Analysis of the Breaking Weight of Toothpicks by Brand and Water Duration

Brileigh Cates, Katie Harr, and Quinlin Neuhaus

STAT 5353

Dr. Olbricht

12/17/2025

Introduction

The 3^2 Completely Randomized Factorial Experiment conducted sets out to find the breaking weight (g) of toothpicks among three different brands and three different water soak durations. The first factor, brand type, consisted of three different levels: Great Value Brand, Dollar Tree Brand, and Diamond Brand. The second factor, water soak duration, also consisted of three different levels: 0 hours, 2 hours, and 24 hours. To give effect hypotheses sufficient power, 10 replicates $\left(n_{ij} = 10, \forall i, j\right)$ were assigned to each brand/water duration combination for a total sample size of 90 toothpicks $(n = 90)$. Multiple controls were put in place to keep our experiment as consistent as possible. Those included: testing all toothpicks on the same day and in the same location (to minimize environmental effects), consistent experimental design, and soaking all (similar sized) toothpicks in the same water type, temperature, and volume.

Experimental Design

Toothpicks that were to be soaked in water were randomly assigned to a treatment combination (Table 1) and placed in separate plastics bags, with the same volume of water in each bag. Because all testing times were predetermined, toothpicks assigned to the 24-hour group were placed in the water 23 hours in advance, while those in the 2-hour group were placed in the water two hours in advance. This ensured that each group soaked for the intended duration and prevented testing overlap, minimizing the chance of over-soaking. Testing proceeded in the following order: all 2-hour toothpicks, then the 24-hour group, and finally the dry (0-hour) group.

	Great Value	Dollar Tree	Diamond
0 hours	5, 12, 24, 54, 61, 66, 72, 75, 81, 88	2, 22, 39, 46, 47, 49, 51, 59, 62, 76	9, 15, 20, 29, 41, 58, 60, 70, 77, 84
2 hours	16, 21, 27, 37, 43, 63, 68, 79, 83, 90	4, 19, 35, 45, 56, 57, 74, 73, 78, 85	14, 18, 25, 26, 32, 40, 53, 65, 67, 69
24 hours	11, 13, 17, 28, 30, 42, 44, 71, 74, 86	7, 8, 10, 23, 31, 34, 48, 55, 80, 82	1, 3, 6, 33, 36, 38, 50, 52, 87, 89

Table 1: Experimental Unit Randomization Scheme

A small contraption consisting of a raised table and hanging plate was designed and constructed to facilitate hanging, weighing, and eventually breaking the toothpicks. The table sits high enough to allow the plate to freely hang suspended in the air, whilst providing a gap for the toothpick to span and weight to be applied (See Figure 1). Several different sizes of galvanized steel nuts were used as weights in the experiment (Large: 48.67 g; Medium: 15.67 g; Hanging Plate: 150 g). All weights, including the weight of the hanging plate were weighed with a kitchen scale with an accuracy of 0.1 grams. A small experiment was conducted prior to the main testing to ensure the weights purchased would be enough to snap the toothpicks and provide sufficiently precise weights.

Figure 1: Experiment Structure

During testing, each toothpick was placed across the table's gap, the plate was hung from the toothpick, and weights were added one at a time until the toothpick snapped. The number of nuts on the plate at the moment of breakage was recorded as the breaking weight. If a toothpick began to bow, several seconds were allowed before adding the next weight. For the dry toothpicks, which had substantially higher breaking weights, a heavier plate was used to ensure testing proceeded in the same manner as the other groups. To mitigate potential effects of over-soaking, the order of testing within each soak duration level was randomized (Table 2).

