Title: Fast Food Marketing Campaign A/B Test (ANOVA, Tukey HSD, Pairwise- 2 sample T-test & Bonferroni, and Linear regression)

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 this [spreadsheet](#) or the image below.

Promotion	location_count	mean_of_sales	stddev_of_sales	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 visualisation)
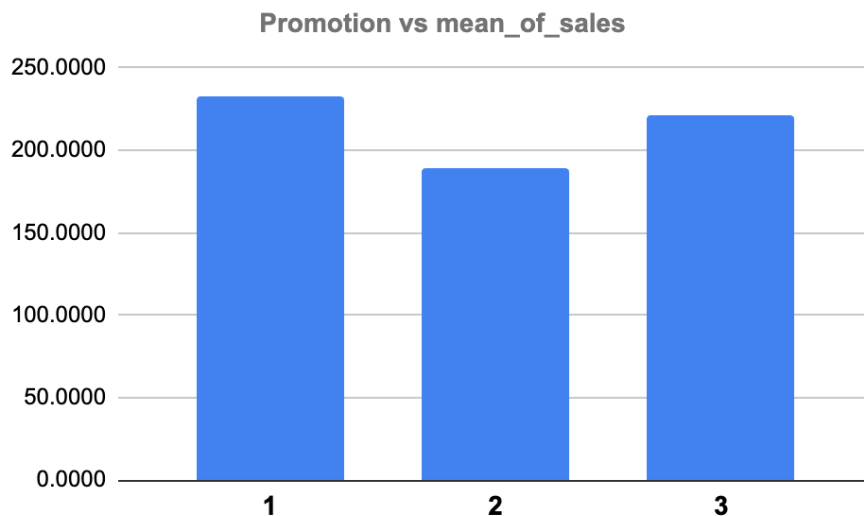
The chart shows that promotion '1' performed the best overall, with the highest average sales.

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. 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: Tukey's HSD (Post-Hoc Test)

Tukey's HSD is specifically designed for **post-hoc tests** after an ANOVA. It compares all pairs of group means and adjusts for the number of comparisons, making it a better choice when one has **three or more groups** and needs to know which pairs are significantly different.
Tukey's HSD uses the **Standardised range distribution** to calculate critical values, which are designed for multiple comparisons.

I used this external calculator to carry out the test.

Compared metrics	Tukey HSD Q statistic	Tukey HSD p-value	Tukey HSD inferfence
1 vs 2	4.6127	0.0010053	** p<0.01
1 vs 3	1.1712	0.6711087	insignificant
2 vs 3	3.5206	0.0370968	insignificant

Result Interpretation

Promotion 1 vs Promotion 2: Promotion 1 is significantly better than Promotion 2 in terms of sales.

Promotion 1 vs Promotion 3: There is no significant difference between Promotion 1 and Promotion 3.

Promotion 2 vs Promotion 3: There is no significant difference between Promotion 2 and Promotion 3.

Test 3: 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.

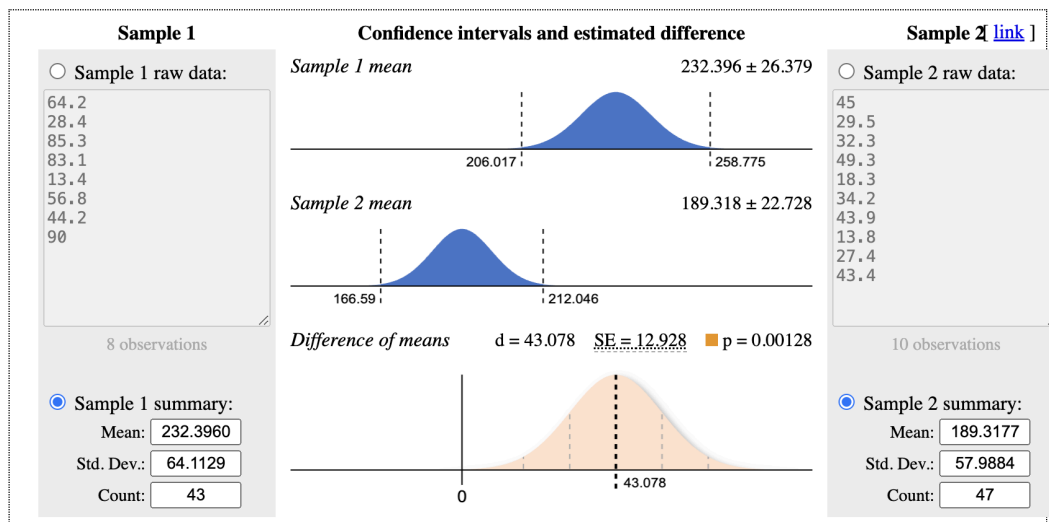I used the data above to carry out my t-tests, using Evan Miller's **2 Sample T-Test** calculator.

