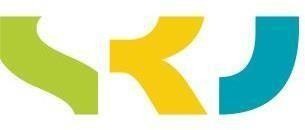
# PROBABILITY AND STATISTICS



**CASE STUDY**

**ON**

# People view about car mileage

|  |  |  |
| --- | --- | --- |
|  | **BY** |  |
| **DVV. SREEKAR REDDY**  **P. AMULYA**  **R. VARSHA**  **S. MALAVIKA**  **B. NAGA RUSHITHA**  **A. SRIVARSHAN** |  | **2203A54037**  **2203A54038**  **2203A54039**  **2203A54040**  **2203A54041**  **2203A54042** |
| **K. SHIVA BHANU** |  | **2203A54043** |

**UNDER THE GUIDENCE OF**

**DR. C BALARAMA KRISHNA**

**SUBMITTED TO**

**Department of mathematics**

SR University, Ananthasagar, Hanamakonda

Question: Can we conclude that the claim of dealer is true?

* **Take a small sample of people who have Maruti Baleno car and collect the year of purchase and the current mileage of their cars.**
* **Study the significant difference between the sample mean and**

Population mean. (Take both 0.01 and

0.05 level of significance)

* **Also, Calculate the least squares regression line which predict the approximate mileage for a given year of purchase.**

Here is the collection of scores of a sample of

data collected the mileage of cars according to their purchase

|  |  |  |
| --- | --- | --- |
| **SERIAL NUMBERS** | **MILEAGE DATA** | **YEAR OF PURCHASE** |
| **1** | **25** | **2017** |
| **2** | **26** | **2018** |
| **3** | **23** | **2018** |
| **4** | **27** | **2019** |
| **5** | **24** | **2020** |
| **6** | **25** | **2019** |
| **7** | **26** | **2020** |
| **8** | **28** | **2021** |
| **9** | **24** | **2017** |
| **10** | **25** | **2018** |
| **11** | **27** | **2020** |
| **12** | **26** | **2021** |

# SIGNIFICANT DIFFERENCE BETWEEN

**SAMPLE MEAN AND POPULATION MEAN**

**Let n be the sample size of the data and μ be the sample mean of the mileage of Maruti Baleno Cars, x̄ be the population mean of the Maruti**

**Baleno Car’s mileage.**

**Given alpha values are 0.01 and 0.05 The Standard Deviation is 1.22**

**Applying Z test = (sample mean – population mean) / (Standard Deviation / square root of sample size)**

**= (25.5-24)/(1.22/sqrt(12)) Zcal =4.259**

**Ztab=1.96 (at 0.05 LOS) Ztab=1.28 (at 0.01 LOS)**

**As Zcal > Ztab**

**Reject the null hypothesis at 0.05 level of significance**

**Reject the null hypothesis at 0.01 level of significance**

# CODE:

import numpy as np

import scipy.stats as stats

import statsmodels.api as sm

mileage\_data = [25, 26, 23, 27, 24, 25, 26, 28,

24, 25, 27, 26]

year\_of\_purchase = [2015, 2016, 2016, 2017,

2018, 2017, 2018, 2019, 2015, 2016, 2018,

2019]

sample\_mean = np.mean(mileage\_data)

sample\_std = np.std(mileage\_data, ddof=1) population\_mean = 24

sample\_size = len(mileage\_data) t\_statistic = (sample\_mean -

population\_mean) / (sample\_std /

np.sqrt(sample\_size))

degrees\_of\_freedom = sample\_size - 1 alpha\_01 = 0.01

alpha\_05 = 0.05

critical\_t\_01 = stats.t.ppf(1 - alpha\_01,

df=degrees\_of\_freedom)

critical\_t\_05 = stats.t.ppf(1 - alpha\_05, df=degrees\_of\_freedom)

print("Test Statistic:", t\_statistic)

