Experimental Analysis Using Data Science

Analyzing Factors Impacting Laptop Battery Life

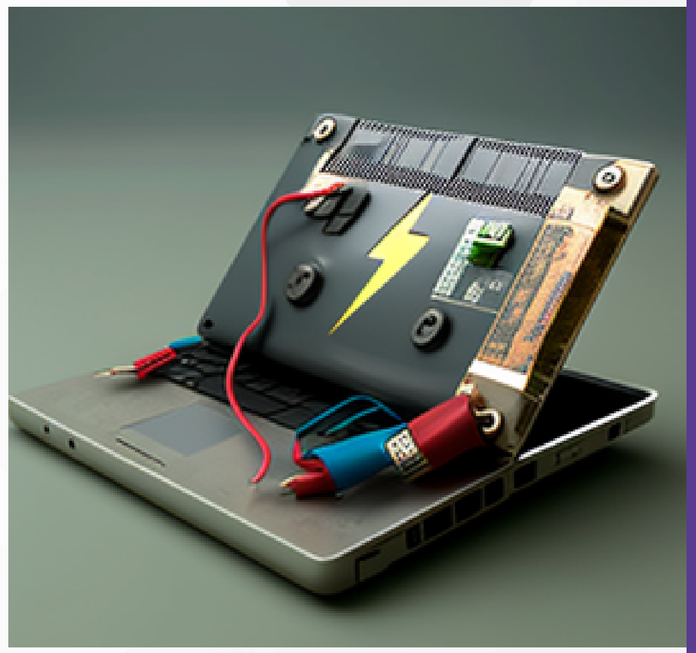
CONTENT

INTRODUCTION

Significance of Battery Life: Understanding battery life is important as it directly affects user experience, productivity, and device portability, influencing consumer choices and technological advancements in energy efficiency.

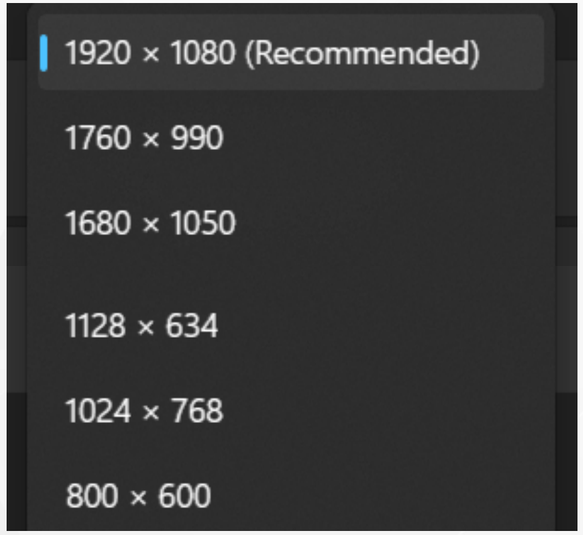
Objectives of the analysis: The primary objective of this analysis is to systematically identify and evaluate the key factors that influence laptop battery life, including hardware specifications, software settings, and usage patterns, to provide actionable insights for users and manufacturers.

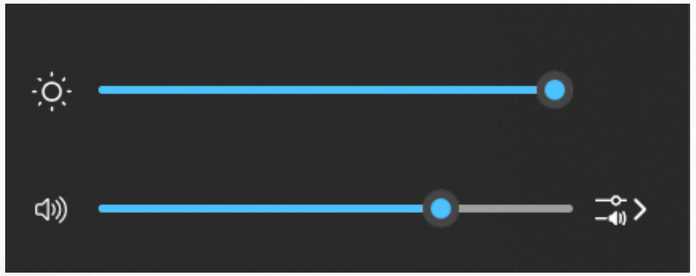
Tools & Methodology used: Utilized advanced statistical software RStudio for data analysis, enabling the application of various regression models and visualization techniques to assess the impact of different factors on laptop battery life.

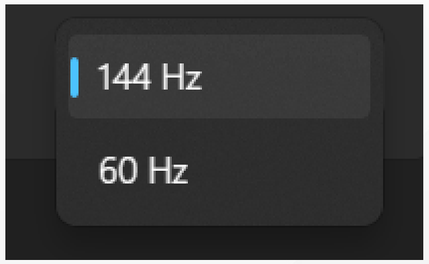


DATA COLLECTION & PREPROCESSING

- Data has been collected using **Pure Battery analytics App**. by testing different combinations of the following factors and measuring how quickly the battery discharged under each condition. For instance:
  - **Refresh Rate** was varied between 60Hz and 144Hz.
  - **Brightness** was tested at low (0%) and high (100%) settings.
  - **Resolution** was either low (900x600 pixels) or high (1920x1080 pixels).
  - **Speaker Volume** was tested as either muted or full volume.







