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

Homework #2

Instructor: Dr. Zafeirakis Zafeirakopoulos *Name:* *Student Id:*
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}.

Solution Approach: i read all from excell file and kept hem on tall array.After that i add 0 1 2 3 4 ones and print them. a,b,c you can see in output.txt. i write the outputs part d i print here.

(b) Estimate λ from the given data.

Solution Approach: I did in this part also from my tall array add one line and / 14 to find one year and i did all of them for 20 years then calculate mean of them.
$\lambda = 0.7$

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

$f(x;\lambda) = \Pr(X{=}x) = \frac{\lambda^x e^{-\lambda}}{x!}$ I find this formula and used it to calculate poisson distribution and I multiplied these values with total number of cases(280) to find the predicted values.We already know that lambda value is 0.7 from the solution of previous question.

Predicted values:

$\frac{0.7^0 e^{-0.7}}{0!}$ * 280 = 139.046 for 0-defect
$\frac{0.7^1 e^{-0.7}}{1!}$ * 280 = 97.33 for 1-defect
$\frac{0.7^2 e^{-0.7}}{2!}$ * 280 = 34.066 for 2-defects
$\frac{0.7^3 e^{-0.7}}{3!}$ * 280 = 7.94 for 3-defects
$\frac{0.7^4 e^{-0.7}}{4!}$ * 280 = 1.39 for 4-defects

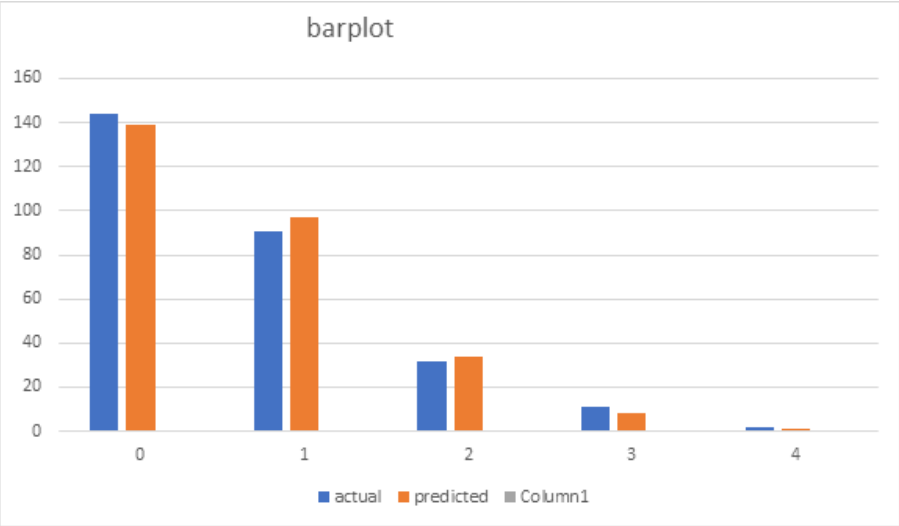
\# of Defects	\# of cases in all company between the years
0	144
1	91
2	32
3	11
4	2

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	139.04
1	91	97.33
2	32	34.06
3	11	7.94
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.



(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.

```
# of casesin all companiesbetween the years | Predicted # of casesin all companiesbetween the years
0             144          |           139.04
1              91          |           97.33
2              32          |           34.06
3              11          |            7.94
4               2          |            1.39

Process finished with exit code 0
```

According to the barplot in (c): it is fit to data when compare (0 differrence = 5, 1 difference 6, 2 difference =2, ,1 difference =3, 0 difference = 0.60)them sometimes Actual cases a little bigger sometimes predicted. I think performance is great.

(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 think, no. Because if we examine the distribution of total number of defects then we see that total number of 0 defect is greater than others.(0 is 144,4 is 2) from 0th to 4th error cars is increasing so we can say that companies don't produce too many defective products.

