Hypothesis Testing

Yong Chen, Jamie Pham, Elvin Xu

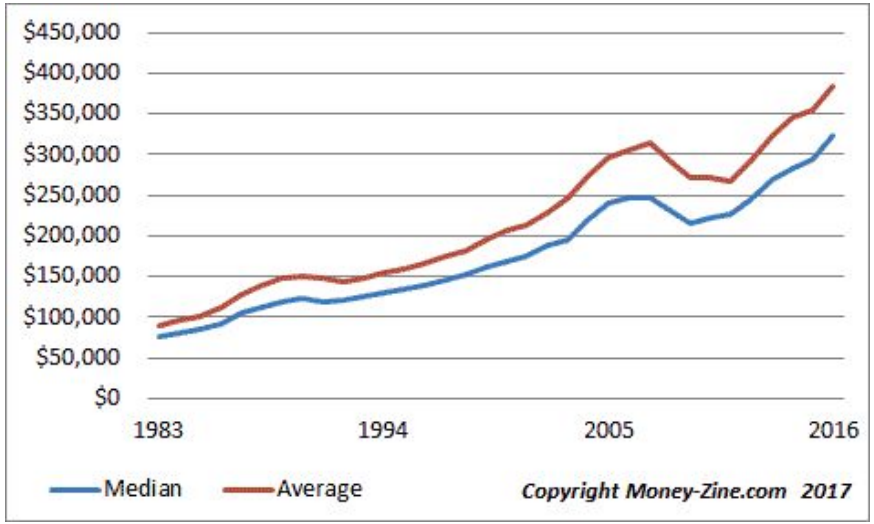
Introduction

- Hypothesis: A premise or a claim that we want to test.
- For hypothesis testing, we need to have:
  - Null Hypothesis: H_0.
    - Also known as the first hypothesis.
    - States that the population parameter is equal to the claimed value.
  - Alternative Hypothesis: H_1
    - Used in hypothesis testing that is contrary to the null hypothesis
    - Usually taken to be that the observations are the result of a real effect

Example: Average Home Price

According to Money-zine.com:

- Average home price in the U.S. is: $384,000.00.
- Median home price in the U.S. is: $332,500.00.
    - This number will not be used.
- We will set the average home price as the null hypothesis later on.



Hypothesized Scenario:

Given the average home price of 50 states along with Washington D.C. (51 sample size), reject or accept the following claim via hypothesising testing .

- According to Money-Zine, the average home price in the U.S. is: $384,000.00.
- Let's assume that confidence level is 95%.
  - 95% CI is the standard for most hypothesis testing.
  - Sometimes 99% CI is used for more precise measurement.

Step 1: Setup the Hypothesis

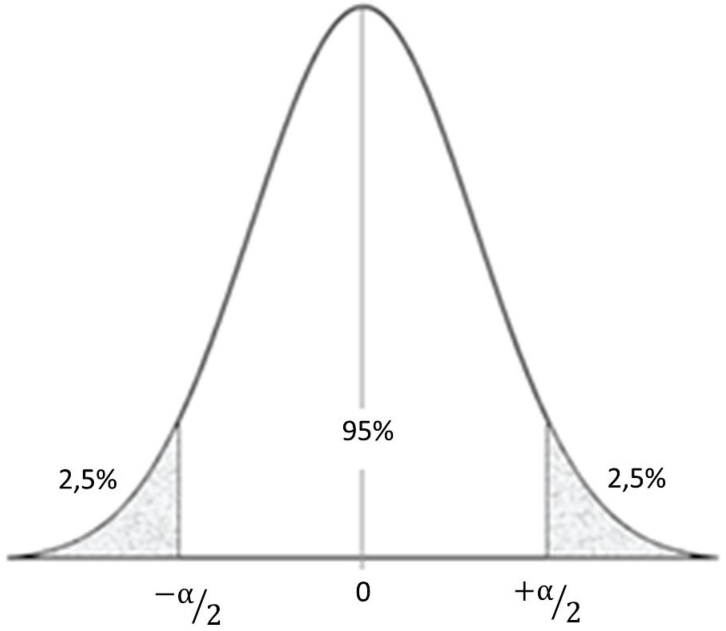
- Testing: Average home price in the U.S. is: $384,000.00.
- Null Hypothesis
  - H_0: $\mu = 384000.$
  - The average home price in the U.S. is $384,000.00.
- Alternative Hypothesis
  - H_1: $\mu \neq 384000.$
  - The average home price in the U.S. is not $384,000.00.
- Confidence interval: 95%.
- We will use **two-tailed test**, since we don't know if the average home price according to sample size is greater than less than $384,000. We just want to check whether the average home price is $384,000 or not.

Step 2: Identify Given Variables

- Sample size: **n = 51**.
  - The data consists of average housing price from 50 states and Washington DC, therefore, the sample size is 51.
- Degree of freedom: **df = 50**.
  - Keep in mind that df = n - 1, df = 51 - 1, so df = 50.
- Level of Significance: α = **0.05**.
  - Since the Confidence Interval = 95%, that means we set the level of significance is 5%.
  - $\alpha = 1 - CI \rightarrow \alpha = 1 - 0.95 \rightarrow \alpha = 0.05$.
- Median income and average home price for all 51 samples.

Step 3: Find the Critical Value.

