Fast-Food Promotions A/B Test Analysis

Hands-on Suggested Solution

This is an analysis of a fast-food campaign dataset, where three different promotions were tested across stores of varying sizes and ages.

Goal of the Test

The main goal of this A/B test is to determine which promotion leads to higher weekly sales. By comparing average sales across the three promotions, we aim to identify the best-performing strategy.

Target Metric

The dataset provides weekly sales in thousands of dollars.

- **Target metric:** Average weekly sales per store.
- This metric directly reflects promotion effectiveness and is suitable for comparison across groups.

Hypothesis Setup

For each pairwise comparison between promotions, we define the hypotheses as follows:

Null Hypothesis (H_0):

- There is no difference in average weekly sales between the two promotions.

$\mu 1 = \mu 2$

Alternative Hypothesis (H_1):

- The average weekly sales differ between the two promotions.

$\mu 1 \neq \mu 2$

Since we are comparing three promotions (Promo 1, Promo 2, Promo 3), we run **three pairwise tests**:

1. **Promo 1 vs Promo 2**
2. **Promo 1 vs Promo 3**
3. **Promo 2 vs Promo 3**

Because the analysis involves multiple pairwise comparisons, we use a **two-tailed t-test** with a **99% confidence level ($\alpha = 0.01$)** to reduce the risk of ***Type I errors*** (false positives).

Why Two-Tailed?

- A two-tailed test (μ1=μ2) is unbiased — it checks for any difference without any assumption in advance which promotion could be better.
- If the difference is significant, we then interpret the direction by comparing the group means.

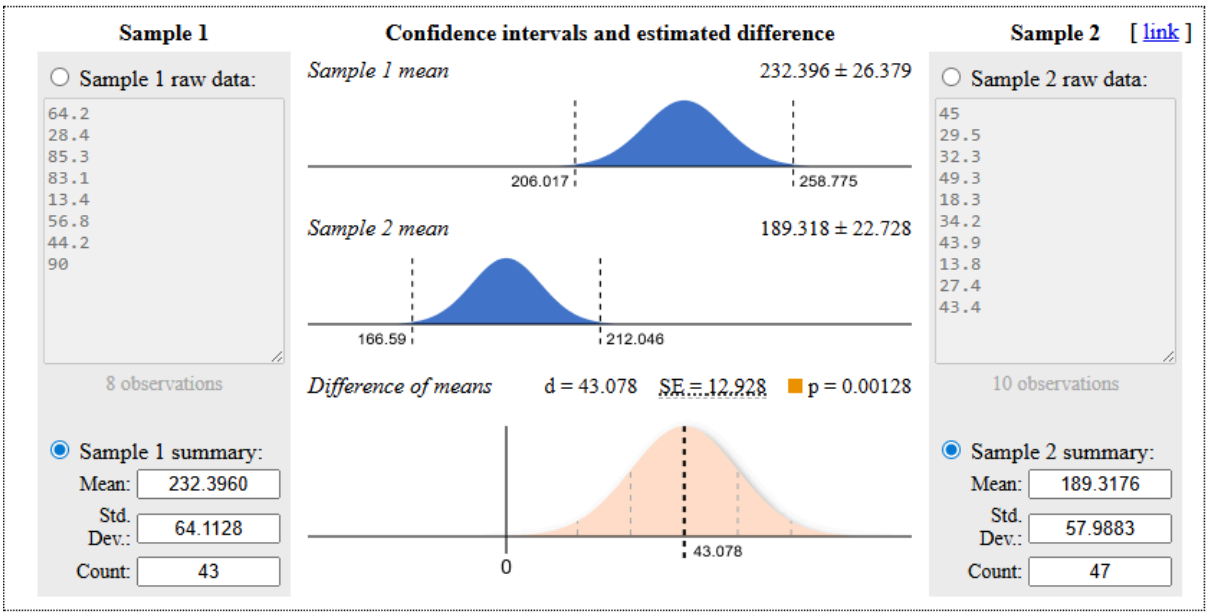
Why be unbiased?

- **It prevents you from assuming the winner before testing.**Starting unbiased ensures you don't design the analysis in a way that favors one promotion, which keeps the results credible.
- **It protects you from false conclusions**.If you assume a direction (e.g., "Promo 1 is better"), you increase the risk of interpreting random noise as a real effect.
- **It makes the test statistically valid and defensible.** A two-tailed, unbiased setup is the standard in A/B testing because it avoids directional bias and ensures the results can be trusted by stakeholders.