The resulting dataset showed how quickly the battery drained under different combinations while operating

- **Crome Browser +**
- Playing video on **VLC Media**

DATA COLLECTION & PREPROCESSING

- A matrix with 16 experimental runs (factor level combinations) and battery discharge rates was created. This ensures data is structured for analysis.
- There are 4 factors and 2 levels of each high on low.Based on different combinations of factor levels, the observations have been made.
- It can be seen in the table that when all the factors are high, the battery drainage is highest and least when all are at low levels.

	A-Refresh Rate		B-Brightness		C-RESOLUTION		D-SPEAKER.		RATE
	60Hz(Low)	144Hz(High)	0 (Low)	100 (High)	900x600p	1920x1080p	NO(MUTE)	YES(Full)	(%DISCHARGE/Hrs.)
1									31.08
a									39.84
b									37.68
ab									59.68
c									31.32
ac									37.32
bc									35.4
abc									56.72
d									34.2
ad									43.08
bd									38.52
abd									61.36
cd									32.76
acd									39
bcd									37.08
abcd									60.04

INDICATORS :

LOW LEVEL	HIGH LEVEL

DATA PREPROCESSING & TRANSFORMATION

```
> design.matrix
      I  A  B AB  C AC BC ABC  D AD BD ABD CD ACD BCD ABCD  Rate
(1)   1 -1 -1  1 -1  1  1  -1 -1  1  1  -1  1  -1  -1    1 31.08
a     1  1 -1 -1 -1 -1  1   1 -1 -1  1   1  1   1  -1   -1 39.84
b     1 -1  1 -1 -1  1 -1   1 -1  1 -1   1  1  -1   1   -1 37.68
ab    1  1  1  1 -1 -1 -1  -1 -1 -1 -1  -1  1   1   1    1 59.68
c     1 -1 -1  1  1 -1 -1   1 -1  1  1  -1 -1   1   1   -1 31.32
ac    1  1 -1 -1  1  1 -1  -1 -1 -1  1   1 -1  -1   1    1 37.32
bc    1 -1  1 -1  1 -1  1  -1 -1  1 -1   1 -1   1  -1    1 35.40
abc   1  1  1  1  1  1  1   1 -1 -1 -1  -1 -1  -1  -1   -1 56.72
d     1 -1 -1  1 -1  1  1  -1  1 -1 -1   1 -1   1   1   -1 34.20
ad    1  1 -1 -1 -1 -1  1   1  1  1 -1  -1 -1  -1   1    1 43.08
bd    1 -1  1 -1 -1  1 -1   1  1 -1  1  -1 -1   1  -1    1 38.52
abd   1  1  1  1 -1 -1 -1  -1  1  1  1   1 -1  -1  -1   -1 61.36
cd    1 -1 -1  1  1 -1 -1   1  1 -1 -1   1  1  -1  -1    1 32.76
acd   1  1 -1 -1  1  1 -1  -1  1  1 -1  -1  1   1  -1   -1 39.00
bcd   1 -1  1 -1  1 -1  1  -1  1 -1  1  -1  1  -1   1   -1 37.08
abcd  1  1  1  1  1  1  1   1  1  1  1   1  1   1   1    1 60.04
```

- The factors were encoded as -1 and 1 for low and high levels (e.g., -1 for low refresh rate, 1 for high refresh rate).
- Rows (Experimental Conditions): Represent specific scenarios or combinations of factor levels.
  - (1) represents the baseline or control condition, where all factors are set to their low levels.
  - a, b, c, d represent respective factors at a high level each.
  - ab, ac, bc, abc, etc. represent conditions where multiple factors are at their high levels .

- Columns (Factors and Interactions): Represent the variables in the experiment and how they might interact with each other.
  - I is the intercept term or baseline of the experiment
  - A, B, C, D: These represent the main factors or variables in the experiment.
  - AB, AC, AD, BC, BD, CD, etc. are interaction terms between the factors which indicate that the effect of one factor depends on the level of another factor
  - ABC, ABD, ACD, BCD, and ABCD are higher-order interaction terms