(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.

```
import java.io.*;
import java.util.Scanner;
// main needed for all
public  static void main(String[] args)throws Exception
        {
            Scanner scanner = new Scanner(new File("C:\\Users\\tangel\\Desktop\\olasil
            int [][] tall = new int [100][100];
```

```
            int [] my = new int [10000];
            int i = 0,j=0;
            while(scanner.hasNextInt())
            {

                //end of one line
                if(j == 16){
                    j=0;
                    i++;
                }

                //copy read
                tall[i][j] = scanner.nextInt();
                //forward on array
                j++;
            }
            //System.out.println(tall[1][15]);

            Read a = new Read();
            //a.write_file();


           // a.question_one(tall);
           // a.calc_lamda(tall);
            a.third_question();
           // a.table();


        }

        \item The code block for (a)\\
                Paste here. Your code should compute the values in Table \ref{tab1} co
\begin{lstlisting}
//to work this you should run my main first
        public void question_one(int tall[][]) {

            //our Defects
            // for zero
            int num = 0;
            int num1 = 0;
            int num2 = 0;
            int num3 = 0;
            int num4 = 0;
            for (int i = 0; i < 20; i++) {
                for (int j = 2; j < 16; j++) {
                    if (0 == tall[i][j])
                        num++;
                    if (1 == tall[i][j])
                        num1++;
                    if (2 == tall[i][j])
                        num2++;
                    if (3 == tall[i][j])
                        num3++;
                    if (4 == tall[i][j])
                        num4++;
                }
            }
```

```
            System.out.println(num);
            System.out.println(num1);
            System.out.println(num2);
            System.out.println(num3);
            System.out.println(num4);


            //for one
            //for two
            //for three
            //for four


        }
        \\ this is also table which i created
            public void table(){

            System.out.println("# of " + "Defectss  | # of cases" + "in all company" +
            System.out.println("0\t\t\t   | 144");
            System.out.println("1\t\t\t   | 91");
            System.out.println("2\t\t\t   | 32");
            System.out.println("3\t\t\t   | 11");
            System.out.println("4\t\t\t   | 2");

            System.out.println("# of cases" + "in all companies" + "between the years
"in all companies" +  "between the years");
            System.out.println("0\t\t\t  144\t\t\t|\t\t\t 139.04");
            System.out.println("1\t\t\t   91\t\t\t|\t\t\t 97.33");
            System.out.println("2\t\t\t   32\t\t\t|\t\t\t 34.06 ");
            System.out.println("3\t\t\t   11\t\t\t|\t\t\t  7.94");
            System.out.println("4\t\t\t    2\t\t\t|\t\t\t  1.39");




    }

            \item The code block for (b)\\

            public void calclamda( int tall[][]){
        float val=0;
        float temp=0;
        for (int i =0;i<20;i++ ) {
            for (int j = 2; j <16 ; j++) {
                val+=tall[i][j];
            }
            val/=14.0;
            temp+=val;
            val=0;
        }

            temp/=20.0;
        System.out.println(temp);
    }
```

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

```
static int factorial(int n) {
        if (n == 0)
            return 1;
        else
            return (n * factorial(n - 1));
    }


public void third_question( ){
        int total=280;
        double e= 2.7182;
        double lamda = 0.7;
        double result;


        double y=Math.pow(e,lamda);

        for(int i=0;i<5;i++){
            double x= Math.pow(lamda,i);
            result= x/(y * factorial(i));
            result*=280;
            System.out.println(result);

        }




}
```

- The code block for (d)

```
            import org.jfree.chart.ChartFactory;
    import org.jfree.chart.ChartPanel;
    import org.jfree.chart.JFreeChart;
    import org.jfree.chart.plot.PlotOrientation;
    import org.jfree.data.category.CategoryDataset;
    import org.jfree.data.category.DefaultCategoryDataset;
    import org.jfree.ui.ApplicationFrame;
    import org.jfree.ui.RefineryUtilities;

    public   class BarChart_AWT extends ApplicationFrame {

        public BarChart_AWT( String applicationTitle , String chartTitle ) {
            super( applicationTitle );
            JFreeChart barChart = ChartFactory.createBarChart(
                    chartTitle,
                    "Category",
                    "Score",
                    createDataset(),
                    PlotOrientation.VERTICAL,
                    true, true, false);

            ChartPanel chartPanel = new ChartPanel( barChart );
            chartPanel.setPreferredSize(new java.awt.Dimension( 560 , 367 ) );
            setContentPane( chartPanel );
        }
```

```
    private CategoryDataset createDataset( ) {
        final String estimated = "ESTIMATED";
        final String real = "REAL";
        final String ford = "FORD";
        final String zero = "zero";
        final String two = "two";
        final String one = "one";
        final String three = "three";
        final String four = "four";
        final DefaultCategoryDataset dataset =
                new DefaultCategoryDataset( );

        dataset.addValue( 139.04 , estimated , zero );
        dataset.addValue(  97.33 , estimated , one );
        dataset.addValue(  34.06 , estimated , two );
        dataset.addValue(   7.94 , estimated , three );
        dataset.addValue(     1.39, estimated , four );

        dataset.addValue( 144 , real , zero );
        dataset.addValue(  91 , real , one );
        dataset.addValue(  32    , real , two );
        dataset.addValue( 11 , real , three );
        dataset.addValue( 2 , real , four);



        return dataset;
    }

    public static void main( String[ ] args ) {
        BarChart_AWT chart = new BarChart_AWT("Car Usage Statistics",
                "Which car do you like?");
        chart.pack( );
        RefineryUtilities.centerFrameOnScreen( chart );
        chart.setVisible( true );
    }
}
```