Statistical Hypothesis Testing

Minor in AI, IIT Ropar

1 Why Traditional Statistics Fail: The Threshold Paradox

The Silent Café Paradox

Story: In a small town, four friend groups face a dilemma - should they try the new café? Each member has an "action threshold":

- Radhika (0): "I'll go immediately - I love new places!"
- Aryan (1): "I'll go if at least 1 friend joins"
- Neha (2): "I need 2 friends to go first"
- Raj (3): "Only if 3 friends commit"

The Plot Twist:

- **Group 2:** [Aryan, Aryan, Neha, Neha, Raj]
  Seems promising: Average threshold = 1.8
  Reality: Café remains empty - no initiators!

- **Group 4:** [Radhika, Aryan, Raj, Raj, Raj]
  Looks average: Mean threshold = 2
  Reality: Only Radhika and Aryan go initially - others stay home!

The Statistical Mystery

- **Surface Analysis:** Group 4's mean (2) matches population average

- **Hidden Truth:** Thresholds create *activation chains*
  → No 0-threshold = No ignition (Group 2)
  → Majority need critical mass = System stalls (Group 4)

Key Insight: "The First Mover Crisis"

- **Trendsetters (0-threshold):** The spark for social change.
  Early adopters (Threshold 0) drive group behavior.

- **Statistical Deception:**
  Mean (μ) = $\frac{\sum x_i}{n}$ ignores activation sequences.
  Median ignores threshold dependencies.
  Traditional statistics (mean/median/mode) can be misleading

- **Solution Path:** Requires **network analysis + hypothesis testing**
  Threshold analysis instead of **basic statistics**
  "Do groups with 0-threshold members have significantly different activation rates?"

2 Hypothesis Testing: The Data Scientist's Toolkit

Problem Statement

"Do Company A and Company B have significantly different salaries?"
Traditional approach: Compare means
Better approach: Statistical hypothesis testing

2.1 Core Concepts

In hypothesis testing, several foundational ideas guide the analysis. The **null hypothesis** (H_0) is the default assumption that there is "no significant difference" between the groups or variables under study, meaning any observed variation is attributed to random chance rather than a true effect. In contrast, the **alternative hypothesis** (H_1) asserts that a "significant difference exists," indicating that the observed data reflect a genuine effect or difference between the groups.

A critical part of this analysis is the **p-value**, which represents the probability of obtaining the observed results, or even more extreme ones, if the null hypothesis were true. This measure helps determine whether the evidence is strong enough to reject H_0 in favor of H_1. To make this decision, a **significance level** (α) is set, commonly at 0.05, signifying a 5% risk of a false positive. If the p-value falls below this threshold, the observed difference is considered statistically significant, leading to the rejection of the null hypothesis.

- **Null Hypothesis (H_0):** "No significant difference"
- **Alternative Hypothesis (H_1):** "Significant difference exists"
- **p-value:** Probability of observing results if H_0 is true
- **Significance Level (α):** 0.05 (5% risk of false positive)

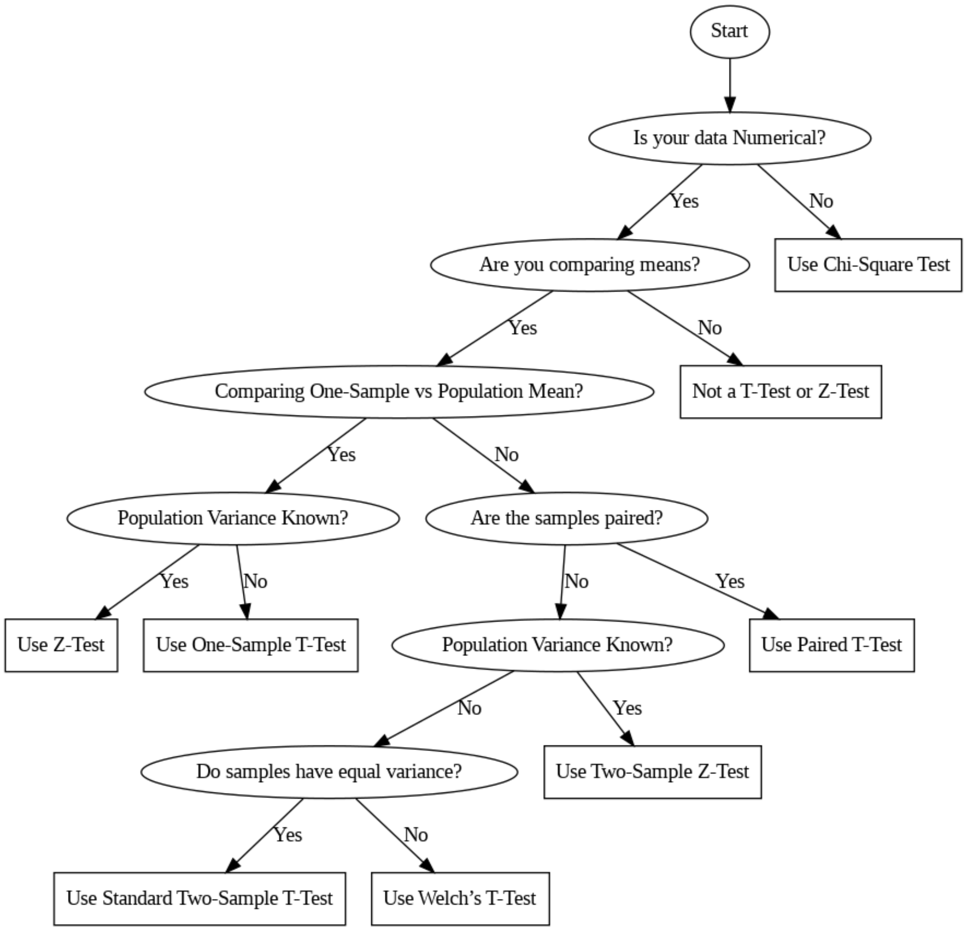
3 Choosing the Right Statistical Test

Test Selection Framework

1. Check variance equality (Levene's Test)
2. Small variance → Use T-test
3. Large variance → Use Welch's Test
4. Categorical data → Chi-square Test

