STAT0029 Report

24076332_24091068_20434451

1 Summary

This study investigates the effects of egg size, storage temperature, and boiling time on the proportion of soft yolk in boiled eggs. A controlled experiment was conducted using eggs of three sizes, as medium size (M), large size (L), very-large size (VL), stored at room temperature (RT) or refrigerated storage (RS). The response variable, the ratio of soft yolk diameter to total yolk diameter, was analyzed using a linear regression model. Results indicate that all three factors significantly influence yolk consistency, with time having the most substantial effect. The final model predicts the soft yolk ratio based on different cooking conditions, showing that refrigerated eggs require longer boiling times to achieve similar yolk textures than those stored at room temperature. The model aligns closely with published culinary data, suggesting a time range of 6–8.5 minutes for achieving a semi-runny yolk. However, confidence intervals reveal considerable variation, indicating that further sampling is necessary for refinement.

2 Introduction

The motivation for studying soft-boiled eggs lies in their role as an excellent source of protein, their high nutritional value, and their widespread popularity in daily diets [1]. Previous studies have shown that M. Lersch et al. [2] found that the refrigeration status of eggs affects their cooking outcome, while P. Roura et al. [3] discovered that the boiling environment influences heat transfer within the egg. It is also hypothesized that egg size may influence the formation of soft yolk. Therefore, the independent variables include whether the egg was previously frozen, egg sizes, and the boiling duration after the water reaches its boiling point (treated as a continuous variable, ranging from 6 minutes to 8.5 minutes, with 30-second intervals). The response variable is the ratio of the soft yolk diameter to the total yolk diameter, defined as:

$$Ratio = \frac{(EL - 2EW)}{EL} \tag{1}$$

Where EL represents the yolk diameter, and EW denotes the thickness of the solid yolk on one side.

The data is then processed. Upon importing the data, it is observed that some ratio values are negative, indicating the presence of outliers, which result from measurement errors. To address these outliers, negative values are set to zero, treating them as fully cooked eggs.

3 Research Method

This section primarily describes the experimental design and the procedures undertaken during the experiment.

3.1 Sample Size Estimation

The primary objective is to determine the required sample size to detect significant effects of two independent categorical variables on the soft yolk ratio: **Temperature** (RS vs. RT) and **Size** (M, L, VL), along with one continuous variable, boiling **Time**. We assume that the data follows a normal distribution and that the variances are equal across all groups, ensuring that the statistical tests used in the analysis are valid. At the same time, as a rough estimate, the interaction between different variables is not considered here. To ensure a balanced experimental design, we randomly selected 6 eggs, covering different egg sizes and storage conditions, at multiple boiling time points. Based on this sampled data, we calculated the standard deviation σ of the soft yolk ratio and estimated the minimum detectable difference δ.

In this estimation, the null hypothesis (H_0) states that **Temperature**, **Size**, and boiling **Time** each have no significant effect on the soft yolk ratio. The alternative hypothesis (H_1) states that each of these factors independently has a significant effect on the soft yolk ratio. A higher power means the experiment is more likely to detect a true effect, reducing the probability of Type II error.

For **Temperature** and **Size**, the required sample size is calculated using the following formula:

$$n \geq \frac{2\sigma^2(z_\alpha + z_\beta)^2}{\delta^2} \tag{2}$$

Where n represents the required sample size per group, and σ was calculated as 0.17 in this experiment. The assumed δ is set to 0.2 cm. Since this study adopts a two-sided test, the significance level $\alpha = 0.05$ is adjusted to $\alpha/2 = 0.025$, and the desired statistical power is set at $1-\beta = 0.7$, making $\beta = 0.3$. Based on these parameters, the required sample size for **Temperature** is $N = 17.6$. For **Size**, although it is a categorical variable with three levels, we apply the formula by treating the comparison as a stepwise pairwise comparison. Specifically, we first assume that M-sized and L-sized eggs have a measurable difference in the soft yolk ratio and calculate the required sample size using the standard two-group power formula. Once this distinction is established, we extend the comparison to L-sized vs. VL-sized eggs, assuming that VL eggs are not positioned between M and L but instead represent the upper extreme of the size spectrum. Based on this approach, the required sample size is $N = 26.4$.

