A/B Test Analysis: Ad Campaign Variant

December 10, 2025

1 Overview

This white paper summarizes an A/B test comparing a control ad campaign to a test variant. The objective was to assess impact on purchases (primary) while monitoring cost efficiency and intermediate conversion rates. Data comprised 29 control rows and 30 test rows; an aligned view (29/29) removes one test-only date to check robustness.

2 Data and Metrics

- **Primary**: Purchases.

- **Guardrail**: Cost per acquisition (CPA = purchases / spend).

- **Supporting**: Purchase per reach, purchase per impression.

- **Spend**: Included as a budget guardrail (not to increase materially).

- Preprocessing: Dropped one all-NA control row; created aligned set with common dates only.

3 Methods

- Tests: Welch two-sample t-test (unequal variances) and Mann–Whitney U as robustness.

- Multiple comparisons: Benjamini–Hochberg FDR on p-values per dataset (full, aligned).

- Uncertainty: Bootstrap 95% CIs for mean differences.

- Interpretation focuses on direction, magnitude, and statistical significance (alpha = 0.05).

4 Results

4.1 Key Point Summary

- Purchases: No significant difference; wide CI includes both lift and decline.

- CPA: No significant difference; cost efficiency inconclusive.

- Conversion rates: Test significantly increases purchase per reach (˜126%) and purchase per impression (˜70%) in both full and aligned analyses; survives FDR.

- Spend: Test spends ˜11% more; increase is statistically significant (guardrail broken).

4.2 Metric Table (Full Dataset: n_test = 30, n_ctrl = 29)

Metric	Test Mean	Ctrl Mean	Abs Diff	% Diff	t p (FDR)
Purchases	521.23	522.79	-1.56	-0.30%	0.976 (0.976)
CPA	0.2066	0.2320	-0.0255	-10.97%	0.280 (0.374)
Purchase / Reach	0.01413	0.00633	0.00780	123.15%	0.0042 (0.0085)
Purchase / Impression	0.00843	0.00500	0.00342	68.48%	0.0020 (0.0080)
Spend (USD)	2563.07	2304.07	259.00	11.2%	0.0071 (−)

Table 1: Welch t-test p-values shown with FDR in parentheses where applied. Spend guardrail not FDR-adjusted. Mann–Whitney U agreed on significance patterns.

4.3 Metric Table (Aligned Dataset: n_test = 29, n_ctrl = 29)

Metric	Test Mean	Ctrl Mean	Abs Diff	% Diff	t p (FDR)
Purchases	512.72	522.79	-10.07	-1.93%	0.8469 (0.8469)
CPA	0.2022	0.2320	-0.0299	-12.87%	0.2052 (0.2736)
Purchase / Reach	0.01434	0.00633	0.00801	126.45%	0.0045 (0.0089)
Purchase / Impression	0.00848	0.00500	0.00348	69.65%	0.0023 (0.0090)
Spend (USD)	2572.24	2304.07	268.17	11.6%	0.0060 (−)

Table 2: Aligned excludes the single test-only date; conclusions unchanged.

4.4 Confidence Intervals for Mean Differences (Full)

- Purchases: [-109.46, 91.21]

- CPA: [-0.0747, 0.0162]

- Purchase / Reach: [0.0036, 0.0135]

- Purchase / Impression: [0.0015, 0.0056]

- Spend: [84.58, 435.92]

5 Interpretation

- The test variant improves conversion efficiency (per reach/impression) with strong statistical support.

- Total purchases show no statistically detectable change; with current sample, both lift and decline remain plausible.

- CPA is inconclusive; CIs cross zero, so no clear efficiency gain.

- Spend increases by ˜11% with significance, breaking the spend guardrail; higher cost accompanies the conversion-rate gains.

6 Limitations

- Small sample size (29–30 rows per group) yields wide CIs on purchases and CPA.

- One test-only date; robustness checked via aligned analysis.

- No revenue/AOV data, so monetary impact beyond CPA cannot be quantified.

7 Recommendation

- Treat current evidence as positive for conversion efficiency but neutral/negative on budget: the guardrail spend increase is significant, while purchases and CPA are inconclusive.

- If budget is constrained, classify as neutral/negative pending more data or cost control.

- If budget can flex, consider extending the experiment to reduce uncertainty on purchases and CPA, and add revenue/AOV to assess ROI and revenue per impression/reach.