Accelerating Burn Speed of Birch Wood Popsicle Sticks

E. Croffoot-Suede, S. Kim

4/25/2022

Introduction

There's a popular idea that dousing something in gasoline will make it burn, but is this premise actually true? Is it possible to douse an object in a flammable liquid so that it will burn faster? To test this, we chose to light popsicle sticks on fire under four different treatments: dry (no accelerant applied), soaked in 75.5% ethanol alcohol, soaked in 91% isopropyl alcohol, and soaked in 100% acetone. We chose these because they were the highest purity of each liquid accelerant we could find.

Both ethanol alcohol and isopropyl alcohol are from the alcohol family of volatile liquids. Common uses for ethanol alcohol are in alcoholic beverages, since ethanol is a byproduct of fermentation of many organic compounds. Common uses for isopropyl alcohol are as a detergent or as an antiseptic in first-aid. Our source for the ethanol alcohol was in Galen's 151-proof[1] charcoal filtered vodka, which was distilled from grain neutral spirits. Our source for the isopropyl alcohol was in a 91:9 diluted solution of isopropyl alcohol to water, distributed and sold by Kroger. The vodka was the highest proof available, and the isopropyl alcohol was the most concentrated solution available.

Conversely, acetone is from the ketone family of volatile liquids. Common uses for acetone are as a solvent (such as in nail-polish remover), and it also breaks down hydrocarbons. The acetone was reported to be 100% pure acetone, sold as nail polish remover. However, it also contained a bittering agent (likely to dissuade children from drinking it). The acetone was distributed and sold also by Kroger.

Curiously, all three solutions had warnings on the bottles warning of the flammability of each liquid, which bode well for our hypothesis that soaking objects in them would increase that object's flammability.

Hypotheses

We predicted that using any liquid accelerant would be faster than no liquid accelerant, and that in comparing individual treatments, there would be no significant difference between the accelerants, but there would be a significant difference between no accelerant and with accelerant.

Methods

Experiment Setup

For the experiment, we used birch wood popsicle sticks. We also needed our treatments, which were the ethanol alcohol, isopropyl alcohol, acetone, and no accelerant. To hold the burning sticks off the ground (so they could be lit from below - flames rise), we used four mason jars, one for each stick within a block. To light the sticks, we used a barbeque lighter, and for timing, we used an Apple iPhone. For safety purposes, we also had extra water.

We began by randomly assigning four treatments to each block using Python. We then filled plates with 100 milliliters of each of the three liquid accelerants, as well as drawing a line on the popsicle sticks at 90 millimeters. The line at 90 millimeters would indicate where the stick should be placed on the plate.

For each block, we dipped the popsicle sticks in the indicated accelerant-filled plates based on the assigned treatments for two minutes. After two minutes passed, we took the sticks out and put them on the mason jars so they were elevated. We then lit the end of the popsicle stick that was soaked in the accelerant on fire, holding the lighter for 10 seconds on each popsicle stick. The lighter was held approximately one inch away from the end of the popsicle stick angled from the bottom up so the flame would rise toward the stick. We then watched all four sticks burn until they all burnt out, then measured the distance from the unburnt end of each popsicle stick to the closest burn line, logging that as "distance" to indicate how much of the stick is left **unburnt**, in millimeters.

---

[1]Proof is a measure of alcohol content by volume, corresponding to twice the percentage of alcohol by volume. So, a 100-proof spirit contains 50% alcohol by volume.

Data Description

The population under study consists of a sample of popsicle sticks made of birch wood. We cannot guarantee the randomness of the popsicle sticks out of the population of popsicle sticks that exist in the world (this would be impossible without exorbitant costs), but because of the relative similarity between popsicle sticks and consistency in runs, we believe that our experiment results will generalize to similar popsicle sticks, especially other sticks from the same package.

This gives us a 4! experimental design, since we have four treatment effects and no other factors.

