Minor in AI

Hypothesis Testing with Z-Test

1 The Chocolate Factory Dilemma

Real-World Motivation

A renowned chocolate factory markets its bars as "precisely 50g." During routine quality checks, you randomly select 10 bars and find an average weight of 52g with a standard deviation of 2g. **Is this 2g difference meaningful**, or could it occur by random chance in a properly calibrated production line?

Why This Matters:

- **Cost Implications**: Halting production for adjustments costs $20,000/hour
- **Regulatory Compliance**: Products must stay within ±3g of claimed weight
- **Brand Trust**: Consistent weight maintains customer loyalty

Hypothesis Testing Approach:

1. **Define Thresholds**: Establish a 95% confidence level (5% error tolerance)
2. **Quantify Uncertainty**: Calculate how much variation is expected from random sampling
3. **Statistical Proof**: Determine if 52g is statistically different from 50g using the Z-test

Key Insight: A 2g difference in a small sample (n=10) carries more uncertainty than in larger samples. The Z-test helps distinguish *meaningful discrepancies* from *random fluctuations* using probability theory.

2 ABC of Z-Testing

2.1 Statistical Hypotheses

Every test begins with two mutually exclusive claims:

- **Null Hypothesis** (H_0): Assumes no effect/difference
  (Factory claim: $\mu = 50$g)
- **Alternative Hypothesis** (H_1): Challenges the status quo
  (Actual mean $\neq 50$g)

The Z-test mathematically compares these using sample data. A key requirement is either:

- Large sample size ($n \geq 30$), or
- Known population standard deviation

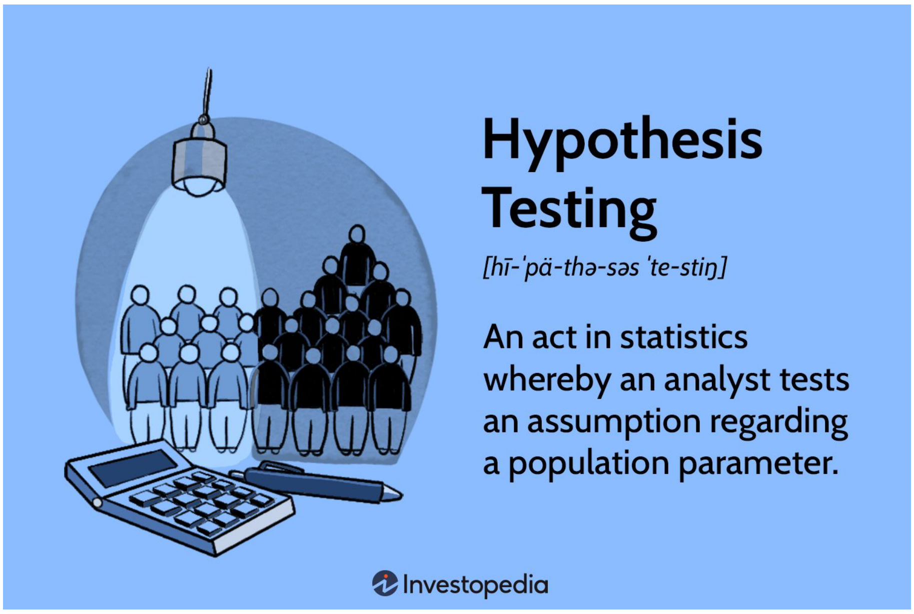

Figure 1: Source: Investopedia, *Hypothesis testing*.

2.2 The Z-Score Formula

Central Equation

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Where:

- $\bar{X}$: Observed sample mean (52g)
- μ: Hypothesized population mean (50g)
- σ: Standard deviation (2g)
- n: Number of observations (10)

This Z-score represents how many standard errors the sample mean deviates from the claimed mean. Larger absolute values indicate stronger evidence against H_0.

2.3 Decision Framework

Hypothesis testing follows a structured workflow:

1. **Define Significance Level** (α): Typically 5% (0.05)
2. **Calculate Critical Value**: Z = ± 1.96 for 95% confidence
3. **Compute Test Statistic**: Z-score from sample data
4. **Compare and Conclude**:
   - Reject H_0 if $|Z| >$ Critical Value
   - Fail to reject H_0 otherwise

3 Getting Hands-In

3.1 Visualizing Critical Regions

The following Python code visualizes critical regions in a normal distribution using the SciPy and Matplotlib libraries. It plots the sampling distribution and highlights the rejection zones based on a 95% confidence level.

