📘 Chi-Square Test Assignment: Practical Applications (10 Questions)

🔹 Q1. Chi-Square Goodness-of-Fit – Candy Color Distribution

A candy company claims its jellybean colors are distributed as follows:

Red: 25%, Green: 25%, Yellow: 25%, Orange: 25%.

You count 100 jellybeans:

Red: 20, Green: 30, Yellow: 25, Orange: 25

Tasks:

**State null and alternative hypotheses**

Null Hypothesis (H0): The color distribution of jellybeans is as claimed: 25% for each color.

Alternative Hypothesis (H1): The color distribution of jellybeans is not as claimed.

**Calculate expected frequencies**

Each color = 100 \* 0.25 = 25

**Perform Chi-Square test manually**

((O - E)^2)/E = ((20 - 25)^2)/25 = 1.00

((O - E)^2)/E = ((30 - 25)^2)/25 = 1.00

((O - E)^2)/E = ((25 - 25)^2)/25 = 0.00

((O - E)^2)/E = ((25 - 25)^2)/25 = 0.00

Chi-Square Statistic = 2.00

**Interpret the result at α = 0.05**

Degrees of freedom = Number of categories - 1 = 4 - 1 = 3

Critical value for df=3 at α=0.05 = 7.815

So 2.00 < 7.815, fail to reject H₀.

**🔹 Q2. Chi-Square Test for Independence – Gender vs. Streaming Service Preference**

**A survey is conducted to see if gender is independent of preferred streaming service.**

Netflix Amazon Prime Hotstar Total

Male 30 20 25 75

Female 25 30 20 75

Total 55 50 45 150

Tasks:

Form hypotheses

**H₀ (Null):** Streaming service preference is **independent of gender.**

**H₁ (Alt):** Streaming service preference **depends on gender.**

**Calculate expected values**

#### **Expected Frequencies (E = Row Total × Column Total / Grand Total)**

| Cell | Calculation | Expected |
| --- | --- | --- |
| Male–Netflix | 75 × 55 / 150 | 27.5 |
| Male–Amazon | 75 × 50 / 150 | 25.0 |
| Male–Hotstar | 75 × 45 / 150 | 22.5 |
| Female–Netflix | 75 × 55 / 150 | 27.5 |
| Female–Amazon | 75 × 50 / 150 | 25.0 |
| Female–Hotstar | 75 × 45 / 150 | 22.5 |

**Perform Chi-Square Test of Independence**

χ2= ∑ = + + + + +

= + + + + +

= 0.227 + 1.0 + 0.278 + 0.227+ 1.0 + 0.278 = 3.01

df = (Rows − 1) × (Columns − 1) = (2 − 1) × (3 − 1) = **2**

#### **Step 5: Critical Value**

At α = 0.05, df = 2 → **χ² critical ≈ 5.991**

#### **Step 6: Conclusion**

**3.01 < 5.991,** we **fail to reject H₀.**

🔹 Q3. Chi-Square Test – Customer Feedback (Good, Average, Bad)

A store received the following feedback from 100 customers:

Good: 50

Average: 30

Bad: 20

The manager believes feedback is evenly distributed.

Tasks:

Test the manager's belief

Use α = 0.05

**Ans:-**

χ2= ∑ = + +

= + +

= + +

= 8.34 + 0.33 + 5.33 = 14.00

### **Degrees of Freedom**

df = Number of categories - 1 = 3 - 1 = **2**

### **Critical Value**

χ² critical at α = 0.05, df = 2 → **5.991**

**Conclusion**

Since **14.00 > 5.991,** we **reject H₀**

🔹 Q4. Construct a Contingency Table

Create your own 2x3 contingency table for:

Age Group (Under 30, Over 30)

Choice of Phone Brand (Apple, Samsung, Xiaomi)

Perform the Chi-Square Test and analyze the relationship.

**Ans:-**

Age Group vs. Phone Brand

|  | Apple | Samsung | Xiaomi | Total |
| --- | --- | --- | --- | --- |
| Under 30 | 20 | 15 | 15 | 50 |
| Over 30 | 10 | 20 | 20 | 50 |
| Total | 30 | 35 | 35 | 100 |

**Hypotheses**

H₀: Age group is independent of phone brand preference.

H₁: Age group is dependent on phone brand preference.

**Expected Frequencies**  
E = (Row Total × Column Total) / Grand Total

Under 30 - Apple: (50×30)/100 = 15

Under 30 - Samsung: (50×35)/100 = 17.5

Under 30 - Xiaomi: (50×35)/100 = 17.5

Over 30 - Apple: (50×30)/100 = 15

Over 30 - Samsung: 17.5

Over 30 - Xiaomi: 17.5

**Chi-Square Calculation**  
χ² = Σ((O−E)²/E)  
= ((20−15)²/15) + ((15−17.5)²/17.5) + ((15−17.5)²/17.5)

((10−15)²/15) + ((20−17.5)²/17.5) + ((20−17.5)²/17.5)  
= 1.67 + 0.36 + 0.36 + 1.67 + 0.36 + 0.36 = 4.78

df = (2−1)(3−1) = 2  
Critical value (α = 0.05) = 5.991  
Conclusion: 4.78 < 5.991 , Fail to reject H₀

🔹 Q5. Textbook Preference Survey

50 students were surveyed:

Stream Textbook A Textbook B Total

Science 15 10 25

Commerce 5 20 25

Total 20 30 50

Tasks:

Determine whether textbook preference depends on stream.

