Course Code: **STAT 537**

Course Title: Statistical Methods for Applied Researchers II

Anton Phipps, Sanjir Inam Salsabil

Project Final Report

The Effects of Temperature and Pressure on the Espresso Brewing Process

Dr. Keumhee Chough

Submission due by December 10th, 2021

TABLE OF CONTENT

Background

Espresso is not only a beverage itself, but a base for many other beverages enjoyed the world over. According to the Paulig Barista Institute, the fundamental aspect of an espresso brewing recipe is the Dose (the amount of dry coffee being used), the Yield (the amount of coffee in the cup after brewing), and the Brew Time. For a more refined recipe adjustments are made to Temperature, Pressure Profiling, and Extraction time (2021, Korhonen, J.)

Once brewed, espresso consists of 3 parts, the heart; the body; and the crema. The crema is made up of Carbon Dioxide bubbles, surrounded by water and oils. The ratio of crema to the other liquids influences the flavour profile, the details of which will be discussed in a later section.

In 2010 a new method of Brewing espresso named Caffe Firenze was developed. This method uses a sealed extraction chamber where the water and gases raise the pressure of the system above that of the traditional brewing method (Parenti, Guerrini, Masella, Dainelli, & Spugnoli 2014.) The goal of our data analysis is to examine if the new brewing method can be used to attain a specific crema ratio.

Review of Methodology

Our study is based on “A new method for Espresso Coffee brewing: Caffe Firenze,” by Parenti, Guerrini, Masella, Dainelli, and Spugnoli, published in the Journal of Agricultural Engineering (2014) The data which we will be using is simulated data which shares the same mean and standard deviation. The dependent variables are temperature, and pressure. The dependent variable is the ratio of Crema to liquid in percentage. Our null hypothesis is that the temperature and pressure have no effect on the crema ratio. Our alternative hypothesis is that either – or both – of pressure and temperature influence the ratio of crema.

Of course, this raises the question, if the crema ratio can be controlled using pressure and pressure what would be the ideal ratio? To answer this question two baristas were consulted as subject matter experts from Block 1912 Café. During an interview conducted on 2021-11-07, Baristas Praveer, and Mackenzie provided the following insight:

- Generally, Crema volume is associated with freshness of coffee grounds
- More Crema suggests a stronger flavour
- Blended drinks are easier to make with less crema, due to ease of mixing and delicate undertones
- In simple Espressos, americanos, and cappuccinos more Crema is viewed as better
- In the end, it is subjective

With this insight, we can generally say that for blended drinks having less crema would be ideal, while for strait espresso, or unblended drinks a larger crema ratio is desirable.

The data has the 3 different temperatures: 75; 85; and 90 degrees Celsius, and 2 different pressures: 15; and 20 bar. This results in 6 different treatments, each of which is replicated 9 times making a total of 54 samples. This satisfies the rule of thumb for regression analysis of having 10 samples for each factor involved in the model. If a relationship between pressure, temperature and crema ratio can be established, baristas could use this information to create the ideal espresso depending on what type of drink they are making. With this motivation we will attempt to build a linear recession model.

Design Critique

The experimental process to collect the values from the original study are quite thorough. Measurements are carefully taken, and the factors outside of temperature and pressure are held constant. Specifically, a digital balance was used with accuracy of 0.1 grams was used to measure the coffee grounds. The final brew weight was 25.7 ± 0.6 g averaged over all the samples (Parenti et al., 2013, p262.) It is apparent that careful attention was paid to the experimental procedure. The analysis however was not performed using the standard statistical analysis tools.

First, there should have been data provided for the traditional method for analysis. Second, simply comparing the average of the methods without checking for significance is meaningless. There is no indication that L.I.N.E. assumptions were checked, or that any of the other residual checks were performed. As such, we will proceed with a more through data analysis, relying on the data which we were provided.

