# **Title: Fast Food Marketing Campaign A/B Test (ANOVA and Pairwise- 2 sample T-test & Bonferroni)**

**Scenario**

A fast-food chain is preparing to introduce a new menu item but has yet to decide which of three potential marketing campaigns to use for promotion. The new item is launched in multiple locations across randomly selected markets to identify which campaign will drive the most sales. Each location implements a different campaign, and weekly sales of the new item are tracked for the first four weeks.

**Goal**

Evaluate A/B testing results and decide which marketing strategy works the best.

**Note:**

The dataset is aggregated by LocationID, PromotionID and week. However, I aggregated by LocationID and PromotionID before conducting the statistical tests.

Aggregating by **LocationID** and **PromotionID** before conducting statistical tests simplifies the data, reducing noise from weekly variations. This helps focus the analysis on differences between promotions across locations, leading to clearer insights and more reliable comparisons.

# **Experiment Design**

## 

## **Task Variables**

The variables provided in the dataset include:

* MarketID: unique identifier for market
* MarketSize: the size of the market area by sales
* LocationID: unique identifier for store location
* AgeOfStore: age of store in years
* Promotion: one of three promotions that were tested
* Week: one of four weeks when the promotions were run
* SalesInThousands: sales amount for a specific LocationID, Promotion, and week

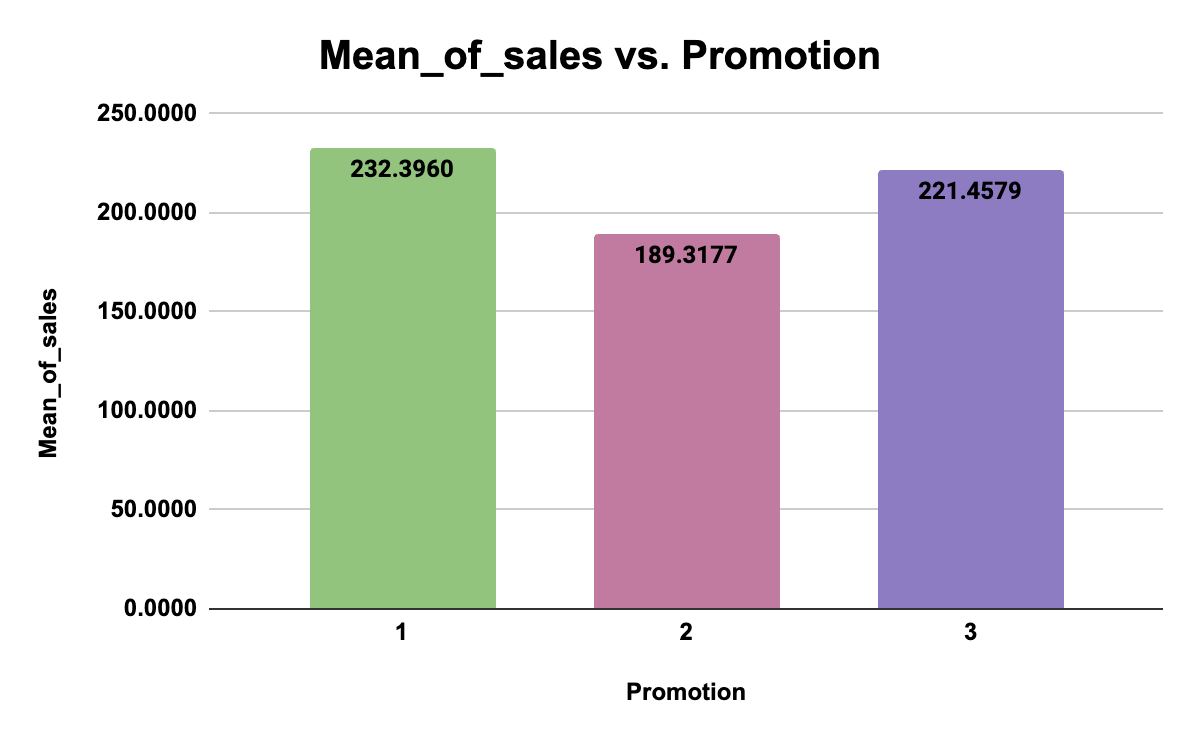
## **Metrics Choice**

I used BigQuery to run queries and generate key metrics required for this analysis.

