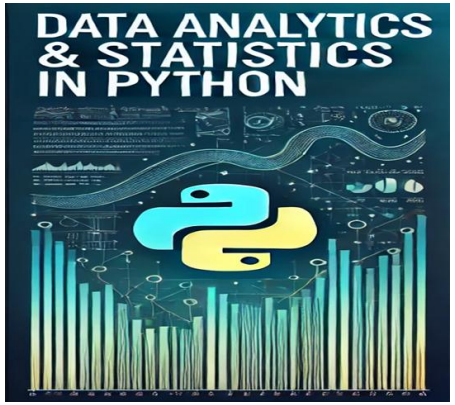
Data Analytics & Statistics in Python

Hypothesis Testing in simple words



Metropolia
University of Applied Science

Learning data-driven decision-making with Pytho

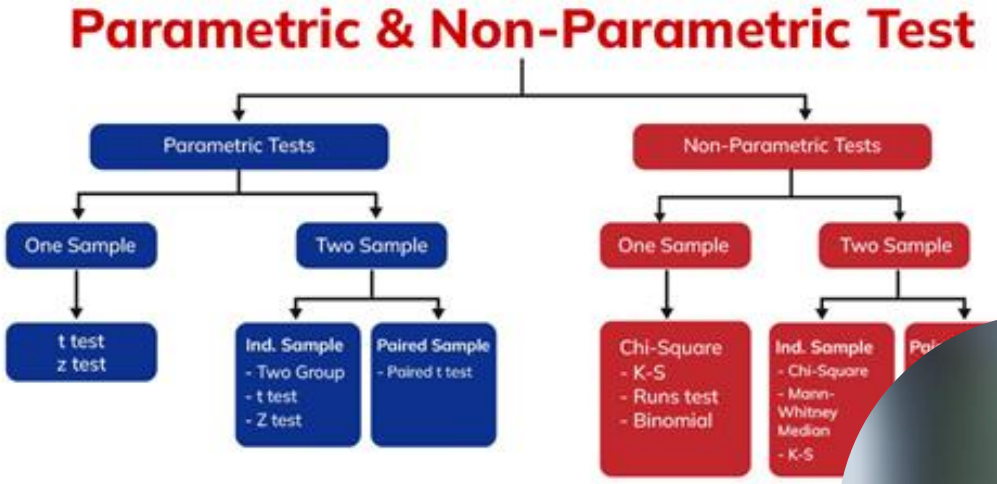
- **Instructor:** Hamed Ahmadinia, Ph.D.
- **Email:** hamed.ahmadinia@metropolia.fi

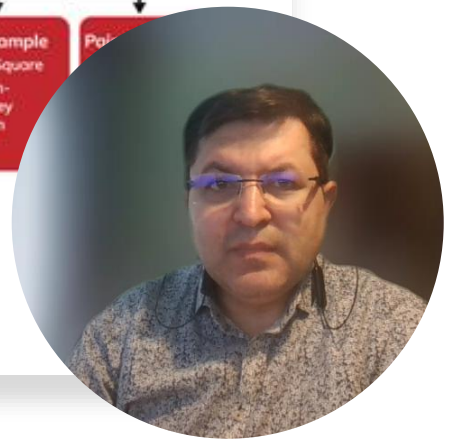
Hypothesis Testing in simple words

- Using Bitcoin & Ethereum price data as examples
- Helps make data-driven decisions
- Common questions:
  - "Is today's Bitcoin price significantly different from last week?"
  - "Do Bitcoin & Ethereum have similar price trends?"

Types of Hypothesis Tests

- **Parametric Tests**: Used when data follows a normal distribution.
- **Nonparametric Tests**: Used when data is not normal or sample size is small.



