

Homework #2

Instructor: Dr. Zafeirakis Zafeirakopoulos
 Assistant: Gizem Süngü

Name: Aygün Bayır

Student Id: 161044119

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 L^AT_EX. 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(λ) 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}.

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

Table 1: Actual cases

(b) Estimate λ from the given data.
 $\lambda = 0.7$

(c) Update Table 1 in Table 2 with Poisson predicted cases with the estimated λ .

\# of Defects	\# of cases in all companies between the years	Predicted \# of cases in all companies between the years
0	144	139.04
1	91	97.33
2	32	34.07
3	11	7.95
4	2	1.39

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 defects. You should put the figure.

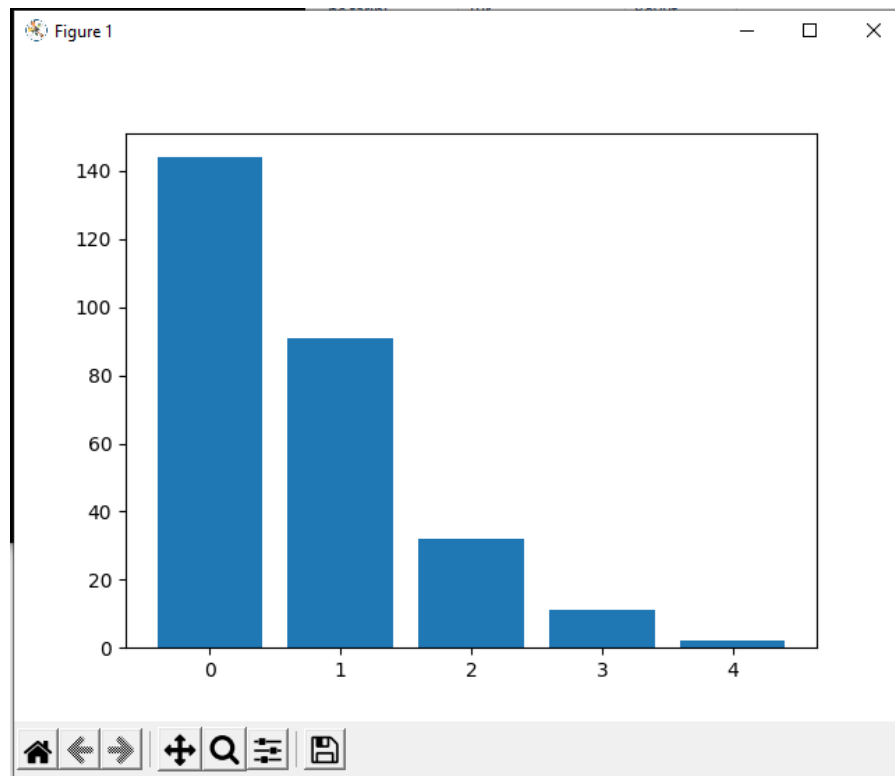


Figure 1: Number of defects

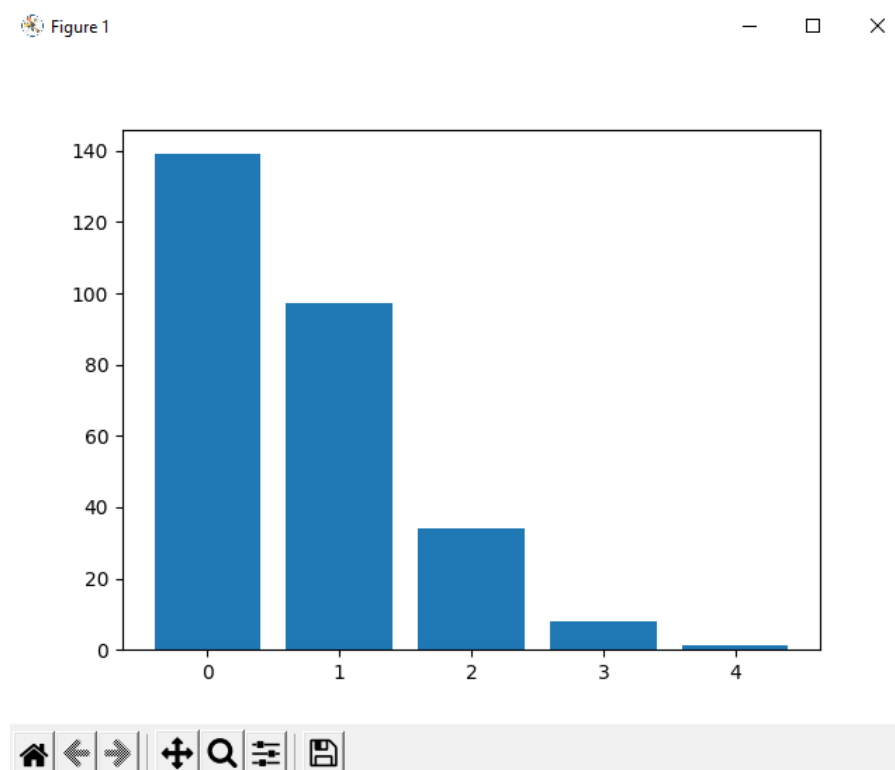


Figure 2: Predicted number of defects

(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.

Predicted number of cases which we get by using poisson distribution are very close to our data. I can say it fit the data very well. For example, for zero, number of cases is 144 and predicted number is 139.04 which is very close to number of cases. I honestly say that poisson distribution give us numbers that are approximately 5 percent more or less to our data numbers which is very fine.

(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.

I can say that no, road transportation is not dangerous for us. Since we know that every year, car companies produce millions of cars and we see in our data there are only tens or hundreds of manufacturing defects are happenning. It is very small numbers for these kind of big number of productions. So, margin of error is very low. Therefore I can say that road transportation is very safe.

(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.

Listing 1: Common part

```
# import csv library for reading text file
import csv
with open('manufacturing-defects.txt', 'r') as file:
    reader = csv.reader(file, delimiter='\t') # all values are seperated with tab
    data = list(reader)

# import matplotlib library for drawing barplot
import matplotlib.pyplot as plt

# Function to calculate factorial of number n
def factorial(n):
    if n == 0:
        return 1
    else:
        return n * factorial(n-1)
```

- The code block for (a)
Paste here. Your code should compute the values in Table 1 column 2.

Listing 2: Part A

```
# Part A
# Calculation
a_counts = [0 for x in range(5)]
for x in range(20):
    for y in range(2,16):
        if int(data[x][y]) == 0:
            a_counts[0] = a_counts[0] + 1
        elif int(data[x][y]) == 1:
            a_counts[1] = a_counts[1] + 1
        elif int(data[x][y]) == 2:
            a_counts[2] = a_counts[2] + 1
        elif int(data[x][y]) == 3:
            a_counts[3] = a_counts[3] + 1
        elif int(data[x][y]) == 4:
            a_counts[4] = a_counts[4] + 1

# Printing result
print()
```

```

print("*****Part A*****")
print("#_of_Defects\t#_of_defects_in_all")
print("\t\tcompanies_between_the_years")
print("_____")
for x in range(5):
    print(x, "\t\t", a_counts[x])

total = 0.0
for x in range(5):
    total = total + float(a_counts[x])

```

- The code block for (b)
Paste here. Your code should compute λ .

Listing 3: Part B

```

# Calculating mean
mean = 0.0
for x in range(20):
    for y in range(2,16):
        mean = mean + float(data[x][y])
mean = mean / total

```

- The code block for (c)
Paste here. Your code should compute the values in Table 2 column 3.

Listing 4: Part C

```

euler = 2.7182818

# Calculating poisson predicted cases
poisson_counts = [0.0 for x in range(5)]
for x in range(5):
    poisson_counts[x] = (mean**x)*(1/(euler**mean)) / float(factorial(x)) * total

# Printing results
print()
print("*****Part B*****")
print("Lambda=", mean)
print()
print("*****Part C*****")
print("#_of_Defects\t#_of_defects_in_all\tPredicted #_of_defects")
print("\t\tcompanies_between\tin_all_companies")
print("\t\tthe_years\t\tbetween_the_years")
print("_____")
for x in range(5):
    print(x, "\t\t", a_counts[x], "\t\t\t", float('%0.2f' % (poisson_counts[x])))

```

- The code block for (d)
Paste here. Your code should draw the barplot.

Listing 5: Part D

```

# Part D
print()
print("*****Part D*****")
print("Barplot_drawing(s)_will_be_shown_in_new_window(s).")
# Common in both figure
langs = [0,1,2,3,4]

# Barplot: Number of Defects
plt.bar(langs, a_counts)
plt.show()

# Barplot: Predicted number of Defects
plt.bar(langs, poisson_counts)
plt.show()

```
