Factorial Experiment on Removal Rate of Water Filter*

Yanzun Jiang

2025-04-09

1 Introduction

Access to clean water is a global priority, and the efficiency of water filtration is critical. This study investigates the factors influencing the contaminant removal rate of a household water filter using a 2^3 full factorial design with 3 replications. The experiment uses a dataset with three factors: filter type (sand vs. activated carbon), contaminant concentration (10 mg/L vs. 20 mg/L), and retention time (5 mins vs. 15 mins). Each was evaluated at 2 levels. By analyzing these factors and their interactions, this study aims to find the significant factors that affect the efficiency of water filtration, thereby contributing actionable insights for improving water filtration technology.

The full factorial design was selected to evaluate all main effects and the interactions. With 8 experiment runs and 3 replications for each of them (24 runs in total), the design ensures statistical power to detect significant effects while accounting for experimental variability. Replication minimizes the influence of uncontrolled factors, enhancing the reliability of the results. The dataset was obtained from Kaggle. This experiment seeks to address the following question: **Which factor or interaction has significant influence on removal rate?**

*Code and data supporting this paper is available at: https://github.com/Stary54264/STA305-A4

2 Data Analysis

The table below shows the treatment and the outcome for each run. The means and variances for the outcome are also listed in the table. Using the data, it is easy for us to get the main effects and the interactions.

$$ME(A) = 16.40; ME(B) = 3.75; ME(C) = 7.04$$
$$INT(A, B) = -0.58; INT(A, C) = -4.69; INT(B, C) = 1.32$$
$$INT(A, B, C) = 2.43$$

Assume the effects are independent, and the replications are also independent, we could get the pooled variance and the variance for any effect, which is about 28.938 and 4.823 respectively. By assuming a student-t distribution for the effects, and with the variance of them, we could get 95% confidence intervals for them, which is also listed in the table below.

Run	A	B	C	y_{i1}	y_{i2}	y_{i3}	$\bar{y}_i$	s_i^2	95% CI
1	-1	-1	-1	62.48	57.65	54.94	58.36	14.62	(53.70, 63.02)
2	-1	-1	1	69.31	72.71	71.57	71.20	3.00	(66.54, 75.86)
3	-1	1	-1	68.24	62.68	60.46	63.79	16.05	(59.13, 68.45)
4	-1	1	1	82.62	72.67	67.94	74.41	56.12	(69.75, 79.07)
5	1	-1	-1	78.83	81.21	87.33	82.46	19.22	(77.80, 87.12)
6	1	-1	1	83.83	75.43	83.87	81.05	23.61	(76.38, 85.70)
7	1	1	-1	89.90	73.38	82.34	81.87	68.40	(77.21, 86.53)
8	1	1	1	93.84	87.19	82.88	87.97	30.49	(83.31, 92.63)

From Figure 1 we could see that the effect of A (filter type) is at the upper-right corner of the graph, and it is obviously lying above the diagonal. However, Figure 2 suggests that none of the effects are significant, since none of the heights of the lines exceed the ME threshold.

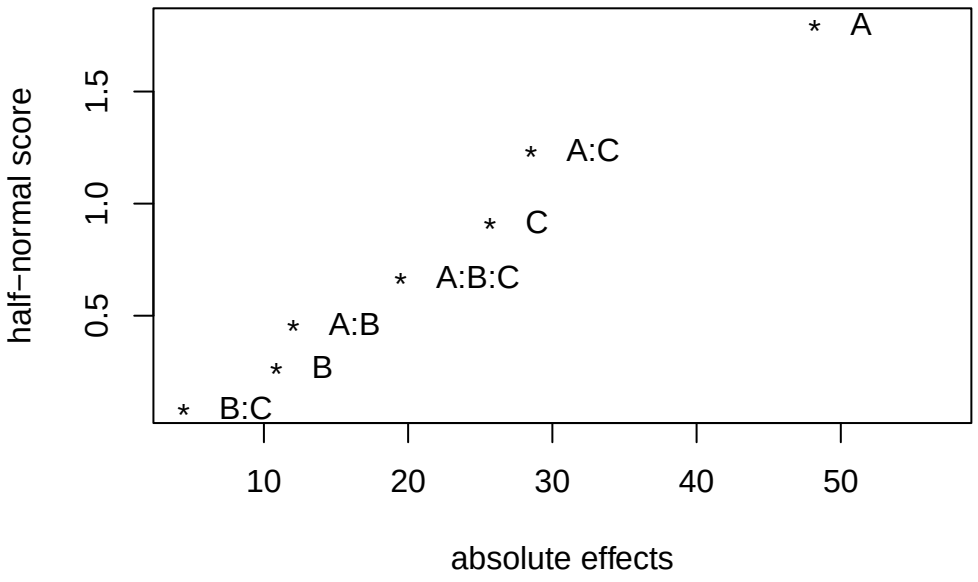

Figure 1: Half-Normal Plot of Effects From Water Filter Study

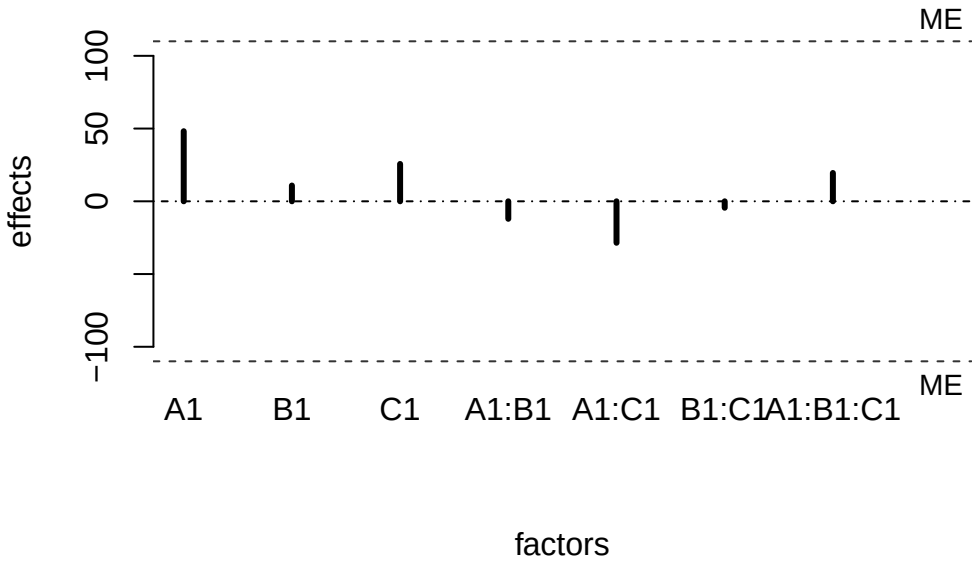

Figure 2: Lenth Plot of Effects From Water Filter Study

3 Conclusion

The experiment found no statistically significant effects with α=0.05 among these factors. This might be due to the need to avoid type II error. However, A (filter type) showed the largest effect compared to B (contaminant concentration) and C (retention time), suggesting it might still influence the efficiency of water filtration but does not meet the threshold. While no definitive conclusions could be drawn, filter type warrants further investigation with refined levels or increased replication to clarify its role.