Data Analysis

Linearity

As mentioned earlier, the independent variables available for analysis are temperature in Celsius, and pressure in Bars. Our linear model will follow one of the following 3 formats, which all have linear relationships:

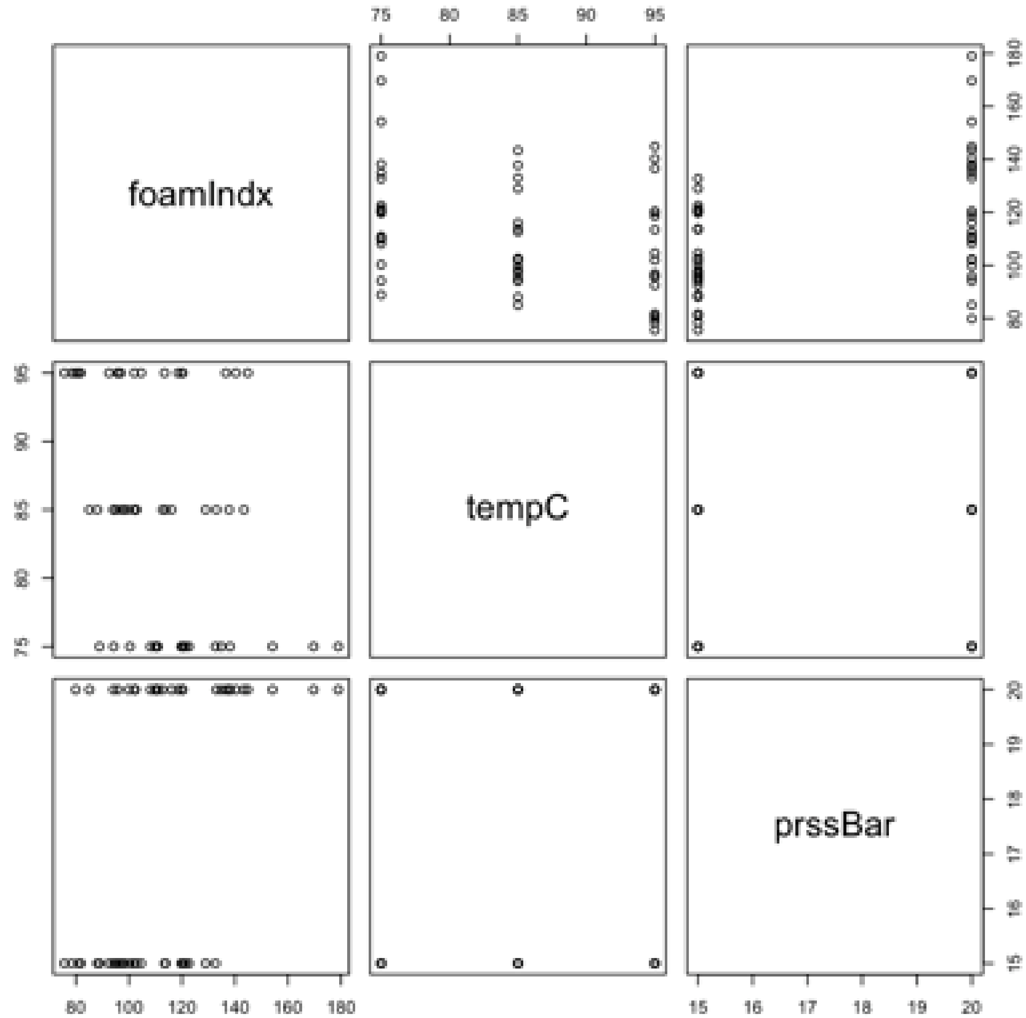
$$Y_{Crema} = \beta_0 + \beta_{temp} \cdot X_{temp} + \beta_{press} \cdot X_{press}$$

$$Y_{Crema} = \beta_0 + \beta_{temp} \cdot X_{temp}$$

$$Y_{Crema} = \beta_0 + \beta_{press} \cdot X_{press}$$

Independence

To perform the checks for Independence, we performed the checks both visually and based on our full model. It can be seen that there is no linear relationship between our independent variables.




```
# Evaluate Collinearity
sqrt(vif(lm_full)) > 2 # variance inflation factors
```

espresso$tempC: FALSE **espresso$prssBar:** FALSE

Normality, and Equal Variance

To test for normality, we first setup the three potential models mentioned earlier:

```
lm_full <- lm(espresso$foamIndx ~ espresso$tempC + espresso$prssBar)
lm_temp <- lm(espresso$foamIndx ~ espresso$tempC)
lm_press <- lm(espresso$foamIndx ~ espresso$prssBar)
```

We then graphed the relevant plots for each of the models:

Model: Crema ~ temp + pressure

Model: Crema ~ temp

Model: Crema ~ pressure

As following:

