 1 in Table 2 with Poisson predicted cases with the estimated \lambda.$

| \# of Defects | \# of cases in all company between the years |
|---|---|
| 0 | 144 |
| 1 | 91 |
| 2 | 32 |
| 3 | 11 |
| 4 | 2 |

Table 1: Actual cases

| \# of Defects | \# of cases in all companies between the years | Predicted \# of cases in all companies between the years |
|---|---|---|
| 0 | 144 | 139.0438850615947 |
| 1 | 91 | 97.33071954311627 |
| 2 | 32 | 34.06575184009069 |
| 3 | 11 | 7.948675429354495 |
| 4 | 2 | 1.3910182001370366 |

Table 2: Actual vs. Predicted Cases

**(d)** Draw a barplot for the actual cases (Table 2 in column 2) and the predicted cases (Table 2 column 3) with respect to # of defecrs. You should put the figure.
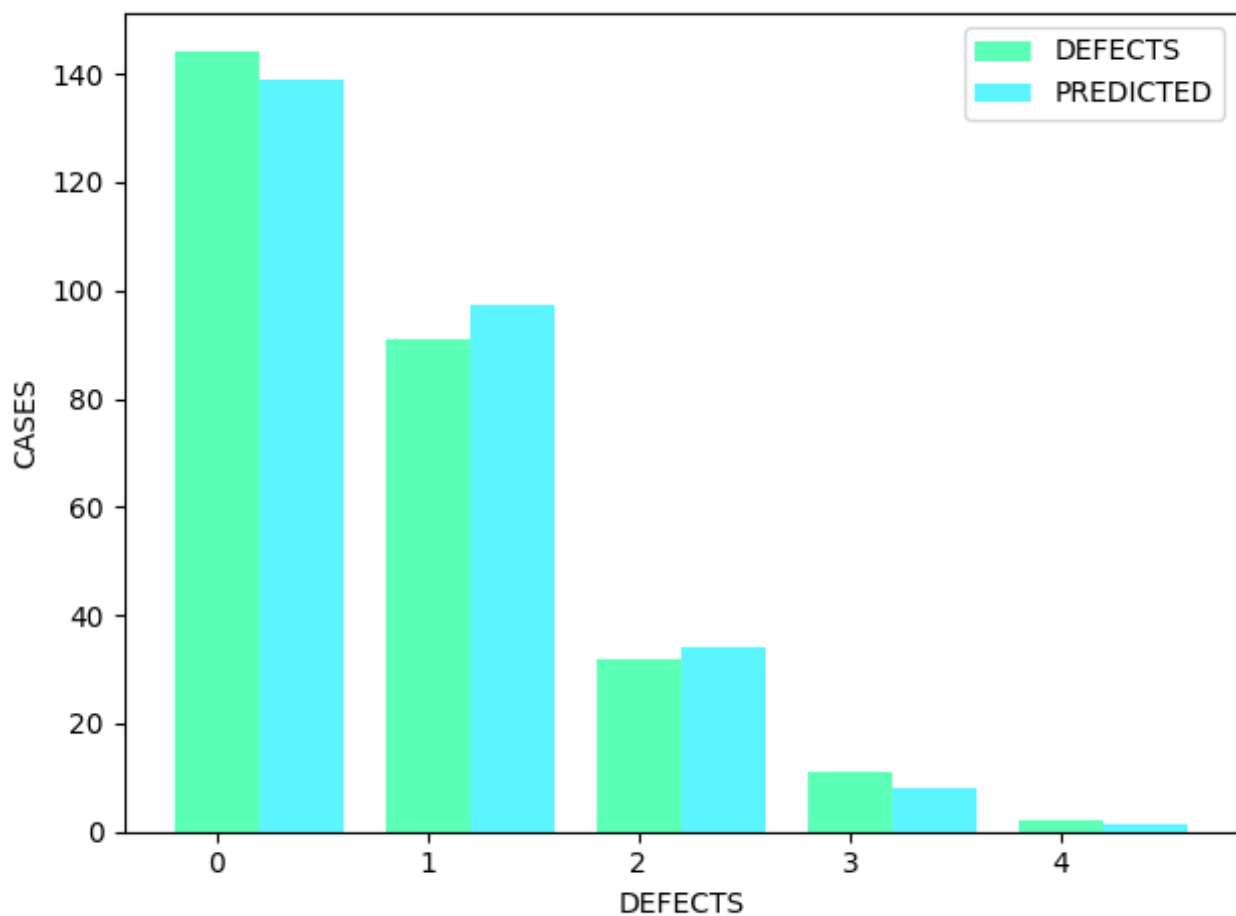