Replicate	1	2	3	4	5	6	7	8	9	10
0 hours	G, D, DT	D, G, DT	G, DT, D	G, DT, D	DT, D, G	G, D, DT	G, D, DT	G, DT, D	G, DT, D	D, G, DT
2 hours	G, D, DT	D, DT, G	D, G, DT	D, DT, G	DT, D, G	DT, G, D	G, DT, D	D, DT, G	D, DT, G	DT, G, D
24 hours	DT, D, G	G, DT, D	D, DT, G	G, DT, D	G, D, DT	G, D, DT	D, DT, G	G, D, DT	G, D, DT	DT, G, D

G = Great Value, D = Diamond, DT = Dollar Tree

Table 2: Testing Order Randomization Scheme

Statistical Model

To analyze the effect of brand and water duration on breaking weight (g), a Two-way ANOVA model, denote as:

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk};\ \ i = 1,2,3;\ j = 1,2,3;\ k = 1,\dots,10 \tag{1}$$

was utilized. The response variable, denoted y_{ijk}, is the breaking weight of the k^{th} toothpick from brand i soaked in water duration j. The term μ is the overall average breaking weight of all toothpicks, the terms α_i and β_j represent the main effects of brand and water duration respectively, and the term $(\alpha\beta)_{ij}$ is the interaction effect between the brand and water duration. The error term, denoted ϵ_{ijk}, is assumed to be independent and identically distributed with a constant variance σ^2 (That is, $\epsilon_{ijk} \overset{iid}{\sim} N(0,\sigma^2)$). To restrict overparameterization, the zero-sum constraint is assumed, that is $\sum\alpha_i = \sum\beta_j = \sum(\alpha\beta)_{ij} = 0$.

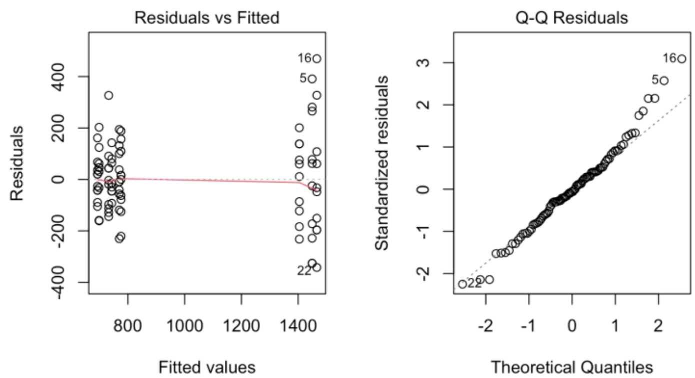

Figure 2: Residuals vs Fitted (left), Q-Q Residuals (right)

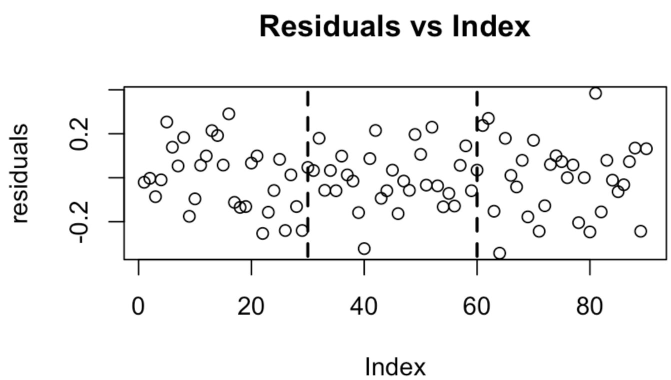

Figure 3: Residuals vs Index

Looking at the residuals vs fitted plot (Figure 2, left), megaphone behavior is observed, indicating that the constant variance assumption is not met. It is also of note that the sample variance ratio, $\frac{s_{max}^2}{s_{min}^2} = 12.9667 > 3$, supports the claim that the constant variance assumption is violated. The Q-Q residuals plot (Figure 2, right) supports the claim that the normality assumption is relatively met, as the graph follows the linear line $y = x$ well, noting skewness in the tails. The behavior observed in the residual vs fitted and Q-Q residuals plots indicate that a transformation could be helpful, which will be investigated further. The residuals vs index plot (Figure 3) has no obvious trend, indicating the independence assumption is met.

To remedy the assumptions violations, a Box-Cox transformation was investigated. From Figure 4, we see that a natural log transformation of the response could be beneficial. Using $\lambda = 0$, the Two-way ANOVA model becomes:

$$\ln y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}; \;\; i = 1, 2, 3; \; j = 1, 2, 3; \; k = 1, \dots, 10 \qquad (2)$$

where each term on the right-hand side is the same, but now the response is the log breaking weight for the k^{th} toothpick from brand i soaked in water duration j.

