Stat 541 Experimental Design Project 1

Ben DeVries, Eric Folsom, Ryan Kardoes, Alex Trauner

INTRODUCTION

For this project, we aim to explore the effect of adding salt to water on the time it takes to boil the water. We will be utilizing a randomized complete block design to perform our analysis. The following write-up details the experimental procedure, including a discussion of variables and the results of a pilot study.

Experimental Design

The following lists our proposed experimental design to answer our research question. Please note that the "Pilot Study" step both holds discussion for the study itself, and the results of the pilot.

1. Define the objectives: The objective of this experiment was to compare boiling times of water with and without salt to determine which boils faster.
2. Identify response and sources of variation.
    a. Treatment factors and levels: One treatment factor (salt) with 2 levels: unsalted, salted.
    b. Response: Time until boiling, in seconds. True difference is defined as 30 seconds.
    c. Sources of variation: Pot material, stove used, burner used, time to preheat burner, amount of water used, amount of salt used, order of boiling (pot temperature).
    d. Blocking factor: Pot + burner combination
3. Choose a rule for assigning the experimental units to trt (design).
    a. We would choose a complete block design, where each pot/burner combination received multiple repetitions of salted vs. unsalted.

4. Specify measurements to be made, experimental procedure, and the anticipated difficulties.

    a. Measurements:

    Seconds until water is at a rolling boil (determined by one group member, timed by phone timer to the nearest 1/100th of a second).

    b. Procedure:

    All unsalted pots will be boiled first, then salted pots, to avoid residual salt in pots and measuring cups that may contaminate unsalted pots. Burners and pots should be preheated to boiling, then the first pot should be dumped out once temp is equilibriated. 2 cups of water should be added to pots. Unsalted pots can be boiled as is, salted pots should have 2 tablespoons of salt added to water prior to adding to pot. Time should start when the pot is placed on the burner. Time should be recorded when the pot is at a rolling boil - determined by one group member to reduce subjectivity.

    c. Anticipated difficulties:

    Human subjectivity in identifying boiling point, ensuring pots and water are at same temp for different repetitions. Differences in ambient temperature due to heat from stovetop burners.

5. Pilot experiment:

    We completed a pilot experiment to determine which size pot (small or large) would have less variability and therefore give us more power in determining boiling time differences. To complete this pilot study small and large pots were brought to boil with 2 cups of water. The first rep was dumped out to equilibriate temperature, then the 3 following repetitions were recorded for boiling time in seconds.

    Our pilot experiment yielded the following results:

```
pre_dat <- data.frame(pot = rep(c("small", "large"), each = 3),
                      seconds = c(170, 162, 184, 112, 126, 119))
sig <- pre_dat |> group_by(pot) |> summarise(s = sd(seconds)) |> pull()
pre_dat
```

```
    pot seconds
1 small     170
2 small     162
3 small     184
4 large     112
5 large     126
6 large     119
```

```r
sig
```

```
[1]  7.00000 11.13553
```

```r
power.t.test(delta = 30, sd = sig[1], sig.level = 0.05, type = "two.sample",
             power = 0.90, alternative = "one.sided")
```

```
     Two-sample t test power calculation

              n = 2.143766
          delta = 30
             sd = 7
      sig.level = 0.05
          power = 0.9
    alternative = one.sided

NOTE: n is number in *each* group
```

```r
power.t.test(delta = 30, sd = sig[2], sig.level = 0.05, type = "two.sample",
             power = 0.90, alternative = "one.sided")
```

```
     Two-sample t test power calculation

              n = 3.33303
          delta = 30
             sd = 11.13553
      sig.level = 0.05
          power = 0.9
    alternative = one.sided

NOTE: n is number in *each* group
```

Therefore, we will use small pots in our experiment, as the SD was 7 seconds for small pots vs. 11.14 seconds for large pots, indicating less variability in small pots.

6. Specify the model:

    a. Seconds to boiling = constant + effect of salt + effect of burner block + error

b. $Y_{ijt} = \mu + \alpha_i + \beta_j + \epsilon_{ijt}$

Where:

Y_{ijt} is the response (boiling time in seconds)

μ is a constant

α_i is main effect of treatment (salted vs. unsalted)

β_j is main effect of block (burner)

ϵ_{ijt} is the error term

7. Outline the analysis:

   a. Compare differences between salted and unsalted water for time until boiling. Determine differences (confidence intervals) in boiling time for each treatment, and determine which treatment is best (fastest/least time until boiling).

