# 📘 Detailed Guide: Chi-Square Tests (with Formulas and Concepts)

## 🔹 1. Chi-Square Test Overview

The **Chi-Square Test** is a **non-parametric statistical test** used for:

* Comparing **observed** data with **expected** data.
* Evaluating relationships between **categorical variables**.

### Types of Chi-Square Tests:

1. **Goodness-of-Fit Test** — Is one categorical variable’s distribution equal to a hypothesized distribution?
2. **Test of Independence** — Are two categorical variables independent?

## 🔹 2. Hypothesis Formation

* **Null Hypothesis (H₀)**: There is **no difference** between observed and expected frequencies.
* **Alternative Hypothesis (H₁)**: There **is a difference** between observed and expected frequencies.

For Test of Independence:

* **H₀**: Variables are independent.
* **H₁**: Variables are dependent (associated).

## 🔹 3. Chi-Square Goodness-of-Fit Test

### ✅ When to Use:

* Single categorical variable
* Compare **observed counts** with a **theoretical distribution**

### 🔧 Steps:

1. **Collect Observed Frequencies (Oᵢ)**
2. **Determine Expected Frequencies (Eᵢ)**

- Uniform: Eᵢ = Total Observations / k  
 - Unequal: Eᵢ = Total × Expected Proportion

1. **Apply Chi-Square Formula:**
2. **Degrees of Freedom:**
3. Compare χ² to critical value or use p-value
4. **Decision:**
   * p < α → Reject H₀
   * p ≥ α → Fail to reject H₀

## 🔹 4. Chi-Square Test of Independence

### ✅ When to Use:

* Two **categorical variables**
* Determine whether they are **associated** or **independent**

### 🔧 Steps:

**1. Construct Contingency Table (Oᵢⱼ)**

**2. Calculate Expected Frequencies:**

**Eᵢⱼ = (Row Total × Column Total) / Grand Total**

**3. Apply Chi-Square Formula:**

**χ² = Σ((Oᵢⱼ - Eᵢⱼ)² / Eᵢⱼ)**

**4. Degrees of Freedom: df = (r-1)(c-1)**

**5. Compare χ² to critical value or use p-value**

**6. Decision: p < α → Reject H₀; p ≥ α → Fail to reject H₀**

## 🔹 5. Interpreting Results

| Method | Interpretation |
| --- | --- |
| Critical Value Method | χ²\_calc > χ²\_critical → Reject H₀ |
| p-value Method | p < α → Reject H₀; p ≥ α → Fail to reject H₀ |

## 🔹 6. Key Concepts Explained

| Concept | Explanation |
| --- | --- |
| **Observed Frequency (Oᵢ)** | Actual count from data |
| **Expected Frequency (Eᵢ)** | Theoretically predicted count |
| **Degrees of Freedom (df)** | Goodness-of-fit: df = k–1; Independence: df = (r–1)(c–1) |
| **Significance Level (α)** | Probability threshold (commonly 0.05 or 0.01) |
| **p-value** | Probability of observing data as extreme under H₀ |
| **Reject H₀** | If result is significant → evidence against H₀ |
| **Fail to Reject H₀** | If result not significant → no evidence against H₀ |

## 🔹 7. Chi-Square Table Quick Reference

|  |  |  |
| --- | --- | --- |
| df | 0.05 Critical Value | 0.01 Critical Value |
| 1 | 3.84 | 6.63 |
| 2 | 5.99 | 9.21 |
| 3 | 7.81 | 11.34 |
| 4 | 9.49 | 13.28 |
| 5 | 11.07 | 15.09 |

## 🔹 8. Applications from Tasks

|  |  |  |
| --- | --- | --- |
| Task | Test Type | Purpose |
| Q1 Candy Colors | Goodness-of-Fit | Match observed with claimed distribution |
| Q2 Gender vs Streaming | Test of Independence | Association between gender and streaming preference |
| Q3 Customer Feedback | Goodness-of-Fit | Even distribution of feedback |
| Q4 Age vs Phone Brand | Test of Independence | Relationship between age group and phone brand |
| Q5 Stream vs Textbook | Test of Independence | Textbook preference by stream |
| Q6 Calculator Practice | Both | Tool use and result comparison |
| Q7 Political vs Education | Test of Independence | Preferences linked to education level |
| Q8 Dice Rolls | Goodness-of-Fit | Fairness of dice roll outcomes |
| Q9 Browser Preference | Goodness-of-Fit | Uniformity of browser preference |
| Q10 Titanic Dataset | Test of Independence | Gender and survival relationship |

## 🔹 9. Python Tools

### Goodness-of-Fit:

from scipy.stats import chisquare  
chisquare(f\_obs=[O1, O2, ...], f\_exp=[E1, E2, ...])

### Test of Independence:

from scipy.stats import chi2\_contingency  
data = [[row1], [row2], ...]  
chi2\_contingency(data)