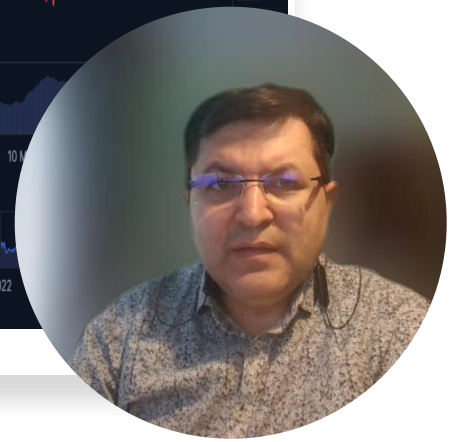


One-Sample Test (Testing Against a Known Value)

Metropolia
University of Applied Sciences

MONERO

- Example: Is today's Monero price significantly different from the past 30-day average?
- **Null Hypothesis (H_0):** Today's price = Past 30-day average
- **Alternative Hypothesis (H_1):** Today's price is different



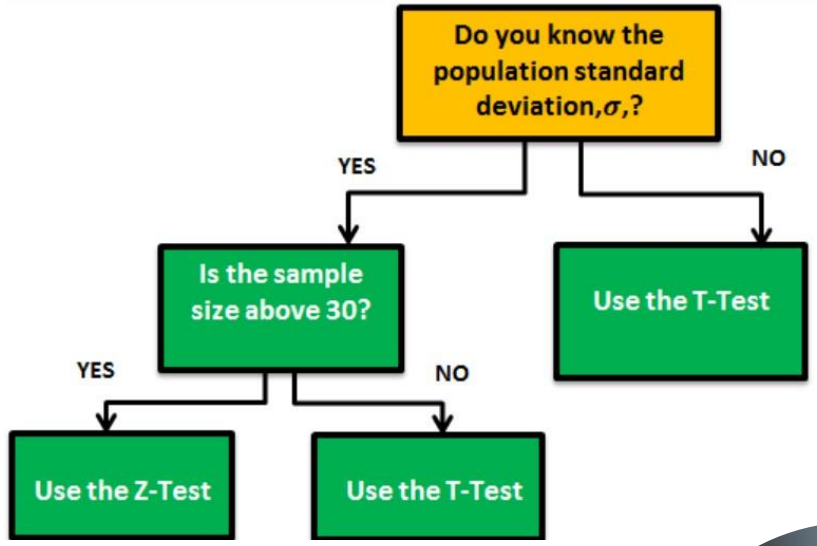
One-Sample Test - Methods

- **A) If Monero Prices Are Normally Distributed**: Use a **t-test**

t-Test Formula

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

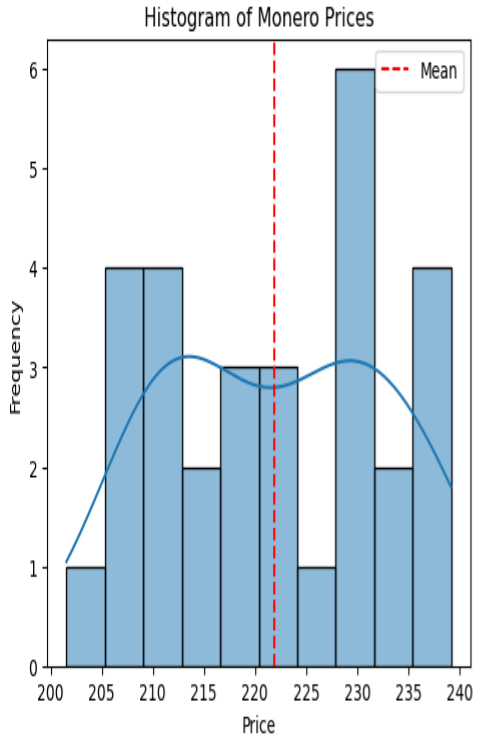
$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

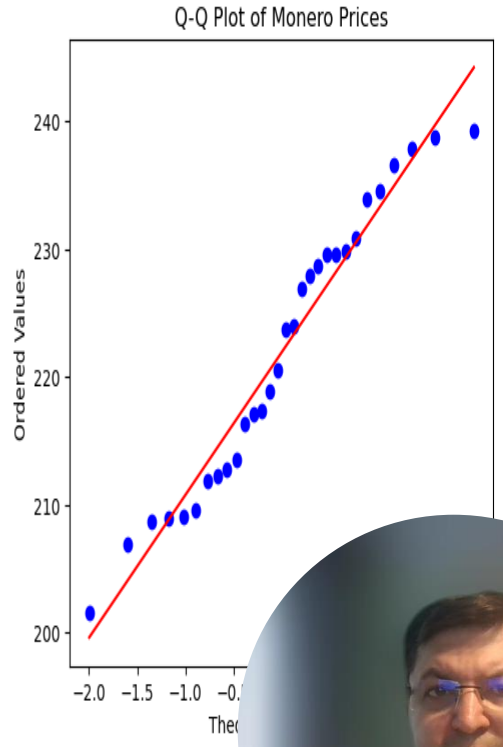


One-Sample Test - Methods

- **Example Calculation:**
  - Past 30-day average price = $221,000
  - Today's price = $209,000
  - Standard deviation = $10,93
  - Sample size = 30
- **We perform a one-sample t-test**
- **Result Interpretation:**

1. If **p-value < 0.05**, we reject H_0 and say today's price is significantly different.
2. If **p-value > 0.05**, no significant difference.



