STA 305: Design and Analysis of Experiments

Final Project – Group 19

Final Report

Due Date: August 23rd, 2021, at 11:59 PM EDT

Examining factors affecting Buffon's Needle Experiment

Christopher Chen | **Roshan Ravishankar**

Bart Harvey | **Rachata Tuangsitthisombat**

Conrad Jureczek | **Shahrukh Zahid**

Shamsun Nila | **Xiang Zhuang**

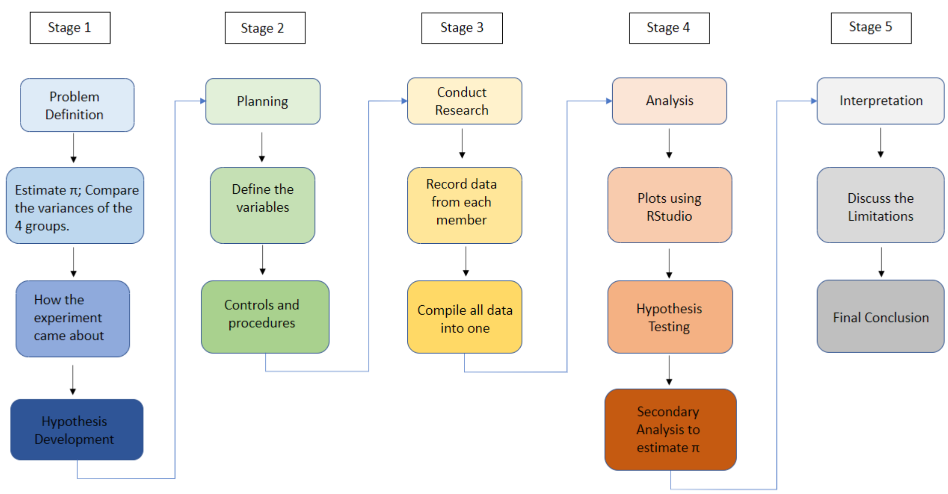
Description of Case Study

Buffon's Needle Experiment (1) was described in the 18th Century as a method to estimate π. It is conducted by dropping a specified number (n) of 'needles' (or 'sticks') of length *l* onto a surface with parallel lines separated by a distance, *t* (where $t > l$) and then counting the number (*c*) of needles that cross one of the lines. This enables the probability *c/n* to be determined. It has been shown that this probability is equal to $2l/t\pi$, which, when rearranged, provides the estimate of π as $2ln/tc$. The target population involves all possible drops of needles on a surface with parallel lines. The experimental unit is each iteration of dropping needles of length *l* onto a surface with parallel lines separated by a distance, *t.* The objective of the proposed experiment is to examine if the value of n or the ratio of *l:t* affects the estimation of the probability *c/n*, and hence π. The study will explore deeper into what variables in the experiment affect the estimation of π.

Description of Experimental Design

- Number_of_Drops is one of the dependent factors, with two levels: 10 or 20 needles dropped per trial.
- Needle_to_Width_Ratio is the length of needles used in a trial in relation to width of the parallel lines on the surface that the needles were dropped onto. It is the second dependent factor, also with two levels: 0.9 or 0.5 (e.g., needles of length 1.5" being dropped onto a surface with parallel lines separated by 3.0").
- Number_Crossed_Line is the number of needles crossing a parallel line on the landing surface, as determined by the group member conducting that trial.
- The **response variable** is Cross_Line_Proportion calculated by dividing the Number_Crossed_Line variable by the Number_of_Drops.

The experiment was carried out as a two-by-two factorial design. For convenience, each of the members of our group volunteered to collect data for one of the l:t ratios (1:2 or 9:10) by conducting 25 iterations for each of n = 10 and n = 20 and recording for each trial how many needles/sticks crossed a line (i.e., c). The data collected by the eight group was pooled to create the data set for analysis. The data collected by the eight group was pooled to create the data set for analysis. The experiment is a Between-Subjects group experiment as each group member separately conducted trials of dropping needles. There was no potential nuisance variable identified and each group member conducted 50 trials of dropping either 10 or 20 needles while assigned to only one ratio (i.e., 0.5 or 0.9). Thus, there was no randomization and the design incorporated replication through multiple trials. Initially, the experimental design was set up to include each group member as block. However, the latter understanding is that there could be a lack of significant effect on the response variable. Thus, we did not incorporate blocking in the experimental design for the simplicity of the design and to achieve efficiency. Further, each group member followed identical instructions of setting up the experiment and determining if a needle crossed the line. In times where a member is unable to determine if a needle indeed crossed the lines, the instructions outline that we look from above to check if there is a discontinuity in the parallel line where we believe the line crossed (Appendix C, Figure 1). If there is a discontinuity, then this counts as the needle crossing. Thus, we incorporated control in our design. For convenience, each member chose a needle to width ratio to conduct the experiment; thus, we do not have randomization in our design.