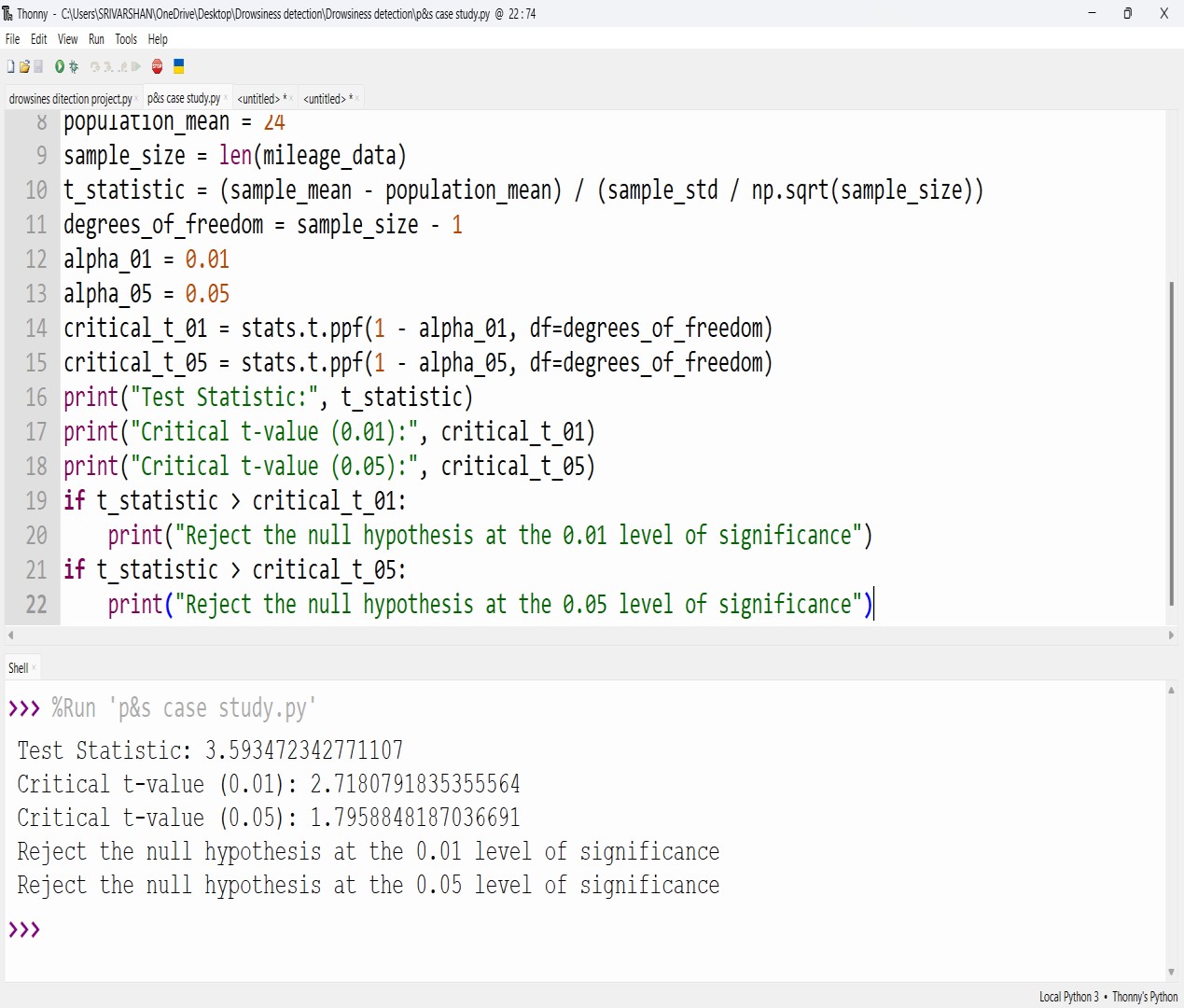
print("Critical t-value (0.01):", critical\_t\_01) print("Critical t-value (0.05):", critical\_t\_05) if t\_statistic > critical\_t\_01:

print("Reject the null hypothesis at the 0.01 level of significance")

if t\_statistic > critical\_t\_05:

print("Reject the null hypothesis at the 0.05 level of significance")

# OUTPUT:



**LEAST SQUARES REGRESSION LINE**

|  |  |  |  |
| --- | --- | --- | --- |
| **x** | **y** | **X^2** | **xy** |
| **2017** | **25** | **4068289** | **50425** |
| **2018** | **26** | **4072324** | **52468** |
| **2018** | **23** | **4072324** | **46414** |
| **2019** | **27** | **4076361** | **54513** |
| **2020** | **24** | **4080400** | **48480** |
| **2019** | **25** | **4076361** | **50475** |
| **2020** | **26** | **4080400** | **52520** |
| **2021** | **28** | **4084441** | **56588** |
| **2017** | **24** | **4068289** | **48408** |
| **2018** | **25** | **4072324** | **50450** |
| **2020** | **27** | **4080400** | **54540** |
| **2021** | **26** | **4084441** | **52546** |