Steps

To ensure that we could find statistically significant results, we ran a power calculation. We assumed that we would use six time blocks, and we chose an alpha level for this of $\alpha = 0.05$, which is the probability of a Type I error (false rejection). We chose a Type II error probability of 0.1. Given a $k = 4$ (given four treatment), the calculated sample size was 20, but given a $k = 6$ (given six time blocks), our calculated sample size was 24. Given our two-way layout, we should take the larger N calculated out of $k = 4, 6$. Based on this, we used $N = 24$, meaning that we should use 6 blocks of 4 observations each.

To generate the order in which we burnt our sticks, we used Python to generate an empty dataframe that had the treatments for 24 runs. We then randomly assigned each of the rows to one of six blocks. This gives us an experimental design with six blocks with four treatments per block.

This method of block generation, however, means that it is possible to have multiple iterations of one treatment and none of other treatments in the same block. For example, it is possible to have "acetone", "acetone", "isopropyl", and "dry" in a single block, meaning that "ethanol" would not be tested in this block. This gives us arbitrary unbalanced data.

Statistical Analysis

Data were cleaned using Python version 3.7 (Python, Wilmington, DE) and pandas version 1.0.3. Data were analyzed using R version 4.0.3 (The R Project for Statistical Computing, Indianapolis, IN). The Skillings-Mack procedure was run in R using the NSM3 package version 1.16. The Wilcoxon rank-sum tests were run using the stats package in R.

Analysis

We began by examining the data. To do so, we generated a boxplot of distances, as well as a boxplot of treatment against distance.
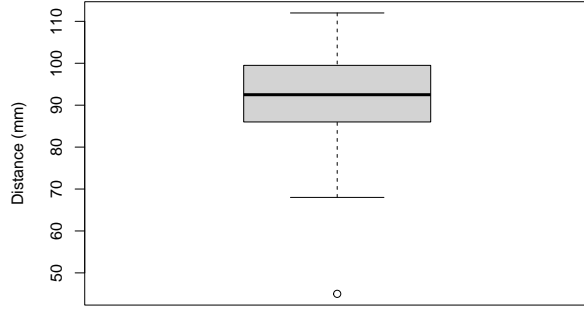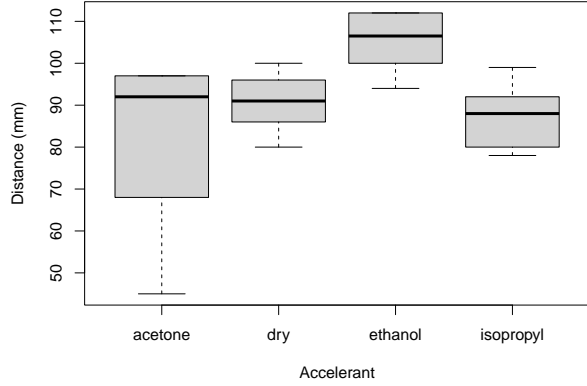
Figure 1: Unburnt distance


Figure 2: Unburnt distance by accelerant

Figure 1 shows the range of response values, but to get a better view of the treatment effects, we also made Figure 2, a plot of unburnt distance by accelerant. We are not using the blocking factor as a factor of interest, so the only plot to be made using a factor against response is with treatment, as shown in Figure 2.

We begin with the model

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

with $i = 1, 2, 3, 4$ and $j = 1, 2, 3, 4, 5, 6$, assuming that the conditions for an ANOVA are met.

Because of its higher relative asymptotic efficiency, we began by running an analysis of variance on the data. We believe that the independence condition is met because each popsicle stick should not affect another popsicle stick during the burn. However, when we checked the conditions for an analysis of variance, we found that equal variance condition and the normality condition were not met for a parametric test.