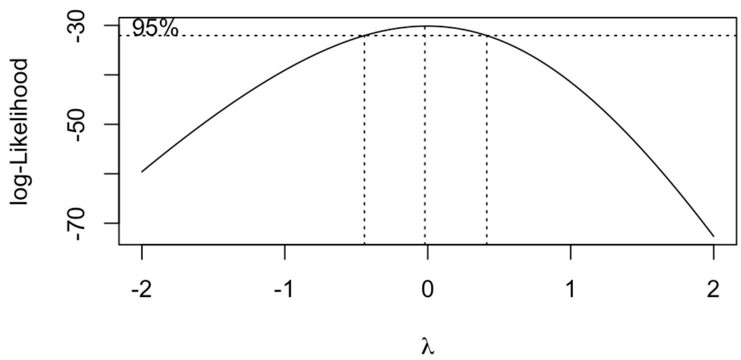

Figure 4: Box-Cox Transformation

Upon transformation, the assumptions were checked again. It is notable that the normality assumption is met, as there is not much deviation from the line $y = x$ (Figure 5, right). The residuals vs fitted plot indicates the constant variance assumption is relatively met (Figure 5, left). The sample variance ratio, $\frac{s_{max}^2}{s_{min}^2} = 3.894 > 3$, implies that there might still be issues with the constant variance assumption, so the Modified-Levene test was conducted. That is:

$$H_0\colon \sigma_{11}^2 = \sigma_{12}^2 = \sigma_{13}^2 = \sigma_{21}^2 = \sigma_{22}^2 = \sigma_{23}^2 = \sigma_{31}^2 = \sigma_{32}^2 = \sigma_{33}^2$$

$$H_a\colon \text{at least one } \sigma_{ij}^2 \text{ differs}$$

which yielded a p-value of $0.5698 > 0.05 = \alpha$, indicating that the assumption that the variance is constant is appropriate. The residuals vs index plot (Figure 6) still supports the claim that the independence assumption is met, as there are not obvious trend or patterns.

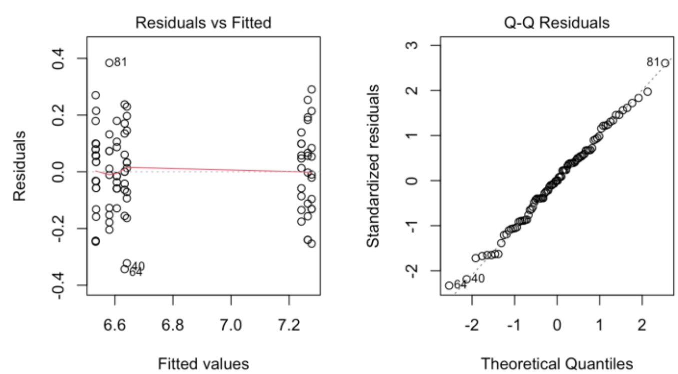

Figure 5: Residuals vs Fitted (left), Q-Q Residuals (right)

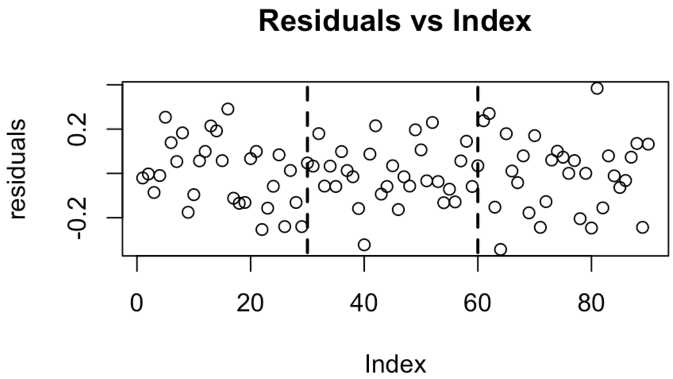

Figure 6: Residuals vs Index

Statistical Methods/Results

After confirming that the assumptions were met for model (2), the interaction plot was interpreted, and the Two-way ANOVA was conducted. The interaction plot (Figure 7) indicates there is no interaction effect. As water duration increases, the log breaking weight for each brand of toothpick decreases. There does appear to be an intersection between brands Dollar Tree and Great Value, which was investigated further. Table 3 displays the result of the ANOVA analysis. First, the interaction effect was investigated, and the hypothesis test

