Statistical Inference & A/B Testing
Graded Task
By Augustas Eskertas

A/B/C Test - Marketing Campaigns

Goal

The purpose of this A/B/C test was to identify which of the three marketing campaigns would most effectively **boost sales for the new product**.

Hypothesis Testing

Null Hypothesis (H_0)

There is no statistically significant difference between the means of sales revenue generated by the three marketing campaigns.

Pairwise comparison:

All of the following arguments must be met should the null hypothesis remain true.

- $H_0: \mu_A = \mu_B$
- $H_0: \mu_A = \mu_C$
- $H_0: \mu_B = \mu_C$

Alternative Hypothesis (H_1)

There is a statistically significant difference in sales revenue means generated by at least one of the three marketing campaigns.

To reject the null hypothesis, at least one of the arguments above must be proven false.

Confidence Level

A confidence level of **99% (α = 0.01)** will be used in this conducted test analysis to minimize the risk of a Type I error.

Target Metric

To assess revenue performance across all three marketing campaigns, **Sales Revenue** has been selected as the primary metric for this experiment. This metric will facilitate a comparative analysis of the monetary effectiveness of each campaign, thereby supporting the achievement of the experiment's objective.

Sales Revenue was extracted from the "sales_in_thousands" column in the dataset.

Population

The experiment's population included all locations across all of the markets of the fast-food chain.

Sample

- **Promotion 1** had a sample size of 43 locations (tested in 10 markets);
- **Promotion 2** had a sample size of 47 locations (tested in 9 markets);
- **Promotion 3** had a sample size of 47 locations (tested in 10 markets).

Duration

The test recorded results over a four-week period following the test launch.

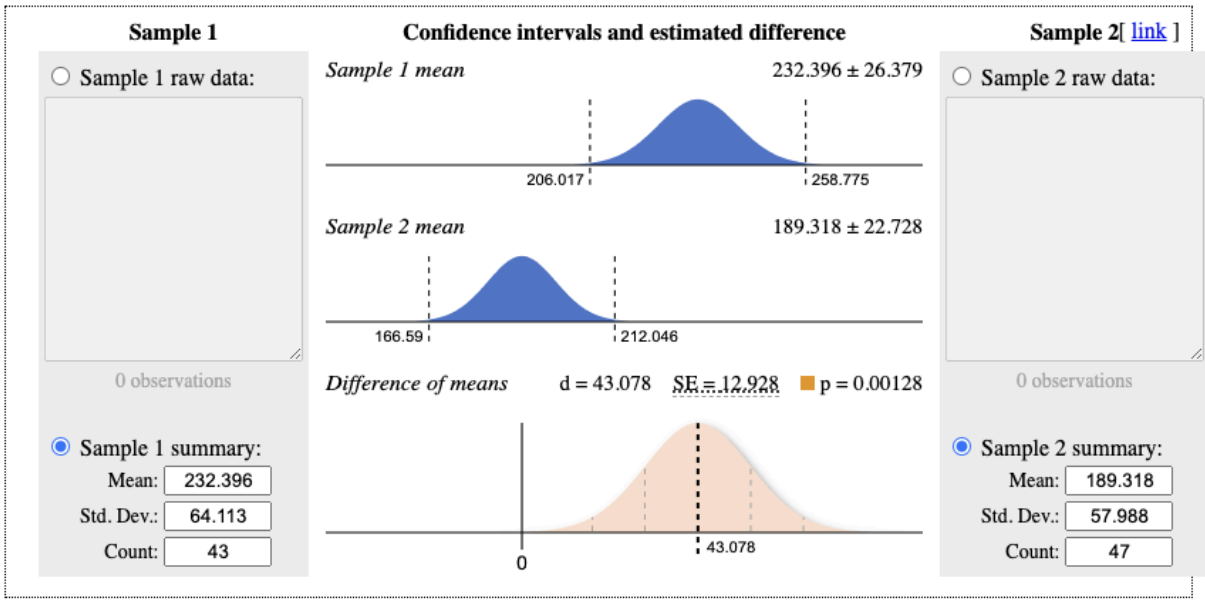
Calculations

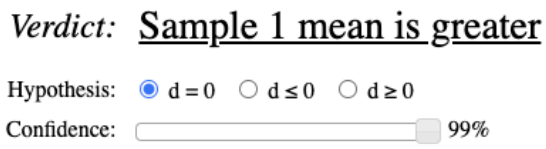
Firstly, I have gathered the necessary data for performing t-tests using the information extracted from BigQuery. The table below depicts the compiled results. The revenue data has been aggregated by promotion, including the sample size, mean, and sample standard deviation.

