Z-Test, P-Value, and Critical Value: Explained Intuitively

# 🔍 Introduction

This document walks you through the concepts of Z-test, p-value, and critical value in a simple, intuitive way. It includes a real-world example using OLS (Ordinary Least Squares) regression where you're testing whether an independent variable (X) significantly impacts the dependent variable (Y).

# 📊 Real-World Scenario: Does X Affect Y?

Imagine you're running a regression to understand if a marketing variable X (say, advertisement budget) has an impact on Y (sales).  
  
You formulate hypotheses:  
H₀: X has no effect on Y (coefficient = 0)  
H₁: X has an effect on Y (coefficient ≠ 0)  
  
You run an OLS regression and get a t-statistic of 2.2 and a p-value of 0.03.

# 📉 What Does P-Value Mean?

Assume H₀ is true — i.e., X has no effect on Y. Then the p-value answers:  
“What’s the probability of observing a result this extreme (t = 2.2 or higher) just by chance?”  
  
A p-value of 0.03 means there's only a 3% chance you'd see this extreme result if the null hypothesis were true.

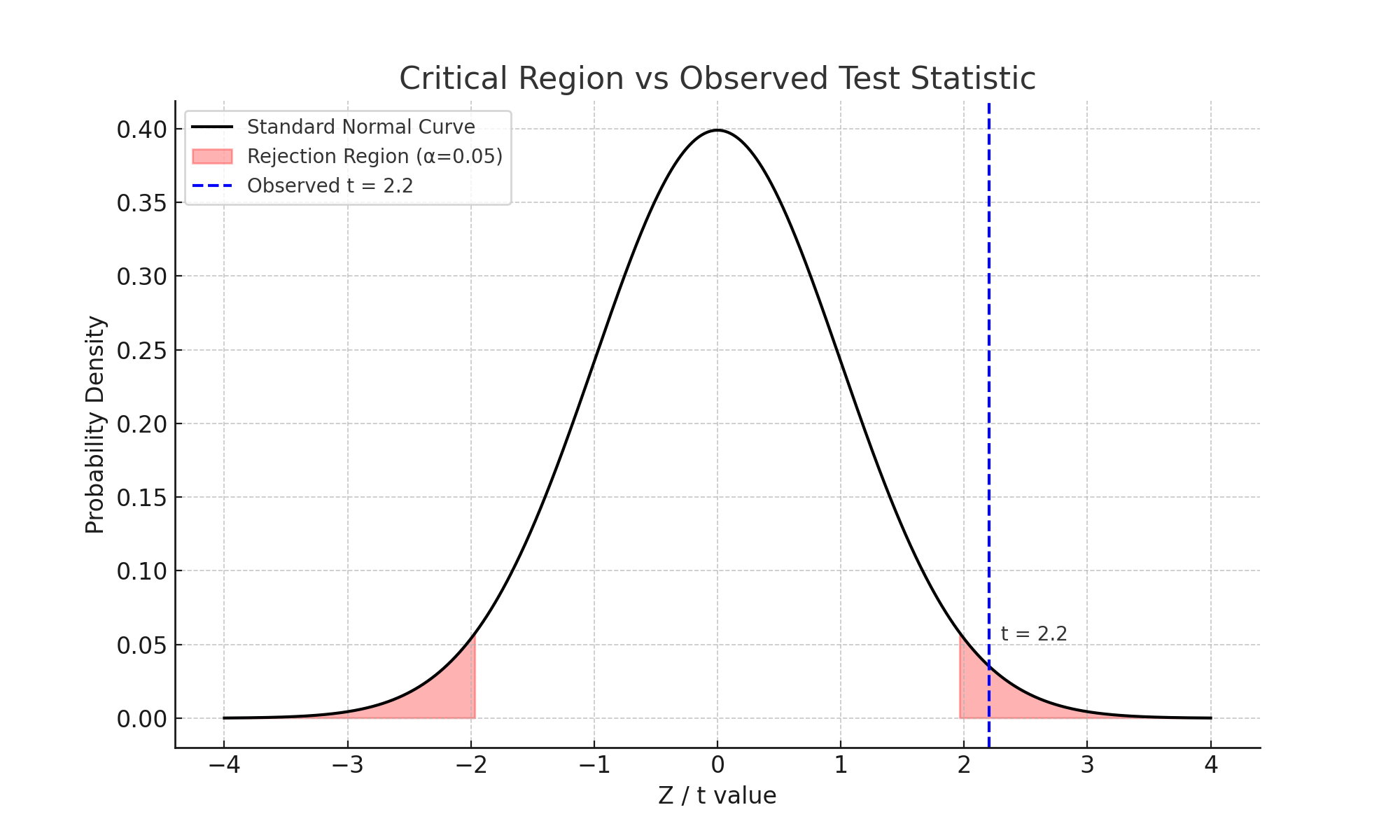
# 📏 What is Critical Value?

The critical value defines the threshold beyond which results are considered rare/extreme enough to reject H₀.  
For a 95% confidence level (α = 0.05), the critical z-values are ±1.96 for a two-tailed test.  
  
If your test statistic is greater than +1.96 or less than -1.96, it falls in the rejection region.

# 🔗 Connection Between P-Value and Critical Value

Both are two sides of the same coin:  
- If test statistic > critical value → reject H₀  
- If p-value < α → reject H₀  
  
In our case: p-value (0.03) < α (0.05) → reject H₀. The result is statistically significant.

# 📊 Visual: Critical Region vs P-Value



# ✅ Summary

• A p-value tells you the probability of seeing such an extreme result assuming H₀ is true.  
• A critical value is a threshold based on your α level that determines whether to reject H₀.  
• In both methods, you're checking whether the result is rare/extreme enough to disbelieve the null.  
• In regression, you use these values to check if a feature (X) significantly impacts your target (Y).

# 🧪 What is a Z-Test?

A Z-test is a type of hypothesis test used to determine whether there is a significant difference between sample data and a known population value (mean or proportion), assuming the population standard deviation is known.  
  
It is most commonly used when:  
• The population standard deviation is known.  
• The sample size is large (n > 30).  
• The sampling distribution is approximately normal.  
  
Z-Test Formula for Mean:  
z = (x̄ - μ) / (σ / √n)  
  
Where:  
• x̄ = sample mean  
• μ = population mean  
• σ = population standard deviation  
• n = sample size  
  
We compare the calculated z-statistic with the critical z-value or use the corresponding p-value to make decisions.

# 📉 How is P-Value Calculated Visually?

To calculate the p-value from the test statistic (e.g., z = 2.2), we calculate the probability of observing a value at least this extreme — i.e., in the tails of the standard normal distribution.  
  
This is done by integrating the area under the normal curve beyond the observed statistic.  
• For a two-tailed test, we calculate the area to the right of +z and to the left of -z and sum them.  
• For a one-tailed test, we only consider one side.  
  
For z = 2.2:  
p-value = P(Z ≥ 2.2) + P(Z ≤ -2.2)  
This area represents how likely we are to see a value this extreme under the null hypothesis.