DATA EXPLORATION AND ANALYSIS

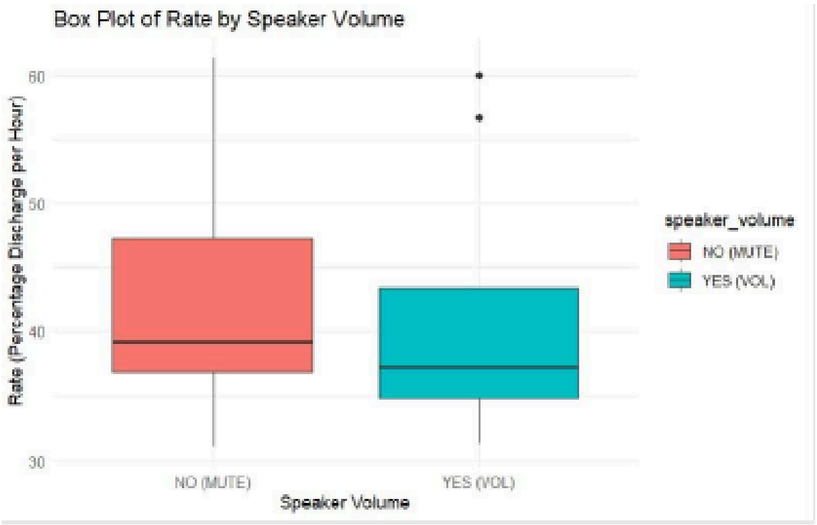
```
> eff=cbind(Ieff,Feff)
> eff
             I      A      B     AB      C     AC     BC
Rate 42.1925 14.875 12.235 7.405 -1.975 -0.745 -0.025

   ABC     D     AD     BD    ABD     CD    ACD    BCD   ABCD
0.605 2.125 0.355 -0.245 0.265 -0.095 0.115 0.715 0.085
```

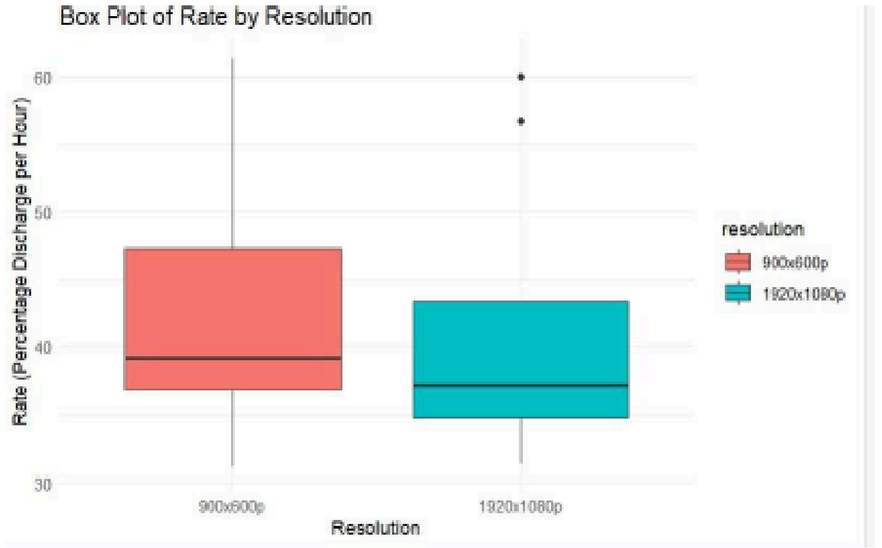
>> FROM ABOVE RESULT WE CAN SEE WE HAVE BOTH +VE AND -VE EFFECT OF FACTORS ON RATE OF DISCHARGE

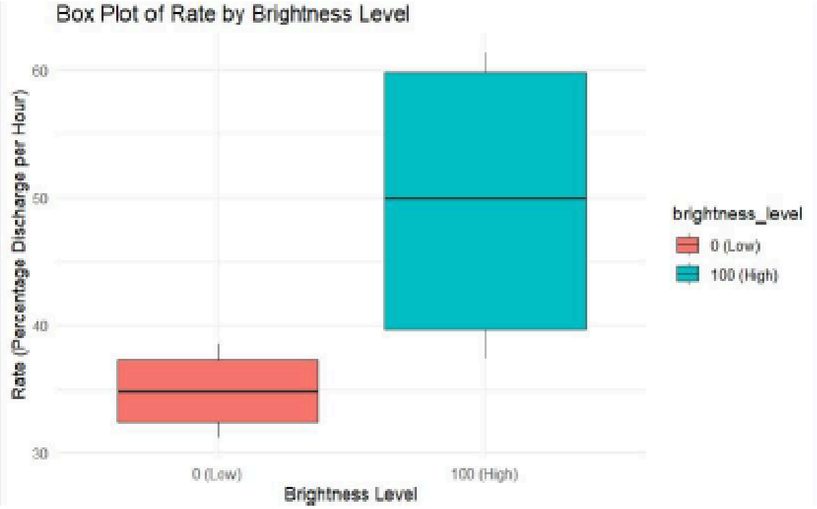
- By looking at the magnitude of these effects, we can determine which factors (or combinations of factors) are the most influential on the outcome.
- factor A has a relatively high effect of 14.875, making it a key player in influencing the response.
- The intercept (I) captures the average response rate when no factors are influencing the system. It is the starting point around which all other effects (main effects and interactions) are compared.