Data Analysis

The experiment's null hypothesis is that there is no significant difference in the mean estimation of π between the different treatment groups. Our alternative hypothesis states that there is a significant difference in the mean estimation between the different treatment groups. The treatment groups are:

Group 1: 100 trials of dropping 10 needles with a needle to width ratio of 0.5.

Group 2: 100 trials of dropping 20 needles with a needle to width ratio of 0.5.

Group 3: 100 trials of dropping 10 needles with a needle to width ratio of 0.9.

Group 4: 100 trials of dropping 20 needles with a needle to width ratio of 0.9.

$$H_0: \mu_{(0.5,10)} = \mu_{(0.5,20)} = \mu_{(0.9,10)} = \mu_{(0.9,20)}$$

$\mu_{(0.5,10)}$: mean of cross proportion of Group 1

$\mu_{(0.5,20)}$: mean of cross proportion of Group 2

$\mu_{(0.9,10)}$: mean of cross proportion of Group 3

$\mu_{(0.9,20)}$: mean of cross proportion of Group 4

The primary data analysis will be a two-way analysis of variance. A secondary analysis will be pairwise comparisons of the four treatment groups' variances be to estimate π. These will be carried out using F tests with each α level being adjusted by Holm's procedure (2). An alpha level of 0.05 will be used.

From the summary statistics, the mean proportion of needles crossing a line is 40.0% (Appendix B, Figure 2). The proportion of needles found to cross a line differed when assessed by the needle-to-line width ratio: 29.4% for the length-to-width ratio of 0.5 and 51.0% for 0.9 (Appendix A, Figure 1). A similar proportion of needles were found to cross a line when assessed by the number of needles dropped per trial: 39.1% for 10 needles and 41.4% for 20, as summarized in the (Appendix A, Figure 2). Consistent with first two boxplots, Appendix A Figure 3 boxplot shows that the proportion of needles crossing a line is dependent on the needle length-to-line width ratio. The interaction plot (Appendix A, Figure 4) suggests some degree of interaction between the two levels of "number of needles dropped" and the two levels of "needle length-to-line width ratio".

A two-way ANOVA was conducted using the linear model with interaction and a model was fit in R (Appendix B, Figure 2).

$$\mathbf{Yijk = \mu T + \alpha j + \beta k + (\alpha\beta)jk + \epsilon ijk}$$

Each member of the group collected data separately, to pool a dataset together for analysis, which confirms each observation is independent from another. Most data points can be seen falling on the black diagonal line, and increasing positively and concentrated in the middle, which confirms the assumption that error terms in the model are distributed randomly (Appendix

A, Figure 5. We can confirm the assumption of equal variances of error terms hold as the grouped points can be seen gathered for all groups, meaning the groups differ only by their means, but have the same variance (Appendix A, Figure 6).

A two-way analysis of variance was run to assess the effects the effects of the two independent variables, plus their interaction, on the response variable, proportion of needles crossing a line. The results confirm that the ratio of the needle length to the width of the lines has the greatest effect on the proportion of needles found to cross a line ($P < .001$) (Appendix B, Figure 10). Additionally, the interaction between number of needles dropped and the ratio of the needle length to the width of the lines also found to have a statistically significant effect ($P = .042$).

While the analyses above address the primary question of this experiment, whether any of the dependent variables affect the proportion of needles crossing a line, they do not assess whether either of the independent variables affect the precision with which that proportion is measured. As described above, to assess this the variances of the four groups were compared by pair-wise F tests with each α level being adjusted by Holm's procedure (2). The mean of the four groups respectively were found to be 0.298, 0.2905, 0.483, 0.5365 (Appendix B, Figure 11). The variances of the four groups respectively were found to be 0.02120808, 0.01634823, 0.03657677, 0.01524520 (Appendix B, Figure 12). Adjusted alpha-levels seen in Appendix B, Figure 13.