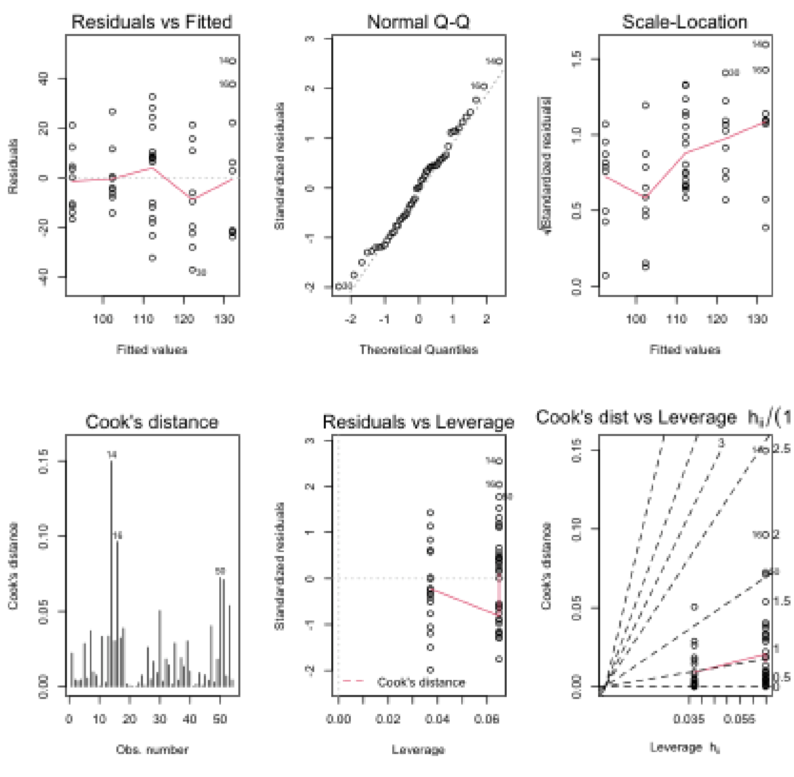


Figure 1: Model: Crema ~ temp + pressure

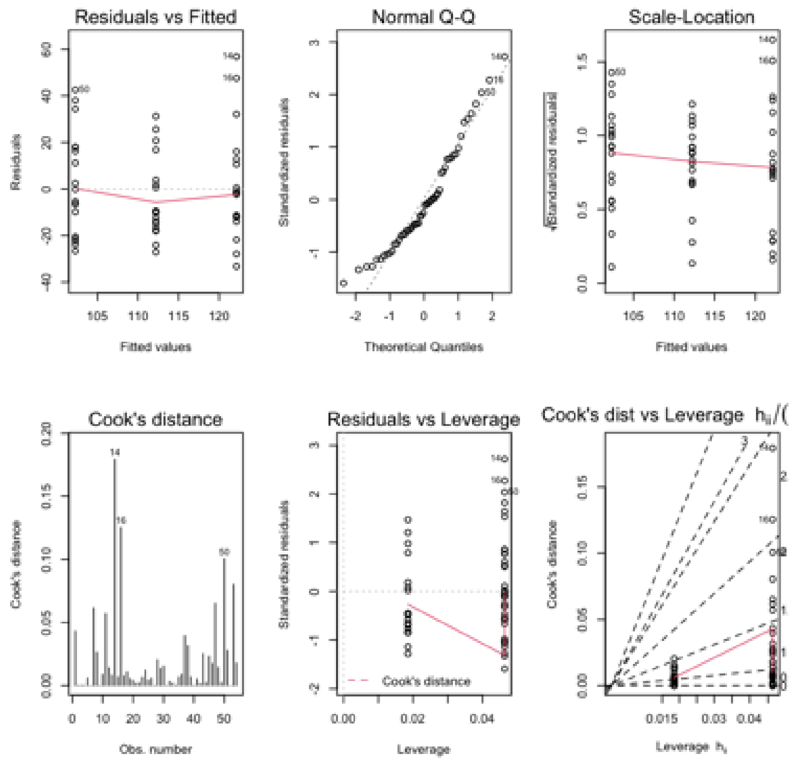


Figure 2: Model: Crema ~ temp

```
hat values (leverages) are all = 0.03703704
 and there are no factor predictors; no plot no. 5
```