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 (H_0): The retention rate is the same for both groups (no difference). I.e

$$H_0 : p1=p2$$

- Alternative hypothesis (H_a): The retention rates are different between the two groups. i.e
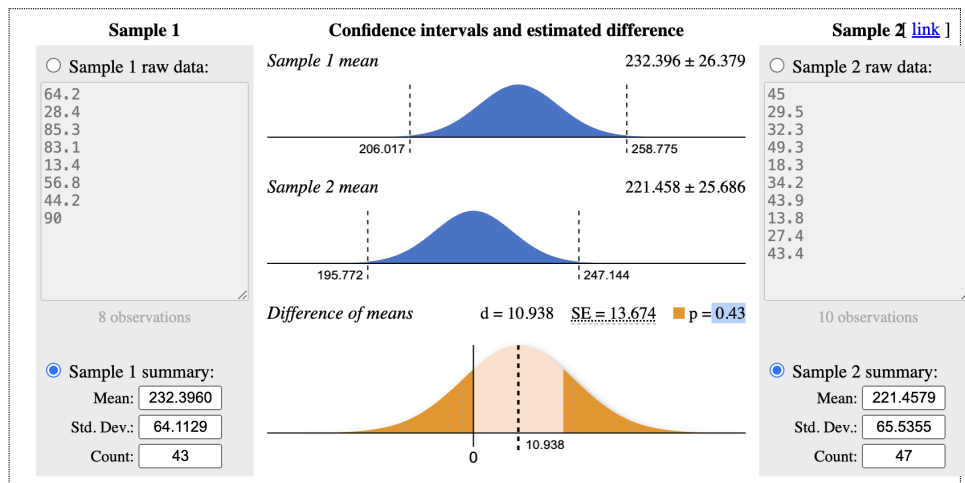
$$H_a : p1 \neq p2$$

Promotion 1 vs 2

Sample 1	Confidence intervals and estimated difference	Sample 2 [link]
○ Sample 1 raw data:	*Sample 1 mean* 232.396 ± 26.379	○ Sample 2 raw data:

Sample 1 raw data:
```
64.2
28.4
85.3
83.1
13.4
56.8
44.2
90
```
8 observations

Sample 2 raw data:
```
45
29.5
32.3
49.3
18.3
34.2
43.9
13.8
27.4
43.4
```
10 observations

Sample 1 mean 232.396 ± 26.379   206.017   258.775

Sample 2 mean 189.318 ± 22.728   166.59   212.046

Difference of means   d = 43.078   SE = 12.928   ■ p = 0.00128   0   43.078

◉ Sample 1 summary:
Mean: 232.3960
Std. Dev.: 64.1129
Count: 43

◉ Sample 2 summary:
Mean: 189.3177
Std. Dev.: 57.9884
Count: 47

Verdict: <u>Sample 1 mean is greater</u>

Hypothesis: ◉ d = 0   ○ d ≤ 0   ○ d ≥ 0

Confidence: 99%

Promotion 1 vs 3

Sample 1	Confidence intervals and estimated difference	Sample 2 [link]

Sample 1

○ Sample 1 raw data:

```
64.2
28.4
85.3
83.1
13.4
56.8
44.2
90
```

8 observations

◉ Sample 1 summary:

Mean: 232.3960
Std. Dev.: 64.1129
Count: 43

Confidence intervals and estimated difference

Sample 1 mean   232.396 ± 26.379

206.017   258.775

Sample 2 mean   221.458 ± 25.686

195.772   247.144

Difference of means   d = 10.938   SE = 13.674   ■ p = 0.43

0   10.938

Sample 2 [link]

○ Sample 2 raw data:

```
45
29.5
32.3
49.3
18.3
34.2
43.9
13.8
27.4
43.4
```

10 observations

◉ Sample 2 summary:

Mean: 221.4579
Std. Dev.: 65.5355
Count: 47

Verdict: No significant difference

Hypothesis: ◉ d = 0   ○ d ≤ 0   ○ d ≥ 0
Confidence: 99%

Promotion 2 vs 3

Sample 1

○ Sample 1 raw data:

```
64.2
28.4
85.3
83.1
13.4
56.8
44.2
90
```

8 observations

◉ Sample 1 summary:

Mean: 189.3177
Std. Dev.: 57.9884
Count: 47

Confidence intervals and estimated difference

Sample 1 mean   189.318 ± 22.728

166.59   212.046

Sample 2 mean   221.458 ± 25.686

195.772   247.144

Difference of means   d = −32.14   SE = 12.764   ■ p = 0.0136

−32.14   0

Sample 2 [link]

○ Sample 2 raw data:

```
45
29.5
32.3
49.3
18.3
34.2
43.9
13.8
27.4
43.4
```

10 observations

◉ Sample 2 summary:

Mean: 221.4579
Std. Dev.: 65.5355
Count: 47

Verdict: No significant difference

Hypothesis: ◉ d = 0   ○ d ≤ 0   ○ d ≥ 0
Confidence: 99%

Pairwise comparison (α = 0.01) and Bonferroni correction (α = 0.0033)				
Compared metrics	mean difference (d)	standard error	p-value	statistical sigificant?
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 (H_0)**.

Test 4: Linear regression (Additional analysis, not required by Turing).

I used the linear regression model to assess the impact of the other different variables (e.g., `market_size`, `age_of_store`, and `week`) along with the `location` on sales.

I used the XLMiner Analysis ToolPak Add-on in GoogleSheet

Dependent Variable: `sales` (the outcome variable)

Independent Variables: `market_size`, `age_of_store`, `week`, and `location_id` (the predictors or features).

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 99%	Upper 99%
Intercept	29.1719	2.8518	10.2291	0.0000	23.5699	34.7739	21.8001	36.5436
market_size_num	15.0583	1.0196	14.7681	0.0000	13.0553	17.0612	12.4226	17.6940
location_id	-0.0195	0.0021	-9.0650	0.0000	-0.0237	-0.0153	-0.0250	-0.0139
age_of_store	0.1140	0.0910	1.2521	0.2111	-0.0648	0.2928	-0.1213	0.3493
week	-0.1645	0.5326	-0.3088	0.7576	-1.2107	0.8818	-1.5412	1.2123

Coefficients and Interpretations:

1. **Intercept (29.17)**:
   ○ This represents the baseline sales when all the independent variables are zero. If all variables are set to zero, the predicted sales would be 29.17 units. This is the base sales level.
2. **Market Size (15.06, p = 0.000)**:
   ○ The coefficient is 15.06, meaning for every increase in market size (from small to medium or medium to large), sales increase by about 15 units.
   ○ The p-value is highly significant (p = 0.000), indicating that **market size** is a strong predictor of sales.
3. **Location ID (-0.0195, p = 0.000)**:
   ○ The coefficient is negative, suggesting that as the **location ID** increases, sales decrease slightly. For each unit increase in location ID, sales drop by 0.0195 units.
   ○ While the effect size is small, the p-value is highly significant (p = 0.000), indicating that **location ID** plays a statistically significant role in predicting sales. This could imply that certain locations (with higher IDs) perform worse in sales, which may warrant further investigation into location-based factors.
4. **Age of Store (0.114, p = 0.211)**:
   ○ The coefficient is positive, suggesting that for each additional unit increase in the **age of the store**, sales increase by 0.114 units. However, the p-value (p = 0.211) shows this effect is not statistically significant, meaning the **age of the store** does not strongly influence sales in this model.
5. **Week (-0.164, p = 0.758)**:
   ○ The coefficient for the **week** is slightly negative, meaning sales decrease by 0.164 units as the week progresses. However, the p-value (p = 0.758) indicates that this is not statistically significant. Therefore, the time factor (weeks) does not have a notable impact on sales in this analysis.