3.1 Levene's Test: Variance Checker

To determine whether two groups have equal variances, we use Levene's test. A high p-value (typically above 0.05) suggests that the variances are not significantly different, while a low p-value indicates that they are.




```python
from scipy.stats import levene  # Importing Levene's test from scipy.
    stats

# Sample salary data for two groups
salaries_A = [55,60,62,58,63,59,61,60,57,64,62,56,59,58,65]
salaries_B = [52,55,57,53,58,54,56,55,51,59,57,50,54,53,60]

# Performing Levene's test to check variance equality
levene_stat, p_value = levene(salaries_A, salaries_B)
print(f"Variance Check: p={p_value:.4f}")
```

3.2 Independent T-Test

An independent t-test compares the means of two groups under the assumption that their variances are equal. If the p-value is below a chosen significance level (e.g., 0.05), we reject the null hypothesis and conclude that there is a significant difference between the two means.

```python
from scipy.stats import ttest_ind  # Importing t-test function

# Performing an independent t-test assuming equal variance
t_stat, p_value = ttest_ind(salaries_A, salaries_B, equal_var=True)
print(f"T-Test Results: t={t_stat:.2f}, p={p_value:.4f}")
```

3.3 Welch's T-Test (Unequal Variance)

Welch's t-test is similar to the independent t-test but does not assume equal variances. It is useful when the two groups have significantly different variances.

```
# Performing Welch's t-test assuming unequal variances
t_stat, p_value = ttest_ind(salaries_A, salaries_B, equal_var=False)
print(f"Welch's T-Test Results: t={t_stat:.2f}, p={p_value:.4f}")
```

3.4 Chi-Square Test (Categorical Data)

The chi-square test is used to determine if there is a significant association between categorical variables. The test compares observed frequencies with expected frequencies under the assumption of independence.

```
from scipy.stats import chi2_contingency  # Importing chi-square test
    function

# Observed frequency table (Male, Female) for different brands
observed = [[30,40],  # Apple
            [50,30],  # Samsung
            [20,30]]  # OnePlus

# Performing chi-square test
chi2_stat, p_value, dof, expected = chi2_contingency(observed)
print(f"Chi-Square Test Results: X2={chi2_stat:.2f}, p={p_value:.4f}")
```

4 Real-World Case Studies

4.1 Case 1: Student Sleep Hours (Z-Test)

A Z-test is useful when comparing a sample mean to a known population mean, assuming the population standard deviation is known. In this case, we analyze student sleep hours to test if the mean sleep duration differs significantly from 8 hours.

```
import scipy.stats as stats  # Importing statistical functions from
    SciPy
import numpy as np  # Importing NumPy for array operations

# Generating sample data with a normal distribution (mean=7.5, std dev
    =1.5, sample size=40)
sample_data = np.random.normal(loc=7.5, scale=1.5, size=40)

# Computing the Z-score manually
z_score = (np.mean(sample_data) - 8.0) / (1.5 / np.sqrt(40))  # Standard
     error adjustment

# Calculating the p-value for a one-tailed test
p_value = stats.norm.cdf(z_score)
print(f"Z-Test Results: Z={z_score:.2f}, p={p_value:.4f}")
```

4.2 Case 2: Battery Life Claim (Two-Tailed Z-Test)

A two-tailed Z-test is used when we want to check whether a sample mean is significantly different (higher or lower) than a claimed population mean. This is useful for verifying manufacturer claims about product performance.

```
# Calculating the p-value for a two-tailed test
p_value = 2 * (1 - stats.norm.cdf(abs(z_score)))  # Multiply by 2 for two-tailed significance
print(f"Two-Tailed Z-Test p-value: {p_value:.4f}")
```

5 Conclusion: Essential Insights

Cheat Sheet

- Always check variance before choosing T-test/Welch
- $p < 0.05 \rightarrow$ Reject null hypothesis
- Threshold analysis needs special handling
- Type I Error: False positive (5% risk)
- Type II Error: False negative

Remember!

- Z-test: Large samples ($n > 30$)
- T-test: Small samples with normal distribution
- Chi-square: Categorical frequency analysis
- Levene's Test: Gatekeeper for variance checks

6 The Human Spark in Numbers

The Unseen Pulse of Groups

Behind every statistic lies a human story. Our café tale reveals a beautiful truth: **Change begins with just one brave soul.**

- Averages tell us *where* groups stand, but miss *how* they move
- The magic isn't in the middle – it's in those who dare to start
- Next time you see a trend, ask: "Who lit the first candle?"

This isn't just about numbers – it's about understanding the invisible threads that connect us all. Whether launching a product, spreading ideas, or building communities, remember: **the right spark can ignite what perfect averages never could.**