MATH 118: Statistics and Probability **(Due: 07/06/21)**

Homework #2

Instructor: Dr. Zafeirakis Zafeirakopoulos *Name:Mustafa Tokgöz* *Student Id:171044077*
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(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}.

(b) Estimate λ from the given data.

0.7

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

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

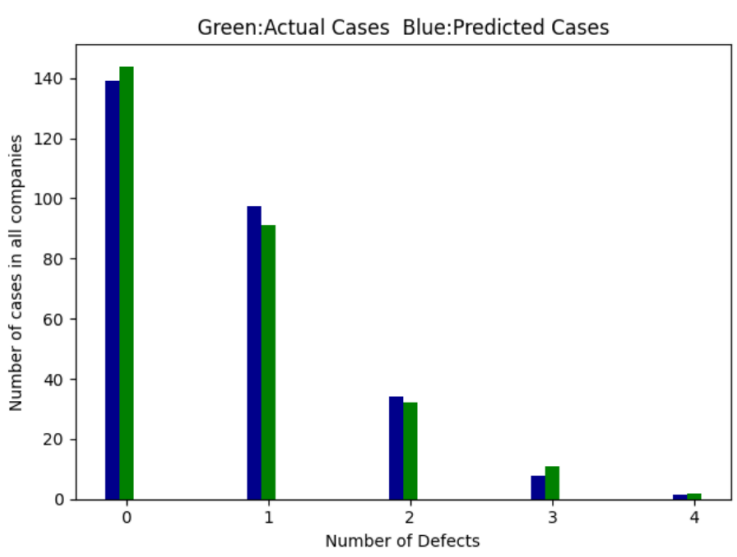
Table 1: Actual cases

\# of Defects	\# of cases in all companies between the years	Predicted \# of cases in all companies between the years
0	144	139.04388506159466
1	91	97.33071954311626
2	32	34.06575184009068
3	11	7.948675429354496
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.

Answer of d)



(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.

Answer of e:
I think poisson distribution fits the data well. Values between actual cases and poisson predicted cases are close to each other. For instance,For case 0 actual case is 144 and poisson predicted case is 139.0. So They are close to each other when we look other cases.

(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?

Answer of f:
I think poisson distribution is predicting well because actual case is so close to predicted case with respect to table 2 and bar plot. I think there is a risk for road transportation according to table 2 and bar plot and It is dangerous for us. Because there is a lot of defects greater and equal to 1. So I think It is dangerous.

(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.

```
        name = "manufacturing_defects.txt"
        //opens file
        file = open(name)
        data = []
        landa = 0 // actually lambda but python doesn't let to name it
        cases = [0, 1, 2, 3, 4] //cases
        //reads file and splits them
        for line in file:
            values = []
            for i in line.split():
                values.append(int(i)) //case values
            if len(values) != 0:
                data.append(values)


```

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

```
    print("Question a) ")
        cases_of_companies = []
        //to campute the cases for all company
        for i in range(len(cases)):
            case = 0
            for j in range(len(data)):
                for k in range(len(data[j])-2):
                    if data[j][k+2] == i:
                        case = case + 1
            cases_of_companies.append(case)
        print(cases_of_companies)

```

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

```
    print("Question b) ")
        i = 0
        // this computes mean value because lambda is mean
        for i in range(len(data)):
            for j in range(len(data[i]) - 2):
                landa = landa + data[i][j + 2]
        landa = landa / (len(data) * (len(data[0]) - 2))
        print(landa)


```

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

```
    print("Question c) ")
        defects = 0
        pre_case = []
        //to compute the poisson distrubution
        for k in range(len(cases_of_companies)):
            defects = defects + cases_of_companies[k]

```

```
8
9         //using python library
10        for i in range(len(cases_of_componies)):
11            pre_case.append(defects * scipy.stats.poisson.pmf(i, landa))
12        print(pre_case)
13
```

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

```
1         //to show in barplot
2         ax = plt.subplot(111)
3         w = 0.1
4         show_cases = []
5         //to seem good for graphic
6         for i in range(len(cases)):
7             show_cases.append(cases[i]-w)
8
9         ax.bar(show_cases, pre_case, color ='darkblue', align = 'center',width=w)
10        ax.bar(cases,cases_of_componies, color='green', align = 'center', width = w)
11        plt.xlabel("Number of Defects")
12        plt.ylabel("Number of cases in all companies")
13        plt.title("Green:Actual Cases  Blue:Predicted Cases")
14        plt.show()
15
```