Statistical Analysis: Long-term Effectiveness of the Educational Strategy "I Prefer Plain Water"

Carlos Polanco

November 22, 2025

Executive Summary

This document complements the main manuscript with the complete statistical analysis, mathematical formulas, and numerical results of the study on the effectiveness of the "I Prefer Plain Water" strategy.

1 Sample Size Calculation

The sample size was calculated using the formula for comparison of means:

$$n = \frac{(Z_{1-\alpha/2} + Z_{1-\beta})^2 \times 2\sigma^2}{\Delta^2} \tag{1}$$

Where:

- $Z_{1-\alpha/2} = 1.96$ (type I error of 5%)
- $Z_{1-\beta} = 1.28$ (90% power)
- $\sigma = 399$ ml (standard deviation)
- $\Delta = 80.7$ ml (minimum detectable difference)

Substituting the values:

$$n = \frac{(1.96 + 1.28)^2 \times 2 \times (399)^2}{(80.7)^2} = 257 \text{ participants} \tag{2}$$

2 Beverage Consumption Analysis

2.1 Baseline Consumption Pattern

Baseline consumption results are presented in Table 1:

Table 1: Baseline Beverage Consumption by Group (ml/day)

Beverage	Group 1	Group 2	Group 3	Total
Plain water	485	512	503	500
Sugar-sweetened beverages	215	201	208	208
Drinkable dairy	340	325	334	333
Other beverages	45	52	48	48

2.2 Repeated Measures ANOVA Analysis

The repeated measures ANOVA model is specified as:

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij} \tag{3}$$

Where:

- Y_{ij}: Water consumption in group i at time j
- μ: Overall mean
- α_i: Effect of group i
- β_j: Effect of time j
- $(\alpha\beta)_{ij}$: Group-time interaction effect
- ϵ_{ij}: Random error

3 Main Results

3.1 Plain Water Consumption

Changes in plain water consumption are shown in Table 2:

Table 2: Changes in Plain Water Consumption (ml/day)

Group	T1 (Baseline)	T2 (Post)	T3 (1-year)
Group 1	485	625	510
Group 2	512	698	545
Group 3	503	712	685

Repeated measures ANOVA showed:

$$F_{\text{time}} = 15.32, \quad p < 0.001 \tag{4}$$

$$F_{\text{group} \times \text{time}} = 4.87, \quad p = 0.008 \tag{5}$$

$$F_{\text{group}} = 2.14, \quad p = 0.118 \tag{6}$$

3.2 Sugar-Sweetened Beverage Consumption

Changes in sugar-sweetened beverages are presented in Table 3:

Table 3: Changes in Sugar-Sweetened Beverage Consumption (ml/day)

Group	T1 (Baseline)	T2 (Post)	T3 (1-year)
Group 1	215	180	205
Group 2	201	165	190
Group 3	208	158	145

4 Post-hoc Analysis and Multiple Comparisons

Bonferroni correction was used for multiple comparisons:

$$p_{\text{adjusted}} = p_{\text{original}} \times k \tag{7}$$

Where k is the number of comparisons.
Significant differences were found in:

- Group 3 vs Group 1: $p = 0.003$
- Group 3 vs Group 2: $p = 0.012$
- T3 vs T1 in Group 3: $p < 0.001$

5 Analysis of Covariance (ANCOVA)

ANCOVA was performed adjusting for baseline variables:

$$Y_{\text{post}} = \beta_0 + \beta_1 X_{\text{group}} + \beta_2 Y_{\text{baseline}} + \epsilon \tag{8}$$

Results showed:

$$\beta_{\text{Group3}} = 132.5 \text{ ml}, \quad p < 0.001 \tag{9}$$

$$R^2 = 0.342 \tag{10}$$

6 Non-Parametric Tests

For non-normal variables, the Friedman test was used:

$$\chi_F^2 = \frac{12}{Nk(k+1)} \sum_{j=1}^{k} R_j^2 - 3N(k+1) \tag{11}$$

Where:

- N: number of subjects
- k: number of time points
- R_j: sum of ranks for time j

Friedman test results:

$$\chi^2 = 18.65 \tag{12}$$

$$p = 0.0003 \tag{13}$$

7 Effect Size Analysis

Effect size (Cohen's d) was calculated:

$$d = \frac{\bar{X}_1 - \bar{X}_2}{s_{\text{pooled}}} \tag{14}$$

$$s_{\text{pooled}} = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \tag{15}$$

Effect sizes obtained:

- Group 3 (T3 vs T1): $d = 0.82$ (large effect)
- Group 2 (T3 vs T1): $d = 0.35$ (small effect)
- Group 1 (T3 vs T1): $d = 0.18$ (very small effect)

Statistical Conclusion

The statistical analysis confirms that the three-year intervention (Group 3) produces a sustained and statistically significant increase in plain water consumption ($p < 0.001$), with a large effect size ($d = 0.82$). Groups with shorter interventions showed temporary effects that were not maintained in the long-term follow-up.