Figure 3: Residuals vs Fitted
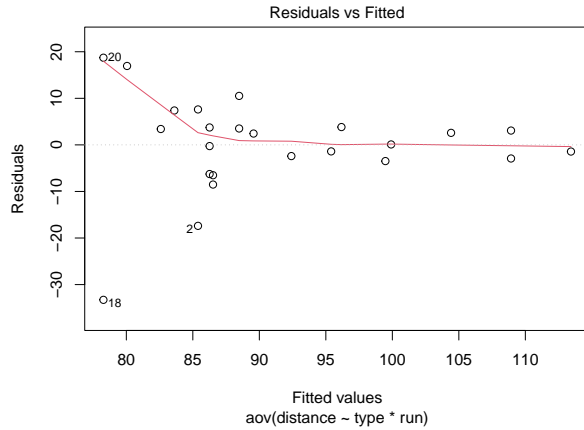
Figure 4: Normal Q–Q
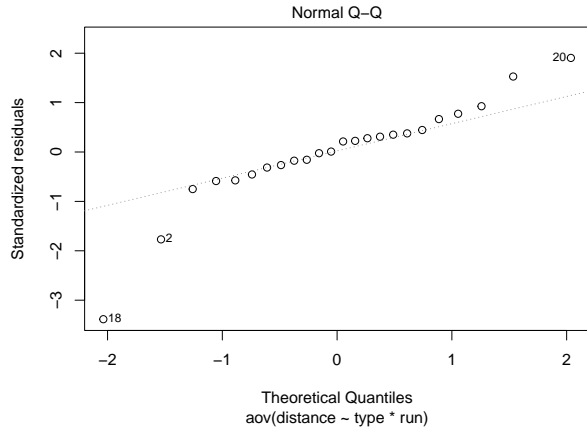
In Figure 3, it is clear that there is a serious issue with the equal variance condition. There is a distinct "funnel" shape from left to right, where the variance at the fitted values around 80 to 90 are much larger than the variance at 105 to 110.

In Figure 4, there are flaired tails that indicate an overdispersed data distribution compared to a normal distribution. That is the distribution was "squished" toward the center compared to the normal distribution.

We attempted to fix the data shape using a data transformation, including the y^2, y^3, $\sqrt{y}$, and $\ln y$ transformations. However, none of these transformations changed the data enough to change the dispersion or the shape of the distribution to be more normal. Instead, we chose to use a non-parametric method instead. For

4

the type of data, we chose to use the Skillings-Mack procedure. The Skillings-Mack procedure is a distribution free test for general alternatives for data from an arbitrary incomplete block design. This means that it can be used where there are missing observations in a single block. Its parametric cousin is the two-way ANOVA, and it can be used as a proxy for an overall F-test where the conditions are not met and the data are from an arbitrary incomplete block design.

Thusly, our model is updated to

$$y_{ij} = \theta + \alpha_i + \beta_j + \epsilon_{ij}$$

with $i = 1, 2, 3, 4$ and $j = 1, 2, 3, 4, 5, 6$, to accommodate this new procedure, given that the conditions for the ANOVA are not met.

The Skillings-Mack procedure requires that the N random variables are mutually independent. That is, $X_{ij1}, ..., X_{ijc_{ij}}$, $i = 1, ..., n$, and $j = 1, ..., k$ are mutually independent. We believe this to be true for our data given the collection method. It also requires that for each fixed (i, j), the c_{ij} random variables are a random sample from a continuous distribution function F_{ij}. We also believe this to be true because we used a random sample out of the population of popsicle sticks in the world. Lastly, the treatment response should be a continuous variable such that "less than", "equal to", and "greater than" can be placed into context. We believe that "length" in millimeters as a response variable satisfies this condition.

The hypotheses for the Skillings-Mack procedure are as follows:

$$H_0 : [\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4],$$

$$H_A : [\alpha_1, \alpha_2, \alpha_3, \alpha_4 \text{ not all equal}],$$

when $c_{ij} \equiv 1$, for $i = 1, 2, 3, 4$ and $j = 1, 2, 3, 4, 5, 6$.