Listing 1: Distribution Plotting in Python

```python
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import norm

# Parameters
population_mean = 50
sample_std = 2
sample_size = 10

# Calculations
standard_error = sample_std / np.sqrt(sample_size)
critical_z = 1.96  # 95% confidence level
lower_bound = population_mean - critical_z * standard_error
upper_bound = population_mean + critical_z * standard_error

# Visualization
x = np.linspace(46, 54, 500)
y = norm.pdf(x, population_mean, standard_error)
plt.figure(figsize=(10,4))
plt.plot(x, y, label='Sampling Distribution')
plt.axvline(lower_bound, color='red', linestyle='--',
            label='Rejection Boundary')
plt.axvline(upper_bound, color='red', linestyle='--')
plt.fill_between(x, y, where=(x < lower_bound)|(x > upper_bound),
                 color='orange', alpha=0.3, label='Rejection Zone')
plt.legend()
plt.title('Normal Distribution with 95% Confidence Interval')
plt.xlabel('Sample Mean (grams)')
plt.ylabel('Probability Density')
plt.show()
```

3.1.1 Code Explanation

This code snippet visualizes a normal distribution along with critical regions by computing confidence intervals and plotting them. Below is a breakdown of each section:

- **Importing Libraries:** The necessary libraries such as NumPy, Matplotlib, and SciPy are imported for numerical computation and visualization.
- **Defining Parameters:**
  - $\mu = 50$ (Population Mean)
  - $\sigma = 2$ (Sample Standard Deviation)
  - $n = 10$ (Sample Size)

- **Computing Standard Error:**

$$SE = \frac{\sigma}{\sqrt{n}}$$

  This gives an estimate of the variability of the sample mean.

- **Determining Critical Values:** Using the standard normal table, the critical z-score for a 95% confidence level is 1.96. The lower and upper rejection boundaries are calculated as:

$$\text{Lower Bound} = \mu - Z_{critical} \times SE$$

$$\text{Upper Bound} = \mu + Z_{critical} \times SE$$

- **Plotting the Distribution:**
  - The probability density function (PDF) of the normal distribution is plotted.
  - Vertical dashed red lines indicate the rejection boundaries.
  - The rejection regions (where the sample mean is unlikely to fall) are shaded in orange.

- **Displaying the Plot:** The final plot provides a visual representation of the sampling distribution and critical regions, helping in hypothesis testing.

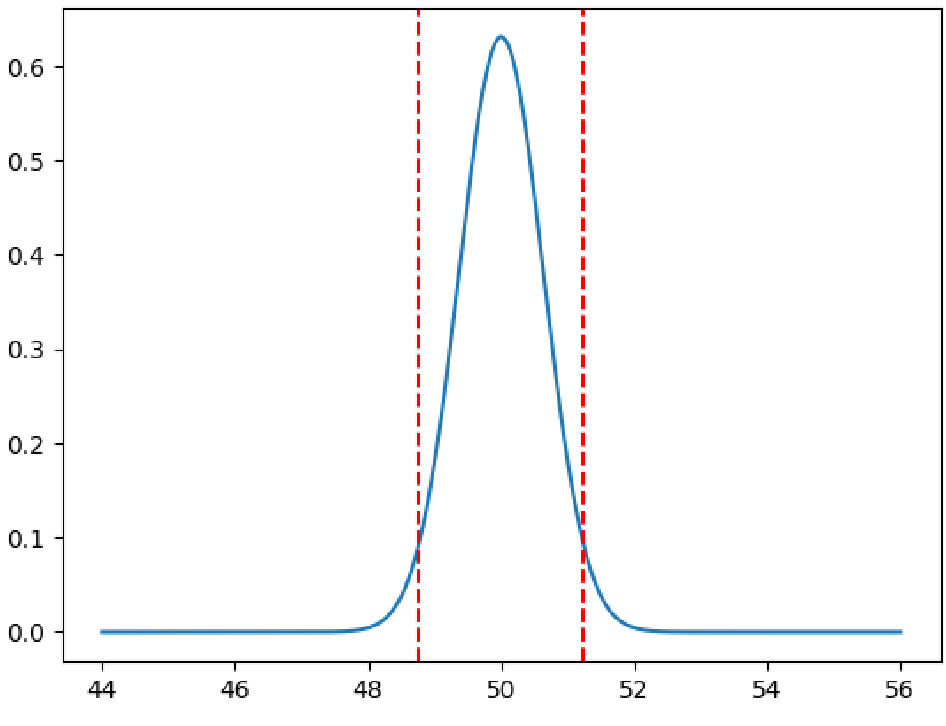

Figure 2: Source: Investopedia, Visual guide to decision boundaries: Sample means beyond red lines (51.24g/48.76g) would reject H_0

3.2 Automated Hypothesis Testing

The following Python code performs a Z-test for hypothesis testing. It calculates the Z-score and compares it with a critical value to determine whether to reject the null hypothesis (H_0).

