**Factors Influencing the Cooking time of Eggs**

**University of Toronto Mississauga**

**STA305H5 Winter 2023 LEC0102**

Simranjeet Bilkhu, Zhengyang (John) Fei, Abdur Imtiyas, Raima Mehareen, Jing Mo, Marving Roopchan, Yuxin Zhang

Mar 26, 2023

# **Introduction**

Although the consumption of eggs is common worldwide and considered a staple in many households due to its richness in protein, the cooking time of an egg is widely disputed. The question of how to maximise the cooking time of an egg often crosses the mind but is often neglected due to the lack of consensus worldwide. It is this exact question that we set out to investigate and shed light onto with our experimental study.

In particular, we aim to answer the question whether factors such as water temperature and water salinity influence the internal temperature of an egg given a controlled 10 minutes of boiling time. The results that can be obtained from this study are essential as it will reveal what factors impact the cooking time of an egg and help individuals boil their eggs to perfection.

## Methodology

### Variables

In the experiment factors influencing the cooking time of an egg three variables were considered two explanatory or independent variables and one response or dependent variable. The explanatory variables were the salinity and starting temperature of the water in which the egg was boiled. Both of these variables were qualitative in nature. The salinity of the water was binary, with two possible levels: 0 grams/litre or 50 grams/litre, making it a categorical variable. Similarly, the starting temperature of the water was categorical, with two possible levels: 10 degrees Celsius or 20 degrees Celsius. The response variable which was measured was the temperature of an egg in degrees celcius after being boiled, which is a quantitative variable. To measure the salinity of the water, 37 grams of salt was measured with a cooking scale and then added to 750ml of water, otherwise the water was left unsalted. To measure the starting temperature of the water, a digital thermometer was used to ensure that it was either 10 degrees Celsius or 20 degrees Celsius. Finally, after boiling the egg for 10 minutes, we measured its temperature after resting for thirty seconds using the same thermometer.

### Observations and Treatments

The total number of observations in this experiment was 40; there were 40 experimental units, each experimental unit being a large egg. Each explanatory variable has two levels, this experiment is crossed thus all possible combinations of factors and their levels were possible so there were four possible treatments that were used. Treatment one was unsalted water with a starting temperature of 10 degrees celsius, treatment 2 was unsalted water with a 20 degrees celsius starting temperature, treatment 3 was water that was salted with 50 grams/litre with a 10 degree celsius starting temperature, and finally the fourth treatment was salted water with 50 grams/litre with a 20 degree celsius starting temperature. The starting temperature of water is simply the temperature of water prior to boiling. The 40 experimental units were numbered one to forty and using a random number generator in R without replacement a random number from this range was selected and the egg which corresponded to that number was selected. The selected eggs were assigned treatment 1 then 2 and 3 and finally the fourth treatment this cycle was repeated until all eggs were assigned a treatment.

### The Experiment

To conduct this experiment, several steps were required, and three of the four experimental design principles were incorporated. Forty large eggs were randomly selected and boiled one at a time in a pot containing 750ml of water with four possible treatments, as mentioned earlier. After placing each egg in the pot of water, the heat was turned on, and a ten-minute timer was started. Once the timer was up, the egg was removed from the pot and allowed to rest for thirty seconds in a bowl at room temperature. The temperature was then measured using a digital thermometer. Since the temperature fluctuated constantly, the temperature was recorded for two minutes and the maximum reading was taken as the result for that specific treatment. To prevent nuisance factors, randomization, replication, and control were employed in the experiment. As all experimental units were homogeneous, blocking was not applicable for this study. Randomization was used as all forty eggs were numbered one to forty and using a random number generator in R without replacement a number was selected from this range. Once the number was selected that corresponding egg was assigned a treatment. The randomly selected eggs were assigned treatment one then two then three and finally four then the cycle was repeated, once this process was over there were ten randomly selected eggs in each treatment group. This is also demonstrative of treatment level replication as each treatment group contained 10 randomly selected experimental units or eggs, which in turn reduces the total variability within treatment groups. Finally, a multitude of control variables were considered and kept constant throughout the experiment. These consisted of using the same pot and initial pot temperature, stove top and stove top temperature, constant volume of water at 750 ml per egg, egg size and brand of egg, constant egg starting temperature, boiling one egg at a time, checking the egg was not expired, cooking time at ten minutes, placing the egg in the same location in the pot, type of salt, and finally the thermometer used was the same throughout.

## 

## 

## 

## Analysis/Discussion

### **Model Assumptions**

### **Normality**

As there are only ten observations per group, we check model normality for each group separately using the Shapiro-Wilk test of normality, testing at the 0.05 significance level.

{include r output} !!!

Groups 1, 2, 3 pass the Shapiro - Wilk test of normality (we fail to reject the null hypothesis for each test at the 0.05 significance level). Group 4 fails the test at the 0.05 significance level with a p-value 0f 0.00154148.

**Constant Variance**

To test for constant variance between groups, we proceed with the Bartlett test for homogeneity of variance. We note that the results of the Bartlett test may be suspect due to the violation of normality indicated by the Shapiro - Wilk test.

{inlcude r output} !!!

With a p-value of 0.905 reported from the Bartlett test, we fail to reject that the variances between groups are not the same, and we satisfy the assumption of homogeneous variance between groups.