The most extreme pair of variances are for the two treatment groups with a ratio between the length of the needles and the width of lines equal to .9 (i.e., 0.03657677 for 10 drops per trial and 0.01524520 for 20 drops). This F-ratio is equal to 2.4 with 99 degrees of freedom in the numerator and denominator, with a P-value < 0.001 so these two variances would be considered statistically significantly different (i.e., P-value < Holm's adjusted alpha level of 0.00833).
The next most extreme pair of variances are for the treatment group with a ratio between the length of the needles and the width of lines equal to .9 and 20 drops per trial (i.e., 0.01524520) and the other group (ratio = 0.5) with 10 drops (i.e., 0.02120808). This F-ratio is equal to 1.39, again with 99 degrees of freedom in the numerator and denominator, with a P-value = 0.05 so

these two variances would not be considered statistically significantly different because the observed P-value of 0.05 is greater than the Holm's adjusted alpha level of 0.01.

The estimated means and variances for the four groups can be used to construct 95% confidence intervals as mean ± t.025, df=99 * SEM gives the table seen in Appendix B, Figure 14. As noted, π can be estimated through the equation: *2ln/tc*, where *l/t* equals the length-to-width ratio and *c/n* equals 1/the estimated proportion as such the estimated values of π (i.e., π_hat) corresponding to the mean and 95% confidence limit values listed above would be $\hat{\pi}_{(o.5,10)} = 3.53$ to 3.20, $\hat{\pi}_{(o.5,20)} = 3.88$ to 3.58, $\hat{\pi}_{(o.9,10)} = 3.60$ to 3.30, and $\hat{\pi}_{(o.9,20)} = 3.43$ to 3.28. Using the ANOVA table, we calculated the effect size where, $\widehat{\omega}^2$=0.346 and the corresponding F-value is equal to 0.727 (Appendix B, Figure 15). The results confirmed that the observed result is a large effect. Similarly, we calculated the sample size needed to detect a medium, effect (i.e., F-value = 0.25) with 80% power. The results confirm that sample size of 100 in each of the four treatment groups would be sufficient to detect a Cohen's Effect Size of 0.166 (i.e., between low and medium effect) (Appendix B, Figure 17).

Limitations

Each member collected data separately using different landing sites for dropping the needles and subjectively determined the height that the needles were dropped from; this adds variability in the data collection, but it is unknown if any systematic error was introduced as a result. Additionally, for convenience, each group member chose the needle-to-width ratio to conduct data collection, which was not randomly assigned. We understand that could also limit the accuracy of experimental results. Number of sticks only differed by 10 (20 vs 10) so may not have been sufficiently different to detect a difference. In future studies would recommend a greater difference, perhaps 10 vs 40, be examined.

Conclusion

The needle_to_width_ratio has a strong statistical effect on the estimate of the proportion of sticks that cross one of the parallel lines, with the 0.9 treatment group having a statistically significantly higher proportion (i.e., ~ 50%), compared to the 0.5 treatment group (i.e., ~ 30%).

In addition, a modestly significant interaction was found between the two studied factors (see Figure 7), arising because the proportion of needles/sticks crossing a line was higher for the 20 trials treatment group compared to the 10 trials treatment group, but only for the 0.9 treatment group (i.e., ~54% vs 48%), while quite similar for the 0.5 treatment group (i.e., 29% vs. 30%).

References

1. Buffon's Needle Experiment. In: Rosenthal JS. Struck by Lightning. London, UK: Granta Publications, 2005: 187.
2. Holm S. A simple sequentially rejective multiple test procedure. Scand J Statistics 179; 6:65-70.

