A/B Testing Case Study: Improving Click-Through Rate via Design Optimization

1. **Background & Problem Statement**

   Our team set out to improve the click-through rate (CTR) on a product page by experimenting with UI changes. Specifically, we wanted to see if a new design variant (Group B) could outperform the existing version (Group A). The hypothesis was that even small UX tweaks can drive measurable improvement in user engagement.

2. **Hypothesis**

- Null Hypothesis (H_0): There is no difference in the click-through rate between Group A and Group B ($p_1 = p_2$).
- Alternative Hypothesis (H_1): Group B has a higher click-through rate than Group A ($p_2 > p_1$).

3. **Experiment Design**

   We designed a controlled A/B test, randomly assigning users to either:

- Group A (Control): Baseline design with a historical click rate of 12%.
- Group B (Treatment): New design with an expected improvement to 15%.

Parameters:
Significance Level (α): 0.05
Power (1−β): 0.80
Sample Size per Group: 2031, users (determined via power analysis). We simulated binary click behavior using a binomial distribution to mimic real-world user interaction.

4. **Power & Sample Size Calculation**

   To ensure the test had enough statistical strength to detect a real difference between the two versions, I performed a power analysis using the statsmodels library in Python. The goal was to compare a baseline conversion rate of approximately 11.9% (Group A) with an improved rate of 14.2% (Group B). Using a standard significance level of 0.05 and a desired statistical power of 80%, the analysis determined that a minimum of 2,031 users per group was required. This sample size gives the test a strong chance of detecting a meaningful improvement in user engagement if one truly exists.

5. **Results Summary**

After simulating the test, we computed the actual click-through rates:

	Group	Clicks	Total Users	Conversion Rate
0	A	242	2031	11.92%
1	B	288	2031	14.180000000000001%

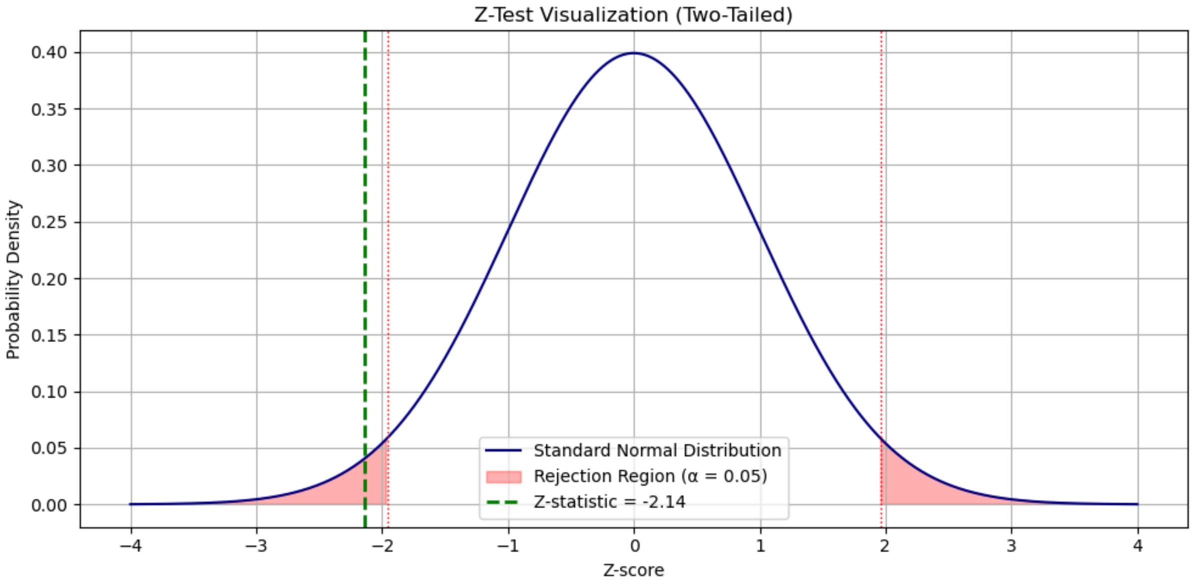
Z-Test Results:

```
        Z-Test Results:
Z-statistic: -2.143
P-value: 0.0321
Statistically significant Group B performs better than Group A.
```

Interpretation: Group B performs significantly better than Group A.

6. **Visualization**

The Z-distribution below shows the rejection regions (shaded red), with the observed z-statistic marked as a dashed green line:



7. Type I & Type II Error Discussion

- Type I Error (False Positive):
  Risk of concluding B is better when it's not — here, 5% risk ($\alpha = 0.05$).
- Type II Error (False Negative):
  We controlled for this by choosing a power of 80%, which gives a 20% chance of missing a true effect.

8. Business Conclusion & Recommendations

The A/B test showed a statistically significant improvement in click-through rate with the new design (Group B). Group A had a conversion rate of 11.92%, while Group B reached 14.18%, representing a relative increase of over 18%. With a p-value of 0.0321, this result is unlikely due to chance and supports adopting the new design.

Recommendations:

- Roll out the Group B design to all users to benefit from the improved performance.
- Monitor post-launch metrics to confirm the effect remains consistent in production.
- Explore further A/B tests on additional UI elements to continue improving engagement.