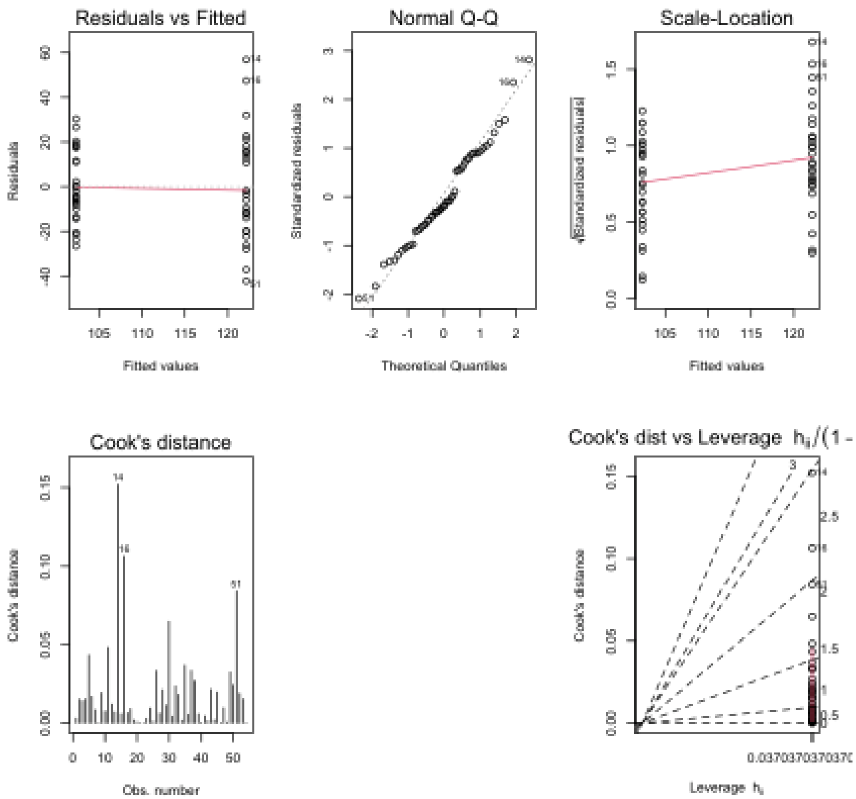


Figure 3: Model: Crema ~ pressure

In all the models, the normality based on the Normal Q-Q graphs suggest normality. Also, In the Residual vs Fitted, the distribution above and below 0 are relatively balanced suggesting equal variance. It should also be mentioned that the Cook's distance is less than 1, meaning that the observations should not have undue influence on our results in any of the models.

```
influencePlot(lm_full, main="Influence Plot",
              sub="Circle size is proportial to Cook's Distance" )
```

A data.frame: 4 × 3

	StudRes	Hat	CookD
	<dbl>	<dbl>	<dbl>
2	0.4398527	0.06481481	0.0045414318
10	0.1495810	0.06481481	0.0005270045
14	2.6980232	0.06481481	0.1497335213
16	2.1092556	0.06481481	0.0962707641

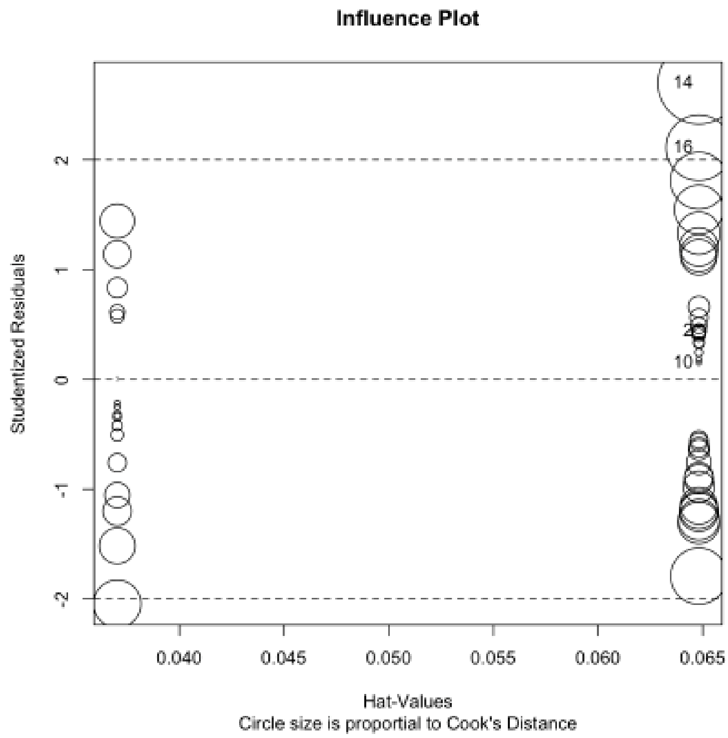


Figure 4:Influence Plot

Development of Models

In the case of this study, there are a limited number of permutations for the models. We will first look at creating the models, and then comparing them pairwise.

```
Call:
lm(formula = espresso$foamIndx ~ espresso$tempC + espresso$prssBar)

Coefficients:
     (Intercept)    espresso$tempC  espresso$prssBar
        127.1629           -0.9925            3.9667


Call:
lm(formula = espresso$foamIndx ~ espresso$tempC)

Coefficients:
   (Intercept)  espresso$tempC
      196.5795         -0.9925


Call:
lm(formula = espresso$foamIndx ~ espresso$prssBar)

Coefficients:
     (Intercept)  espresso$prssBar
          42.800             3.967
```

```
anova(lm_full)
```

