Assignment: Chi-Square Test for Association between Device Type and Customer Satisfaction

1. Background

Mizzare Corporation collected customer satisfaction counts for two smart home device types (Smart Thermostats and Smart Lights) across five satisfaction categories. The firm wants to determine whether device type is associated with the level of customer satisfaction.

2. Data (contingency table)

Satisfaction Level	Smart Thermostat	Smart Light	Row total
Very Satisfied	50	70	120
Satisfied	80	100	180
Neutral	60	90	150
Unsatisfied	30	50	80
Very Unsatisfied	20	50	70
Column totals	240	360	600

3. Hypotheses

- **Null hypothesis (H_0):** Device type and customer satisfaction are independent (no association).
- **Alternative hypothesis (H_1):** Device type and customer satisfaction are not independent (there is an association).

4. Test statistic and distribution

We use the Pearson Chi-Square test for independence. The test statistic is:

$$\chi^2 = \sum_{i}\sum_{j} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

where O_{ij} are observed counts and E_{ij} are expected counts under independence.
Degrees of freedom: $(r-1)(c-1) = (5-1)(2-1) = 4$.

Significance level: $\alpha = 0.05$.

5. Expected counts

Expected count for cell (i,j) is:

$$E_{ij} = \frac{(\text{row total}_i) \times (\text{column total}_j)}{\text{grand total}}$$

Using the table above, the expected counts matrix is:

Satisfaction Level	Smart Thermostat (E)	Smart Light (E)
Very Satisfied	48.0	72.0
Satisfied	72.0	108.0
Neutral	60.0	90.0
Unsatisfied	32.0	48.0
Very Unsatisfied	28.0	42.0

6. Calculations (numeric results)

Using the contingency table, the computed values are:

- Observed (\chi^2) statistic: **5.638** (rounded to 3 decimals)
- Degrees of freedom: **4**
- p-value: **0.2278** (rounded)
- Critical value at (\alpha = 0.05,\ df=4): (\chi^2_{0.95,4} \approx) **9.488**

Because the computed (\chi^2 = 5.638 < 9.488) and (p = 0.2278 > 0.05), we **fail to reject** the null hypothesis.

7. Decision & Interpretation

At the 5% significance level, there is **no statistically significant evidence** of an association between device type (Smart Thermostat vs. Smart Light) and customer satisfaction level. In practical terms, the observed differences in counts across satisfaction categories can reasonably be attributed to random variation rather than a systematic association with device type.

8. Assumptions & Notes

- The Chi-Square test assumes that observations are independent and that expected cell counts are sufficiently large (a common rule of thumb: expected counts ≥ 5). All expected counts here are ≥ 28, so the approximation is appropriate.
- The data are counts drawn from independent customers (assumed by the study design).
- The test does not measure strength or direction of any association — it only tests for presence/absence of association. If you need effect size, consider Cramér's V.

9. Recommendation

Since no association was detected, management cannot conclude that device type leads to differences in satisfaction from this sample alone. If business decisions depend on small differences, consider:

- Collecting a larger, possibly stratified sample for more power.
- Investigating additional variables (e.g., installation quality, price, usage patterns) that might influence satisfaction.

- Computing effect-size measures (Cramér's V) to quantify association magnitude, even if not statistically significant.

10. Python code (reproducible)

You can run the following Python code to reproduce the calculations. This uses scipy.stats for convenience; if SciPy is not available you can compute the statistic manually with NumPy.

```
import numpy as np

from scipy.stats import chi2_contingency, chi2


# Observed contingency table (rows = satisfaction levels, columns = device types)

obs = np.array([

    [50, 70],   # Very Satisfied

    [80,100],   # Satisfied

    [60,90],    # Neutral

    [30,50],    # Unsatisfied

    [20,50]     # Very Unsatisfied

])


chi2_stat, p_value, dof, expected = chi2_contingency(obs, correction=False)

crit_value = chi2.ppf(0.95, df=dof)


print("Chi-square statistic:", round(chi2_stat, 3))

print("Degrees of freedom:", dof)

print("p-value:", round(p_value, 4))

print("Critical value (alpha=0.05):", round(crit_value, 3))

print("\nExpected counts:\n", expected)
```

Reproduction of results (values obtained)

- Chi-square statistic: 5.638
- Degrees of freedom: 4
- p-value: 0.2278
- Critical value (α = 0.05): 9.488

Short academic conclusion

A Pearson Chi-Square test was performed to assess whether device type (Smart Thermostat vs. Smart Light) and customer satisfaction are associated. With (\chi^2(4) = 5.638), (p = 0.2278), we fail to reject the null hypothesis at the 5% significance level;

therefore, there is no statistical evidence of an association between device type and satisfaction in this sample.

Code used:

```
import numpy as np
from scipy.stats import chi2_contingency, chi2

# Observed contingency table (rows = satisfaction levels, columns = device types)
obs = np.array([
    [50, 70],   # Very Satisfied
    [80,100],   # Satisfied
    [60,90],    # Neutral
    [30,50],    # Unsatisfied
    [20,50]     # Very Unsatisfied
])

chi2_stat, p_value, dof, expected = chi2_contingency(obs, correction=False)
crit_value = chi2.ppf(0.95, df=dof)

print("Chi-square statistic:", round(chi2_stat, 3))
print("Degrees of freedom:", dof)
print("p-value:", round(p_value, 4))
print("Critical value (alpha=0.05):", round(crit_value, 3))
print("\nExpected counts:\n", expected)
```