Check the spreadsheet link attached to my submission or the image below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Promotion** | **Location\_count** | **Mean\_of\_sales** | **Std\_dev** | **Min\_value** | **Max\_value** |
| **1** | 43 | **232.3960** | 64.1129 | 151.14 | 380.36 |
| **2** | 47 | **189.3177** | 57.9884 | 111.36 | 332.63 |
| **3** | 47 | **221.4579** | 65.5355 | 124.74 | 354.31 |

**Interpretation of metrics (with visualization)**



The chart shows that promotion ‘1’ performed the best overall, with the highest average sales.

To gain deeper insights from the A/B testing results and verify the best marketing strategy, I carried out a various test:

**Test 1: ANOVA (Analysis of Variance)**

To carry out this test, I used the XLMiner Analysis ToolPak Add-on in Google Sheets/Microsoft Excel. The ANOVA test determined if there are statistically significant differences between the effects of promotions on sales. ANOVA is ideal for comparing multiple groups and can be a preliminary test before further pairwise comparisons.

**Hypotheses**

1. H0: M1 = M2 = M3 There is no difference between promotion averages
2. H1: M1 = M2 = M3 At least one of the promotion averages is different

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **ANOVA** | | | | | | |
| ***Source of Variation*** | ***SS*** | ***df*** | ***MS*** | ***F*** | ***P-value*** | ***F crit*** |
| Between Groups | 45796.69716 | 2 | 22898.34858 | 5.845791932 | 0.0037 | 4.767125046 |
| Within Groups | 524886.7469 | 134 | 3917.065275 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 570683.444 | 136 |  |  |  |  |

From the ANOVA analysis: **p = 0.0037**

The p-value is lower than the α value of 0.01, which indicates a highly significant result. This result suggests that there are significant differences between the sales means for the different promotions (Promotion 1, Promotion 2, and Promotion 3). In other words, **at least one promotion is performing significantly differently from the others** in terms of its effect on sales.

**Result:** Since the p-value (0.0037) is **less** than the significance level (α = 0.01), we **reject the null hypothesis (H₀)**.

**Test 2: Pairwise Comparisons and Bonferroni Correction.**

**Pairwise comparisons** are statistical analyses used to compare the means of different groups two at a time. These comparisons help identify which specific pairs of group means are significantly different from each other when conducting tests like ANOVA. I conducted this test alongside the Bonferroni correction.

The Bonferroni correction adjusts the significance threshold to account for the increased risk of Type I errors (false positives) when performing multiple tests. This helps to ensure that the probability of incorrectly rejecting a null hypothesis remains controlled.

**Using pairwise comparisons along with Bonferroni correction** ensures that one can examine all possible differences while also adjusting for the increased risk of false positives due to multiple tests.

I used the data above to carry out my t-tests, using [Evan Miller’s **2 Sample T-Test** calculator.](https://www.evanmiller.org/ab-testing/t-test.html)

