

Assignment – II

Answer Questions 1 to 8 by Writing Code in Python.

1. The number of customer returns in a retail chain per day follows a Poisson distribution at a rate of 25 returns per day. Write Python code to answer the following questions:

- a) Calculate the probability that the number of returns exceeds 30 in a day.

```
from scipy import stats
1 - stats.poisson.cdf(30, 25)
```

```
0.1366911308473363
```

- b) If the chance of fraudulent return is 0.05, calculate the probability that there will be at least 2 fraudulent returns in any given day.

```
1 - stats.poisson.cdf(2-1, 25*0.05)
```

```
0.35536420706457217
```

2. A student is applying for Masters course in 8 US Universities and believes that she has in each of the eight universities a constant and independent 0.42 probability of getting selected. Write code to answer the following questions:

- a) What is the probability that she will get call from at least 3 universities?

```
p = 0.42
n = 8
pmf = list(stats.binom.cdf(range(0,3), n, p))
1 - sum(pmf)
```

```
0.6251756532117759
```

- b) What is the probability that she will get calls from exactly 4 universities?

```
stats.binom.pmf(4, n, p)
```

```
0.2464944673605121
```

3. The time-of-failure of a machine follows exponential distribution with mean time between failures (MTBF) estimated to be 85 hrs. Write code to answer the following questions:

- a) Calculate the probability that the system will fail before 85 hrs.

```
stats.expon.cdf(85,  
                loc = 1/85,  
                scale = 85)
```

0.6320696377349755

- b) Calculate the probability that it will not fail up to 150 hrs.

```
1 - stats.expon.cdf(150,  
                    loc = 1/85,  
                    scale = 85)
```

0.17126084522766882

4. As per a survey on use of pesticides among 1000 farmers in grape farming for around 10 acres of grape farmland, it was found that the grape farmers spray 38 litres of pesticides in a week on an average with the corresponding standard deviation of 5 liters. Assume that the pesticide spray per week follows a normal distribution. Write code to answer the following questions:

- a) What proportion of the farmers is spraying more than 50 liters of pesticide in a week?

```
1 - stats.norm.cdf(50, 38, 5)
```

0.008197535924596155

- b) What proportion of farmers is spraying less than 10 liters?

```
stats.norm.cdf(10, 38, 5)
```

1.0717590258310887e-08

- c) What proportion of farmers is spraying between 30 liters and 60 liters?

```
stats.norm.cdf(60, 38, 5) - stats.norm.cdf(30, 38, 5)
```

0.9451952957565343

5. A bottle filling machine fills water into 5 liters (5000 cm^3) bottles. The company wants to test the null hypothesis that the average amount of water filled by the machine into the bottle is at least 5000 cm^3 . A random sample of 60 bottles coming out of the machine was selected and the exact contents of the selected bottles are recorded. The sample mean was $4,998.1 \text{ cm}^3$. The population standard deviation is known from the experience to be 1.30 cm^3 . Assume that the population is normally distributed with the standard deviation of 1.30 cm^3 . Write code to test the hypothesis at α of 5%. Explain the results.

```
import math
sample_mean = 4998.1
pop_mean = 5000
pop_std = 1.30
sample_size = 60
z_score = (sample_mean - pop_mean)/(pop_std * math.sqrt(sample_size))
print("z-score: ", z_score)
print("p-value: ", stats.norm.cdf(z_score))

#As the p-value is more than 0.05, the null hypothesis is retained & also
#z-score value is higher than -1.64

z-score:  -0.1886838040459662
p-value:  0.4251703230372859
```

6. A fabric manufacturer would like to understand the proportion of defective fabrics they produce. His shop floor staff have been stating that the percentage of defective is not more than 18%. He would like to test whether the claim made by his shop floor staff is correct. He picked up a random sample of 100 fabrics and found 22 defectives. Use $\alpha = 0.05$ and write code to test the hypothesis that the percentage of defective components is less than 18%.

```
meu = 0.18 * 100
1 - stats.poisson.cdf(22-1, meu)

#since the test is two tailed, this probability is compared with 0.025(i.e alpha/2)
#here 0.201 > 0.025 so the null hypothesis is retained

0.20087639192515905
```

7. Suppose that the makers of ABC batteries want to demonstrate that their battery lasts an average of at least 60 min longer than their competitor brand. Two independent random samples of 100 batteries of each kind are selected from both the brands, and the batteries are used continuously. The sample average life of ABC is found to be 450 min. The average life of competitor batteries is 368 min with the sample standard deviation of 82 min and 78 min, respectively. Frame a hypothesis and write the code to test ABC's claim at 95% significance.

```
import math
sample_mean = 368
pop_mean = 450
pop_std = 82
sample_size = 100
z_score = (sample_mean - pop_mean)/(pop_std * math.sqrt(sample_size))
print("z-score: ", z_score)
print("p-value: ", stats.norm.cdf(z_score))

#As the p-value is more than 0.05, the null hypothesis is retained & also
#z-score value is higher than -1.64

z-score:  -0.1
p-value:  0.460172162722971
```

8. A training institute would like to check the effectiveness of their training programs and if the scores of the people trained improve post the training. A sample of 30 students was taken and their scores were calculated before and after the training. File *trainingscores.csv* contains scores of the students before and after the training. Frame and write the code to test the hypothesis of training effectiveness for the training institute.

```
import pandas as pd
trainingscore_data = pd.read_csv("/trainingscores.csv")
trainingscore_data.drop(4, inplace = True, axis = 0)
trainingscore_data = trainingscore_data.astype({"Score Before Training": int, "Score After Training": int})
stats.ttest_rel(trainingscore_data['Score Before Training'],
                trainingscore_data['Score After Training'])

#As a below result, p-value is 0.09 which is more than 0.05 value, we conclude that
#they are part of same distribution

Ttest_relResult(statistic=-1.736432649785126, pvalue=0.09347987807999412)
```