PROJECT : Introduction to Data Science

Experimental Analysis Using Data Science

CODE LINK : R CODE
DATASHEET : G-SHEET

Members & Contribution
Raghav Vijayvergia (B21ES017) : Coding , Exploration
Aditya Anand (B21ES003) : Coding , Exploration
Mayank Srivastava (B21BB019) : Data collection , ideation
Praveen Kumar (B21CI032) : Data collection , ideation
Mamta Kumari (B21CI025) : Analysis and Interpretation

Problem Statement:

Due to our frequent use of laptops for various activities, including data science projects and everyday tasks, we have noted significant variations in battery performance. This frequently results in disruptions to our workflow, particularly during travel or when away from a power source. We considered identifying the factors that significantly influence battery life, which we can modify to extend the longevity of my batteries. As we suspected that certain factors like the display refresh rate, brightness levels, sound settings, and screen resolution might be playing a significant role in draining the battery. However, we don't have clear data to confirm how much each factor contributes to the issue.

The purpose of this analysis is to understand the impact of these four factors on laptop battery drain. By identifying which settings cause the most significant power consumption, our aim is to optimize our laptop's battery life and make informed decisions about the best configurations for different use cases. This analysis will also help others who face similar battery issues in improving their laptop’s battery performance.

Motivation for the Project:

We want to maximize our screen on time on battery on a single charge. The fast battery drain of high-performance laptops, especially gaming laptops, is a common frustration among users. Gaming laptops, designed to handle high processing demands, tend to discharge rapidly due to resource-heavy tasks such as rendering high-resolution graphics or running at high refresh rates. This issue becomes more pronounced when the laptop is used in performance modes, with high brightness, refresh rate, and speaker volume. The motivation for this project is to understand and analyze the factors contributing to rapid battery discharge in laptops and provide insights on optimizing battery life by adjusting specific parameters such as refresh rate, brightness, and resolution.

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.

Objective:

Analyzing various factors and their impact on LAPTOP battery life.

Tools Used:

- RStudio
- Battery Analytics Application
- Stopwatch
- Hp Laptop

Factors Considered:

A = Refresh Rate
B = Brightness
C = RESOLUTION
D = SPEAKER VOL.

A .Data Collection :

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**

Dataset with different Parameters:

- 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.

INDICATORS :

LOW LEVEL	HIGH LEVEL

	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

B .Data Preprocessing & Transformation :

1 .OUR COLLECTED DATA / WITH LEVEL

```
> data.rate
      A  B  C  D  Rate
(1)  -1 -1 -1 -1 31.08 d    -1 -1 -1  1 34.20
a     1 -1 -1 -1 39.84 ad    1 -1 -1  1 43.08
b    -1  1 -1 -1 37.68 bd   -1  1 -1  1 38.52
ab    1  1 -1 -1 59.68 abd   1  1 -1  1 61.36
c    -1 -1  1 -1 31.32 cd   -1 -1  1  1 32.76
ac    1 -1  1 -1 37.32 acd   1 -1  1  1 39.00
bc   -1  1  1 -1 35.40 bcd  -1  1  1  1 37.08
abc   1  1  1 -1 56.72 abcd  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).

2. DESIGN MATRIX GENERATED