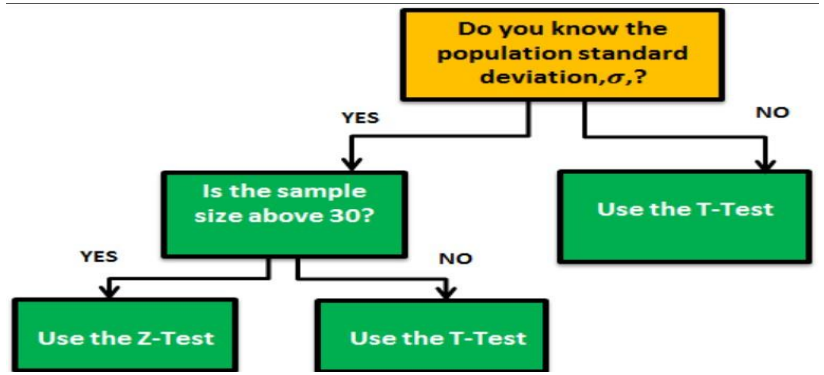


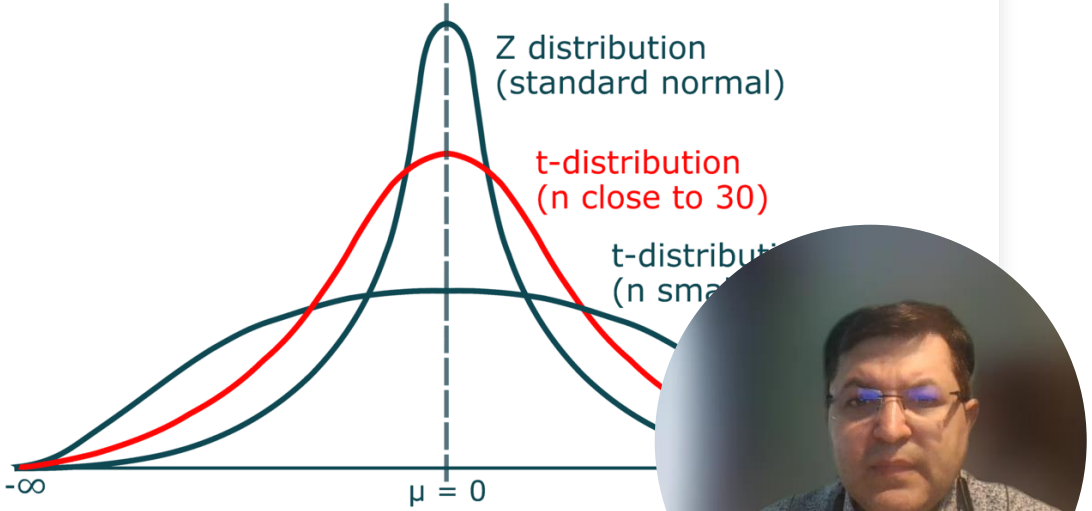
One-Sample Test - Methods

- **B) If Sample Size > 30 Days**: Use a **z-test**
- when we have a large dataset (e.g., 100+ days of Monero prices).

Z Test Statistics Formula

$$\text{Z Test} = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$





One-Sample Test - Methods

- **C) If Prices Are Not Normally Distributed → Use Kolmogorov-Smirnov or Chi-Square Test**
- **Kolmogorov-Smirnov test** checks if today's price follows the same distribution as the past.
- **Chi-square test** can be used if we categorise prices (e.g., above/below $50,000).