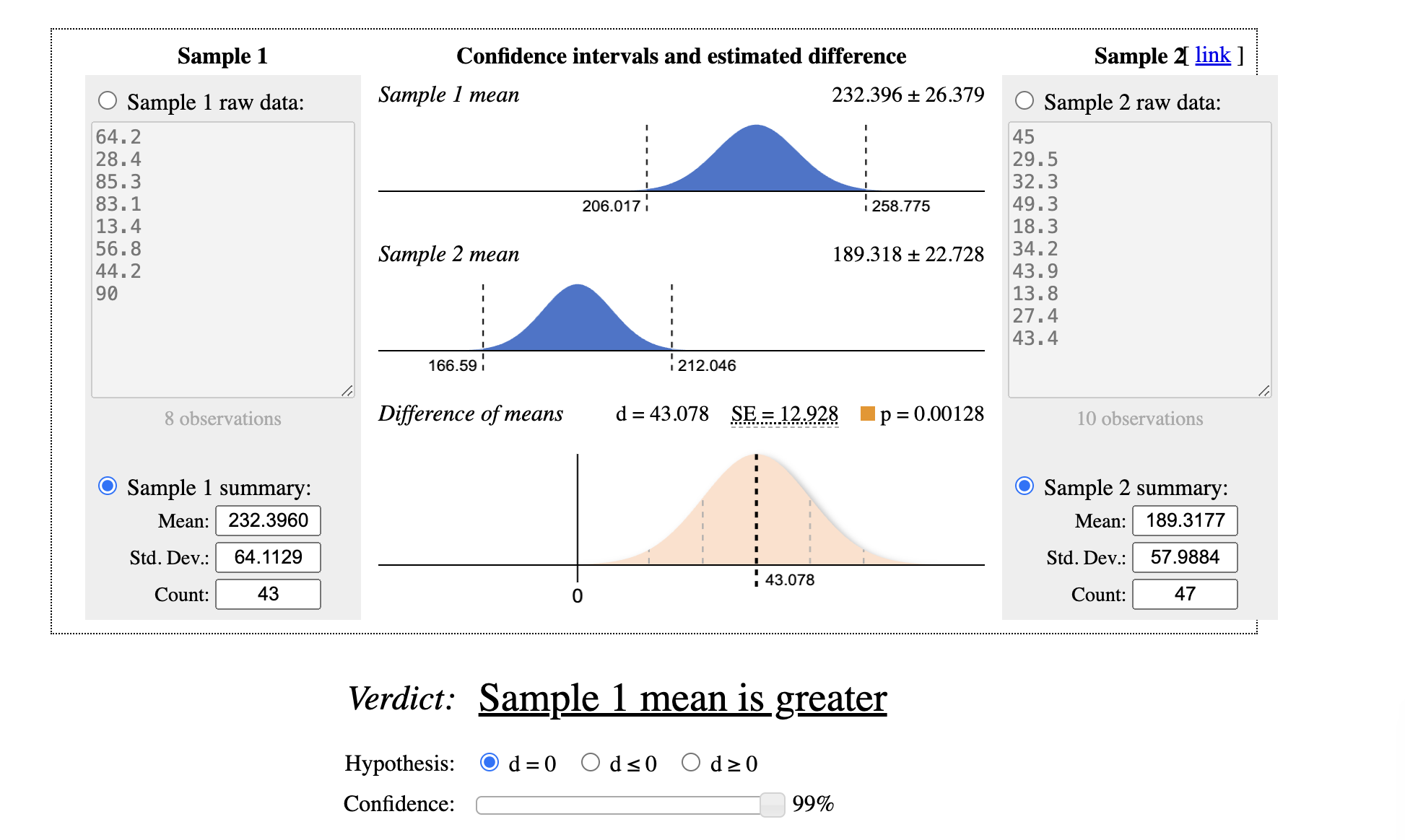
### **Bonferroni-corrected significance level:**

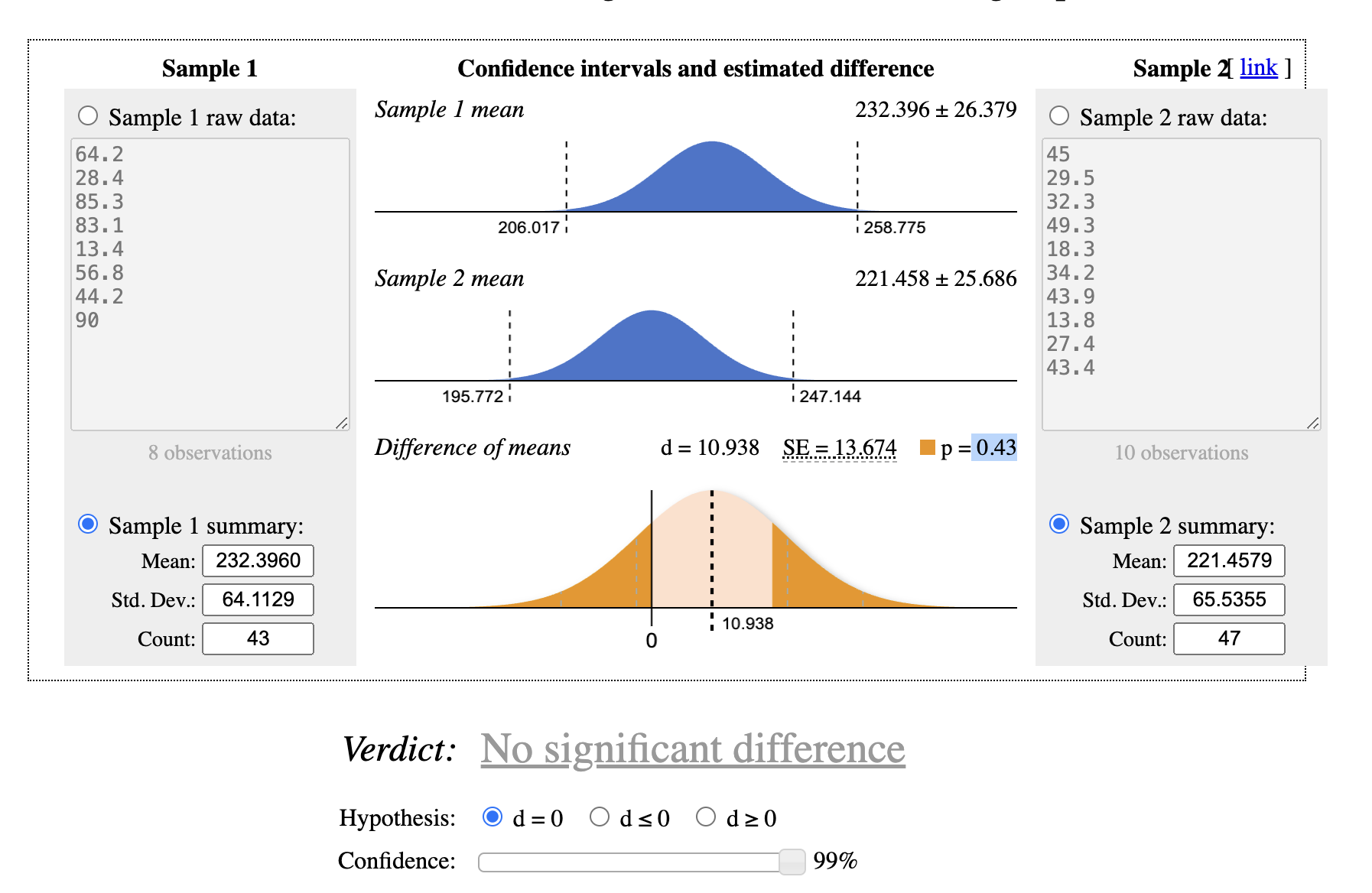
* The usual significance level (α) is 0.01.
* Since there 3 comparisons, I divided the significance level by the number of tests (3):
  + Corrected significance level = α/number of tests
  + α=0.01
  + Corrected α=0.01/3 ≈ 0.0033