```
> 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
```

- 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**
- 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 high levels** .

3, FACTOR ESTIMATES

	I	A	B	AB	C	AC	BC
Rate	42.1925	14.875	12.235	7.405	-1.975	-0.745	-0.025

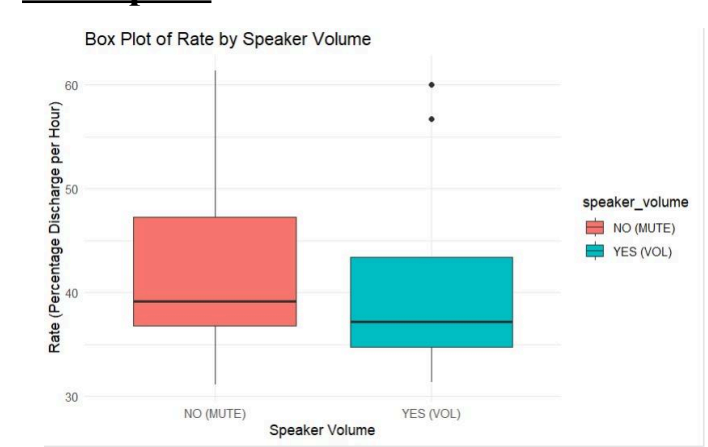
ABC	D	AD	BD	ABD	CD	ACD	BCD	ABCD
0.605	2.125	0.355	-0.245	0.265	-0.095	0.115	0.715	0.085

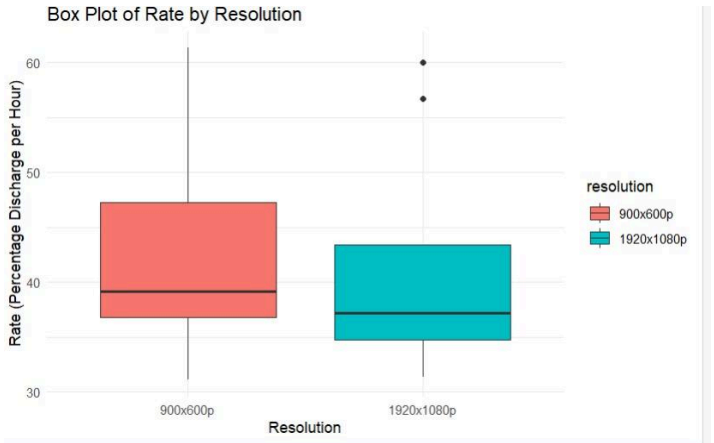
>> **FROM ABOVE RESULT WE CAN SEE that : WE HAVE BOTH +VE AND -VE EFFECT OF FACTORS ON RATE OF DISCHARGE**

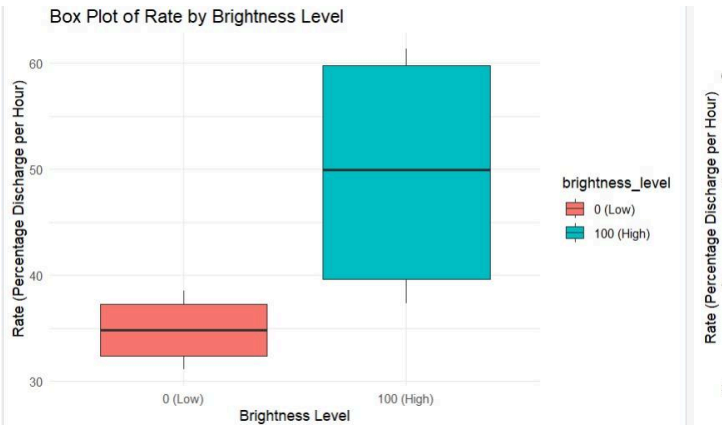
- By looking at the magnitude of these effects, we can determine which factors (or combinations of factors) are the most influential on the outcome.
- **factor A has a relatively high effect of 14.875, making it a key player** in influencing the response.
- **The intercept (I) captures the average response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

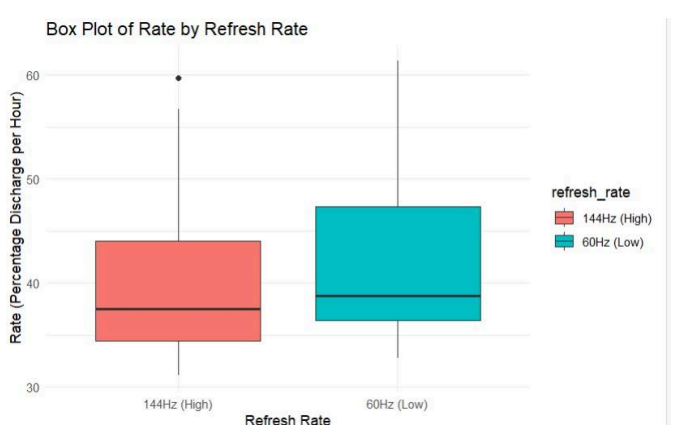
C .Data Exploration and Analysis :

1.Box plots









>>PLOT ANALYSIS

1. The variation in the Box Plot signifies that if we take one factor constant(example brightness level that is low) and change other factors , what would be the range of the rate of percentage discharge per Hour

2. Points Outside the Box Plot (Outliers):

- **The points outside the box (indicated by dots in some of the plots) are called outliers. These are data points that are significantly different from the rest of the data.**
- **Significance of the two outliers:**

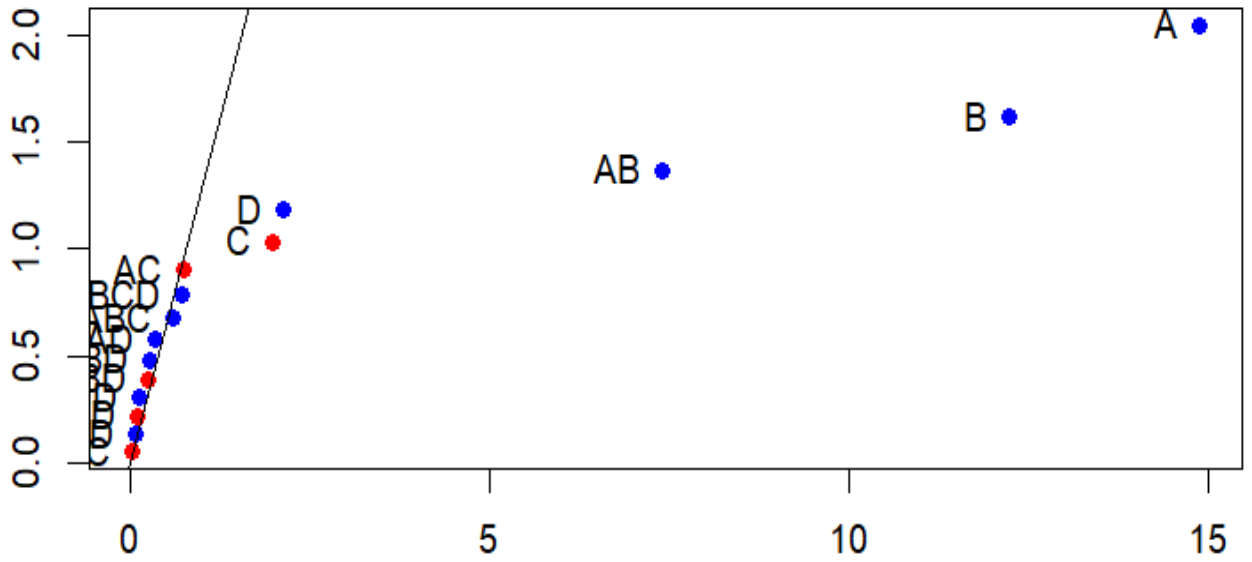
  In the box plot of refresh rate, the outliers represent unusual percentage discharge rates. These values fall far from the typical data range, beyond the "whiskers" (which usually represent 1.5 times the interquartile range).

  These outliers might indicate exceptional cases where the discharge rate was abnormally high or low due to external factors or errors in measurement.

3. No Line in Graph 1 and 2:

- **The Speaker Volume and Resolution box plots show no median line inside the box. This happens when the median value of the data is the same as one of the quartiles, making it overlap with the edge of the box. Essentially, the 25th, 50th, and 75th percentiles are closely aligned.**
- **Significance:**
  - **In the case of Speaker Volume, this likely means that for one of the volume settings (NO or YES), the data points are very tightly clustered, resulting in almost no variation. Hence, the quartiles overlap.**
  - **For Resolution, a similar interpretation applies. The data for one resolution (likely 900x600p) is concentrated, with little variation in the percentage discharge rate.**

2. HALF NORMAL PLOTS



Points that **fall near the straight line** represent effects that are **likely insignificant** (i.e., their variation is due to random chance).
Points that **deviate significantly from the line** represent **significant effects**. These are the factors that have a real impact on the outcome and need attention.

>>PLOT ANALYSIS

- 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**
- 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 .

D.Data Statistical analysis :