The e.d.f. $\hat{F}_n$ is defined for all real numbers x to be

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} I[X_i \leq x],$$

The one-sample Kolmogorov-Smirnov statistic with respect to model F_0 is

$$M_{KS} = \sqrt{n}\max_x |\hat{F}_n(x) - F_0(x)|,$$

The Formula for Chi Square Is

$$\chi_c^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

where:

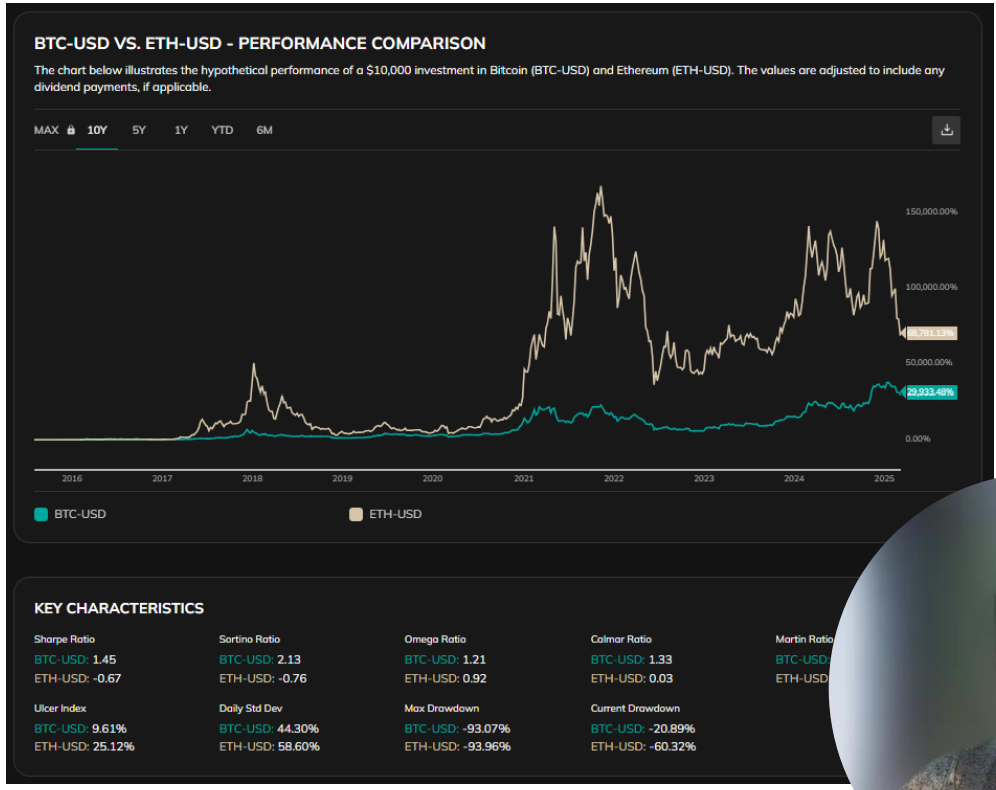
c = degrees of freedom

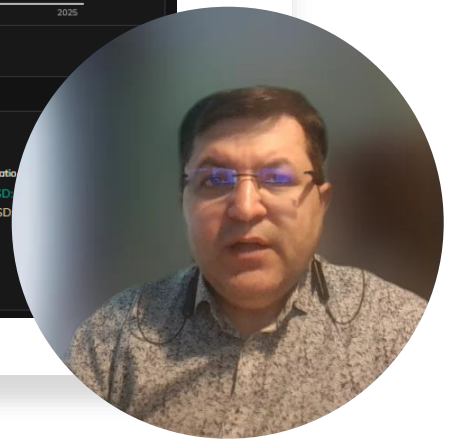
O = observed value(s)

E = expected value(s)

Two-Sample Test (Comparing Two Groups)