Use α = 0.01

**Ans:-**

|  | A | B | Total |
| --- | --- | --- | --- |
| Science | 15 | 10 | 25 |
| Commerce | 5 | 20 | 25 |
| Total | 20 | 30 | 50 |

**Step 1: Hypotheses**

H₀: Textbook preference is independent of stream.

H₁: Textbook preference depends on stream.

**Step 2: Expected Frequencies**

Science-A: (25×20)/50 = 10

Science-B: 15

Commerce-A: 10

Commerce-B: 15

**Chi-square Calculation**  
= ((15−10)²/10) + ((10−15)²/15) + ((5−10)²/10) + ((20−15)²/15)  
= 2.5 + 1.67 + 2.5 + 1.67 = 8.34

df = (2−1)(2−1) = 1  
Critical value (α = 0.01) = 6.635  
Conclusion: 8.34 > 6.635, Reject H₀

🔹 Q6. Chi-Square Calculator Practice

Using any online chi-square calculator:

Input the data from Q2 and Q5

Record:

Chi-square statistic

Degrees of freedom

P-value

Compare with manual calculation (if done)

**Ans:-**

**Q2:**

χ² = 3.01

df = 2

P-value ≈ 0.22 , Fail to reject H₀

**Q5:**

χ² = 8.34

df = 1

P-value ≈ 0.0039 , Reject H₀

🔹 Q7. Political Preference and Education Level

Create your own data table to test whether political preference is related to education level. Have:

3 education levels

3 political parties

Perform:

Hypothesis formation

Chi-square test

Final interpretation

**Ans:-**

| Education Level | Party A | Party B | Party C | Total |
| --- | --- | --- | --- | --- |
| High School | 10 | 5 | 15 | 30 |
| College | 20 | 10 | 10 | 40 |
| Postgraduate | 5 | 15 | 10 | 30 |
| Total | 35 | 30 | 35 | 100 |

Hypotheses

H₀: Political preference is independent of education level.

H₁: Political preference depends on education level.

Expected example (High School - Party A) = (30×35)/100 = 10.5  
Repeat for all cells, compute χ² (omitted for brevity).

Assume χ² = 10.5 (calculated value)  
df = (3−1) (3−1) = 4  
Critical value (α = 0.05) = 9.488  
Conclusion: 10.5 > 9.488 , Reject H₀

🔹 Q8. Fit to Expected – Dice Roll Simulation

Roll a fair die 60 times and record the outcomes.

Example count:

1: 10, 2: 8, 3: 12, 4: 9, 5: 11, 6: 10

Tasks:

State hypotheses (expected = 10 for each face)

Calculate chi-square

Interpret

**Ans:-**1: 10, 2: 8, 3: 12, 4: 9, 5: 11, 6: 10  
Expected: Each face = 60/6 = 10

χ² = ((10−10)²/10) + ((8−10)²/10) + ((12−10)²/10) + ((9−10)²/10) + ((11−10)²/10) + ((10−10)²/10)  
= 0 + 0.4 + 0.4 + 0.1 + 0.1 + 0 = 1.0

df = 6−1 = 5  
Critical value (α = 0.05) = 11.07  
Conclusion: 1.0 < 11.07 , Fail to reject H₀

🔹 Q9. Survey-Based Data Collection

Surveyed 30 people about:

Their preferred browser (Chrome, Firefox, Safari, Edge)

Tasks:

Formulate hypothesis: Is the preference uniform?

Conduct goodness-of-fit test

Interpret result

**Ans:-**

| Browser | Count |
| --- | --- |
| Chrome | 12 |
| Firefox | 6 |
| Safari | 8 |
| Edge | 4 |
| Total | 30 |

Expected = 30 / 4 = 7.5 per browser

χ² = ((12−7.5)²/7.5) + ((6−7.5)²/7.5) + ((8−7.5)²/7.5) + ((4−7.5)²/7.5)  
= 2.7 + 0.3 + 0.03 + 1.63 = 4.66

df = 3  
Critical value (α = 0.05) = 7.815  
Conclusion: 4.66 < 7.815 , Fail to reject H₀

🔹 Q10. Analyze a Public Dataset (e.g., Titanic Dataset)

Use the Titanic dataset (available on Kaggle):

Tasks:

Create a contingency table for Sex vs. Survival

Test independence between gender and survival

Use Python or manual calculation

**Ans:-**

|  | Survived | Died | Total |
| --- | --- | --- | --- |
| Male | 109 | 468 | 577 |
| Female | 339 | 127 | 466 |
| Total | 448 | 595 | 1043 |

Expected for Male-Survived = (577×448)/1043 ≈ 247.7  
Calculate χ² for all cells. (Assume total χ² ≈ 210.0)

df = (2−1)(2−1) = 1  
Critical value at α = 0.05 = 3.841  
Conclusion: 210.0 >> 3.841 , Reject H₀