**Z-Test and T-Test: Definitions, Types, Examples, and Step-by-Step Solutions**

Both **Z-tests** and **T-tests** are statistical methods used to determine whether there is a significant difference between sample means and population means or between two sample means. The choice between them depends on sample size and whether the population standard deviation is known.

**1. Z-Test**

**When to Use:**

* When the **sample size is large (n ≥ 30)** and the **population standard deviation (σ) is known**.
* Used for **normally distributed** data.

**Types of Z-Test:**

1. **One-Sample Z-Test** – Compares a sample mean to a known population mean.
   * Example: Testing if the average height of a sample of 50 students differs from the national average height (known σ).
2. **Two-Sample Z-Test** – Compares means of two independent samples.
   * Example: Testing if the average test scores of two different classes (each n ≥ 30) are significantly different (known σ for both populations).
3. **Z-Test for Proportions** – Tests proportions instead of means.
   * Example: Testing if the proportion of voters supporting a candidate differs from a claimed value (e.g., 50%).

**2. T-Test**

**When to Use:**

* When the **sample size is small (n < 30)** and the **population standard deviation (σ) is unknown**.
* Uses the **sample standard deviation (s)** and follows the **t-distribution** (which has heavier tails for smaller samples).

**Types of T-Test:**

1. **One-Sample T-Test** – Compares a sample mean to a known or hypothesized population mean.
   * Example: Testing if the average weight of 20 apples differs from the claimed 100g (σ unknown).
2. **Independent Two-Sample T-Test** – Compares means of two unrelated groups.
   * Example: Testing if the mean test scores of two different groups of students (n₁ = 15, n₂ = 20) are different.
3. **Paired T-Test** – Compares means from the same group at different times (dependent samples).
   * Example: Testing if a training program improves student scores by comparing before-and-after results.

**Step-by-Step Solution (6 Steps)**

**For Z-Test:**

1. **State Hypotheses:**
   * Null Hypothesis (H₀): μ = μ₀ (no difference)
   * Alternative Hypothesis (H₁): μ ≠ μ₀ (two-tailed) or μ > μ₀ / μ < μ₀ (one-tailed)
2. **Choose Significance Level (α):** Typically 0.05 or 0.01.
3. **Calculate Test Statistic (Z-score):**

Z=Xˉ−μσ/n*Z*=*σ*/*n*​*X*ˉ−*μ*​

* + Xˉ*X*ˉ = sample mean, μ*μ* = population mean, σ*σ* = population SD, n*n* = sample size.

1. **Find Critical Value:**
   * Use **Z-table** for the chosen α (e.g., Z = ±1.96 for α = 0.05 two-tailed).
2. **Compare Z-score with Critical Value:**
   * If |Z| > Z-critical → **Reject H₀**.
3. **Conclusion:**
   * State whether there is enough evidence to reject H₀.

**For T-Test:**

1. **State Hypotheses:** (Same as Z-test)
2. **Choose Significance Level (α):** Typically 0.05.
3. **Calculate Test Statistic (T-score):**

t=Xˉ−μs/n*t*=*s*/*n*​*X*ˉ−*μ*​

* + s*s* = sample standard deviation.

1. **Find Critical Value:**
   * Use **t-table** with **degrees of freedom (df = n - 1)** and α.
2. **Compare T-score with Critical Value:**
   * If |t| > t-critical → **Reject H₀**.
3. **Conclusion:**
   * State whether the result is statistically significant.

**How Critical Values are Found**

**For Z-Test:**

* Critical values come from the **standard normal distribution (Z-table)**.
* Example: For α = 0.05 (two-tailed), critical values are **±1.96** because:
  + P(Z > 1.96) = 0.025 (and P(Z < -1.96) = 0.025).

**For T-Test:**

* Critical values come from the **t-distribution table**, which depends on:
  + **Degrees of freedom (df = n - 1)**
  + **Significance level (α)**
* Example: For df = 24 and α = 0.05 (two-tailed), the critical t-value is **±2.064**.