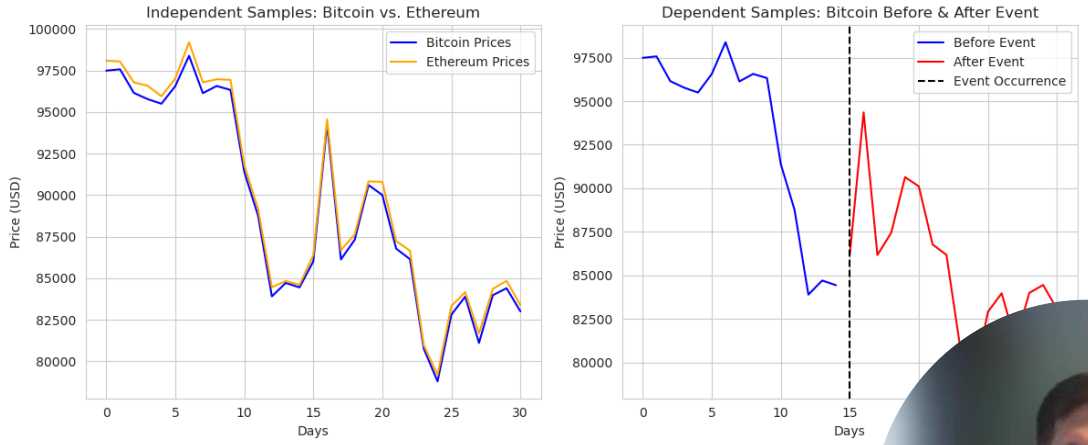
- Example: Do Bitcoin and Ethereum prices have significant differences?
- **Null Hypothesis (H_0):** Prices are the same.
- **Alternative Hypothesis (H_1):** Prices are different

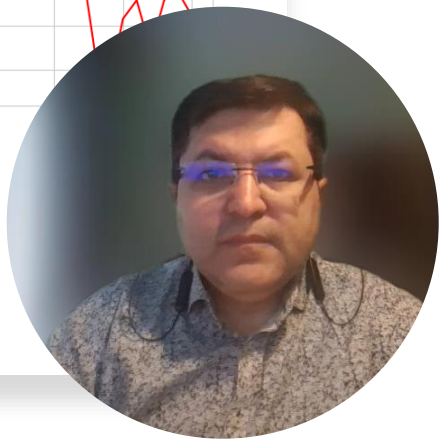




Two-Sample Test - Methods

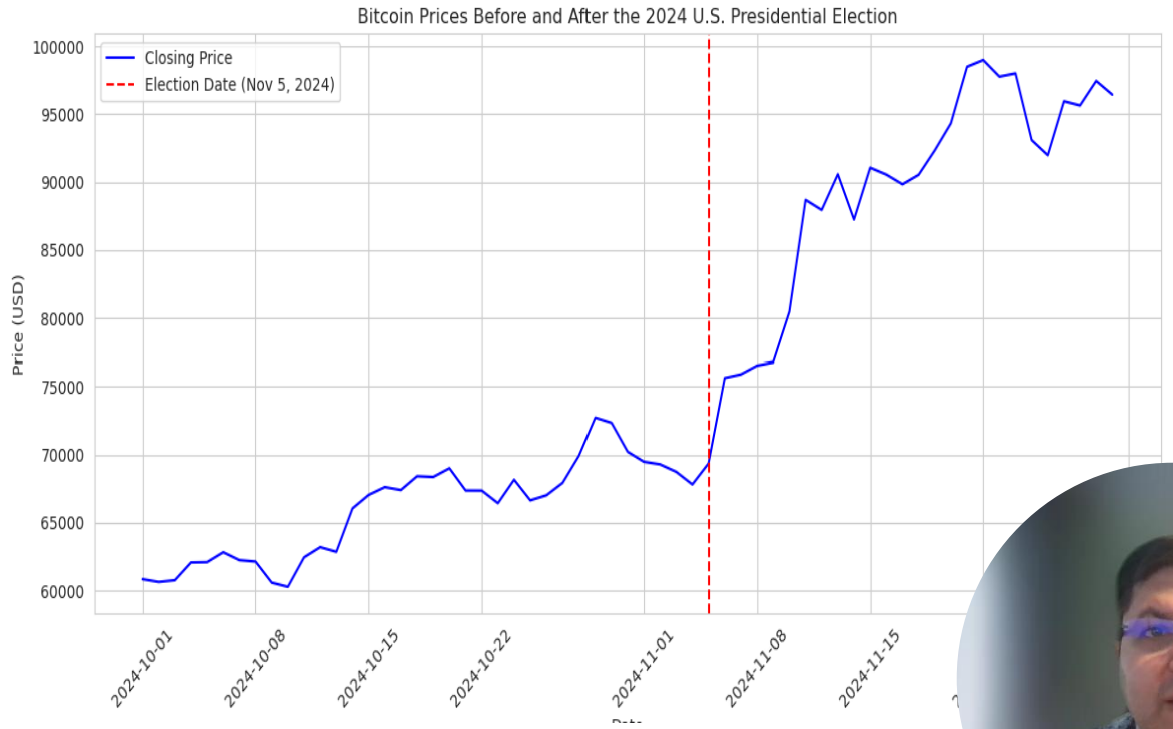
- **If Normally Distributed**: Use a **Two-group t-test**
- **If Not Normally Distributed**: Use **Mann-Whitney Test**
- **If Comparing Categorical Data**: Use **Chi-Square Test**