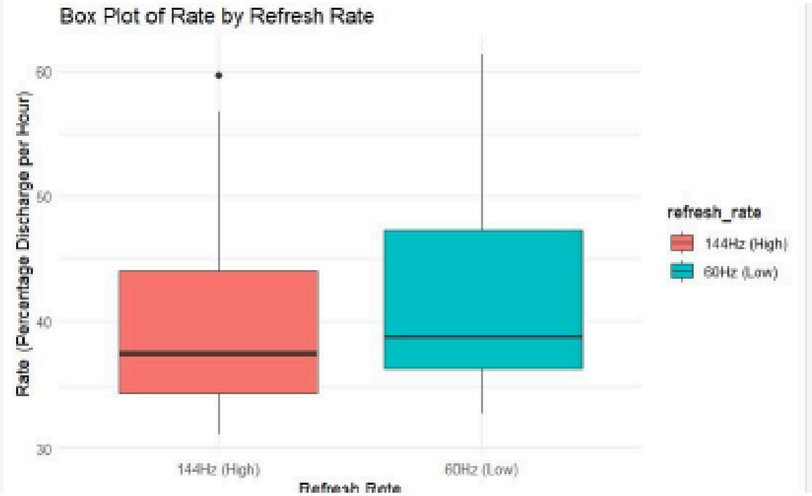
DATA EXPLORATION AND ANALYSIS

1.Box plots



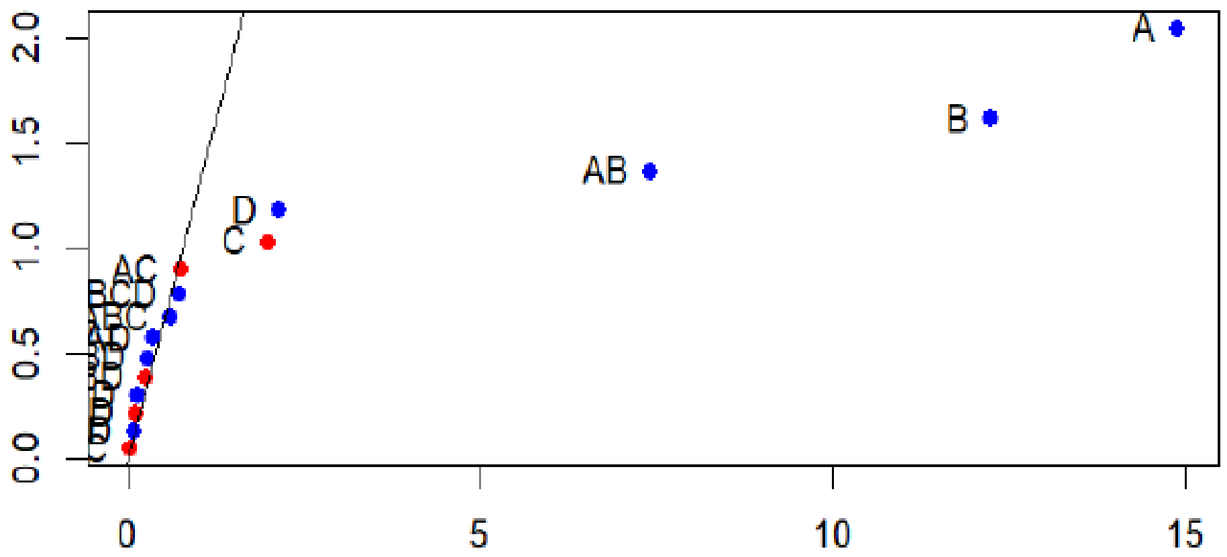






DATA EXPLORATION AND ANALYSIS

2. HALF NORMAL PLOTS



- From the obtained plot, we get that the main factor effects **A and B are highly significant.** Others C & D are less significant
- Also, The **Interaction Effect AB** is also far from the line to be considered as significant Here

- The **red points** in the plot show that the factors have Negative effects.
- The **blue points** in the plot show that the factors have Positive effects.
- The factors far apart from the line are more significant.
- Factor **C(Resolution)** has a Negative effect Surprising. Means better resolution is good for battery life. May be due to inbuilt internal optimization

DATA STATISTICAL ANALYSIS

>Residuals are the differences between the observed values and the predicted values from the regression model that represent the errors or unexplained variance

>>e(x) = yactual – ypredicted

```
> residuals=mod$res
> residuals
    1     2     3     4     5     6     7     8     9    10    11    12    13
-1.26  0.03  0.51  0.23 -1.02 -2.49 -1.77 -2.73  1.86  3.27  1.35  1.91  0.42
   14    15    16
-0.81 -0.09  0.59
```