### **Hypothesis**

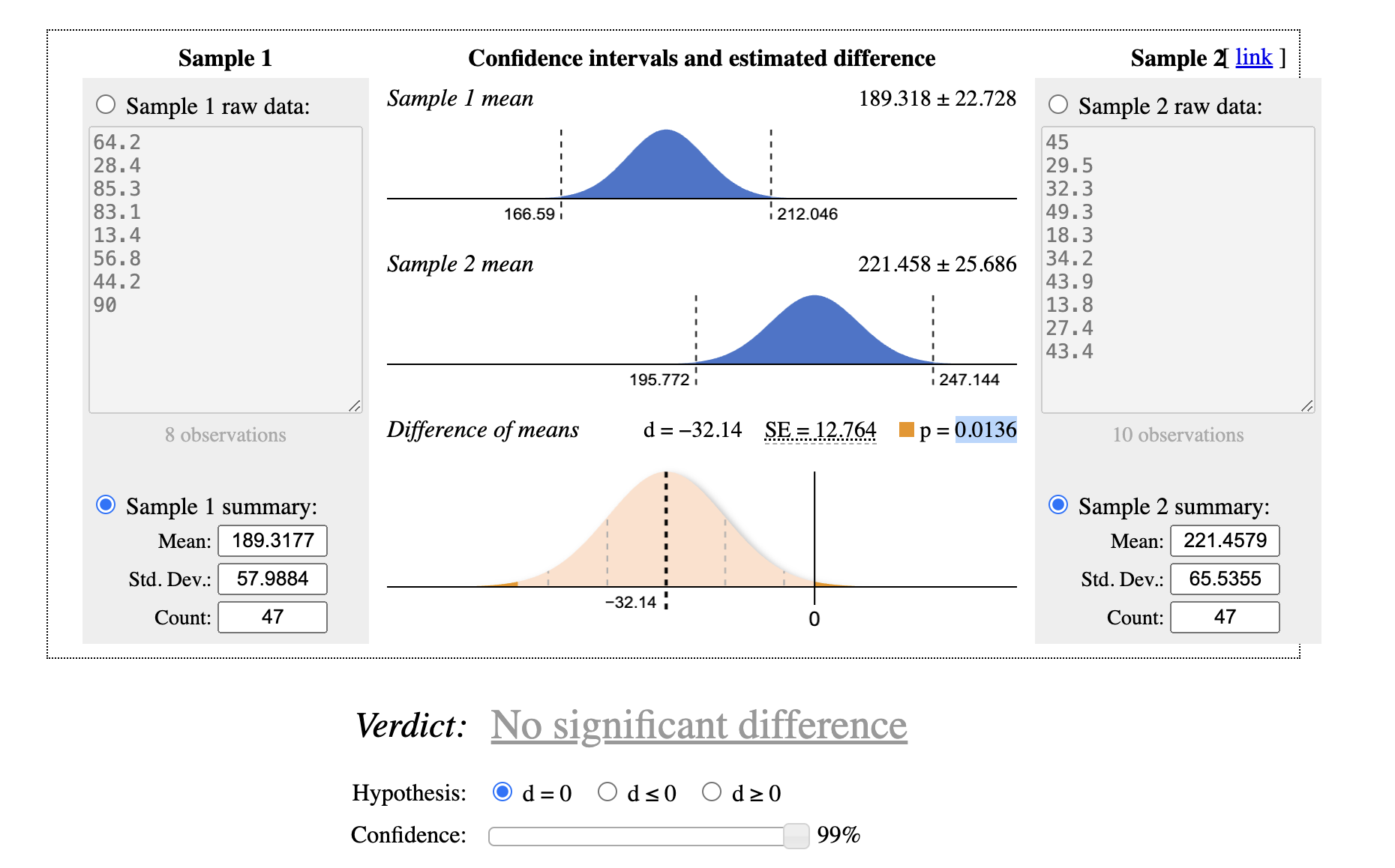
* Null hypothesis (H0): The retention rate is the same for both groups (no difference). I.e H0 ​: p1​=p2​
* Alternative hypothesis (Ha​​): The retention rates are different between the two groups. i.e Ha​ : p1≠p2

