STA305 L0201 Final Project, W2022

Investigate CPU Usage using Factorial Experiment

Yuchen Zeng, 1006101825, Statistic Major

08/04/2022

Introduction

Computers have been an unavoidable part of our life. As students studying data analysis, we have to use computers almost every day. Sometimes when I use my computer, it becomes very slow and the screen is even frozen. I found that it is the result of high CPU usage. CPU is the physical part responsible for processing tasks in a computer(Britannica, 2018), and people can monitor CPU usage to prevent the system from crashing. I am curious to know which task I run on my computer causes high CPU usage. Therefore, the goal of this study is to find out how CPU usage is affected when different tasks are running on the computer.

Materials and Methods

Experimental Design and Data

The most frequent tasks I use a computer to do are browsing online, playing video games, and recording my screen. Hence, I used a 2^3 factorial design to investigate CPU usage, with two qualitative factors - the state of the screen recorder**(R)** and the state of the video game**(G)**. I also used a qualitative factor - the number of pages browsed**(B)**. This design sufficiently includes the major tasks I use a computer to do, so analyzing this experiment aligns with my objective. Each factor has two levels, coded as -1 and +1, which are shown in the following table.

Factor	Name	Context	Level 1(-1)	Level 2(+1)
1	B	number of pages browsed	10 Pages	30 Pages
2	G	state of video game	not running	running
3	R	state of screen recorder	not running	running

To investigate the CPU usage, my response variable for this experiment is the percentage of CPU occupied by user tasks**(y)**, from 0 to 100. For each setting of factors, I have 2 replicated runs. Therefore, there is a total of 8 *2=16 observations.

To collect the data, I use virtual computers hosted by Azure that has the same version as my computer(Windows 10). Virtual computers operate similar to real computers - they also have CPUs(Sheldon, 2021). Thus, I can record observations independently from different virtual computers while I do not need to occupy too many real computers. For each observation, the setting

remains for 15 minutes before the CPU usage is recorded, to make sure all the tasks are fully operating. For example, if $S = running$ $G = not\ running$ and $P = 30$, I will open 30 pages in the browser and turn on the screen recorder, and close any other tasks on the computer. Then, I keep the virtual computer in this state for 15 minutes before I check the CPU usage.

Data collection was completed on Monday, April 3. The following is the data I recorded.

Run	R	G	B	y
1	1	1	1	81
2	1	1	-1	76
3	1	-1	1	61
4	1	-1	-1	56
5	-1	1	1	71
6	-1	1	-1	63
7	-1	-1	1	46
8	-1	-1	-1	32
9	1	1	1	89
10	1	1	-1	83
11	1	-1	1	69
12	1	-1	-1	53
13	-1	1	1	69
14	-1	1	-1	57
15	-1	-1	1	39
16	-1	-1	-1	20

Statistical Analysis

In the analysis, I used a significance level of $\alpha = 0.05$. I first generated the cube plot of my factorial experiment, which shows the value of ys that are the average of two replicated runs for different combinations of R, G, B. Then, I generated a linear model of the factorial experiment according to the course note(Taback, n.d.)

Let Browser be x_1, Game be x_2, Record be x_3, the Usage from the i^{th} run be y_i, then we have

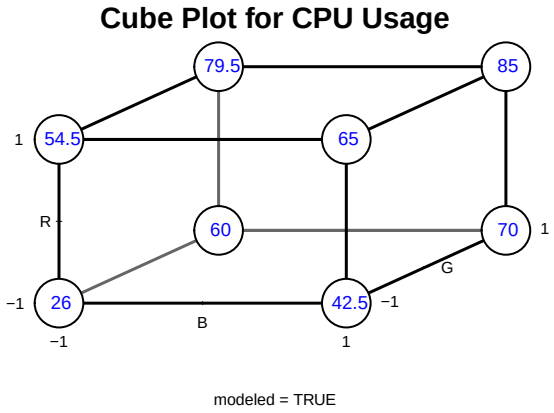
$$x_{i1} = \begin{cases} +1, & \text{if B = Running} \\ -1, & \text{if B = Not running} \end{cases} \quad x_{i2} = \begin{cases} +1, & \text{if G = Running} \\ -1, & \text{if G = Not running} \end{cases} \quad x_{i2} = \begin{cases} +1, & \text{if G = 30} \\ -1, & \text{if G = 10} \end{cases}$$

The linear model for this design is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i1} x_{i2} + \beta_5 x_{i1} x_{i3} + \beta_6 x_{i2} x_{i3} + \beta_7 x_{i1} x_{i2} x_{i3} + \epsilon_i$$

where expression with multiple xs means the interaction, such as $x_{i2}x_{i3}$ means interaction between the state of the screen recorder and the state of the game(Taback, n.d.)
For the effect of each factor, including interactions, there is a hypothesis test. For example, for number of pages browsed**(B)**, $H_0 : \beta_1 = 0$ vs. $H_1 : \beta_1 \neq 0$, where β_1 corresponds to the sample difference in main CPU usage between 10 pages and 30 pages browsed. Assume that the observations are independent and normally distributed, we have $effect/se(effect) \sim t_8$. I then analyzed the p-values of the tests. I calculate main effects and interaction effects from estimates. Confidence intervals are also generated to investigate the true values of effects. Lastly, I created interaction plots and half normal plot to check if they are consist with my result in the linear model.