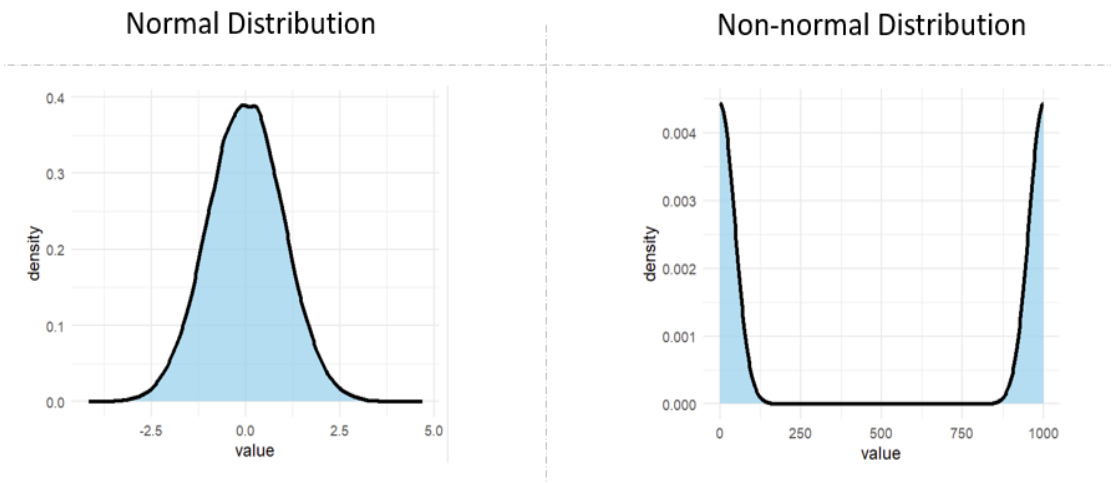
Paired Samples (Before & After a Market Event)

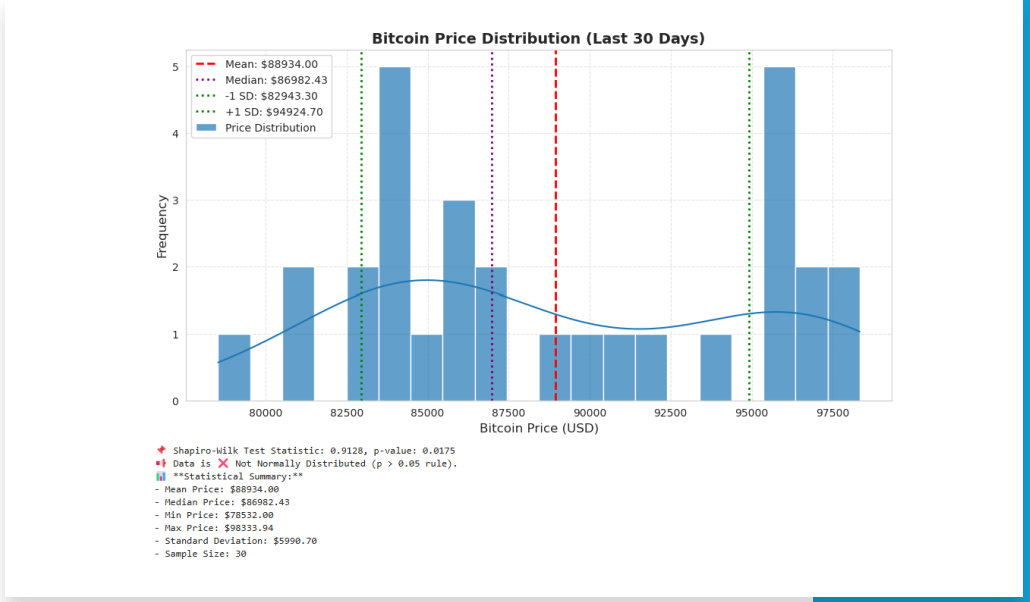
- Example: Did Bitcoin price change after an event (e.g., Trump's presidential election win)?
- **Null Hypothesis (H_0):** No effect.
- **Alternative Hypothesis (H_1):** Event affected price.