**(e)** According to the barplot in (c), does the poisson distribution fit the data well? Compare the values of the actual cases and the values of the poisson predicted cases, and write your

opinions about performance of the distribution.
As we can see in the table the real and predicted values are close to each other. We can say that the distribution data almost fit. So we can make close predictions using poisson distribution.

**(f)** According to your estimations above, write your opinions considering your barplot and Table 2.Do you think that road transportation is dangerous for us? Whether yes or no, explain your reason.
There is not defects half of them but still we can not ignore the other half so we can not say its not dangerous. since the predicted values almost satisfy the real ones.

**(g)** Paste your code that you implemented for the subproblems above. Do not forget to write comments on your code.
Example:

- The common code block for all subproblems
  Paste here. Your code should read the file and compute other things which the following subproblems need.

- The code block for (a)

- 
```python
from numpy import printoptions
import matplotlib.pyplot as plt
import math
import pandas as pd
nl_count=0
#count line number
file = open("manufacturing_defects.txt", "r")
for line in file:
    if line != "\n":
        nl_count += 1
file.close()
#count column number
tab_count=0
for ch in line:
    if ch == "\t":
        tab_count+=1
#creating a dictionary and add the defects number for different number defects
total_each = dict()
file = open("manufacturing_defects.txt", "r")
for line in file:
    value = line.split('\t')
    for i in value[2:]:
        index = int(i)
        if index in total_each:
            total_each[index]+=1
        else:
            total_each[index] =1
file.close()
print("Give a table how many cases occur for all companies between 2000 and 2019 for each number
     of defects")
#printing table
print("# of Defects |# of cases in all company between the years")
print("-----------------------------------------------------")
for i in range(0, len(total_each)):
    print(str(i),"          |",str(total_each[i]))
```

Paste here. Your code should compute the values in Table 1 column 2.

- The code block for (b)

```python
print(" Estimate  from the given data.")
total1=0
total2=0
#calculating lamda as indicated part b
for i in range(0, len(total_each)):
    total1+= total_each[i]*i
for i in range(0, len(total_each)):
    total2+= total_each[i]
print(total1/total2)
lamda=total1/total2
```

Paste here. Your code should compute $\lambda$.

- The code block for (c)

```python
#collecting predicted values in another dictionary
each_predict= dict()
print("Update Table 1 in Table 2 with Poisson predicted cases with the estimated .")
#Poisson formula for each one
for i in range(0, len(total_each)):
    each_predict[i] = pow(math.e, -1*lamda)* pow(lamda, i)
    each_predict[i]/= math.factorial(i)
    each_predict[i]*= total2

print("# of Defects |# of cases in all company between the years|")
print("--------------------------------------------------------")
for i in range(0, len(total_each)):
    print(str(i),"        |",str(total_each[i]),"    |",each_predict[i])
```

Paste here. Your code should compute the values in Table 2 column 3.

- The code block for (d)

```python
print("Update Table 1 in Table 2 with Poisson predicted cases with the estimated .")
#creating values from list of predicted and actual values
total_each_l= []
each_predict_l = []
for i in total_each:
    total_each_l.append(total_each[i])
    each_predict_l.append(each_predict[i])

defects_var = []
for i in total_each:
    defects_var.append(i)

size = len(total_each)
arangee = numpy.arange(size)
opacity = 0.8
#graph x and y values
pltx = plt.bar(arangee,total_each_l, 0.40, alpha=opacity, color='#33FFA4', label='DEFECTS')
plty = plt.bar(arangee+0.40, each_predict_l, 0.40, alpha=opacity,color='#33F3FF',
    label='PREDICTED')
#graph labels
plt.xlabel('DEFECTS')
plt.ylabel('CASES')
plt.xticks(arangee)
```

```
plt.legend()

plt.tight_layout()
plt.show()
```

Paste here. Your code should draw the barplot.