E-Commerce Website A/B Testing and Causal Inference

Background:

An e-commerce company wants to try a new website design to increase the percentage of users who make a purchase. Data from almost 300,000 users was collected on whether the user was assigned to the new website or the old website, how much time the user spent on the website, the country of the user, and whether or not they made a purchase.

Experiment Goal:

Run an A/B test to evaluate whether a redesigned landing page increases user conversion relative to the control page.

Hypothesis:

Users exposed to the treatment page will have a higher probability of conversion than users exposed to the control.

Results:

1. Verified balanced assignment of groups with sample ratio mismatch (SRM)

2. Conducted an A/B test using a two-sample proportion z-test
   a. The observed proportion of conversion is 12.04% for the control group and 11.88% for the treatment group
   b. The p-value is 0.207 and the z statistic is 1.26.
   c. The 95% confidence interval is (0.1187, 0.1220) for the control group and (0.1172, 0.1205) for the treatment group.

3. Calculated effect size of 0.0049

4. Determined the minimum detectable effect (MDE) is 0.030 with a power of 0.8

5. Performed causal inference by building a logistic regression model with covariates
   a. The LLR p-value is 0.27, and all coefficients had p-values greater than 0.05.

Interpretation:

For the A/B test, the null hypothesis is that there is no difference in the proportion of users who convert between the control and treatment groups. The alternative hypothesis is that there is a difference in proportions. Using an alpha significance level of 0.05, the null hypothesis will be rejected if the p-value is less than 0.05 or the z statistic is greater than 1.96. Since the p-value is 0.21 and the z statistic is 1.26, we fail to reject the null hypothesis, indicating that there is no statistically significant difference between the two proportions.

The observed effect size was 0.0049, which is substantially smaller than the minimum detectable effect of 0.03 required for 80% power. As a result, the experiment was not sufficiently sensitive to detect effects of this magnitude.

The logistic regression did not detect a statistically significant treatment effect after controlling for covariates

This suggests insufficient evidence to conclude that the redesigned page causally impacts conversion within the observed sample.

Recommendation:

Given the minimal observed effect size and lack of statistical evidence, the redesigned page should not be rolled out globally at this time. Future experiments could explore alternative design iterations or target high-engagement user segments where exploratory effects appeared larger.