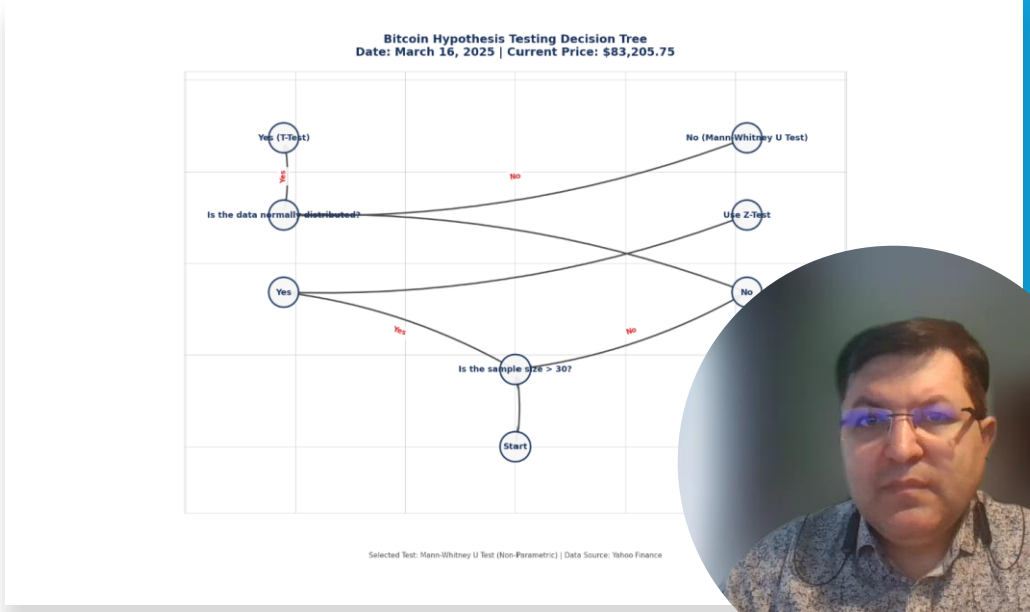


Paired Sample Test - Methods

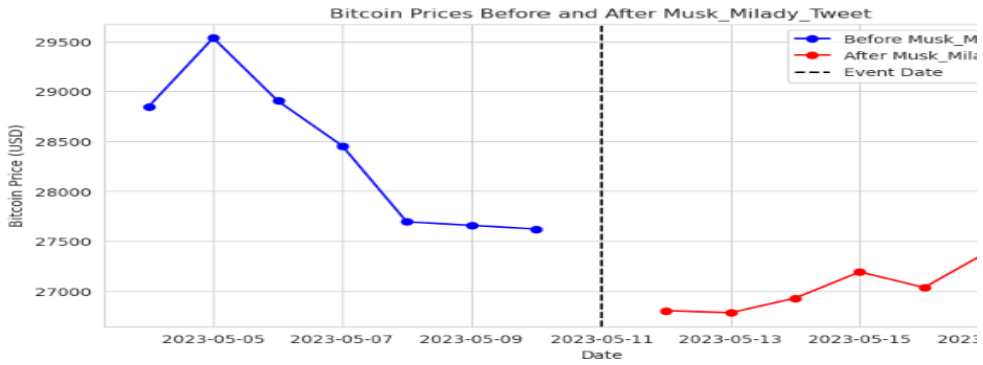
- **If Normally Distributed**: Use a **Paired t-test**
- **If Not Normally Distributed**: Use **Wilcoxon Test**
- **If Using Yes/No Data**: Use **McNemar Test**