**Example Problems**

**Z-Test Example:**

**Problem:** A factory claims batteries last 100 hours (σ = 10). A sample of 50 batteries lasts 98 hours. Test at α = 0.05.

**Solution:**

1. H₀: μ = 100, H₁: μ ≠ 100
2. α = 0.05 → Z-critical = ±1.96
3. Z=98−10010/50=−1.41*Z*=10/50​98−100​=−1.41
4. Since |-1.41| < 1.96 → **Fail to reject H₀**
5. Conclusion: No evidence that battery life differs from 100 hours.

**T-Test Example:**

**Problem:** A sample of 10 students has an average score of 75 (s = 8). Test if this differs from the population mean of 70 (α = 0.05).

**Solution:**

1. H₀: μ = 70, H₁: μ ≠ 70
2. α = 0.05, df = 9 → t-critical = ±2.262
3. t=75−708/10=1.98*t*=8/10​75−70​=1.98
4. Since |1.98| < 2.262 → **Fail to reject H₀**
5. Conclusion: No significant difference from 70.

**Summary Table: Z-Test vs. T-Test**

| **Feature** | **Z-Test** | **T-Test** |
| --- | --- | --- |
| **Sample Size** | Large (n ≥ 30) | Small (n < 30) |
| **σ Known?** | Yes | No (use sample SD, s) |
| **Distribution** | Normal (Z-distribution) | t-distribution (df = n-1) |
| **Critical Value** | Z-table | t-table (depends on df) |

**Final Notes:**

* Use **Z-test** when σ is known or n ≥ 30.
* Use **T-test** when σ is unknown and n < 30.
* **Paired T-test** is for dependent samples (before-after studies).
* **Critical values** depend on α and df (for t-tests).

This structured approach ensures accurate hypothesis testing in real-world scenarios! 🚀

also explain chi square test

**Chi-Square Test: Definition, Types, Examples, and Step-by-Step Solution**

The **Chi-Square (χ²) test** is a non-parametric statistical test used to analyze categorical data. It determines whether observed frequencies differ significantly from expected frequencies under a given hypothesis.

**When to Use the Chi-Square Test?**

* Tests **goodness-of-fit** (how well sample data matches a distribution).
* Tests **independence** between two categorical variables.
* Works with **nominal (categorical) data**, not numerical data.

**Types of Chi-Square Tests**

**1. Chi-Square Goodness-of-Fit Test**

* Checks if sample data matches an expected distribution.
* **Example:** Testing if a die is fair (all six outcomes have equal probability).

**2. Chi-Square Test of Independence**

* Checks if two categorical variables are independent.
* **Example:** Testing if gender (Male/Female) is related to voting preference (Party A/Party B).

**3. Chi-Square Test of Homogeneity**

* Checks if different populations have the same distribution of a single categorical variable.
* **Example:** Comparing the preference for three brands (A, B, C) across two cities.

**Step-by-Step Solution (6 Steps)**

**1. State Hypotheses**

* **Null Hypothesis (H₀):** No difference between observed and expected frequencies (for goodness-of-fit) OR no association between variables (for independence).
* **Alternative Hypothesis (H₁):** Significant difference or association exists.

**2. Choose Significance Level (α)**

* Common choices: **0.05 (5%) or 0.01 (1%)**.

**3. Calculate Expected Frequencies**

* **Goodness-of-Fit:**

Ei=N×pi*Ei*​=*N*×*pi*​

* + N*N* = Total observations, pi*pi*​ = Expected probability for category i*i*.
* **Test of Independence:**

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

**4. Compute Chi-Square Statistic (χ²)**

χ2=∑(Oi−Ei)2Ei*χ*2=∑*Ei*​(*Oi*​−*Ei*​)2​

* Oi*Oi*​ = Observed frequency, Ei*Ei*​ = Expected frequency.

**5. Find Critical Value & Compare**

* **Degrees of Freedom (df):**
  + **Goodness-of-Fit:** df=k−1*df*=*k*−1 (where k*k* = number of categories).
  + **Test of Independence:** df=(r−1)(c−1)*df*=(*r*−1)(*c*−1) (where r*r* = rows, c*c* = columns).
* **Use Chi-Square Table** to find critical value for given df*df* and α*α*.
* **Decision Rule:**
  + If χ2>χcritical2*χ*2>*χcritical*2​ → **Reject H₀**.
  + Else, **fail to reject H₀**.