When we ran the Skillings-Mack procedure, we found an asymptotic upper-tail probability of $0.0536 < 0.1 = \alpha$, and a Skillings-Mack SM test statistic of 7.6587 on 3 degrees of freedom. Given this, we concluded that there is a difference in at least one of the groups.

In order to find which groups have a significant difference, we need to use an associated pairwise comparison procedure that can also be used against arbitrary incomplete block designs. To determine this, we created a new pairwise multiple comparisons procedure that can be used on arbitrary incomplete block designs.

1. Run a Wilcoxon rank-sum test on the data for all $6 = \binom{4}{2}$ treatment pairwise comparisons.
2. Keep the block and response values together, but randomly assign a new treatment for each observation, then calculate the Wilcoxon rank sum for all $6 = \binom{4}{2}$ treatment pairwise comparisons. This will spread the block effect across all treatments. Repeat this r times, and store each of the pairwise comparisons for each of these tests.
3. For each of the six pairwise comparisons groups, compare the 5th and 95th percentiles of the r simulated random-treatment Wilcoxon rank-sum test statistics against the test statistic from each of the pairwise Wilcoxon tests from the actual data. If the actual Wilcoxon rank-sum test statistic falls inside of the $(0.05, 0.95)$ percentile from the random-treatment distribution, there is a significant difference between the two groups.

This procedure is used to determine that given two groups (in this case, each pair of treatments), is there strong statistical evidence to support the hypothesis that one of the groups is significantly different from the other? Because we are using a nonparametric method, we are comparing the medians. This diminishes the effect of using non-normal data. The Wilcoxon rank-sum test calculates distance from the medians using the sum of the rank, not the actual value's difference from the median, which also helps to diminish the effect of extraordinarily large or small outliers.

We chose $r = 1000$ to ensure that we created a suitable distribution of random-treatment test statistics to compare the actual test statistics against.

Results

We got the following comparisons table, which contains the cutoff values at an α level of 0.1.[2] This is essentially a method to calculate the distance from the median and tell whether there is a significant difference between two groups.

Table 1: Comparisons Table

Comparison	Actual	Medians of Simulated U's	$\frac{0.1}{6}$ Percentile	$1 - \frac{0.1}{6}$ Percentile
Ethanol-Dry	2.5	18.5	4.833333	30.66667
Ethanol-Acetone	2.0	18.0	5.000000	30.50000
Ethanol-Isopropyl	35.0	18.0	5.500000	30.50000
Acetone-Dry	17.0	17.5	4.333333	31.16667
Acetone-Isopropyl	19.0	18.0	4.333333	31.16667
Isopropyl-Dry	23.0	18.0	4.000000	30.00000

In Table 1, we found that the simulated U's are not statistically different from the actual U's for the acetone-dry, acetone-isopropyl, and isopropyl-dry combinations. This means that there is no statistically significant difference between the results for isopropyl alcohol, acetone, and no accelerant.

Based on this information, we generated a boxplot of just the unburnt distance by accelerant.