8. Calculate number of observations and time/budget.

   a. N: Based on the our pilot study, we will need to complete at least 3 replicates for each treatment.

   b. Time: Each pot will take ~3 minutes to boil.

   c. Budget: Cost of salt + cost of electricity + cost of water + cost of human time

9. Review and revise. After the experiment, steps could be revised to plan a better future experiment. Step 9 is further expanded upon in the results section.

DATA ANALYSIS

Our data was collected and stored within the "STAT541-Project-Data.xlsx" file attached with this report. The following code chunk details the exact analysis steps that were taken, including the reading of the data, fitting a linear model with the lm() function, performing an ANOVA analysis, and generating pairs of contrasts to answer our research question.

Collected Data

```
#Read and format data
BoilDat <- readxl::read_excel("STAT541-Project-Data.xlsx") |> mutate(
    Treatment = as.factor(Treatment),
    Burner = as.factor(`Burner/pot`)
) |> select(-`Burner/pot`)
BoilDat |> split(BoilDat$Treatment) |> map(knitr::kable)
```

```
$Fresh
```

Treatment	Boiling Time	Burner
Fresh	158.44	Back
Fresh	203.45	Front
Fresh	150.96	Back
Fresh	171.22	Front
Fresh	144.12	Back
Fresh	159.68	Front
Fresh	128.55	Back
Fresh	169.08	Front
Fresh	181.26	Back
Fresh	183.43	Front
Fresh	144.92	Back
Fresh	162.54	Front

```
$Salt
```

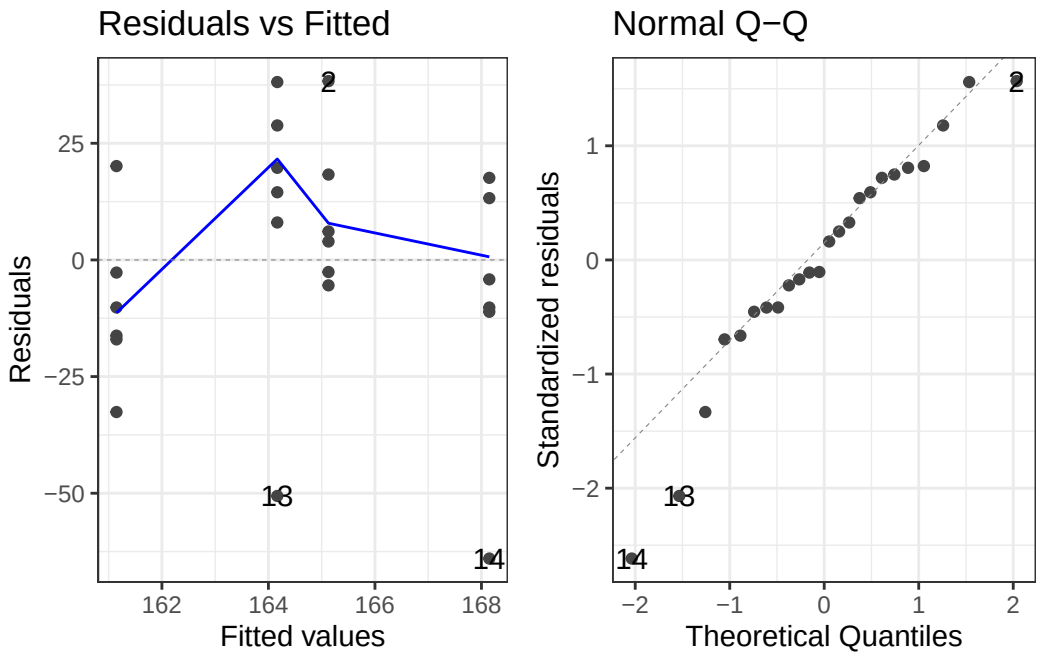
Treatment	Boiling Time	Burner
Salt	113.58	Back
Salt	104.15	Front
Salt	202.29	Back
Salt	181.39	Front
Salt	178.67	Back
Salt	157.95	Front
Salt	183.91	Back
Salt	163.98	Front
Salt	192.99	Back
Salt	185.74	Front
Salt	172.19	Back
Salt	157.04	Front

```r
#Fit model with main effects of treatment and burner
fit.Boil <- lm(`Boiling Time` ~ ., data = BoilDat)
summary(fit.Boil)
```