```
Residuals:
   Min     1Q Median     3Q    Max
 -2.73  -1.08   0.13   0.78   3.27

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  42.1925     0.4595  91.822  < 2e-16 ***
A             7.4375     0.4595  16.186 1.62e-09 ***
B             6.1175     0.4595  13.313 1.51e-08 ***
A:B           3.7025     0.4595   8.058 3.49e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.838 on 12 degrees of freedom
Multiple R-squared:  0.9768,    Adjusted R-squared:  0.9709
F-statistic: 168.1 on 3 and 12 DF,  p-value: 4.585e-10
```

>>Result ANALYSIS :

Estimated discharge rate As : The regression model is given by

$$yp = \beta0 + \beta1x1 + \beta2x2 + \beta3x3 + \beta4x1x4 + \beta5x1x4 \ldots$$

In this case

yp= 42.1925 + 7.4375 **x**1 + 6.1175 **x**2 + 3.7025 **x**1.**x**2

Here x1, x2 are levels of A and B respectively for each exp.

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
```

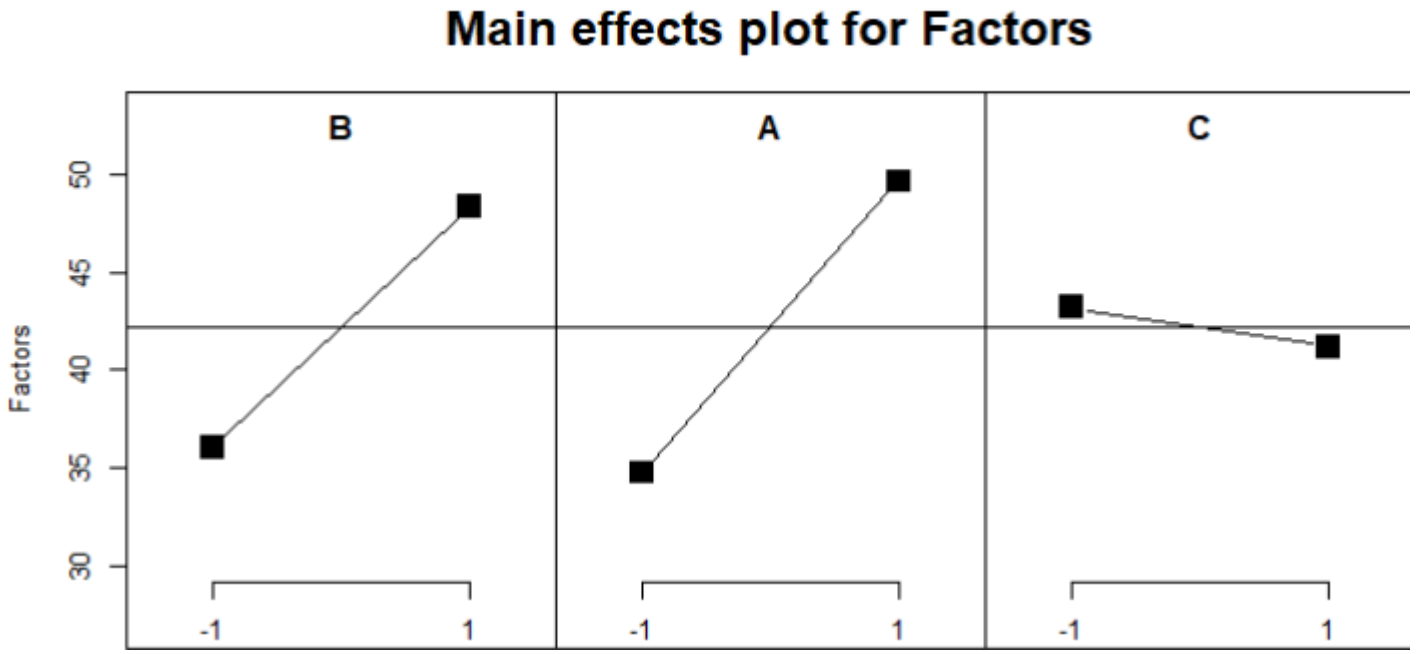
- Positive residuals: When the actual value is higher than the predicted value.
- Negative residuals: When the actual value is lower than the predicted value.
- Residuals close to zero indicate that the model's predictions are accurate

E.Data Effect analysis

1. MAIN EFFECTS PLOTS

A **Main Effect Plot** displays how the mean response of a dependent variable changes as the levels of an independent factor change. It shows the effect of each factor on the response when the other factors are ignored.