Listing 2: Python Z-Test Code

```
from scipy.stats import norm

def z_test(pop_mean, sample_mean, sample_std, n, alpha=0.05):
    # Standard error calculation
    se = sample_std / (n ** 0.5)

    # Z-score computation
    z = (sample_mean - pop_mean) / se

    # Critical value (two-tailed)
    critical_z = norm.ppf(1 - alpha/2)

    # Decision logic
    if abs(z) > critical_z:
        print(f"Reject H0 (Z = {z:.2f}, Critical Z = +/-{critical_z:.2f
            })")
    else:
        print(f"Fail to reject H0 (Z = {z:.2f})")

    return z

# Chocolate factory test
z_test(pop_mean=50, sample_mean=52, sample_std=2, n=10)
```

3.2.1 Code Explanation

This function automates hypothesis testing using the Z-test, a statistical test to determine if a sample mean significantly differs from a population mean.

- **Importing SciPy's Norm:** The function uses `norm.ppf()` from SciPy to compute the critical Z-value.
- **Function Parameters:**
  - `pop_mean`: Population mean (μ)
  - `sample_mean`: Sample mean ($\bar{x}$)
  - `sample_std`: Sample standard deviation (s)
  - `n`: Sample size
  - `alpha`: Significance level (default = 0.05 for a 95% confidence level)
- **Standard Error Calculation:**

$$SE = \frac{s}{\sqrt{n}}$$

  The standard error measures the variability of the sample mean.

- **Z-Score Calculation:**

$$Z = \frac{\bar{x} - \mu}{SE}$$

The Z-score measures how many standard errors the sample mean is away from the population mean.

- **Critical Value Calculation:** The critical value for a two-tailed test at 95% confidence is obtained using:

$$Z_{\alpha/2} = \text{norm.ppf}(1 - \alpha/2)$$

This represents the threshold beyond which the null hypothesis is rejected.

- **Decision Rule:**
  - If $|Z| > Z_{\alpha/2}$, reject H_0 (significant difference).
  - Otherwise, fail to reject H_0 (no significant difference).

- **Example Application:** The function is tested with a chocolate factory scenario where:
  - $\mu = 50$ (Expected weight of a chocolate bar)
  - $\bar{x} = 52$ (Observed sample mean)
  - $s = 2$ (Sample standard deviation)
  - $n = 10$ (Sample size)

  The function determines if the observed sample mean significantly differs from the expected weight.

4 Real-World Applications

Case Studies

1. Pharmaceutical Quality Control
A drug manufacturer claims tablets contain 100mg active ingredient. 50 tablets average 98mg (SD=5mg).
Z-test reveals Z = -2.83 (p=0.0047) - reject H_0, indicating underdosing.

2. E-Commerce A/B Testing
Website A: 12% conversion rate (10,000 visitors)
Website B: 13% conversion rate (10,000 visitors)
Z = 5.0 (p $<$ 0.0001) - statistically significant improvement.

3. Clinical Research
New treatment reduces recovery time from 14 to 13 days (SD=2 days, n=100).
Z = -5.0 shows strong evidence for treatment efficacy.

5 Key Insights

Essential Takeaways

- **Confidence vs Significance**: 95% confidence level corresponds to 5% significance ($\alpha = 0.05$).
- **Sample Size Sensitivity**: Larger samples increase test power but require smaller differences to reject H_0.
- **Error Types**:
  - Type I: False positive (rejecting true H_0)
  - Type II: False negative (failing to reject false H_0)
- **Assumptions**: Normally distributed data or large sample (Central Limit Theorem).

References

- **Figure 1 Source:** Investopedia[1]
- **Figure 2 Source:** Investopedia[2]
- **Z-Test Implementation:** Colab Notebook Link

[1] Retrieved from `https://www.investopedia.com/terms/h/hypothesistesting.asp`.
[2] Retrieved from `https://www.investopedia.com/terms/h/hypothesistesting.asp`.