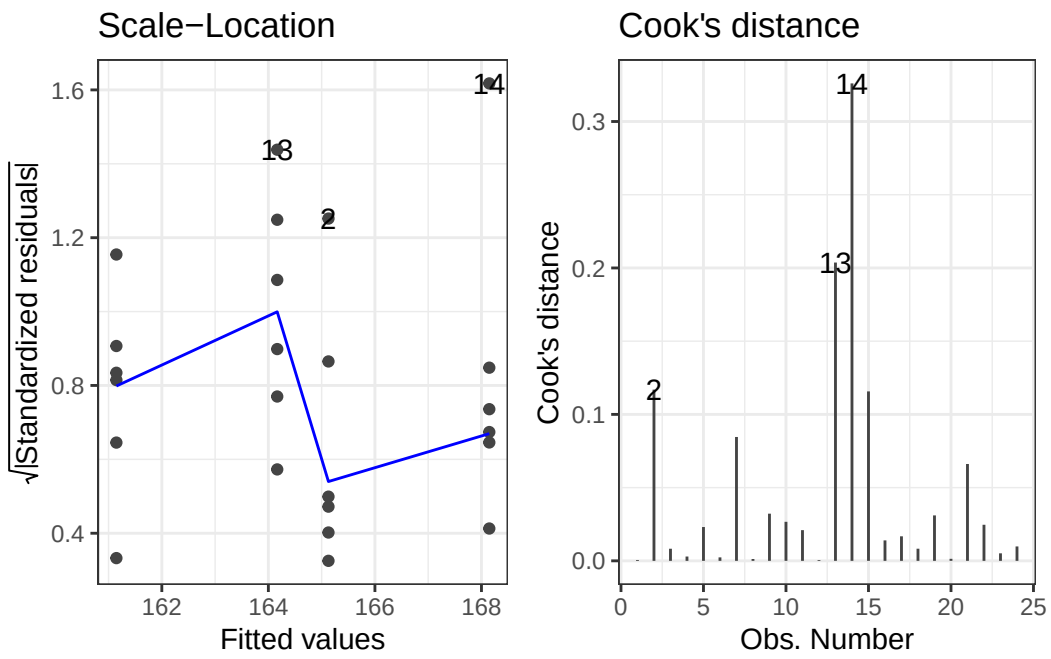
```
Call:
```

```
lm(formula = `Boiling Time` ~ ., data = BoilDat)

Residuals:
    Min      1Q  Median      3Q     Max
-63.997 -10.425   0.682  17.770  38.322

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   161.147      9.244  17.432 5.74e-14 ***
TreatmentSalt   3.019     10.675   0.283    0.780
BurnerFront     3.981     10.675   0.373    0.713
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 26.15 on 21 degrees of freedom
Multiple R-squared:  0.01032,   Adjusted R-squared:  -0.08393
F-statistic: 0.1095 on 2 and 21 DF,  p-value: 0.8968
```