Overview of sales by market size

market_size vs mean_of_sales
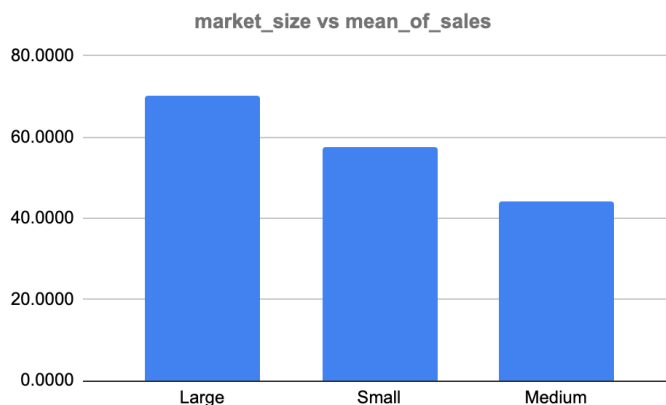
Overall Insight:

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 ($\alpha = 0.01$). This indicates a statistically significant difference between at least one pair of promotions. In other words, not all promotions perform equally in driving sales.

Post-Hoc Tukey's HSD Test:

The **Tukey's HSD Test** further identifies the specific pairs of promotions that have significant differences:

- **Promotion 1 vs Promotion 2**: Statistically significant difference (**p = 0.001**), indicating that Promotion 1 significantly outperformed Promotion 2.
- **Promotion 2 vs Promotion 3**: : No significant difference (**p = 0.671**), suggesting that both performed similarly.
- **Promotion 1 vs Promotion 3**: No significant difference (**p = 0.037**), suggesting that both performed similarly.

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.

Linear Regression Results:

- **Market Size** is the most significant predictor of sales, with larger markets consistently driving more sales. The positive coefficient (15.06) indicates that sales increase by about 15 units for each increase in market size level.
- **Location ID** has a small but statistically significant negative effect on sales, meaning certain locations (with higher IDs) might be underperforming.
- **Age of Store** and **Week** were not significant predictors of sales, suggesting they do not play a critical role in sales performance.

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, and **location-specific factors** also impact outcomes.
- **The age of the store** and **time (week)** do not significantly affect sales.

Recommendations:

1. **Adoption of Promotion 1** for the broader rollout, as it has consistently higher performance compared to Promotion 2 and performs similarly to Promotion 3.
2. Investigatigation of underperforming **locations** (higher location IDs) to determine if there are specific challenges affecting sales at these sites.
3. Focusing marketing efforts on **larger markets** where sales are significantly higher, and consider tailored strategies for **smaller markets**.
4. While the **age of the store** and **time (week)** are not significant in this analysis, there should be continuous monitoring of these variables to ensure long-term consistency in performance.

Appendix

SQL QUERY

```sql
--Key data extraction


WITH

 sales_data AS (

 SELECT

    promotion,

    location_id,

    ROUND(SUM(sales_in_thousands),4) AS sales

  FROM

    `tc-da-1.turing_data_analytics.wa_marketing_campaign`

  GROUP BY

    promotion,

    location_id)

SELECT

 promotion,

 COUNT(*) AS location_count,

 ROUND(AVG(sales),4) AS mean_of_sales,

 ROUND(STDDEV(sales),4) AS std_dev,

 MIN(sales) AS min_value,

 MAX(sales) AS max_value
```

```sql
FROM

  sales_data

GROUP BY

  promotion;




  --Sales by promotion

WITH

  sales_data AS (

  SELECT

    promotion,

    location_id,

    ROUND(SUM(sales_in_thousands),4) AS sales

  FROM

    `tc-da-1.turing_data_analytics.wa_marketing_campaign`

  GROUP BY

    Promotion,

    location_id)

SELECT

  promotion,

  AVG(sales) AS mean_of_sales

FROM
```

```sql
  sales_data

GROUP BY

  promotion,

  sales

ORDER BY

  promotion;




  --Regression analysis

SELECT

  promotion,

  market_size,

  location_id,

  age_of_store,

  week,

  sales_in_thousands AS sales

FROM

  `tc-da-1.turing_data_analytics.wa_marketing_campaign`

GROUP BY

  promotion,

  market_size,
```

```sql
    location_id,

    age_of_store,

    week,

    sales

ORDER BY

  promotion;



  --Sales by market size

WITH

  sales_data AS (

  SELECT

      Promotion,

      location_id,

      market_size,

      sales_in_thousands

  FROM

      `tc-da-1.turing_data_analytics.wa_marketing_campaign`

  GROUP BY

      Promotion,

      location_id,

      market_size,

      sales_in_thousands)
```

```sql
SELECT
  market_size,
  ROUND(AVG(sales_in_thousands),4) AS sales
FROM
  sales_data
GROUP BY
  market_size;
```