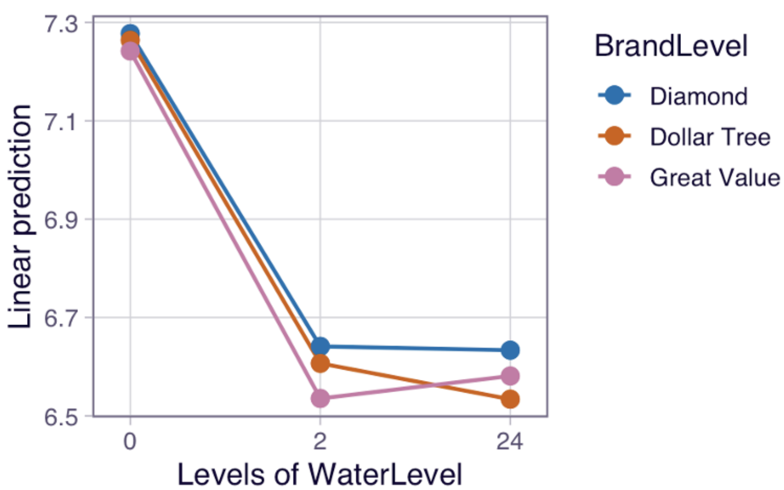

Figure 7: Interaction Plot

$$H_0: (\alpha\beta)_{ij} = 0, \forall i,j \text{ vs } H_a: at\ least\ one\ (\alpha\beta)_{ij} \neq 0 \quad (3)$$

was conducted, yielding an F-Statistics of 0.4791 and respective p-value of $0.7510 > 0.05 = \alpha$. Hypothesis test (3) confirms that there is not sufficient evidence to conclude there is a significant interaction effect between brand type and water duration level on average log breaking weight. As the interaction is not significant, the main effects were tested. That is, the following hypothesis tests were conducted:

$$H_0: \alpha_i = 0, \forall i \text{ vs } H_a: \text{at least one } (\alpha)_i \neq 0 \quad (4)$$

$$H_0: \beta_j = 0, \forall j \text{ vs } H_a: \text{at least one } (\beta)_j \neq 0 \quad (5)$$

From Table 3, the F-Statistics for (4) is 1.4157 with a p-value of 0.2487, and thus there is not sufficient evidence to conclude there is a brand effect of average log toothpick breaking weight. The F-statistics for (5) is 187.5926 with a p-value less than 0.0001, leading to the conclusion that there is a significant water duration level effect on average log breaking weight.

Analysis of Variance Table	df	Sum of Squares	Mean Square	F-Statistic	P-value
BrandLevel	2	0.0683	0.0341	1.4157	0.2487
WaterLevel	2	9.0487	4.5244	187.5926	<0.0001
BrandLevel:WaterLevel	4	0.0462	0.0116	0.4791	0.7510
Residuals	81	1.9536	0.0241		

Table 3: ANOVA Output

Tukey pairwise comparisons (Table 4) were utilized to understand which water duration levels had significantly different average log toothpick breaking weights. The average log breaking weight for toothpicks soaked in water for zero hours was significantly different from the average log breaking weight of both toothpicks soaked in water for two hours and toothpicks soaked in water for 24 hours (p-value <0.0001 for both comparisons). Toothpicks soaked in water for zero hours had the largest average log breaking weight.

Contrast	Tukey Adjusted P-value	Direction
0 hrs vs 2 hrs	<0.0001	0 > 2
0 hrs vs 24 hrs	<0.0001	0 > 24
2 hrs vs 24 hrs	0.9542	2 > 24

Table 4: Tukey Pairwise Comparisons

Conclusion/Discussion

Overall, the experiment's findings suggest that there is not a significant interaction effect between toothpick brand and water duration on average log breaking weight. Continuing to the interpretation of main effects, the experiment suggests there is not a significant brand effect on average log breaking weight, but there is a significant water duration effect on average log breaking weight. Of the specific water duration levels used, the dry toothpicks had significantly higher average log breaking weights, with no significant difference in average log breaking weights between the 2- and 24-hour groups. Error analysis suggests ANOVA assumptions of normality, homoscedasticity, and independence were met when a log transform was applied to the raw breaking weights.