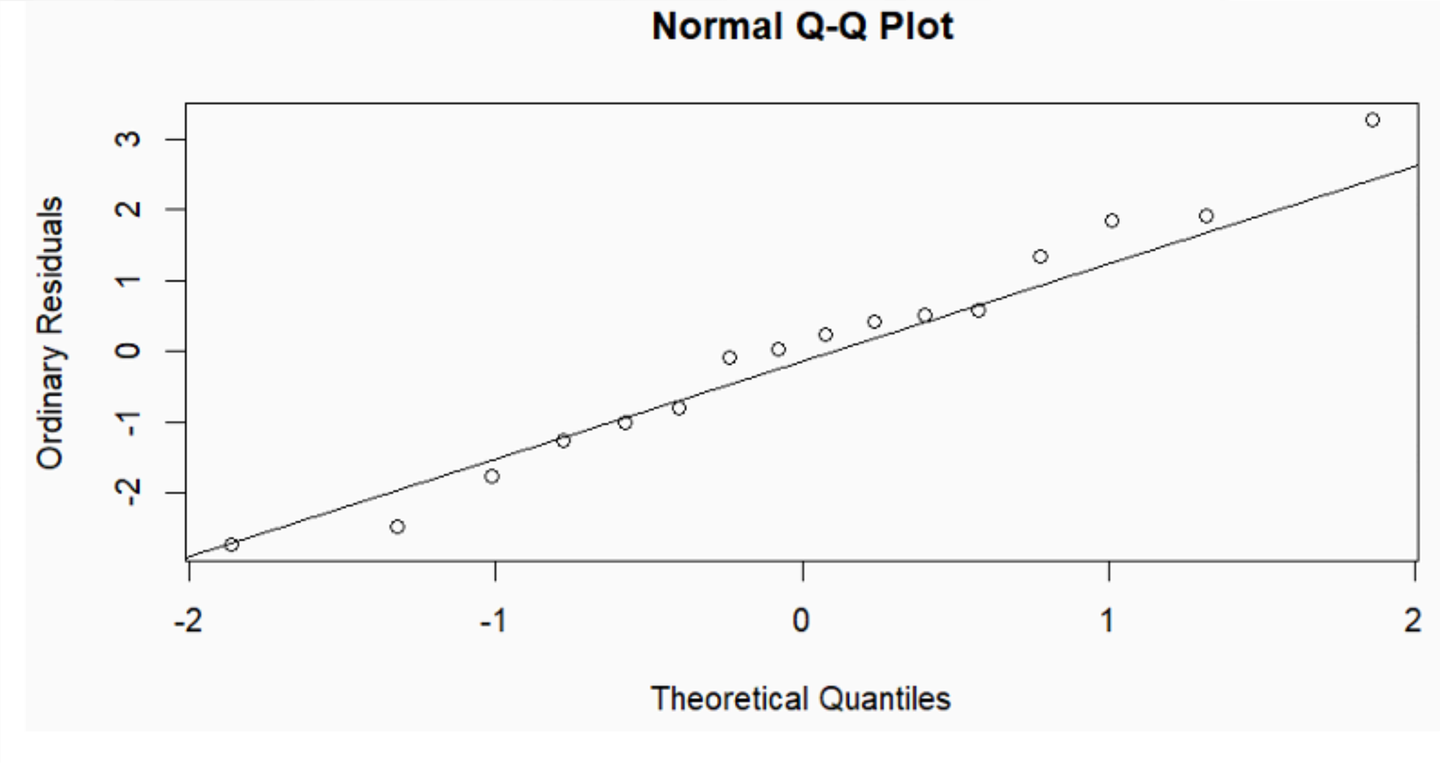
- **Positive residuals:** When the actual value is higher than the predicted value.
- **Negative residuals:** When the actual value is lower than the predicted value.
- **Residuals close to zero indicate** that the model's predictions are accurate

