# CHI-SQAURE TEST

# ASSOCIATION BETWEEN DEVICE TYPE AND CUSTOMER SATISFACTION

Mizzare Corporation has collected data on customer satisfaction levels for two types of smart home devices: Smart Thermostats and Smart Lights. They want to determine if there's a significant association between the type of device purchased and the customer's satisfaction level.

**Data Provided:**

The data is summarized in a contingency table showing the counts of customers in each satisfaction level for both types of devices:

|  |  |  |  |
| --- | --- | --- | --- |
| **Satisfaction** | **Smart Thermostat** | **Smart Light** | **Total** |
| Very Satisfied | 50 | 70 | 120 |
| Satisfied | 80 | 100 | 180 |
| Neutral | 60 | 90 | 150 |
| Unsatisfied | 30 | 50 | 80 |
| Very Unsatisfied | 20 | 50 | 70 |
| **Total** | 240 | 360 | 600 |

**Objective:**

To use the Chi-Square test for independence to determine if there's a significant association between the type of smart home device purchased (Smart Thermostats vs. Smart Lights) and the customer satisfaction level.

**STEP 1:**

**1.STATE THE HYPOTHESIS:**

##HYPOTHESIS

#H0: There is no significant association between the type of smart home device purchased and the customer satisfaction level.

(type of device and satisfaction level are independent).

#H1: There is significant association between the type of smart home device purchased and the customer satisfaction level.

(type of the device and satisfaction level are dependent).

#importing pandas and reading the data from the file.

import pandas as pd

df=pd.read\_csv("chi.csv")

df

|  | **Satisfaction** | **Smart Thermostat** | **Smart Light** |
| --- | --- | --- | --- |
| **0** | Very Satisfied | 50 | 70 |
| **1** | Satisfied | 80 | 100 |
| **2** | Neutral | 60 | 90 |
| **3** | Unsatisfied | 30 | 50 |
| **4** | Very Satisfied | 20 | 50 |

Since there are 2 variables needs to be compared first we have separated the data by performing the chi square test using Satisfaction and Smart Thermostat columns.

In order words expected frequency should be calculated for two types of smart home devices.

**#researchpy is a wrapper which integrates functionalities from several packages like pandas, scipy.stats, numpy, and statsmodels**

import researchpy as rp

table1, results1 = rp.crosstab(df["Satisfaction"], df["Smart Thermostat"],test='chi-square')

table1

|  | **Smart Thermostat** | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Smart Thermostat** | **20** | **30** | **50** | **60** | **80** | **All** |
| **Satisfaction** |  |  |  |  |  |  |
| **Neutral** | 0 | 0 | 0 | 1 | 0 | 1 |
| **Satisfied** | 0 | 0 | 0 | 0 | 1 | 1 |
| **Unsatisfied** | 0 | 1 | 0 | 0 | 0 | 1 |
| **Very Satisfied** | 1 | 0 | 1 | 0 | 0 | 2 |
| **All** | 1 | 1 | 1 | 1 | 1 | 5 |

**Step 2, a.**

|  | **Chi-square test** | **results** |
| --- | --- | --- |
| **0** | **Pearson Chi-square ( 12.0) =** | **15.0000** |
| **1** | **p-value =** | **0.2414** |
| **2** | **Cramer's V =** | **1.0000** |

**results1**

Here the statistic value that is chi-square statistic value for variable Smart Thermostat with Satisfaction Level noted down is 15.0000.

**#calculating chi\_square table value for degrees of freedom=12**

**#chi-square table value (for " Smart Thermostat " with Satisfaction Level)**

**CODE:**

import scipy. stats as stats

chi\_table1 = stats.chi2.ppf (q = 0.95, df = 12)

chi\_table1.round(4)

**OUTPUT:**

21.0261

**Observing the conditions for Smart Light criteria.**

**CODE:**

table2, results2 = rp.crosstab(df["Satisfaction"], df["Smart Light"], test='chi-square')

table2

**OUTPUT:**

|  | **Smart Light** | | | | |
| --- | --- | --- | --- | --- | --- |
| **Smart Light** | **50** | **70** | **90** | **100** | **All** |
| **Satisfaction** |  |  |  |  |  |
| **Neutral** | 0 | 0 | 1 | 0 | 1 |
| **Satisfied** | 0 | 0 | 0 | 1 | 1 |
| **Unsatisfied** | 1 | 0 | 0 | 0 | 1 |
| **Very Satisfied** | 1 | 1 | 0 | 0 | 2 |
| **All** | 2 | 1 | 1 | 1 | 5 |

**Step 2,b**

**results2**

|  | **Chi-square test** | **results** |
| --- | --- | --- |
| **0** | Pearson Chi-square ( 9.0) = | 11.250 |
| **1** | p-value = | 0.259 |
| **2** | Cramer's V = | 0.866 |

#calculating chi\_square table value for degrees of freedom=9

#chi-square table value (for " Smart Light " with Satisfaction Level)

import scipy.stats as stats

chi\_table2 = stats.chi2.ppf(q = 0.95, df = 9)

chi\_table2

**OUTPUT**:

16.918977604620448

**Step 3**

**###obtaining critical value from chi-square distribution**

**CODE:**

import scipy.stats as stats

alpha = 0.05

df = 9

chi\_critical = stats.chi2.ppf(q=1-alpha, df=df)

print("Critical value from chi-square distribution:", chi\_critical)

Critical value from chi-square distribution: 16.918977604620448

If the calculated chi-square statistic is greater than the critical value,

reject the null hypothesis.

If the calculated chi-square statistic is less than or equal to the critical value, not reject the null hypothesis.

In this case, calculated chi-square statistic of 25 and a critical value of approximately 16.92,

We will reject the null hypothesis because the calculated statistic is more than the critical value.

**Step 4**

**#making the decision**

test1=15.0000

test2=11.250

if test1>chi\_table1 and test2>chi\_table2:

print ("H0 is rejected and H1 will be accepted")

else:

print ("H0 is accepted and H1 will be rejected")

**OUTPUT:**

H0 is accepted and H1 will be rejected

**CONCLUSION:**

Hence, we can conclude that we are having strong significance evidence that there is no significant association between smart device and satisfaction levels.

Therefore, the conclusion drawn is there is no significance association between the type of smart device purchased (Smart Thermostat Vs Smart Light) and the Customer Satisfaction Level.