Calculations

The table below summarizes the results of the Two-sample T-tests between the three promotions. All calculations were done using:
https://www.evanmiller.org/ab-testing/t-test.html

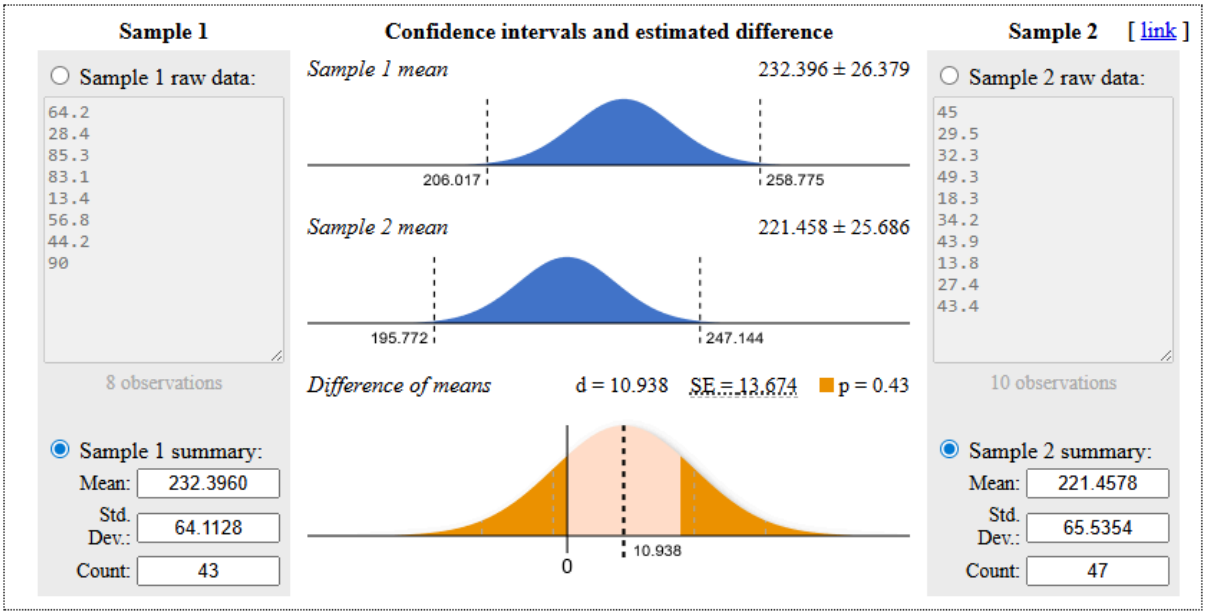
Promotion 1 VS Promotion 2:



Verdict: Sample 1 mean is greater

- **Mean difference:** +43.08
- **p-value:** 0.0013
- **99% CI:** [+26.4, +59.7]
- **Verdict:** Promo 1 significantly better

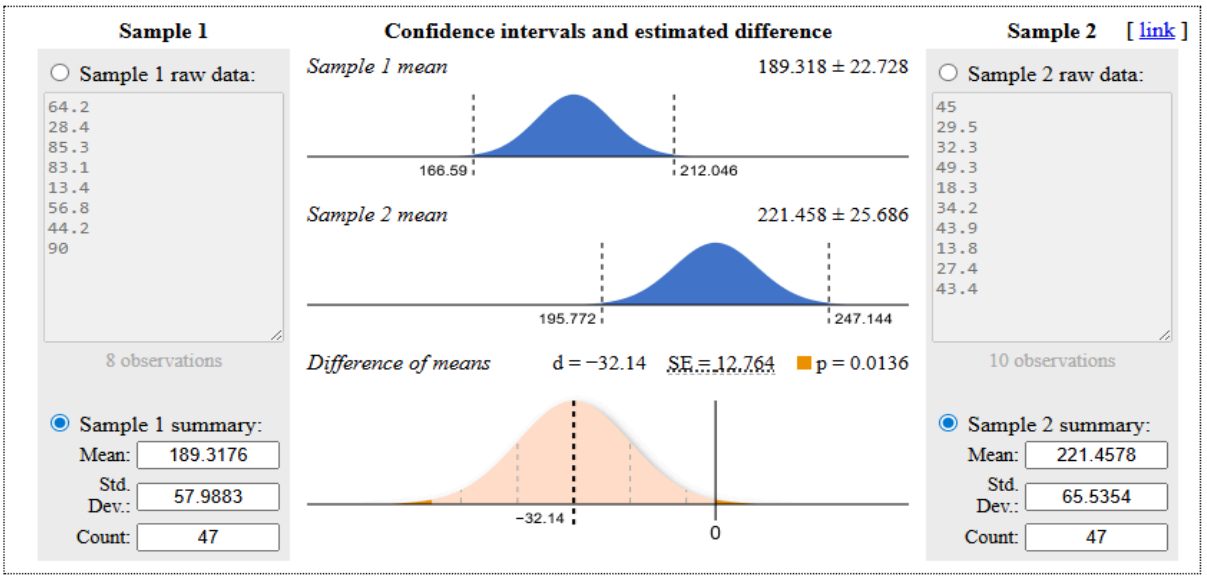
Promotion 1 VS Promotion 3:



Verdict: No significant difference

- **Mean difference:** +10.94
- **p-value:** 0.43
- **99% CI:** [−15.3, +37.2]
- **Verdict:** No significant difference

Promotion 2 VS Promotion 3:



Verdict: No significant difference

- **Mean difference:** −32.14
- **p-value:** 0.0135
- **99% CI:** [−57.8, −6.5]
- **Verdict:** No significant difference

Summary

Comparison	Mean Difference	p-value	99% CI for Difference	Verdict
Promo 1 vs Promo 2	+43.08	0.0013	[+26.4, +59.7]	Promo 1 significantly better
Promo 1 vs Promo 3	+10.94	0.43	[−15.3, +37.2]	No significant difference
Promo 2 vs Promo 3	−32.14	0.0135	[−57.8, −6.5] (approximate)	No significant difference

Decision

Based on the results of the test:

- **Promotion 2** clearly underperforms against **Promotion 1**.
- **Promotion 1 and Promotion 3 are statistically tied overall**, with Promo 3 showing a slight edge in large markets. (see "**Market analysis**" in appendix)
- **Promotion 3 likely beats Promotion 2, but due to strictness of this test, additional testing needed to confirm.**
- **Recommendation**: **Use Promotion 1 or Promotion 3**, depending on strategic priorities (e.g., focus on large markets vs broad rollout). Avoid Promotion 2.

- ***Statistical note: Promo 1's*** *mean exceeds* ***Promo 3*** *by 2.74 units, but with* ***p = 0.12*** *at* ***α = 0.01*** *the evidence is too weak to claim that Promotion 1 is truly better. Promotion 1 has a smaller sample size, thus it has reduced statistical power.* ***Additional data could clarify*** *whether this difference persists.*

Appendix

Query for T-Test:

1. Goal

Determine which of the three **Promotions (1, 2, 3)** leads to the highest **average weekly sales per location**.

2. Target Metric

Since we don't have customer counts, use:

Average Weekly Sales = Total Sales In Thousands / Number of Weeks

Per LocationID per Promotion.

3. SQL Aggregation

```sql
WITH aggregated AS (
  SELECT
    location_id,
    promotion,
    SUM(sales_in_thousands) AS total_sales
  FROM `tc-da-1.turing_data_analytics.wa_marketing_campaign`
  GROUP BY location_id, promotion
)
SELECT
  promotion,
  COUNT(*) AS total_locations,
  SUM(total_sales) AS total_sales,
  AVG(total_sales) AS mean_sales_per_location,
  STDDEV(total_sales) AS stddev_sales
FROM aggregated
GROUP BY promotion
ORDER BY promotion;
```

4. MarketAnalysis

To check whether performance varies by context, we compared **Promotion 1 vs Promotion 3** across market sizes:

Market Size	Promo 1 Avg	Promo 3 Avg	Direction
Large	75.24	77.20	Promo 3 slightly higher
Medium	47.67	45.47	Promo 1 slightly higher
Small	60.16	59.51	Promo 1 slightly higher

Differences are small and not statistically confirmed. Overall, Promo 1 and Promo 3 perform similarly across segments, with Promo 3 slightly ahead in large markets.