Figure 5: Unburnt Distance by Accelerant
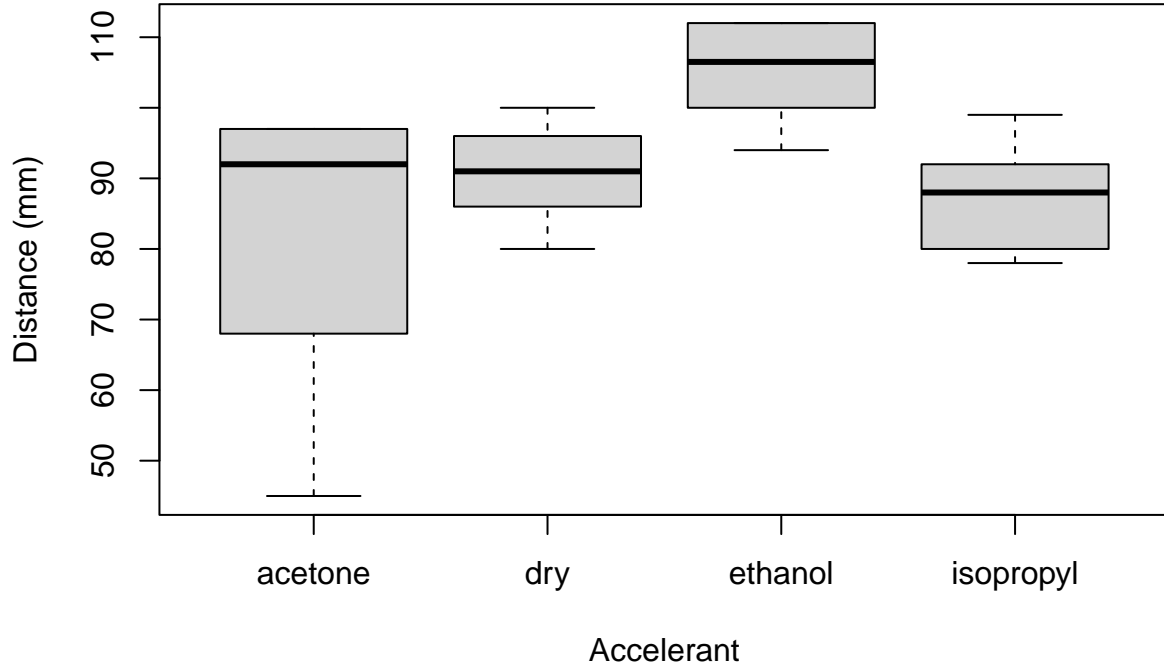
We see that acetone has the lowest box by distance, so we would like to take acetone or anything in acetone's group as calculated by the multiple comparisons procedure.

---

[2]We use the $\frac{0.1}{6}$ and the $1 - \frac{0.1}{6}$, since these are the Bonferroni adjustments given six tests. Not using a Bonferroni adjustment makes the cutoff values for the 90% confidence interval simply $\frac{\alpha}{2}$ and $1 - \frac{\alpha}{2}$.

Discussion

In retrospect, we should have designed our experiment to use a randomized complete block design, instead of an arbitrary incomplete block design. In a randomized complete block design, we would still have six blocks and four treatments, but each block would contain one observation for each of the possible treatments. This would have made our analysis easier. If our data had been collected as a randomized complete block design, we would have used the Friedman procedure for differences in medians using a randomized complete block design, and assuming a significant effect, we would have followed it up with the distribution-free, two-sided, all-treatments Wilcoxon, Nemenyi, McDonald-Thompson multiple comparisons procedure based on Friedman rank sums in a general configuration.

However, we believe that our pairwise multiple comparisons procedure would still be valid given data obtained in an arbitrary incomplete block design. For example, suppose we were looking at data for a factory that produces products in batches, but which does not produce every batch every day. This would generate data in an arbitrary incomplete block design, where a metric of the product is the response, the batch is the treatment, and the day is the block. Using our procedure, we believe that we could diminish the block effect of days in the analysis and come up with an appropriate pairwise multiple comparison between the batches.

We believe that the results can be generalized to a wider population to what accelerants are the most dangerous for increasing flammability in traditionally flammable materials, like wood.

Conclusion

Our goal is to maximize the amount burnt, so the metric for success is to minimize the unburnt distance. We saw in the boxplot from Figure 5 that the best way to decrease unburnt distance is to use acetone. Table 1 tells us that there was no significant difference between acetone, dry, or isopropyl alcohol. Therefore, we recommend that to maximize the amount burnt, to soak birch popsicle sticks in acetone, isopropyl alcohol, or no accelerant at all.

Based on this, we reject our original hypothesis that using any accelerant is better than no accelerant in favor of the theory that 75.5% ethanol alcohol has a negative effect on burning distance.