- A steep slope indicates a strong effect of the factor on the response.
- A flat line indicates little to no effect.



>>PLOT ANALYSIS

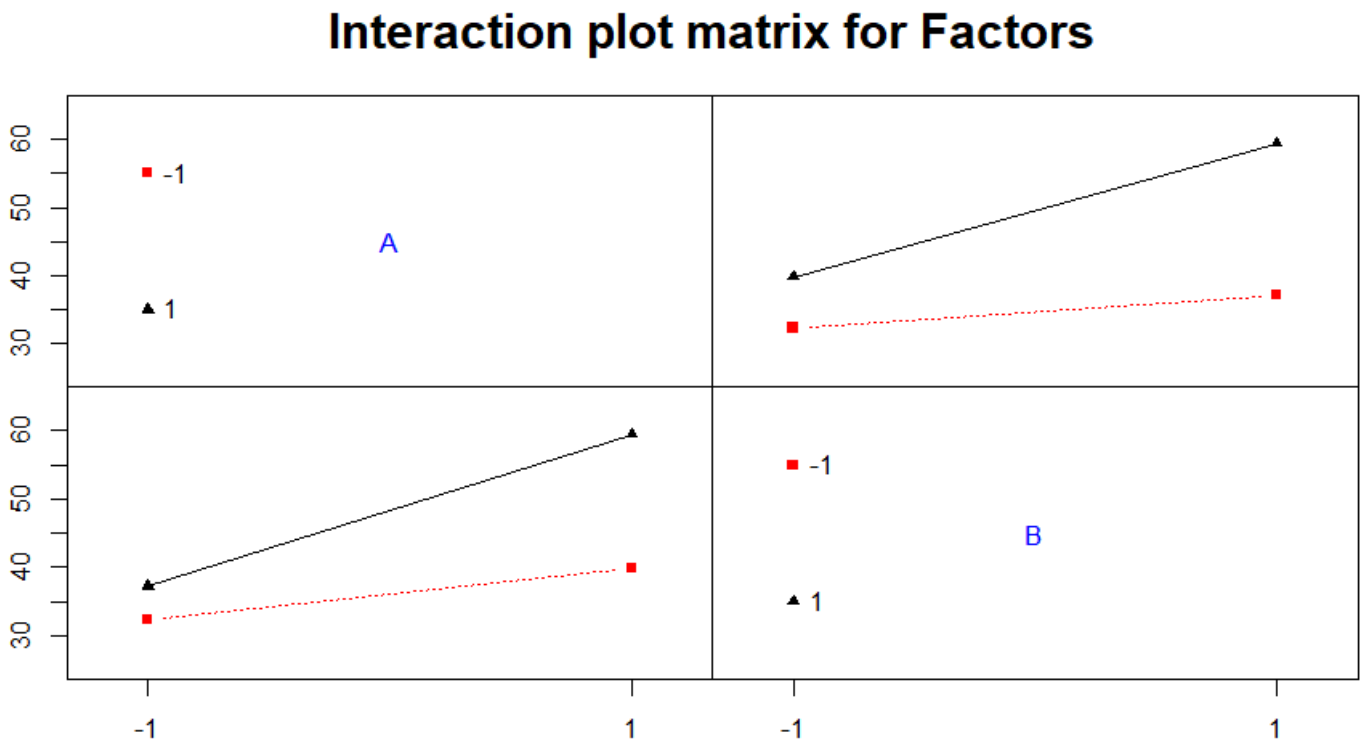
- The plot signifies that for the low level of Refresh Rate (A), the rate of discharging is less, and high for the high level.
- Also , at the low level of Brightness (B), the rate of discharging is less, and high for the high level.
- Compared to A and B, factor C has **a lesser and inverse** effect on battery drainage.