Appendix A

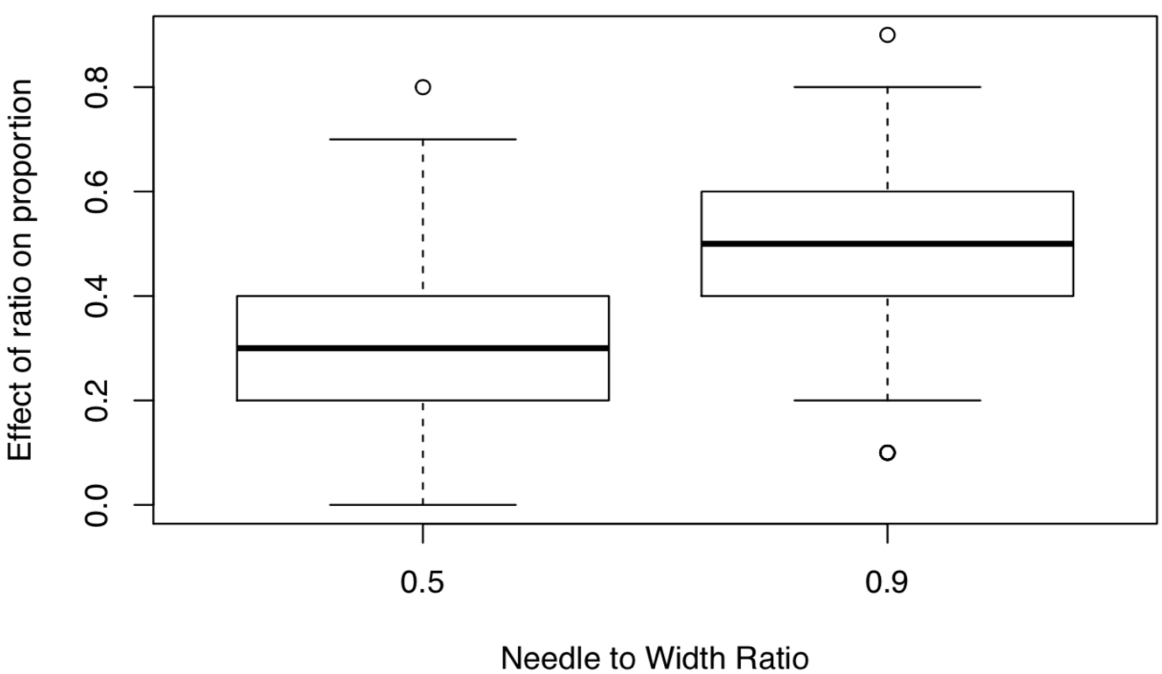


Figure 1: Effect of Ratio on Proportion vs. Needle to Width Ratio boxplot

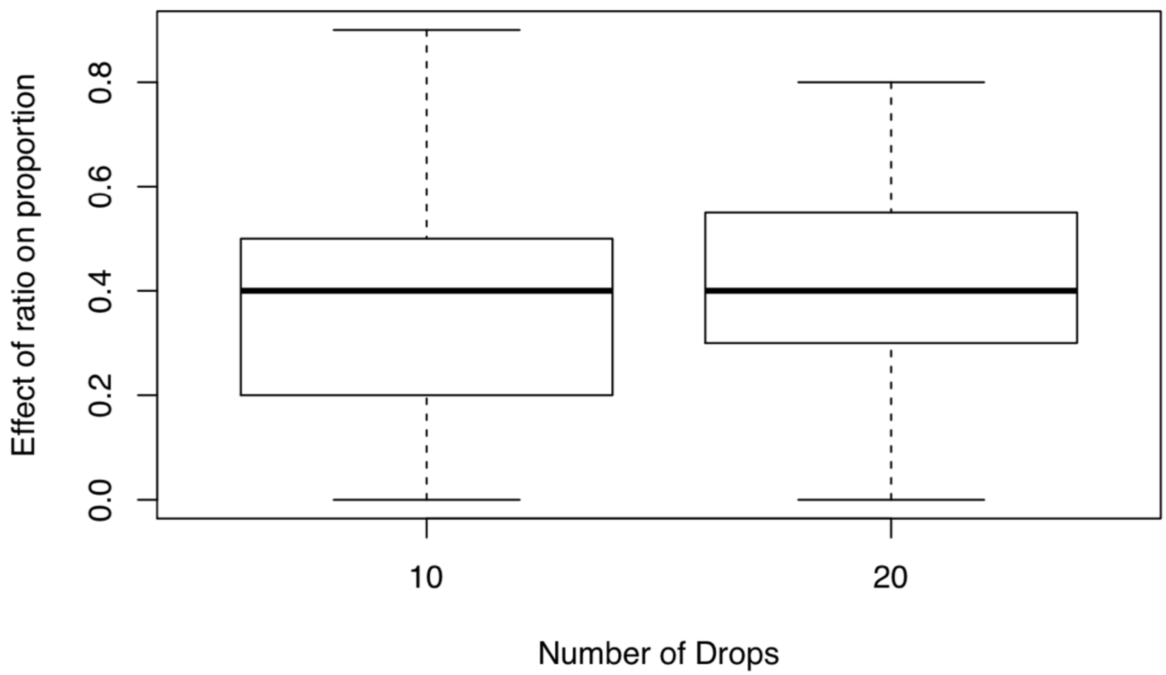


Figure 2: Effect of Ratio on Proportion vs. Number of Drops boxplot

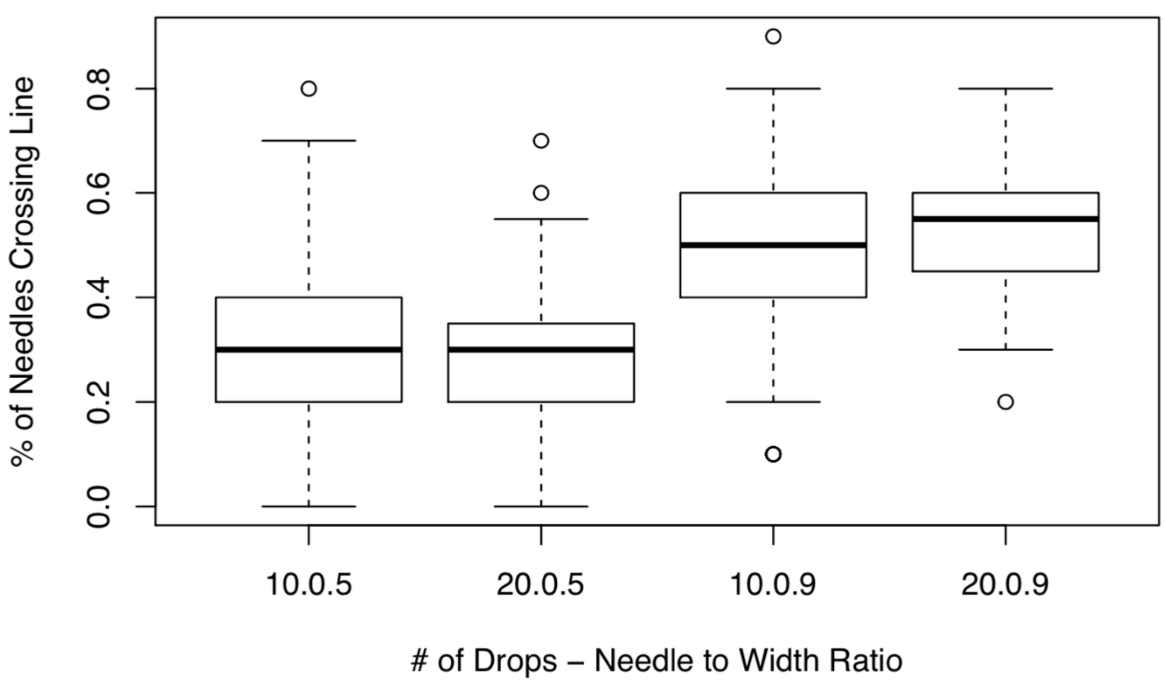


Figure 3: Boxplots of all four groups

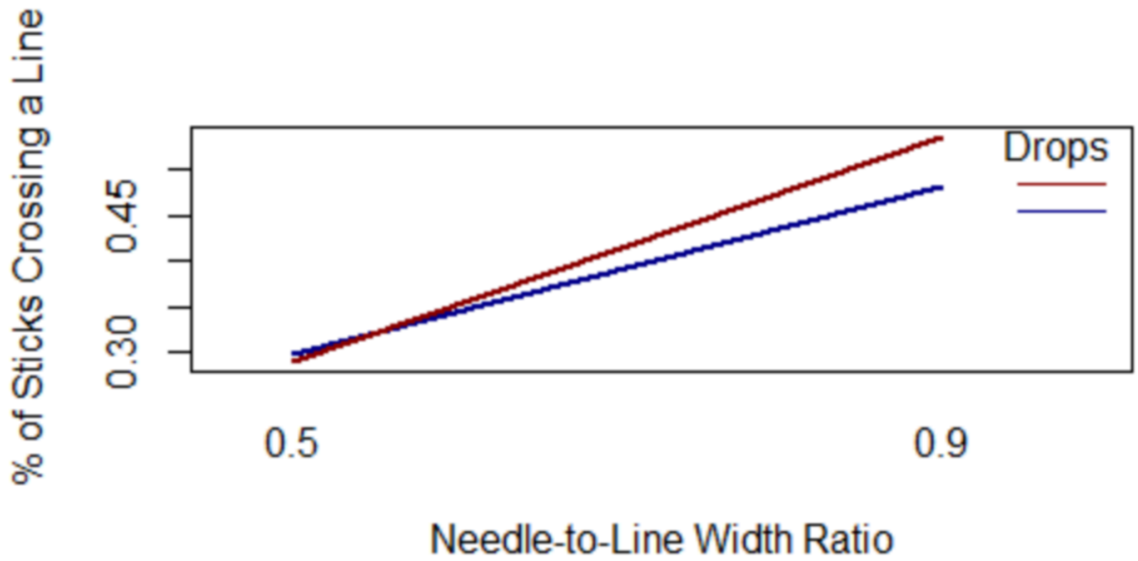


Figure 4: Interaction plot

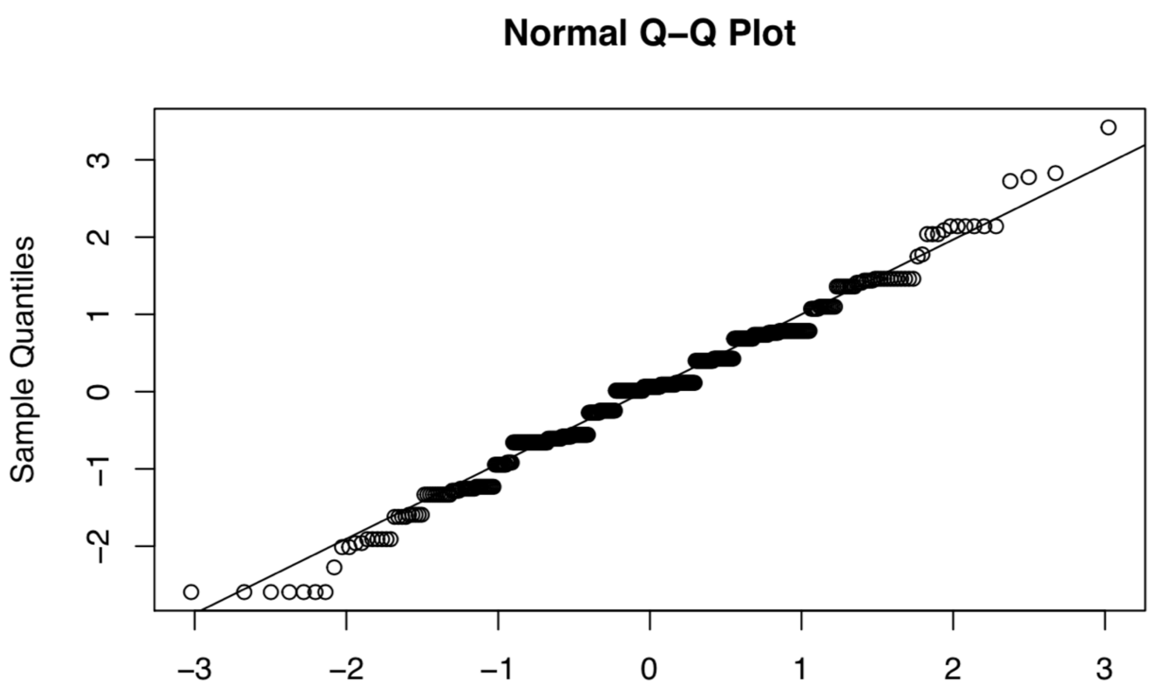


Figure 5: Normal Q-Q Plot

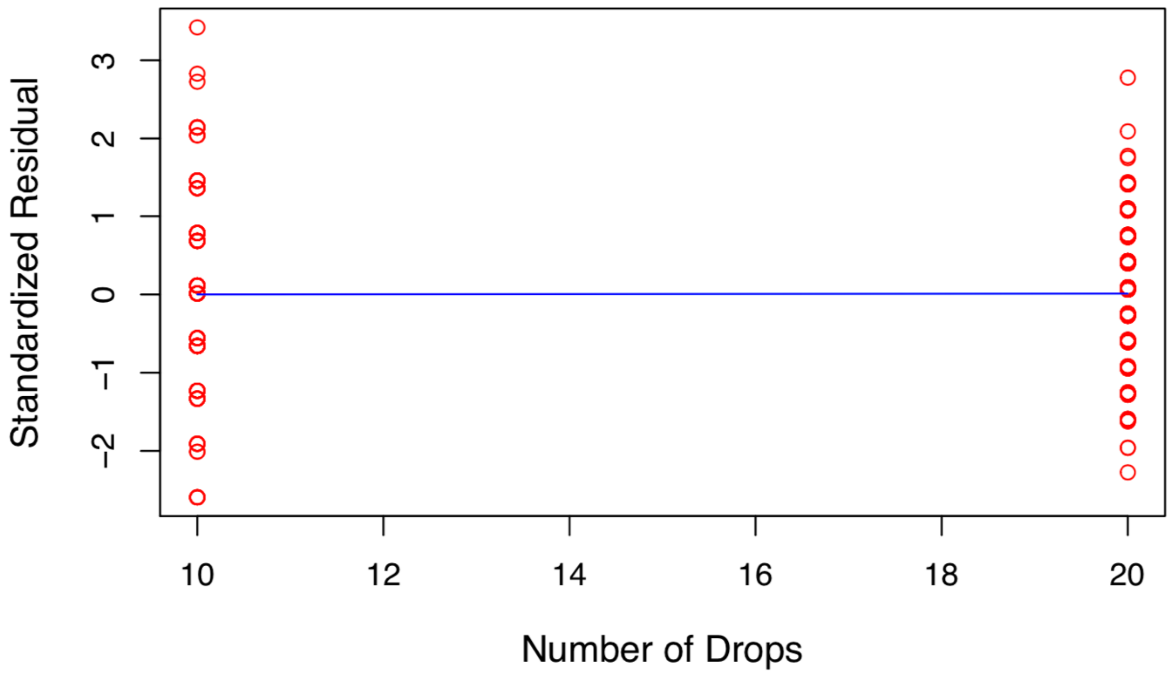


Figure 6: Standardized residuals vs. Number of Drops to check Homoscedasticity

Appendix B

```r
# Loading data
setwd("~/Google Drive/5th Year/STA305")
exp_data <- read_excel("dataset.xlsx")

# cleaning data to remove unncessary N/A
exp_data<- exp_data[, c(2, 3, 4, 5)]
attach(exp_data)
# seperating groups
group1=filter(exp_data,Number_of_Drops=='10',Needle_to_Width_Ratio=='0.5')
group2=filter(exp_data,Number_of_Drops=='20',Needle_to_Width_Ratio=='0.5')