Since **Time** is a continuous variable, its power analysis differs from that of categorical variables. We use regression power analysis to determine the required sample size, based on how well boiling time explains the variation in the soft yolk ratio. The formula for calculating the required sample size is:

$$N = \frac{(z_\alpha + z_\beta)^2}{R^2} \tag{3}$$

In Equation 3, N is the total required sample size, and R^2 represents the proportion of variance in the soft yolk ratio explained by boiling time. We assume $R^2 = 0.2$, meaning that boiling time accounts for 20% of the variation in the soft yolk ratio. The calculated required sample size for cooking time is $N = 33.2$.

We take the largest required sample size as the final experimental sample to ensure sufficient statistical power across all variables. Among the previously determined sample sizes ($N = 17.6$, $N = 26.4$, and $N = 33.2$), we set the final sample size to $N = 36$, slightly exceeding the highest requirement.

3.2 Experimental Procedure

Regarding experimental materials, 6 cartons of eggs were purchased from Marks & Spencer, consisting of 2 cartons of medium-sized eggs, 2 cartons of large-sized eggs, and 2 cartons of very large-sized eggs. Each carton contained 6 eggs, totaling 36 eggs.

A pot was first prepared and placed on an induction cooker for the experimental procedure. The pot was filled with 1000 ml of water, ensuring the water level was sufficient to submerge the eggs entirely. The induction cooker was set to its highest setting, and the water was brought to a boil. Once the water reached boiling, one person placed an egg into the pot while another started the timer. When the set time was reached, the egg was immediately removed from the pot, peeled while wearing gloves, and cut in half along its longitudinal axis. The third person then conducted the necessary measurements. This process was repeated until all required data were collected.

To minimize experimental errors arising from individual differences in procedural execution, the tasks of placing the egg in the pot, timing, and measurement were rotated among three team members. Specifically, for the first egg, team member A placed the egg into the pot, team member B handled the timing, and team member C conducted the measurements. For the second egg, team member B placed the egg, team member C managed the timing, and team member A performed the measurements. This rotation continued throughout the experiment to mitigate systematic errors caused by the same individual performing a specific task repeatedly.

4 Research Result

4.1 Preliminary Data Processing

A linear mixed-effects model was employed to investigate the effects of **Temperature** and **Time** on **Ratio** while accounting for potential intergroup variability due to differences in egg size. Since the experimental data include observations from multiple **Size** groups, which may influence **Ratio**, we opted for a linear mixed-effects model to appropriately handle the random variability across **Size** groups while analyzing **Temperature** and **Time** as fixed effects. The model was fitted using the **lmer** function, yielding an intergroup variance for **Size** of 0.022, indicating a certain degree of variability in **Ratio** among different **Size** groups. Subsequently, the intraclass correlation coefficient (ICC) was calculated, resulting in a value of approximately 0.544, significantly greater than 0.05. This suggests that the random effect of **Size** is substantial, justifying its inclusion in the model.

Furthermore, an interaction plot is generated to examine whether there is an interaction effect between **Size** and **Temperature** and whether this interaction influences **Ratio**, as shown in Figure 1.