This paper will proceed to use data analysis techniques that rely on the assumption of normality even though it has been violated as non parametric alternatives have not been discussed in class

## 

## 

## Discussion/Results

In general, three methods were used: a model with interaction, an additive model, and Tukey's Pairwise comparison.

For the model with interaction, we tested two factors (salt, and water temperature) and their interaction effect.

\includegraphics[width=\textwidth]{interaction plot 2.png}

\includegraphics[width=\textwidth]{interaction plot.png}

The linear regression model and two-way ANOVA table have been computed. The null hypothesis is that there is no interaction effect between the amount of salt and water temperature on the internal temperature of an egg. Assuming an alpha level of 0.05, the p-value from the ANOVA table is 0.9452302, which is larger than 0.05. Therefore, we fail to reject the null hypothesis at the 0.05 significance level, and we conclude that the interaction effect is not significant.

\begin{center}

\begin{tabular}{||c c c c c c||}

\hline

& Df & Sum Sq & Mean Sq & F value & P value \\ [0.5ex]

\hline\hline

salted & 1 & 102.08 & 102.08 & 16.1496 & 0.0002855 \\

\hline

water temp & 1 & 283.56 & 283.556 & 44.8600 & 0.00000008207 \\

\hline

salted:water temp & 1 & 0.03 & 0.0048 & 0.9452302 & \\

\hline

residuals & 36 & 227.55 & 6.321 & & \\

\hline

\end{tabular}

\end{center}

As the interaction effect was not significant, we proceeded to test the main effects with an additive model. In this case, the two factors were added together and tested separately. The null hypothesis for factor A is that there is no salt effect on the internal temperature of the egg. From the two-way ANOVA table, we obtained a p-value of 0.0002345, which is smaller than the alpha level. Therefore, we reject the null hypothesis at the 0.05 significance level and conclude that salt has an effect on the internal temperature of an egg. Similarly, the null hypothesis for factor B is that there is no water temperature effect on the internal temperature of the egg. The p-value for factor B is 0.00000005413, which is also smaller than the alpha level. Hence, we reject the null hypothesis at the 0.05 significance level and conclude that there is a water temperature effect. From this, it is evident that both factor A and factor B are significant.

\begin{center}

\begin{tabular}{||c c c c c c||}

\hline

& Df & Sum Sq & Mean Sq & F value & P value \\ [0.5ex]

\hline\hline

salted & 1 & 102.08 & 102.08 & 16.596 & 0.0002345 \\

\hline

water temp & 1 & 283.56 & 283.556 & 46.100 & 0.00000005413 \\

\hline

residuals & 37 & 227.58 & 6.151 & & \\

\hline

\end{tabular}

\end{center}

Finally, we conducted Tukey pairwise comparisons between the two factors. Based on the plot, we concluded that there is no interaction effect between salt and water temperature.\newline

\includegraphics[width=\textwidth]{tukey CI.png}

In conclusion, based on the results of the above three methods, we conclude that factors such as water temperature and water salinity had an influence on the cooking time of an egg.

**Limitations**

\section{Limitations}

There were various limitations to the study as expected, however there were two main limitations which likely compromised some of the underlying assumptions that come along with any use of regression models.

The first limitation came from equipment failure. The tool that was used to measure the response (an instant read thermometer), failed with 7 observations to go. The thermometer was replaced with another instant read thermometer of a different brand due to local providers no longer carrying the original thermometer that was used. The new instant read thermometer was then used to measure the last 7=8 observations. This may cause an issue as there is little reason to believe that the accuracy of the new thermometer is the same as the old one. This means that it is highly unlikely that the measurement error of the last 8 observations is not the same as the previous observations. In regression terms, this means that constant variance in the error terms within groups is violated:

$$

V(\epsilon\_{i,j}) \neq V(\epsilon\_{i',j}),

\; \; \; i \in 1, 2, ...., 8. \; \; i' \in 9, 10.

$$

**I’ve decided not to include this part as we only have 3-8 pages for this part of the project, and it is questionable to include as a limitation anyways.**

The second limitation came from the experiment itself. One of the factors was water temperature. Since it is reasonable to suggest that temperature is continuous, it is impossible to measure temperatures to exact accuracy, and very difficult to have thermometer temperatures read at the exact temperature desired. To combat this, it was decided that water temperatures were to be measured within finite accuracy. This means that measurements that are "in the neighbourhood" of a desired measurement are considered, for design purposes, to be at the measurement that is desired. In the case of this paper, thermometer readings that were in the range of $[9, 11]$ and $[19, 21]$ were considered to be equivalent to 10 degrees Celsius, and 20 degrees Celsius respectively.

## 

## 

## Conclusion

The primary objective of this experimental study was to examine whether factors such as water temperature and water salinity had an influence on the cooking time of an egg. Our findings suggest that both the factors water temperature and water salinity had an impact on the cooking time of an egg.

However, there were several limitations to this experimental study that may have impacted the overall results. Since the eggs used in this experimental study were from the same brand and size, this might be indicative that our study might only be accurate for the cooking time of eggs that are of the same brand and size. With the assistance of R code, we came to the conclusion that there was no interaction between the two factors, and that both factors on their own were statistically significant. If we had used various different brands and sizes of eggs then we could have yielded a more comprehensive data set to support our findings. On another note, we could have improved the experimental study by increasing the number of eggs used for each treatment which would have yielded more accuracy.

## 

## 

## Appendix