group3=filter(exp_data,Number_of_Drops=='10',Needle_to_Width_Ratio=='0.9')
group4=filter(exp_data,Number_of_Drops=='20',Needle_to_Width_Ratio=='0.9')
```

Figure 1: Setup

```r
summary(exp_data)
```

```
##  Number_of_Drops Needle_to_Width_Ratio Number_Crossed_Line
##  Min.   :10      Min.   :0.5           Min.   : 0.000
##  1st Qu.:10      1st Qu.:0.5           1st Qu.: 3.000
##  Median :15      Median :0.7           Median : 6.000
##  Mean   :15      Mean   :0.7           Mean   : 6.088
##  3rd Qu.:20      3rd Qu.:0.9           3rd Qu.: 8.000
##  Max.   :20      Max.   :0.9           Max.   :16.000
##  Cross_Line_Proportion
##  Min.   :0.000
##  1st Qu.:0.300
##  Median :0.400
##  Mean   :0.402
##  3rd Qu.:0.550
##  Max.   :0.900
```

Figure 2: Summary statistics

```r
boxplot(Cross_Line_Proportion~Needle_to_Width_Ratio, data=exp_data,
        ylab="Effect of ratio on proportion", xlab="Needle to Width Ratio")
```

Figure 3: Effect of Ratio on Proportion vs. Needle to Width Ratio boxplot

```r
boxplot(Cross_Line_Proportion~Number_of_Drops, data=exp_data,
        ylab="Effect of ratio on proportion", xlab="Number of Drops")
```

Figure 4: Effect of Ratio on Proportion vs. Number of Drops boxplot

```r
boxplot(Cross_Line_Proportion~Number_of_Drops+Needle_to_Width_Ratio,
        data=exp_data, ylab="% of Needles Crossing Line",
        xlab="# of Drops - Needle to Width Ratio")
```

Figure 5: Percentage of needles dropped Graph

```r
with(exp_data,interaction.plot(Needle_to_Width_Ratio, Number_of_Drops,
                               Cross_Line_Proportion,col=c("red", "blue"),
                               main="Interaction Plot",
                               xlab="Needle to Width Ratio", ylab="Cross Line
                               Proportion"))
```

Figure 6: Interaction Plot

```r
model = lm(formula=Cross_Line_Proportion~Number_of_Drops*Needle_to_Width_Ratio)
```

```r
summary(model)
```

```
##
## Call:
## lm(formula = Cross_Line_Proportion ~ Number_of_Drops * Needle_to_Width_Ratio)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.38300 -0.09237  0.00950  0.10200  0.50200
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                           0.150500   0.086033   1.749  0.08101 .
## Number_of_Drops                      -0.008375   0.005441  -1.539  0.12456
## Needle_to_Width_Ratio                 0.310000   0.118175   2.623  0.00905 **
## Number_of_Drops:Needle_to_Width_Ratio  0.015250   0.007474   2.040  0.04197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1495 on 396 degrees of freedom
## Multiple R-squared:  0.3512, Adjusted R-squared:  0.3463
## F-statistic: 71.46 on 3 and 396 DF,  p-value: < 2.2e-16
```

Figure 7: Summary of model

```r
resid <- rstudent(model)
fitted <- predict(model)
qqnorm(resid)
qqline(resid)
```