**From the above table:**

**Σx = 24228**

**Σy = 306**

**Σxy = 617827**

**Σx^2 = 48916254**

# NULL HYPOTHESIS:

**H0: x bar= μ2**

**i.e; there is no significant difference between the means of practical and theoretical examination scores**

**We know that equation of linear regression is given by**

**y = a + bx**

# ALTERNATIVE HYPOTHESIS:

**H0: x bar= μ2**

**i.e; there is significant difference**

**between the means of practical and theoretical examination scores**

# STATISTICS:

**According to least squares**

**approximation, normal equations of linear regression are**

**Σy= na + bΣx**

**306 = a(12) + b(24228)**

**306 = 12a + 24228b 1**

**Σxy = aΣx + bΣx^2**

**617827 = a(24228) + b(48916354)**

**24228a + 48916354b = 617827 2**

**Solving equation 1 and equation 2 We get the final a and b values**

**a= -1167.545 b=0.59**

# CODE:

import numpy as np

import scipy.stats as stats

import statsmodels.api as sm

import matplotlib.pyplot as plt mileage\_data = [25, 26, 23, 27, 24, 25, 26, 28,

24, 25, 27, 26]

year\_of\_purchase = [2017, 2018, 2018, 2019,

2020, 2019, 2020, 2021, 2017, 2018, 2020,

2021]

sample\_mean = np.mean(mileage\_data)

sample\_std = np.std(mileage\_data, ddof=1) population\_mean = 24

sample\_size = len(mileage\_data)

t\_statistic = (sample\_mean -

population\_mean) / (sample\_std / np.sqrt(sample\_size))

degrees\_of\_freedom = sample\_size - 1 alpha\_01 = 0.01

alpha\_05 = 0.05

critical\_t\_01 = stats.t.ppf(1 - alpha\_01, df=degrees\_of\_freedom)

critical\_t\_05 = stats.t.ppf(1 - alpha\_05, df=degrees\_of\_freedom)

print("Test Statistic:", t\_statistic)

print("Critical t-value (0.01):", critical\_t\_01) print("Critical t-value (0.05):", critical\_t\_05) if t\_statistic > critical\_t\_01:

print("Reject the null hypothesis at the

0.01 level of significance") if t\_statistic > critical\_t\_05:

print("Reject the null hypothesis at the

0.05 level of significance") x = year\_of\_purchase

y = mileage\_data

x = sm.add\_constant(x)

model = sm.OLS(y, x).fit()

intercept, slope = model.params

print("Least Squares Regression Line:")

print(f"Predicted Mileage = {intercept:.2f} +

{slope:.2f} \* Year of Purchase")

plt.scatter(year\_of\_purchase, mileage\_data, label='Data')

plt.plot(year\_of\_purchase, intercept + slope \* np.array(year\_of\_purchase), color='red',