Row	promotion	sample_size	mean_revenue	std_deviation
1	1	43	232.396	64.113
2	2	47	189.318	57.988
3	3	47	221.458	65.535

$H_0: \mu_A = \mu_B$

Next, using the above values and a pairwise comparison approach, I conducted t-tests in sequence using the Evan Miller A/B Test Calculator to evaluate the three null hypotheses, beginning with the comparison between Campaign A (Sample 1) and Campaign B (Sample 2).

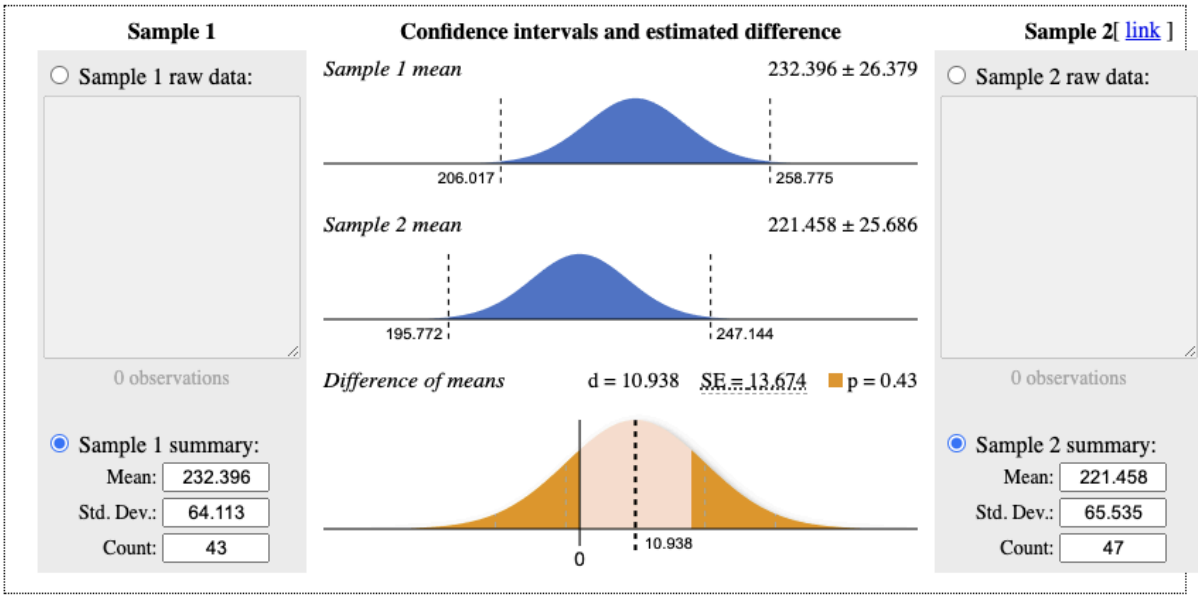




Since the computed p-value of 0.00128 is significantly lower than our significance level of 0.005, it indicates a statistically significant difference between these revenue means. Therefore, the null hypothesis for this portion of the test can be rejected. While we could conclude our analysis at this stage, I will proceed with examining the remaining comparisons.

$H_0: \mu_A = \mu_C$

Following are the calculations for comparing Campaign A (Sample 1) and Campaign C (Sample 2):