Figure 8: QQ plot

```r
plot(resid ~ exp_data$Number_of_Drops, type = "p", xlab = "Number of Drops",
     ylab = "Standardized Residual", cex.lab = 1.2, col = "red")

lines(lowess(exp_data$Number_of_Drops, resid), col = "blue")
```

Figure 9: Homoscedasticity plot

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: Cross_Line_Proportion
##                                      Df Sum Sq Mean Sq  F value  Pr(>F)
## Number_of_Drops                       1 0.0529  0.0529   2.3675 0.12469
## Needle_to_Width_Ratio                 1 4.6440  4.6440 207.8368 < 2e-16 ***
## Number_of_Drops:Needle_to_Width_Ratio   1 0.0930  0.0930   4.1632 0.04197 *
## Residuals                           396 8.8484  0.0223
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 10: Anova table

```r
means
```

```
##        [,1]
## [1,] 0.298
## [2,] 0.2905
## [3,] 0.483
## [4,] 0.5365
```

Figure 11: group means

```r
variances
```

```
##            [,1]
## [1,] 0.02120808
## [2,] 0.01634823
## [3,] 0.03657677
## [4,] 0.0152452
```

Figure 12: group variances

```r
AlphaLevels
```

```
##   Test # Divisor Result
## A 1.0000  6.0000 0.0083
## B 2.0000  5.0000 0.0100
## C 3.0000  4.0000 0.0125
## D 4.0000  3.0000 0.0167
## E 5.0000  2.0000 0.0250
## F 6.0000  1.0000 0.0500
```

Figure 13: Adjusted alpha levels applying Holm’s procedure (2): overall alpha / (total number of tests – test number + 1)

```r
# Group 1 CI
n <- 100
a1 <- mean1
s1 <- sqrt(var1)
error1 <- qt(0.975,df=n-1)*s1/sqrt(n)
left1 <- a1-error1
right1 <- a1+error1

CI1=t(matrix(c(left1,right1)))
```

Confidence intervals:

Group 1

```r
CI1
```

```
##           [,1]      [,2]
## [1,] 0.2691039 0.3268961
```

Group 2

```r
CI2
```

```
##           [,1]      [,2]
## [1,] 0.2616039 0.3193961
```

Group 3

```r
CI3
```

```
##           [,1]      [,2]
## [1,] 0.4541039 0.5118961
```

Group 4

```r
CI4
```

```
##           [,1]      [,2]
## [1,] 0.5076039 0.5653961
```

Figure 14: Confidence levels

```r
pwr.anova.test(k = 4, n = 100, f = 0.727, sig.level= 0.05, power = NULL)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 4
##               n = 100
##               f = 0.727
##       sig.level = 0.05
##           power = 1
##
## NOTE: n is number in each group
```

Figure 15: Power test indicating large effect

```
pwr.anova.test(k = 4, n = NULL, f = 0.25, sig.level= 0.05, power = 0.8)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 4
##               n = 44.59927
##               f = 0.25
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

Figure 16: Power test to calculate the required sample size

```
pwr.anova.test(k = 4, n = NULL, f = 0.166, sig.level= 0.05, power = 0.8)
```

```
##
##      Balanced one-way analysis of variance power calculation
##
##               k = 4
##               n = 99.89504
##               f = 0.166
##       sig.level = 0.05
##           power = 0.8
##
## NOTE: n is number in each group
```

Figure 17: Power test explaining sample size justification

Appendix C:

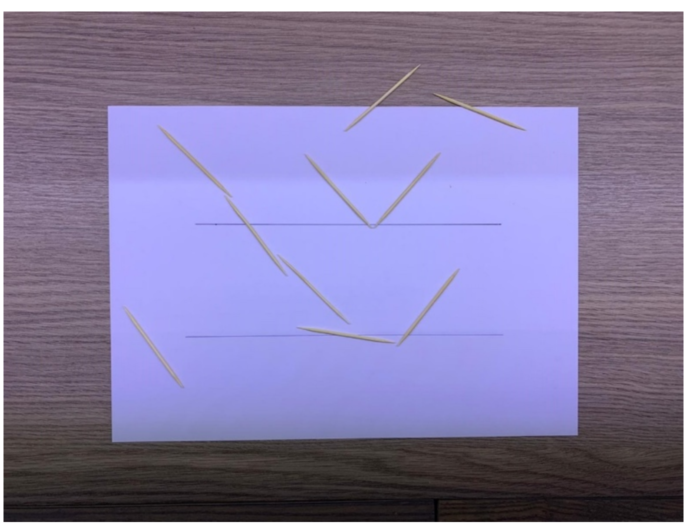


Figure 1: Photograph of a single trial: 10 sticks, 0.9 length_to_width_ratio