---

2. INTERACTION EFFECT PLOTS

An **Interaction Effect Plot** shows how the interaction between two factors influences the dependent variable. It visualizes whether the effect of one factor depends on the level of another factor.

- Parallel lines indicate no interaction.
- Non-parallel lines indicate an interaction between factors. The more the lines diverge, the stronger the interaction effect.



>>PLOT ANALYSIS

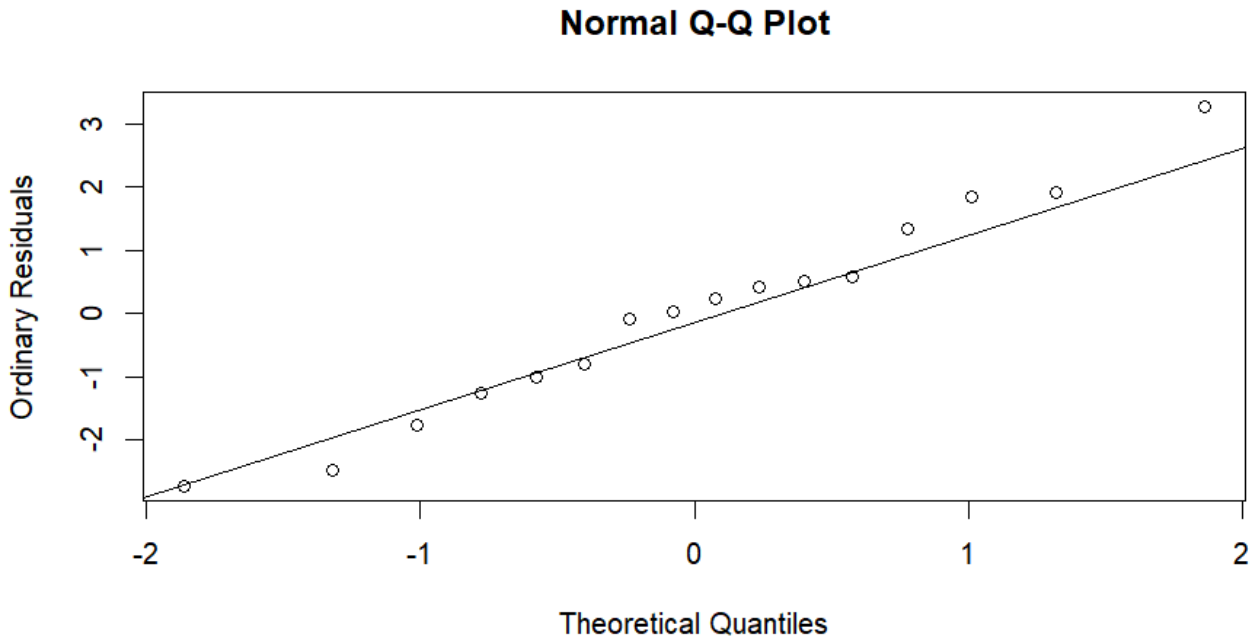
- In GRAPHS, the **black** line signifies the **high level** of factor A and B respectively . The **red** line signifies the **low level** of factor A and B respectively .
- OBSERVATION says as in each block, as the lines are **not parallel to each other**, hence, the interactions between these factors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

---

3. Normal QQ Plot

A **Q-Q plot** is used to assess whether a dataset follows a particular distribution, usually the normal distribution. It compares the quantiles of the sample data to the theoretical quantiles from a specific distribution.If the points fall along the line, the data follows the distribution.Deviations from the line indicate departures from the 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).



Above , The residuals Exactly follow the normal distribution.