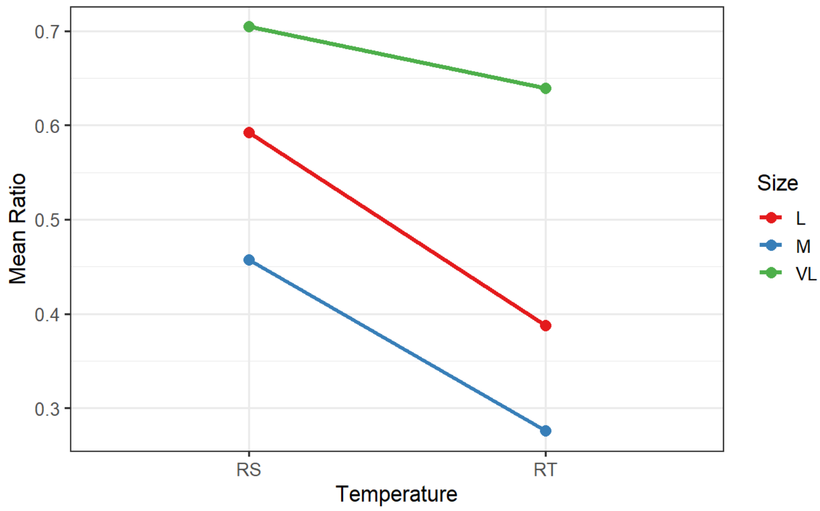

Figure 1: Interaction Effect of Temperature and Size on Ratio

As shown in Figure 1, the lines representing the three different **Size** categories are not parallel. This suggests the presence of an interaction effect between **Size** and **Temperature**, which influences **Ratio**.

Subsequently, interaction effect plots of **Size** vs. **Time** and **Temperature** vs. **Time** were generated, as shown in Figure 2 and Figure 3 .

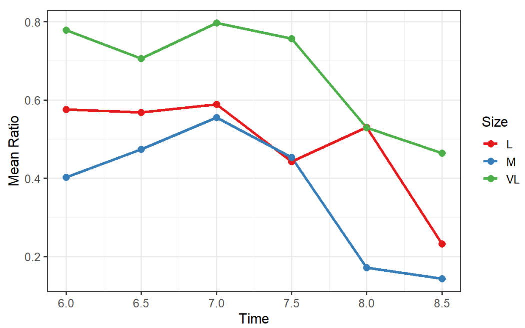

Figure 2: Size vs. Time

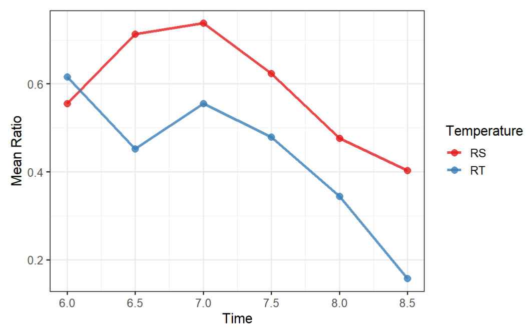

Figure 3: Temperature vs. Time

From Figure 2 and Figure 3, it can be inferred that **Temperature**, **Size**, and **Time** may also exhibit an interaction effect on **Ratio**. Although the overall trend indicates a decrease in **Ratio** over time, the intersecting lines in these plots suggest potential interaction effects among these variables.

4.2 Data Analysis and Regression Result

Model		Df	Sum	Mean Sq	F-value	Pr(>F)
fit1: Time	Time	1	0.392	0.3920	9.922	0.0034 **
fit2: Size	Size	2	0.5665	0.28327	7.997	0.00147 **
fit3: Temperature	Tmperature	1	0.2043	0.20430	4.537	0.0405 *
fit4: Time + Size	Time	1	0.3920	0.3920	16.15	0.000332 ***
	Size	2	0.5665	0.2833	11.67	0.000156 ***
fit5: Size + Temperature	Size	2	0.5665	0.28327	9.397	0.000616 ***
	Temperature	1	0.2043	0.20430	6.778	0.013878 *
fit6: Temperature + Time	Temperature	1	0.2043	0.2043	5.919	0.02057 *
	Time	1	0.3920	0.3920	11.358	0.00193 **
fit7: Size + Temperature + Time	Size	2	0.5665	0.2833	15.34	2.34e-05 ***
	Temperature	1	0.2043	0.2043	11.06	0.00228 **
	Time	1	0.3920	0.3920	21.23	6.60e-05 ***
fit8: Size * Temperature	Size:Temperature	2	0.0337	0.01686	0.543	0.5864
fit9: Time * Size	Time:Size	2	0.0016	0.0008	0.03	0.970112
fit10: Time * Temperature	Time:Temperature	1	0.0264	0.0264	0.758	0.39033