Several challenges were encountered during the experiment. Measurement error in breaking weight could arise from how weights were applied to the toothpick, small variations in soak times, and differences in the angle and placement of the toothpick and hanging string. Galvanized steel nuts were used as weights, but precision could be improved by using a sensitive hanging scale and gradually applying force until the toothpick broke, providing a more accurate reading than the discretized number of nuts.

Consistency in soak times could be improved by performing all trials in parallel, so each toothpick soaks for the exact intended duration, however, this approach is extremely resource intensive. Alternatively, trials could be performed completely serially, with each trial strictly following its soak time, though this approach would be time-consuming. The residuals vs. index plot shows no trend, suggesting that small differences in soak times likely had minimal effect on breaking weight.

Other sources of variation include the angle at which toothpicks spanned the table gap and the placement of the hanging string. To increase precision, the table could be configured so each toothpick rests perpendicular to the gap with a consistent span, and the string could be centered over the toothpick and gap. Overall, improving the measurement method for breaking weight would likely provide the greatest reduction in within-group variation.

Appendix

Brand	Water Duration (hr)	Breaking Weight (g)
Diamond	0	1417.666667
Dollar Tree	0	1424.333333
Great Value	0	1281.666667
Diamond	0	1433.333333
Dollar Tree	0	1840
Great Value	0	1605
Diamond	0	1526.666667
Dollar Tree	0	1714.666667
Great Value	0	1172
Diamond	0	1314.666667
Dollar Tree	0	1511
Great Value	0	1542.333333
Diamond	0	1793
Dollar Tree	0	1730.333333
Great Value	0	1479.666667
Diamond	0	1935
Dollar Tree	0	1276.666667
Great Value	0	1220.666667
Diamond	0	1269.333333
Dollar Tree	0	1526.666667
Great Value	0	1542.333333
Diamond	0	1123.333333
Dollar Tree	0	1220.666667
Great Value	0	1318
Diamond	0	1573.666667
Dollar Tree	0	1123.333333
Great Value	0	1415.333333
Diamond	0	1269.333333
Dollar Tree	0	1123.333333
Great Value	0	1464
Diamond	2	791.6666667
Dollar Tree	2	885.6666667
Great Value	2	650.6666667
Diamond	2	791.6666667
Dollar Tree	2	697.6666667
Great Value	2	760.3333333
Diamond	2	776
Dollar Tree	2	729

Great Value	2	588
Diamond	2	555
Dollar Tree	2	807.3333333
Great Value	2	854.3333333
Diamond	2	697.6666667
Dollar Tree	2	697.6666667
Great Value	2	713.3333333
Diamond	2	650.6666667
Dollar Tree	2	729
Great Value	2	650.6666667
Diamond	2	932.6666667
Dollar Tree	2	823
Great Value	2	666.3333333
Diamond	2	964
Dollar Tree	2	713.3333333
Great Value	2	603.6666667
Diamond	2	713.3333333
Dollar Tree	2	650.6666667
Great Value	2	729
Diamond	2	885.6666667
Dollar Tree	2	697.6666667
Great Value	2	713.3333333
Diamond	24	964
Dollar Tree	24	901.3333333
Great Value	24	619.3333333
Diamond	24	539.3333333
Dollar Tree	24	823
Great Value	24	729
Diamond	24	729
Dollar Tree	24	744.6666667
Great Value	24	603.6666667
Diamond	24	901.3333333
Dollar Tree	24	539.3333333
Great Value	24	635
Diamond	24	807.3333333
Dollar Tree	24	760.3333333
Great Value	24	776
Diamond	24	760.3333333
Dollar Tree	24	729
Great Value	24	588

Diamond	24	760.3333333
Dollar Tree	24	537.6666667
Great Value	24	1059
Diamond	24	650.6666667
Dollar Tree	24	744.6666667
Great Value	24	713.3333333
Diamond	24	713.3333333
Dollar Tree	24	666.3333333
Great Value	24	776
Diamond	24	870
Dollar Tree	24	539.3333333
Great Value	24	823