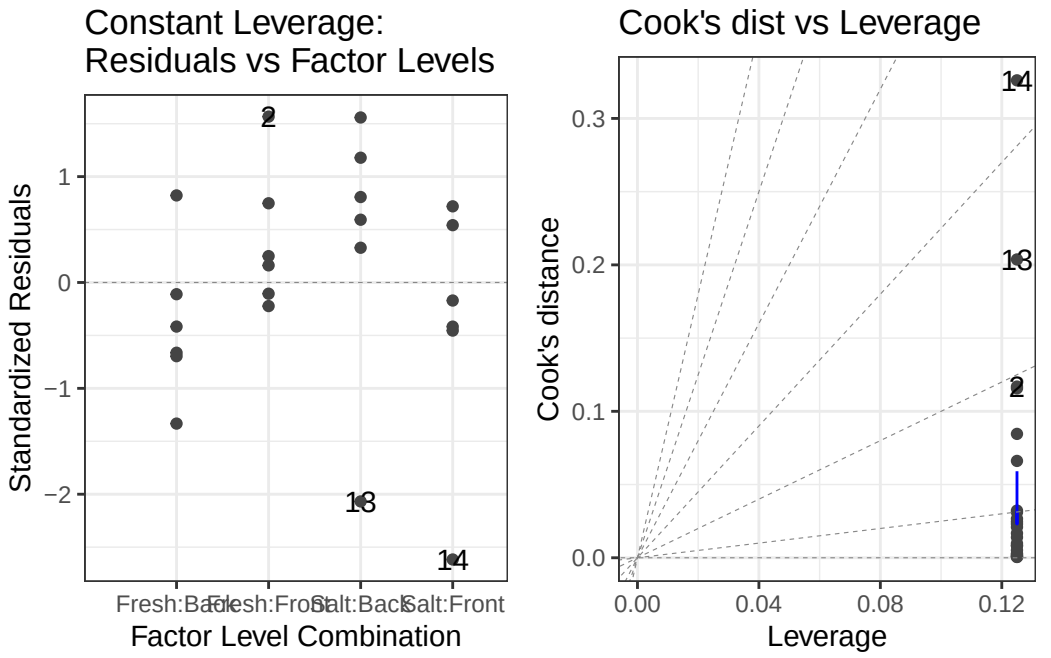
```
#Plots to check assumptions - linearity, normality, constant variance, influential observatio
autoplot(fit.Boil, which=1:2)
```



```
autoplot(fit.Boil, which=3:4)
```




```
autoplot(fit.Boil, which=5:6)
```



```r
#ANOVA and Confidence Intervals
anova(fit.Boil)
```

```
Analysis of Variance Table

Response: Boiling Time
          Df  Sum Sq Mean Sq F value Pr(>F)
Treatment  1    54.7   54.69  0.0800 0.7801
Burner     1    95.1   95.08  0.1391 0.7129
Residuals 21 14357.1  683.67
```

```r
confint(fit.Boil)
```

```
                 2.5 %    97.5 %
(Intercept)   141.92232 180.37185
TreatmentSalt -19.17968  25.21802
BurnerFront   -18.21802  26.17968
```

```r
#Contrasts
emm.Boil <- emmeans(fit.Boil, ~ Treatment, adjust = "tukey")
pairs(emm.Boil, specs = "Treatment")
```

```
 contrast     estimate   SE df t.ratio p.value
 Fresh - Salt    -3.02 10.7 21  -0.283  0.7801

Results are averaged over the levels of: Burner
```

```r
emm.Boil_ByBurn <- emmeans(fit.Boil, ~ Treatment | Burner)
pairs(emm.Boil_ByBurn, specs = "Treatment")
```

```
Burner = Back:
 contrast     estimate   SE df t.ratio p.value
 Fresh - Salt    -3.02 10.7 21  -0.283  0.7801

Burner = Front:
 contrast     estimate   SE df t.ratio p.value
 Fresh - Salt    -3.02 10.7 21  -0.283  0.7801
```

RESULTS

The results indicate that there is no main effect of treatment (salted vs. fresh), nor was there an effect of burner. Based on the results of the ANOVA, there is very weak evidence against the null hypothesis that there is no difference between boiling times for salted and unsalted water (F_{1,21}, p-value=0.7801). The 95% confidence interval for the difference in true mean boiling times in seconds for salted vs. unsalted water was found to be (-19.17968, 25.21802).

Conclusions and Potential Modifications for Future Research

If we were to repeat this experiment, here are some things we might change to improve our design and outcomes:

Repetitions - Greater repetition would give us more power to detect a difference. While the pilot study indicated 3 repetitions was enough power, the pilot study did not take salted vs. unsalted water into account.

Salt mixing methods - We added salt to the measuring cup before dumping water into the pots. If repeating this experiment, it may be more accurate to add salt to the water after it is dumped in the pot. We seemed to have a lot of residual salt that was staying in the measuring cups/not fully emptying into the pot.

Boiling time determination - We used one group member to determine boiling time (visually at a rolling boil). The experiment would be more accurate if multiple thermometers were used to determine boiling point at 100 C.