Verdict: No significant difference

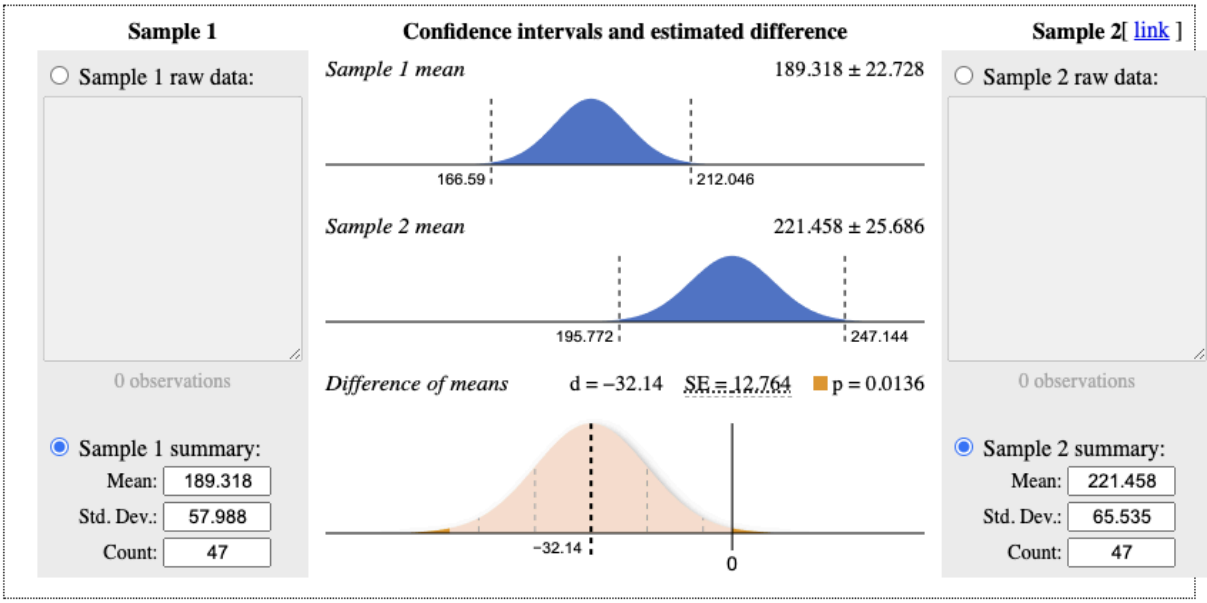
Hypothesis: d = 0 d ≤ 0 d ≥ 0

Confidence: 99%

The p-value of 0.43 exceeds the significance level, indicating that there is no statistically significant difference in this case and here the null hypothesis remains true.

$H_0: \mu_B = \mu_C$

Finally, the comparison between Campaign B (Sample 1) and Campaign C (Sample 2):



The resulting p-value of 0.0136 is above the significance level of 0.005, indicating that there is no statistically significant difference between the revenue means of Campaign B and Campaign C. The null hypothesis remains true..

Decision

Based on the results of the pairwise t-tests, one of the three null hypothesis arguments was rejected, indicating a statistically significant difference in average revenue between Campaign A and Campaign B. However, there were no statistically significant differences when comparing Campaign A to Campaign C, or Campaign B to Campaign C. Although Campaign A had the highest average revenue ($232.40) and showed a significant difference compared to the lowest performer, Campaign B ($189.32), it is premature to declare a clear winning campaign. More data is needed to determine whether Campaign A truly outperformed Campaign C ($221.46), or if Campaign C performed better than Campaign B.

Recommendations

A. The statistically significant difference between Campaign A and Campaign B suggests that Campaign A is performing better. The company could consider increasing investment in Campaign A, as it has a proven track record of generating higher revenue than Campaign B.

B. Although Campaign A did not show a statistically significant difference compared to Campaign C, the revenue means are quite close. With additional data, Campaign C might either outperform or match Campaign A, especially considering sample inequality and that Campaign A had the smallest sample size to begin with during the test. The company could test this further by reallocating some resources from Campaign B to Campaign C to explore its potential more thoroughly.

C. The p-value between Campaign B and Campaign C (0.0136) is very close to the significance level (0.01). This suggests there is likely a meaningful difference between these campaigns, but more data would be needed to confidently draw conclusions. The company should be cautious with Campaign B's performance, as it could be weaker compared to Campaign C.

D. The lack of a clear statistical winner between Campaign A and Campaign C, as well as the marginal significance between Campaign B and Campaign C, indicates that more data is required to make a final decision should the company wish to pick one campaign. However, they should continue gathering data on these campaigns to verify trends and gain better clarity, especially since Campaign C is showing promise and is very close to outperforming Campaign B, while Campaign A's smaller sample size could have skewed the results slightly. Controlled and equal sample sizes are needed for more reliable comparisons.

Appendix

- SQL query (submitted as a separate file on GitHub)
- Two-sample T-test calculator by Evan Miller: https://www.evanmiller.org/ab-testing/t-test.html