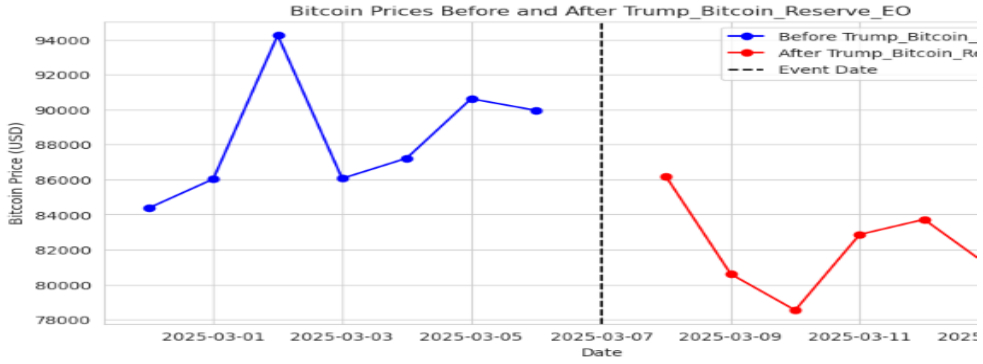




Choosing the Right Hypothesis Test



Event: Musk_Milady_Tweet
T-statistic: 4.1111, P-value: 0.0063

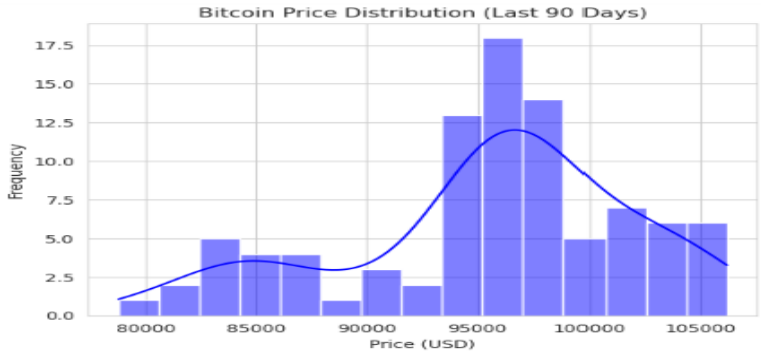


Event: Trump_Bitcoin_Reserve_EO
T-statistic: 2.8562, P-value: 0.0289

Scenario	Test to Use
Compare today's Bitcoin price to the past 30-day average	**t-test** (if normal)
Compare Bitcoin & Ethereum prices	**Two-group t-test** (if normal)
Compare Bitcoin before & after an event	**Paired t-test** (if normal)
Compare Bitcoin & Ethereum if data is not normal	**Mann-Whitney test**
Compare price categories (Above/Below $50K)	**Chi-square test**
Compare today's price to historical patterns	**Kolmogorov-Smi… test**
Compare Bitcoin price trends before & after regulation change	**Wilcoxon … test**
Compare frequency of Bitcoin price increase after news	**McNemar**

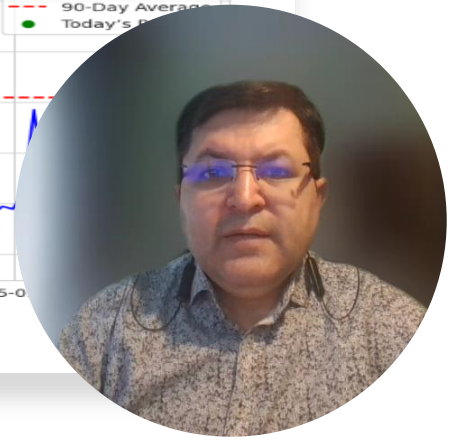
Steps to Conduct Hypothesis Testing

1. **Check if your data is normal** (Use histogram or Shapiro-Wilk test).
2. **Determine the sample type**: One sample, two samples (independent or paired).
3. **Choose the appropriate test** based on distribution & sample type.
4. **Interpret the p-value**:
    1. **p-value < 0.05**: Reject H_0 (significant difference).
    2. **p-value > 0.05**: Fail to reject H_0 (no significant difference).




```
Shapiro-Wilk Test P-Value: 0.0003
Data is NOT normally distributed (use a non-parametric test).

Statistical Test: Wilcoxon Signed-Rank Test
Test Statistic: 33.0000
P-Value: 0.0000
Conclusion: Today's price is significantly different from the 90-day average.
```



Final Takeaways

- The right test depends on **distribution, sample size, and data type**.
- Always check **assumptions** before applying statistical tests.
- **p-value < 0.05** means a statistically significant result.