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.1925     0.4595  91.822  < 2e-16 ***
B             6.1175     0.4595  13.313 1.51e-08 ***
A             7.4375     0.4595  16.186 1.62e-09 ***
B:A           3.7025     0.4595   8.058 3.49e-06 ***
---
```

In this case
yp= 42.1925 + 7.4375 x1 + 6.1175 x2 + 3.7025 x1.x2
Here x1, x2 are levels of A and B respectively for each exp.

DATA EXPLORATION AND ANALYSIS

- **Normal QQ plots**
  - Below , the residuals exactly follow the normal distribution.

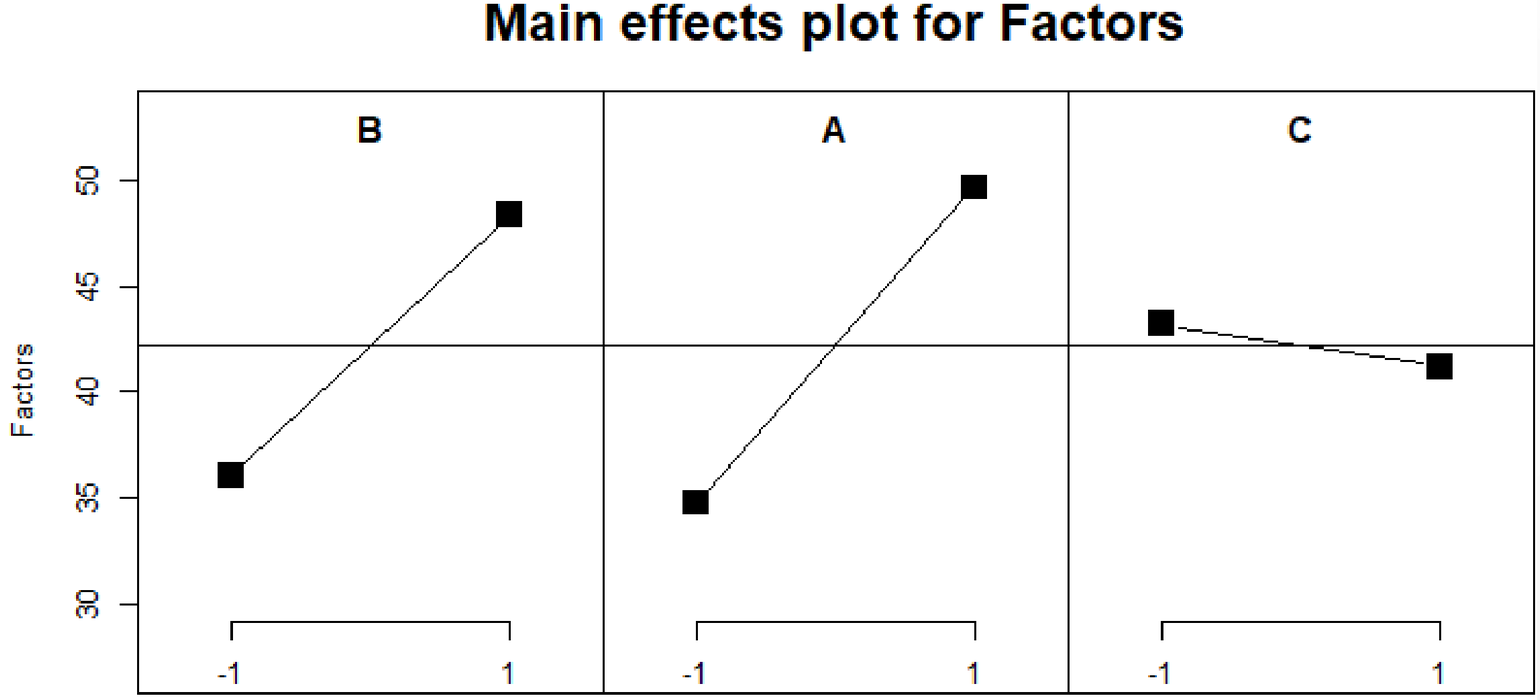


- **Theoretical Quantiles (x-axis)**: They represent the expected values if the data were normally distributed.
- **Ordinary Residuals (y-axis):** Our dataset's residuals (differences between observed and predicted values).

DATA EXPLORATION AND ANALYSIS

MAIN EFFECTS PLOT

- At low refresh rate (A) and brightness (B), the plot shows less discharging than at high.