**Promotion 1 vs 2**

****

**Promotion 1 vs 3**

**Promotion 2 vs 3**



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Pairwise comparison (α = 0.01) and Bonferroni correction (α = 0.0033)** | | | | |
| **Compared metrics** | **Mean Difference (d)** | **Standard Error** | **P-value** | **Statistical significant?** |
| 1 vs 2 | 43.078 | 12.9280 | 0.0013 | YES |
| 1 vs 3 | 10.938 | 13.6740 | 0.4300 | NO |
| 2 vs 3 | -32.14 | 12.7640 | 0.0136 | NO |

**Result:**

Promotion 1 vs 2:

Since the p-value (0.0013) is **less** than the significance level (α = 0.0033), we **reject the null hypothesis (H₀)**.

**Overview of sales by market size**

****

**Overall Insight:**

From the analysis, it is evident that the performance of the three marketing promotions varied significantly across the different locations. Promotion 1 had the highest mean sales, indicating stronger performance compared to the other two promotions. However, a deeper statistical examination reveals additional insights into the significance of these differences.

* **Promotion 1** had the highest average sales and standard deviation, suggesting both higher sales and more variability across locations.
* **Promotion 2** had the lowest mean sales, and its comparison with Promotion 3 shows a statistically significant difference.

### **ANOVA Results:**

The **ANOVA test** resulted in a p-value of **0.0037**, which is below the threshold (α = 0.01). This indicates a statistically significant difference between at least one pair of promotions. In other words, not all promotions perform equally in driving sales.

### **Pairwise Comparison & Bonferroni Correction:**

The **Pairwise Comparison** and **Bonferroni Correction** reveal that:

* **Promotion 1 vs Promotion 2** is the only statistically significant pair under the more stringent Bonferroni correction. This reinforces the finding that Promotion 1 is significantly more effective than Promotion 2.
* There was no significant difference between Promotion 1 and Promotion 3, or between Promotion 2 and Promotion 3, under this correction.

### **Conclusion:**

* **Promotion 1** is the most effective marketing strategy based on pairwise comparisons, significantly outperforming Promotion 2. However, it performs similarly to Promotion 3.
* **Promotion 2** is the weakest performer, as it was significantly outperformed by both Promotion 1 and Promotion 3.
* **Market Size** plays a critical role in the effectiveness of promotions, with larger markets driving higher sales.

### **Summary:**

* **Promotion 1** generated the highest sales and was significantly better than **Promotion 2**, while **Promotion 3** performed similarly to Promotion 1.
* **Market size** is a key factor in sales performance; **location-specific factors** also impact outcomes.

### **Recommendations:**

1. **Adoption76 of Promotion 1** for the broader rollout, as it has consistently higher performance compared to Promotion 2 and performs similarly to Promotion 3.
2. Investigation of underperforming **locations** (higher location IDs) to determine if specific challenges affect sales at these sites.
3. Focusing marketing efforts on **larger markets** where sales are significantly higher, and consider tailored strategies for **smaller markets**.