D. Data Projection ANOVA(by omitting factors C and D)

USING MORE MULTIPLE LEVEL FOR FACTOR A and B for Detailed Analysis

HIGH , MID , LOW

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

>>STRUCTURE

```
> str(data.rate)
'data.frame':   9 obs. of  3 variables:
 $ refresh_rate    : chr  "144Hz" "90Hz" "60Hz" "144Hz" ...
 $ brightness_level: chr  "L1" "L1" "L1" "L2" ...
 $ rate            : num  42.4 36.3 34.1 50.1 42.1 ...
> data.rate
  refresh_rate brightness_level  rate
1        144Hz                L1 42.44
2         90Hz                L1 36.30
3         60Hz                L1 34.10
4        144Hz                L2 50.11
5         90Hz                L2 42.10
6         60Hz                L2 35.70
7        144Hz                L3 63.20
8         90Hz                L3 49.80
9         60Hz                L3 37.89
```

ANOVA

ANOVA (Analysis of Variance) is a statistical method used to compare the means of three or more groups to determine if there are any statistically significant differences between them. It helps identify whether the variation in the data is due to the effect of different treatments (factors) or random chance.

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.

```
> rate.aov <- aov(rate ~ refresh_rate + brightness_level, data = data.rate)
> 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
```

- F-Value: Indicates whether there is a significant effect. A higher F value suggests greater differences between groups relative to random variation.
- P-Value: If $p < 0.05$ (or other significance levels), we reject the null hypothesis, meaning at least one group mean is significantly different.

FROM Above results we find that compared to Brightness (B) << Refresh Rate (A) is more significant as its p-value is below the 0.05 .

E. Data Hypothesis Testing:

1. 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
```

Obtained the rates into groups are as follows into a,b & ab , where alpha = 0.05

>>Result ANALYSIS

- 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

2. Tukey's HSD TEST : Compare all pairs of group means.

```
$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
```

>>Result ANALYSIS

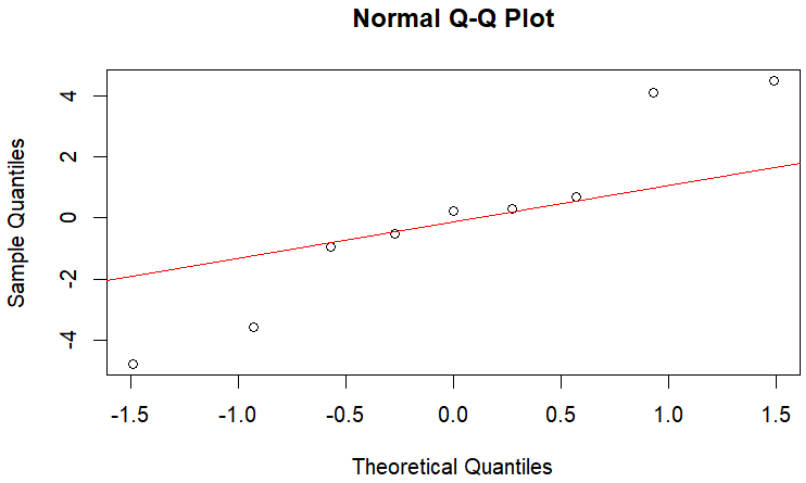
- Above, shows factors A & B have impact on each level where conf. Level = 0.95
- For Factor **A-Refresh rate** : there more difference between mid to high than low to mid shows there is higher variation on high Refresh rate .
- For Factor **B-Brightness level** : there more difference between mid to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

- Rate of Discharging:
  144Hz >> 90Hz > 60Hz
  L3>>L2>L1

ABOVE RESULT implies that using 144Hz or 60Hz is the a crucial factor.

---

3. QQ-PLOT



- A little bit Deviations from the line indicate departures from the expected distribution, such as heavier or lighter tails. Additionally, outliers can be found.
- But overall, The residuals Exactly follow the normal distribution.

F. Conclusion

- Effects of Refresh Rate & Brightness are highly significant.
- Others RESOLUTION & SOUND are less significant .
- Factor C(Resolution) has a Negative effect.
- Means better resolution is good for battery life.
- Compared to Brightness (B) Refresh Rate (A) is more significant.
- Rate of Discharging:
  - 144Hz > 90Hz > 60Hz
  - L3>L2>L1 (but NOT LINEARLY INCREASING)

- **IDEAL CONDITION TO SET FACTORS :**

- REFRESH RATE -60Hz
- RESOLUTION - HIGH (RECOMMENDED)
- BRIGHTNESS - ON LOWER SIDE
- SPEAKER VOL - NOT MATTER MUCH

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.

THANKYOU