- Factor C (resolution) has **a Negative effect (Surprising !) .** Better resolution is good for battery life. May be due to **inbuilt internal optimization** is the expected reason .

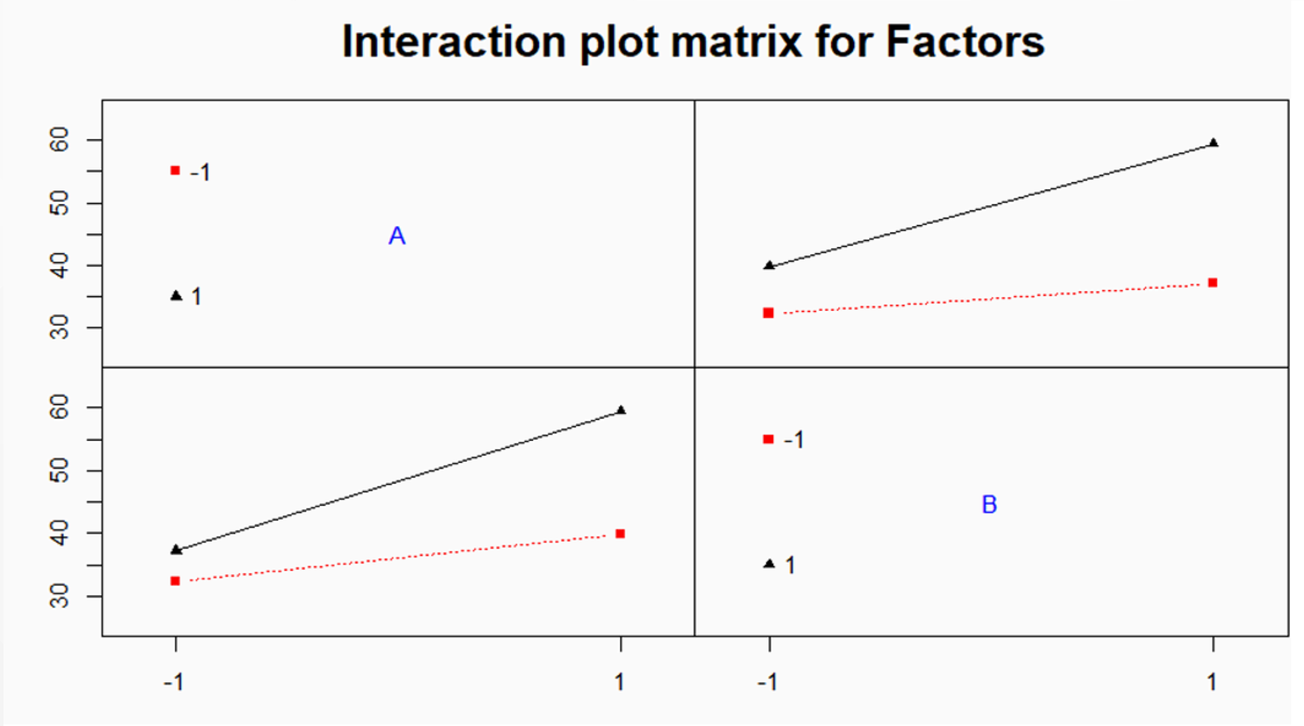


Compared to A and B, factor C has **a lesser and inverse** effect on battery drainage.

DATA EXPLORATION AND ANALYSIS

INTERACTION EFFECTS PLOTS

- In GRAPHS, the **black** line signifies the **high** level of factors A and B, respectively. The **red** line signifies the **low** level of factors A and B, respectively.
- OBSERVATION says , as the lines are not parallel to each other, this **Justify the presence interactions** between these factors A and B .



DATA PROJECTION ANOVA

-omitting factors C and D

USING MULTIPLE LEVEL FOR FACTOR A and B

Multiple levels of significant factors			
A-Refresh Rate	**B-Brightness**		
RATE	L1- 0	L2 -50	L3 -100
144	42.44	50.11	63.2
90	36.3	42.1	49.8
60	34.1	35.7	37.89

FROM THE RESULTS, WE FIND THAT, COMPARED TO BRIGHTNESS (B), REFRESH RATE (A) IS **MORE DOMINANT**. **AS ITS P-VALUE IS BELOW THE 0.05**