A anova: 3 × 5

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
	<int>	<dbl>	<dbl>	<dbl>	<dbl>
espresso$tempC	1	3546.203	3546.2025	9.77970	0.0029133717
espresso$prssBar	1	5310.375	5310.3750	14.64493	0.0003557753
Residuals	51	18493.033	362.6085	NA	NA

```
anova(lm_full, lm_temp)
```

A anova: 2 × 6

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	51	18493.03	NA	NA	NA	NA
2	52	23803.41	-1	-5310.375	14.64493	0.0003557753

```
anova(lm_full, lm_press)
```

A anova: 2 × 6

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	51	18493.03	NA	NA	NA	NA
2	52	22039.24	-1	-3546.203	9.7797	0.002913372

There is a significant difference between the two full models and our potential simplified models. The R library has a simple way of both creating the simplified models and comparing them to the full model for performance, of which we made use. Backwards elimination will be used as it is the recommended procedure.

```
myLm_step_back=step(lm_full, direction="backward", trace=TRUE)
summary(myLm_step_back)
```

```
Start:  AIC=321.15
espresso$foamIndx ~ espresso$tempC + espresso$prssBar

                   Df Sum of Sq   RSS    AIC
<none>                          18493 321.15
- espresso$tempC    1    3546.2 22039 328.63
- espresso$prssBar  1    5310.4 23803 332.78

Call:
lm(formula = espresso$foamIndx ~ espresso$tempC + espresso$prssBar)

Residuals:
    Min      1Q  Median      3Q     Max 
-37.074 -14.187   0.267  10.733  46.881 

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)    
(Intercept)      127.1629    32.6111   3.899 0.000283 ***
espresso$tempC    -0.9925     0.3174  -3.127 0.002913 ** 
espresso$prssBar   3.9667     1.0365   3.827 0.000356 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.04 on 51 degrees of freedom
Multiple R-squared:  0.3238,	Adjusted R-squared:  0.2973 
F-statistic: 12.21 on 2 and 51 DF,  p-value: 4.639e-05
```

After running the analysis, we obtain the same result as in our pairwise comparison that the full model performs best. We proceed to check the confidence interval:

```
confint(lm_full)
```

A matrix: 3 × 2 of type dbl

	2.5 %	97.5 %
(Intercept)	61.693359	192.6323817
espresso$tempC	-1.629650	-0.3553504
espresso$prssBar	1.885745	6.0475885

<u>Our final Model:</u>

$$Y_{Crema} = 127.1629 - 0.9925 \cdot X_{temp} + 3.9667 \cdot X_{press}$$

Overlaying the data with our best fit plane, we get the following 3-dimensional graph:

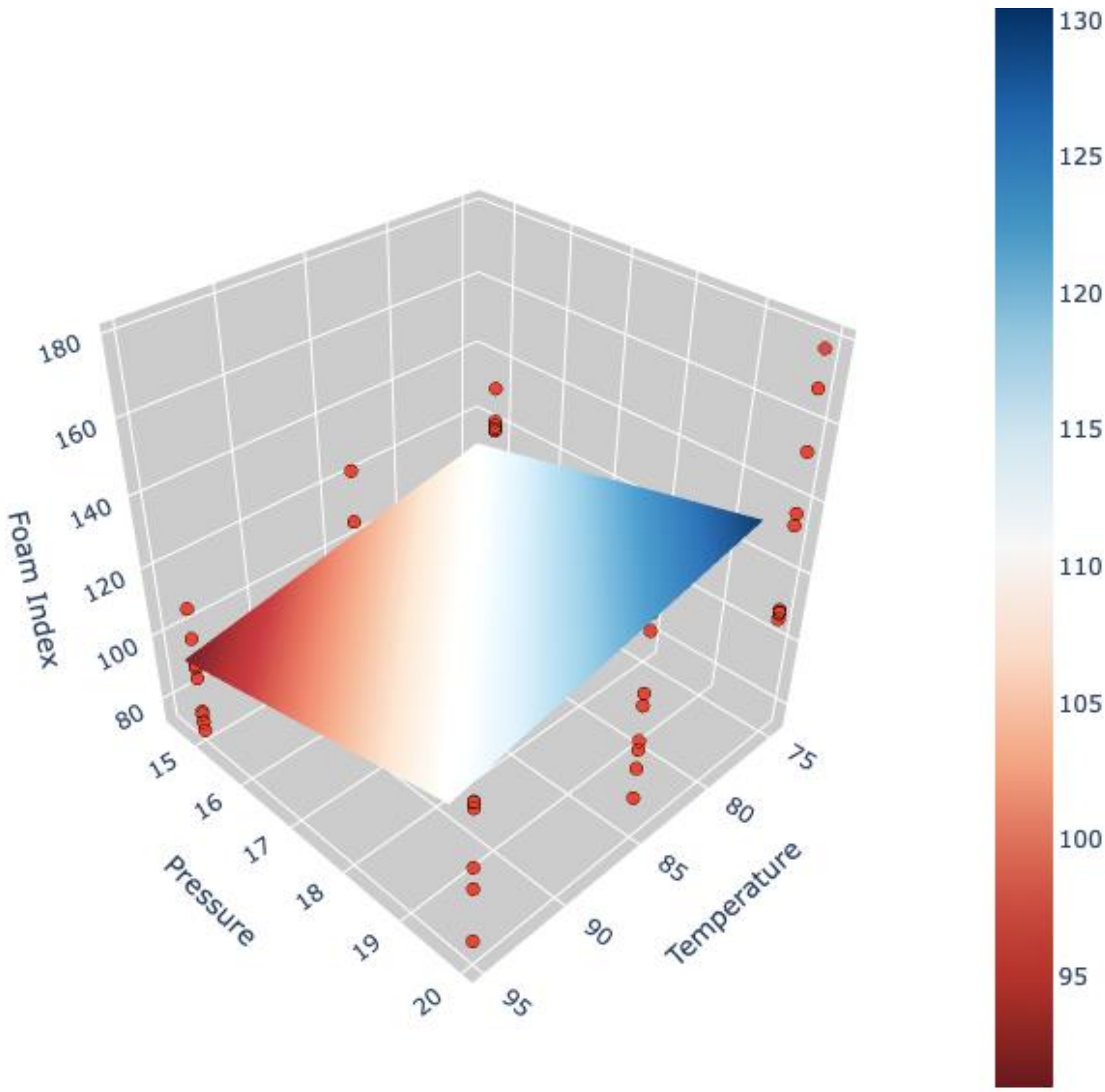


Figure 5: Three-dimensional graph with the best fit

Conclusions

After a detailed analysis of the data, we have decided to reject our null hypotheses that there is no relationship between the pressure and temperature, and the ratio of crema in the espresso brewing process. It must be noted that there is also a large confidence interval in the resultant regression model. This interval will make a consistently accurate prediction impractical at best. It is our suggestion that rather than looking for a specific ratio in the production of crema, the linear relationship should be used as a general guideline for how to adjust the temperature and pressure dependent on what type of drink is being made.

For blended drinks which have a more delicate taste, a latte for example, the temperature could be increased, and the pressure decreased to reduce the crema. This has an added benefit of offsetting the cooling that the beverage will experience as cream is added. On the other hand, in an espresso where no milk will be added, the temperature can be decreased and the pressure increased to produce more crema, and the stronger flavour. In either case, the results are more consistent with gross adjustments and goals rather than fine-tuned adjustments.

References

1. Masella, P., Guerrini, L., Spinelli, S., Calamai, L., Spugnoli, P., Illy, F., Parenti, A. (2014). A new espresso brewing method. Journal of Food Engineering.

2. Block 1912 Café, Baristas Praveer; and Mackenzie interviewed on 2021-11-07

3. Creating the Perfect Espresso Recipe, BLOG // Jori Korhonen // Helsinki // 29.11.2021 et al. https://www.baristainstitute.com/blog/jori-korhonen/november-2021/creating-perfect-espresso-recipe

Division of Work

We as a team of two members concisely planned and conducted this design of experiment and simultaneously contribute as our level best to achieve the optimal result as demonstrated in our final report.

Collaboration Experience

We enjoyed thoroughly and learned a lot while accomplishing this project followed by delivering the presentation at class.