**6. Conclusion**

* State whether there is enough evidence to reject the null hypothesis.

**How Critical Values are Found**

* Chi-square critical values depend on:
  + **Degrees of freedom (df)**
  + **Significance level (α)**
* **Example:** For df=2*df*=2 and α=0.05*α*=0.05, the critical value is **5.991** (from χ² table).

**Example Problems**

**Example 1: Chi-Square Goodness-of-Fit Test**

**Problem:** A die is rolled 60 times. The observed frequencies are:

| **Outcome** | **1** | **2** | **3** | **4** | **5** | **6** |
| --- | --- | --- | --- | --- | --- | --- |
| Observed (O) | 8 | 12 | 9 | 11 | 10 | 10 |
| Test if the die is fair at α=0.05*α*=0.05. |  |  |  |  |  |  |

**Solution:**

1. **H₀:** Die is fair (all outcomes equally likely).  
   **H₁:** Die is not fair.
2. **Expected (E):** Each Ei=60/6=10*Ei*​=60/6=10.
3. **Calculate χ²:**

χ2=(8−10)210+(12−10)210+(9−10)210+(11−10)210+(10−10)210+(10−10)210=1.0*χ*2=10(8−10)2​+10(12−10)2​+10(9−10)2​+10(11−10)2​+10(10−10)2​+10(10−10)2​=1.0

1. **Critical Value:** df=6−1=5*df*=6−1=5, χcritical2=11.070*χcritical*2​=11.070.
2. **Decision:** Since 1.0<11.0701.0<11.070, **fail to reject H₀**.
3. **Conclusion:** No evidence that the die is unfair.

**Example 2: Chi-Square Test of Independence**

**Problem:** A survey asks 200 people about gender and voting preference:

|  | **Party A** | **Party B** | **Total** |
| --- | --- | --- | --- |
| **Male** | 50 | 30 | 80 |
| **Female** | 60 | 60 | 120 |
| **Total** | 110 | 90 | 200 |
| Test if gender and voting preference are independent (α=0.05*α*=0.05). |  |  |  |

**Solution:**

1. **H₀:** Gender and voting preference are independent.  
   **H₁:** They are dependent.
2. **Expected Frequencies:**
   * Male, Party A: E=80×110200=44*E*=20080×110​=44
   * Male, Party B: E=80×90200=36*E*=20080×90​=36
   * Female, Party A: E=120×110200=66*E*=200120×110​=66
   * Female, Party B: E=120×90200=54*E*=200120×90​=54
3. **Calculate χ²:**

χ2=(50−44)244+(30−36)236+(60−66)266+(60−54)254=3.03*χ*2=44(50−44)2​+36(30−36)2​+66(60−66)2​+54(60−54)2​=3.03

1. **Critical Value:** df=(2−1)(2−1)=1*df*=(2−1)(2−1)=1, χcritical2=3.841*χcritical*2​=3.841.
2. **Decision:** Since 3.03<3.8413.03<3.841, **fail to reject H₀**.
3. **Conclusion:** No evidence of association between gender and voting preference.

**Summary Table: Chi-Square Test**

| **Feature** | **Goodness-of-Fit Test** | **Test of Independence** |
| --- | --- | --- |
| **Purpose** | Compare observed vs. expected frequencies | Test association between two categorical variables |
| **Degrees of Freedom (df)** | k−1*k*−1 (number of categories - 1) | (r−1)(c−1)(*r*−1)(*c*−1) (rows-1 × columns-1) |
| **Example** | Testing if a die is fair | Testing if gender affects voting choice |

**Final Notes:**

* **Chi-square tests work only with categorical data.**
* **Small expected frequencies (<5) may require Fisher’s Exact Test instead.**
* **Used widely in surveys, genetics, A/B testing, and quality control.**

This structured approach ensures accurate analysis of categorical data! 📊