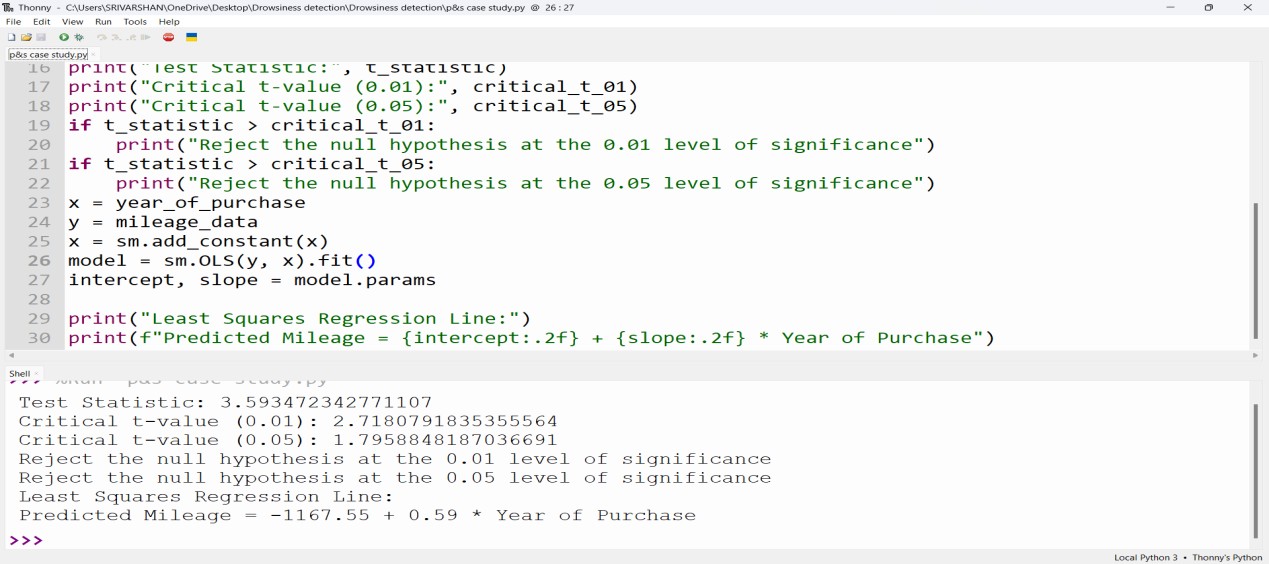
label='Regression Line')

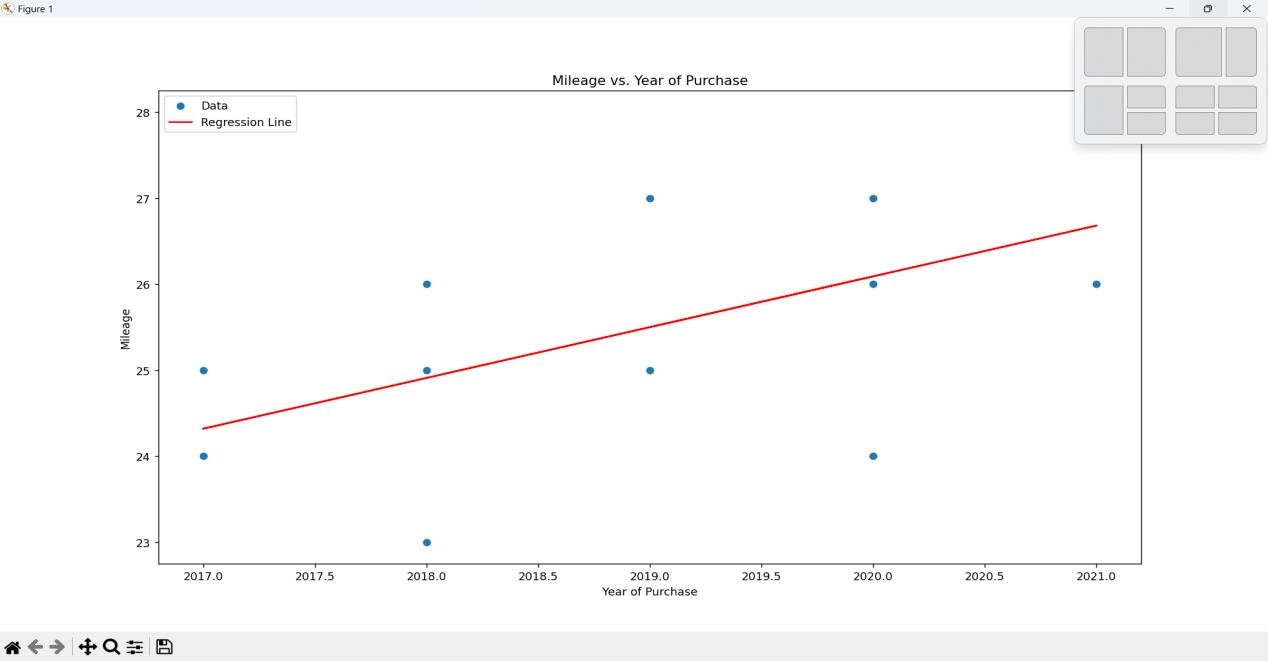
plt.xlabel('Year of Purchase') plt.ylabel('Mileage')

plt.legend()

plt.title('Mileage vs. Year of Purchase') plt.show()

# OUTPUT:





**REFERENCE:**

These are the references which are collected from some people who are using Maruti

Baleno cars in Telangana.