Table 1: Anova Test Results

According to Table 1, the one-way model(fit1, fit2, fit3) shows that **Temperature**, **Size**, and **Time** have significant effects in their model. Compared with the one-way model (fit1, fit2, fit3), the variables in the multiple-way model (fit4, fit5, fit6, fit7) are more significant in the model. Among them, fit7 has the highest significance. The intersection model (fit8, fit9, fit10) shows no significant intersection effect between **Temperature**, **Size**, and **Time**. This refutes the previous hypothesis about the presence of interaction. So, the chosen model is given below:

$$y = \mu + \alpha_i + \beta_j + \gamma x_k + \epsilon_{\mathrm{ijk}} \tag{4}$$

where μ is the total intercept, α_i(i = 1, 2, 3) is the effect of **Size**, β_j(j = 1, 2) is the effect of **Temperature**, x_k is **Time**, γ is the rate of change over time, ϵ_{ijk} is residue.

After obtaining the linear regression results, we further simplified the model:

$$y = c_{\mathrm{ij}} + \gamma x_k \tag{5}$$

where c_{ij} represents the effects of six different conditions(A combination of three egg sizes and whether to refrigerate or not). According to linear regression results, the intercept c_{ij} are 1.328, 1.451, 1.634, 1.177, 1.301, and 1.483, respectively, γ is -0.12221.

It must be noted that the actual range of the response variable ratio is [0,1], which means that this model is only applicable to this interval. When it exceeds this interval, if y is less than 0, all the yolks are dry; if y is greater than 1, it means all the yolks are runny. This model only predicts the process between the beginning of the yolk drying and the complete drying to obtain different degrees of runny eggs.

5 Conclusion

According to the model, to obtain a runny egg with a ratio of about 50%, the required cooking time depends on the size of the egg and whether it is stored in the refrigerator. The estimated times are as follows. From the refrigerator: 6 minutes 47 seconds, 7 minutes 47 seconds, 9 minutes 16 seconds; from room temperature: 5 minutes 37 seconds, 6 minutes 33 seconds, 8 minutes 2 seconds.

Consumers may not always consider egg size when buying eggs. If the egg size is regarded as a random effect, the approximate time to obtain a runny egg of about ratio = 50% should be 8 minutes 16 seconds and 7 minutes 1 second (in and out of the refrigerator, respectively). Comparing the results of this model with data from food websites on the Internet, Moncel's conclusion shows that the time interval for egg yolks to be half thin and half dry is 6-9 minutes, which is similar to the results of the model [4]. If the ratio needs to be adjusted, according to the model, every 1-minute increase or decrease will increase or decrease the ratio by 12.2%.

However, to obtain the corresponding ratio of eggs as accurately as possible, this report suggests that the size of eggs should be considered as much as possible when boiling eggs because the time to obtain the same ratio for eggs of different sizes will vary by 1-2 minutes. For a medium egg that was taken out of the refrigerator and cooked for 6 minutes, with 95% confidence, its ratio range is [0.295 0.895]. This is an extensive range, which means that the experiment needs to wait for more samples to refine the range further.

word count: 1893

References

[1] Yuanhang Shi, Sirui Chen, Yu Liu, Jiayi Liu, Lin Xuan, Guangqi Li, Junying Li, and Jiangxia Zheng. Towards the perfect soft-boiled chicken eggs: defining cooking conditions, quality criteria, and safety assessments. *Poultry Science*, 103(12):104332, 2024.

[2] Martin Lersch. Towards the perfect soft boiled egg, 2017.

[3] P Roura, J Fort, and J Saurina. How long does it take to boil an egg? a simple approach to the energy transfer equation. *European Journal of Physics*, 21(1):95, 2000.

[4] Beth Moncel. How to make soft boiled eggs. `https://www.kaggle.com/datasets/franciscofeng/augmented-china-stock-data-with-fundamentals/data`. Accessed: 2025-02-18.