- Two-tailed test.
- If the t-value is within the shaded area, then we reject null hypothesis.
- $\alpha = 0.05$
- **α / 0.2 = 0.025**
- Critical Value (Upper): $t_{0.025} \approx 2.008559$
  - In R: qt(0.025, df = 50, lower.tail = F)
- Critical Value (Lower): $-t_{0.025} = -2.008559$
  - Due to symmetry.
- Accept H_0 if t value is within [-2.008559, 2.008559].
  - Non-shaded area.
- Reject H_0 otherwise.
  - Shaded area.

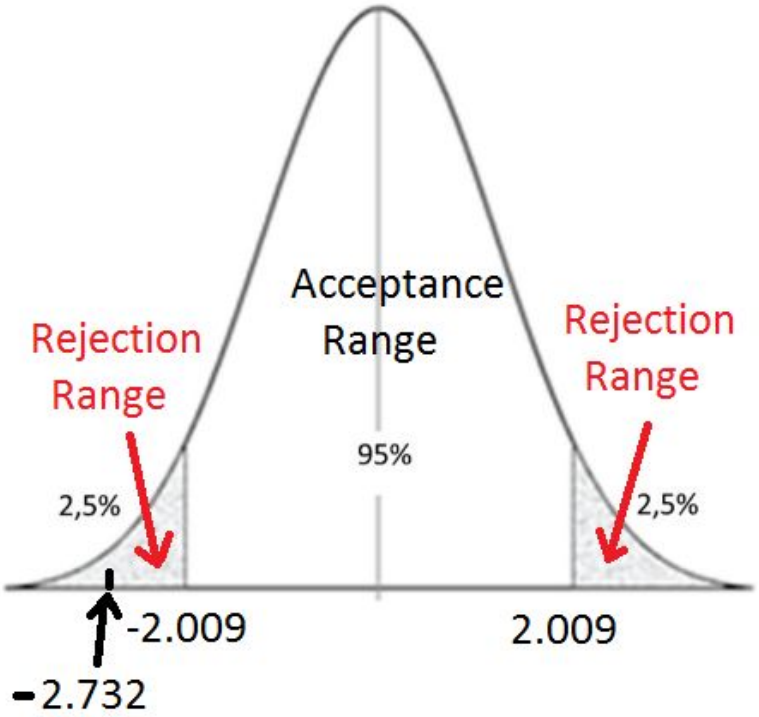


Step 4: Get Variables for the Sample Data

- Population Mean: **μ = 384000.00**.
  - Average home price in the U.S. is: $384,000.00.
- Sample Mean: **X' = 325020.90**.
  - In R: mean(Average_Home_Price).
  - Within these 51 samples, the average cost is **$325020.90**.
- Sample Standard Deviation: **s = 154169.90**.
  - In R: sd(Average_Home_Price).
  - Within these 51 samples, the standard deviation is **$154169.90**.

Step 5: Make Decision Based on T-Value

- Now we need to find the t-value.
  - Formula: T-value = **(X' - μ) / (s / sqrt(n))**
  - **T-value = (325020.90 - 384000.00) / (154169.90 / sqrt(51))**
  - **T-value = -2.73201852**
- Since the T-value is -2.73201852, which is outside of [-2.008559, 2.008559], or the acceptance range. We reject the null hypothesis H0, where the average home price is $384,000.00.
- Conclusion: Reject H_0.



Step 6: Calculate the P-Value

- Alternatively, we can find P-value.
- Since our our t-value end up being on the left side of the rejection range, and our sample mean is less than population mean, we use less than sign in this case.
- P = P(X' < 384000.00) = P(t < -2.73201852)
  - In R: pt(-2.73201852, 50, lower.tail = T)
  - P-Value = **0.004338067**.
- The probability of that the null hypothesis being correct is around **0.43%**, which is less than 5%. Since the probability of the original conclusion being correct is too low, we should reject the null hypothesis.
- If the probability of the null hypothesis being correct is around 0.43%, that means the alternative hypothesis being correct should be **99.57%**.

Step 7: Make Decision Based on P-Value

- Compare P-Value to level of significance.
  - P-Value < level of significance.
  - 0.004338067 < 0.05.
- Since $p < \alpha$, that means we reject null hypothesis H_0 again.
- 0.004338067 is not only less than 0.05, it's also less than 0.01, that means there is a very strong evidence that the null hypothesis is incorrect.
- According to the sample size, the actual average home price is not even close to $384,000.00, which is the average home price that Money-Zine claimed.
- Conclusion: Reject H_0 again.

Summary

- We find a website that gives us the mean value for home price across the United States. We tried to prove the average home price in the U.S. is $384,000 using average home price for each of the 50 states and Washington D.C. (51 samples).
  a. Setup the null hypothesis and alternative hypothesis.
  b. Identify the given variables.
  c. Find the critical values on the T-distribution graph.
  d. Use R to get required variables to calculate the T-value.
  e. If the T-value is within the acceptance range, accept H_0, reject otherwise.
  f. Compute for the P-value.
  g. If P-value > level of significance, accept H_0, reject otherwise.

Source

Average Home Price by State:

- https://www.trulia.com/home_prices/

Average Home Price in the U.S.

- http://www.money-zine.com/financial-planning/buying-a-home/average-home-prices/