```
> summary(rate.aov)
                 Df Sum Sq Mean Sq F value Pr(>F)
refresh_rate      2  387.7  193.86  10.398 0.0260 *
brightness_level  2  244.8  122.39   6.565 0.0545 .
Residuals         4   74.6   18.64
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

DATA HYPOTHESIS TESTING

LSD TEST

It's based on pairwise comparisons and identifies where differences between group means exist.

```
$groups
       rate groups
63.2  63.20      a
50.11 50.11     ab
49.8  49.80     ab
42.44 42.44      b
42.1  42.10      b
37.89 37.89      b
36.3  36.30      b
35.7  35.70      b
34.1  34.10      b
```

WHERE ALPHA = 0.05

Null Hypothesis (H_0): There is no significant difference between the means of the groups.

Alternative Hypothesis (H_1): At least one pair of means is significantly different.

- Group "a" has the highest rate and is distinct from the lowest group.
- Group "b" contains lower rates that are similar to each other but different from the highest rate.
- Group "ab" is a middle group they fall in between the two extremes and can be considered part of both categories

DATA HYPOTHESIS TESTING

TUKEY'S HSD TEST

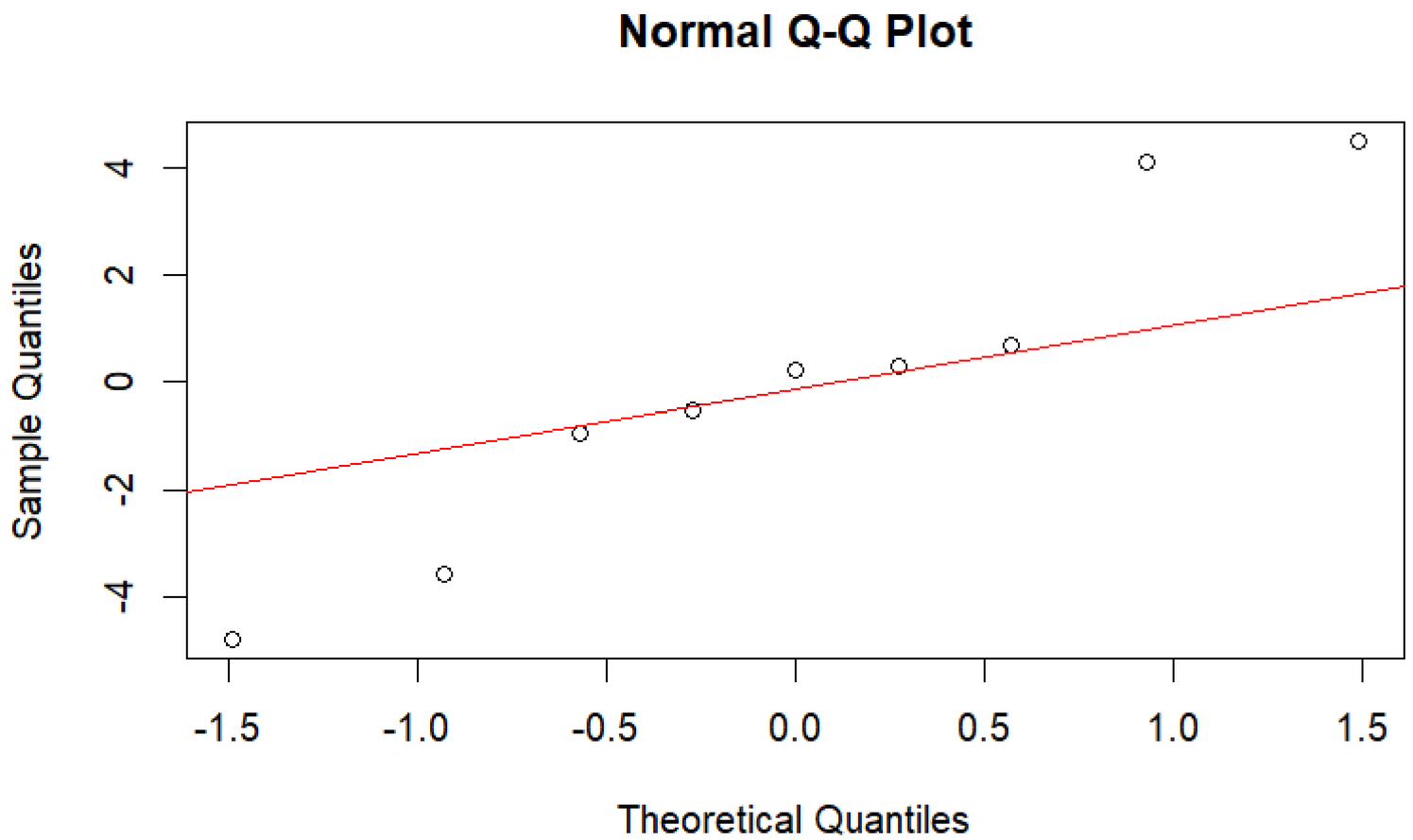
```
$refresh_rate
                diff       lwr      upr     p adj
90Hz-60Hz   6.836667 -5.728199 19.40153 0.2424679
144Hz-60Hz 16.020000  3.455134 28.58487 0.0226806
144Hz-90Hz  9.183333 -3.381532 21.74820 0.1224995

$brightness_level
           diff        lwr      upr     p adj
L2-L1  5.023333 -7.5415323 17.58820 0.4122775
L3-L1 12.683333  0.1184677 25.24820 0.0485634
L3-L2  7.660000 -4.9048657 20.22487 0.1901441
```

COMPARE ALL PAIRS OF GROUP MEANS.

RATE OF DISCHARGING:
144HZ > >90HZ > 60HZ
L3>>L2>L1

- Above shows factors A & B have impact on each level whereconf. Level = 0.95
- For **Factor A-Refresh rate** : there is more difference between mid to high than low to mid, which shows there is higher variation on high Refresh rate .
- For **Factor B-Brightness level** : there is more difference between mid to high than low to mid, which shows there is higher variation on high Brightness . **NOT LINEARLY INCREASING**

DATA EXPLORATION AND ANALYSIS



A little bit Deviations from the line indicate departures from the expected distribution, such as heavier or lighter tails. Additionally, outliers can be found.

The residuals follow the normal distribution

CONCLUSION

1. The most significant factors that determined the amount of time a battery could last were the brightness and the refresh rate.
2. A higher screen resolution had a beneficial impact on the battery life of the laptop, most likely as a result of optimizations made by the laptop.

This project utilized data analysis and statistical methods in order to gain an understanding of how various settings on a laptop affect the battery life of the device. The project found that the refresh rate and brightness rate were the primary factors that contributed to the draining of the battery.

CONCLUSION

- It is concluded that **to maximize battery life:**
  - Refresh Rate should be kept at **60Hz.**
  - Brightness should be on the **lower side**.
  - Resolution can be **kept high** without worrying about battery drain.
  - Speaker Volume did n**ot significantly affect** battery life.

THANK YOU