Results and Discussion



	Estimate	Std. Error	t-value	p-value	2.5%	97.5%	Effect
(Intercept)	60.312	1.279	47.143	0.000	114.725	126.525	120.624
R	10.688	1.279	8.354	0.000	15.475	27.275	21.376
G	13.313	1.279	10.406	0.000	20.725	32.525	26.626
B	5.312	1.279	4.153	0.003	4.725	16.525	10.624
R:G	-2.062	1.279	-1.612	0.146	-10.025	1.775	-4.124
R:B	-1.312	1.279	-1.026	0.335	-8.525	3.275	-2.624
G:B	-1.437	1.279	-1.124	0.294	-8.775	3.025	-2.874
R:G:B	0.188	1.279	0.147	0.887	-5.525	6.275	0.376

The above table shows the estimate, standard error, t-value, p-value, confidence interval, and effect for each factor and their interaction.

Interpreting p-values: Only R, G, B have p-values < 0.05, which means we have significant evidence to reject H_0. Thus, there is evidence that the average CPU usage is different when running the screen recorder or not, running the video game or not, and having 10 pages or 30 pages browsed. All interactions of factors have p-values > 0.05, meaning that we fail to reject H_0 that there is no interaction among the three factors.

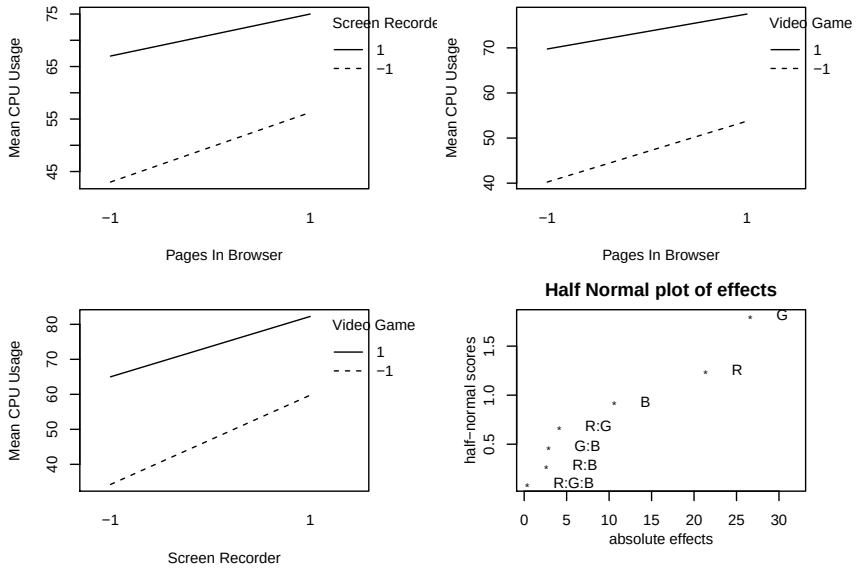
The main effects and interaction effects are calculated by multiplying the estimates by 2.

Interpreting main effects: The mean CPU usage increases by 21.376% when we run the screen recorder, increases by 26.626% when we run the video game, and increases by 10.624% when we browse 30 pages compared to browsing 10 pages.

Interpreting confidence intervals: For true values of effects, the 95% confidence interval is (20.725, 32.525) for running the video game, (15.475, 27.275) for running the screen recorder, and (4.725, 16.525) for browsing 30 pages compared to 10 pages. The confidence intervals of all interactions effects also include 0, meaning that there is no evidence of interactions. For interpretation example, we are 95% confident that the true effect of running the video game on CPU usage lies within the interval between 20.725% and 32.525%.

The standard error for all factorial effects is the same and can be calculated from the standard error of each measurement. Then, $s = 1.279 * 2 = 2.558$, $s^2 = 2.558 * 2 = 6.543$.

The estimate variance of effect for duplicated runs is $Var(effect) = \frac{1}{4}s^2 = 1.636$. The standard error of any factorial effect is $se(effect) = \sqrt{1.636} = 1.279$.



Since in all interaction plots, the two lines are parallel, it indicates no interaction between any pair of the three factors. In the half normal plot, only B, R, and G deviates from the line, meaning that only running screen recorder or not, running video game or not, and having 10 pages or 30 pages browsed are significant factors. This is consistent with our result from the p-value in the linear model.

Conclusion

According to the result of my experiment, there is enough evidence that running the video game, running the screen recorder, and opening 30 pages compared to 10 pages increase CPU usage. Specifically, running the video game raises the CPU usage by 21.38% on average, which is the highest among all three factors. The CPU usage increases by 21.38% when running the screen recorder. Browsing 30 pages instead of 10 pages only raises the CPU usage by 10.625% on average, which is the lowest among all three factors. Also, the experiment suggested that the effect of these three factors are independent. Therefore, it is relatively safe for me to browse many pages at the same time. However, I should avoid running multiple tasks when I am running the video game or the screen recorder, to prevent high CPU usage which may lead to system frozing or crashing.

References

Britannica, T. Editors of Encyclopaedia (2018, March 14).*central processing unit. Encyclopedia Britannica. https://www.britannica.com/technology/central-processing-unit*

Sheldon, R., & Kirsch, B. (2021, March 17). *What is a virtual machine and how does it work? SearchServerVirtualization. https://searchservervirtualization.techtarget.com/definition/virtual-machine*

Taback, N. (n.d.). *Design of experiments and observational studies. 10 Factorial Designs at Two Levels - 2^k Designs. https://scidesign.github.io/designbook/factorial-designs-at-two-levels-2k-designs.html#linear-model-for-a-2k-factorial-design*